LEGISLATIVE COUNCIL

ANIMAL WELFARE COMMITTEE

INQUIRY INTO THE PROPOSED AERIAL SHOOTING OF BRUMBIES IN KOSCIUSZKO NATIONAL PARK

HEARING – Monday 18 December 2023

SUPPLEMENTARY QUESTIONS

Mrs Claire Galea

1. You made a submission to the Federal Inquiry into the 'Impacts and management of feral horses in the Australian Alps'.

    a. Were you invited to give evidence at that Inquiry – if not, why not?

Yes I was invited to give evidence and I presented in person on Monday the 18th of December 20236.

b. Having followed that inquiry and the final report, do you think it was a fair process – if not, why not?

I have not seen the final report and there fire cannot comment on that. However, there was considerable evidence presented questioning the number of wild horses in KNP and it would be in the interest of everyone, including the committee, to question those involved in the collection of the data and subsequent analysis and reporting to provide evidence that has been requested.

c. Do you think the Federal Inquiry listened to the perspectives of experts like yourself, and other groups concerned with protecting brumbies?

I can only speak for myself as I did not hear all of the evidence presented due to work commitments so close to Christmas. The attempt to discredit me as a witness was unexpected and disrespectful. I can only hope that this was not done to other members who were presenting evidence.

2. One argument made by the NSW Government is that the horses are affecting other species of animals in the park - have you seen evidence that brumbies are solely to blame for other species declining or are other issues at play here like the large developments happening at Kosciuszko National Park and climate change?

I am a biostatistician and not an ecologist in terms of the impact of the wild horses on other species or other issues my concern has always been surrounding the methodology used to obtain the population estimates.

However, I do have access to peer- reviewed literature and have attached to this email the following manuscript which demonstrated that 83% of walked line transect in the Eastern Victorian Alps was undisturbed by wild horses and 99% of the Bogong High Plains had no impact of feral horses.

Use of density-impact functions to inform and improve the environmental outcomes of feral horse management

David McKenzie Berman[1], **Jill Pickering**[2], **Deane Smith**[3] and **Benjamin L. Allen**[4,5]

[1]University of Southern Queensland, Toowoomba, QLD, Australia
[2]Australian Brumby Alliance, Victoria Gardens, Richmond, VIC, Australia
[3]NSW Department of Primary Industries, Vertebrate Pest Research Unit, Armidale, NSW, Australia
[4]University of Southern Queensland, Toowoomba, QLD, Australia
[5]Centre for African Conservation Ecology, Nelson Mandela University, Port Elizabeth, South Africa

Correspondence: David McKenzie Berman

WILDLIFE BIOLOGY

Research article

Use of density-impact functions to inform and improve the environmental outcomes of feral horse management

David McKenzie Berman[1], **Jill Pickering**[2], **Deane Smith**[3] **and Benjamin L. Allen**[4,5]

[1]University of Southern Queensland, Toowoomba, QLD, Australia
[2]Australian Brumby Alliance, Victoria Gardens, Richmond, VIC, Australia
[3]NSW Department of Primary Industries, Vertebrate Pest Research Unit, Armidale, NSW, Australia
[4]University of Southern Queensland, Toowoomba, QLD, Australia
[5]Centre for African Conservation Ecology, Nelson Mandela University, Port Elizabeth, South Africa

Correspondence: David McKenzie Berman

Wildlife Biology
2023: e01107
doi: 10.1002/wlb3.01107

Subject Editor: Hideharu Tsukada
Editor-in-Chief: Ilse Storch
Accepted 16 May 2023

The available science often demonstrates the need for feral horse population control but not the degree of control required to achieve environmental conservation objectives. To better manage the influence of feral horses, we must first understand the relationship between feral horse density and environmental impact. We recorded vegetation and soil disturbance, and the sign of potential causes of this impact in two parts of the Australian Alps, the Bogong high plains (BHP) and the Eastern Victorian Alps (EVA). We calculated density-impact functions to assist managers with determining feral horse density targets for control programmes. Minimal sign of feral horse impact was detected on the BHP, with no impact of feral horses observed along 99% of the length of transects. In contrast, impacts assigned to feral horses were significantly higher in the EVA, where a larger, higher-density population of feral horses existed. However, greater than 83% of the walked transect length was still undisturbed by feral horses in the EVA. We detected a threshold of horse impact at ~ 250 horse faecal piles per ha. Above this threshold, a slight increase in horse density resulted in a disproportionately large increase in impact. In this context, a relatively small population control effort may substantially reduce direct horse impact. But where horse densities exist below this threshold, considerably more expense and control effort (resulting from the difficulties related to control at low density) is likely to make very little difference to an already low level of direct impact. The combined impacts associated with the sign of deer, feral pigs, fire and humans were large compared to that of feral horses. Management of feral horses to reduce their direct impact is unlikely to be beneficial without complementary management to reduce the effects of these other agents of impact.

Keywords: brumby, density-impact, environmental impact, *Equus caballus*, mustang, overgrazing, wildlife management

NORDIC SOCIETY OIKOS

www.wildlifebiology.org

Introduction

Effective management of feral horses *Equus caballus* and their environmental impact is challenging due to complex ecological and socio-political issues (Beever et al. 2019, Scasta et al. 2020, Boyce et al. 2021). Terminology complicates the situation further. For instance, the term 'feral' is shunned by some sections of the community because it is considered derogatory, or an undesirable synonym of 'pest' (ITRG 2016, SAP 2020). Others are content with 'pest horse' or 'invasive species' and are sensitive to potential barriers to population control, such as suggestions that feral horses may be beneficial to some economic, environmental or social value. Some use 'free-ranging' or 'wild', but we prefer the term 'feral' because it is accurate given that feral horses are derived from domestic horses. Feral horses have evolved through three phases of selective pressure: the first before domestication, the second requiring adaptation to a domestic state, and the third phase requiring adaptation to living wild, independent of humans once more (Berman 1991). The meaning of the term impact has also become distorted due to the emotion involved and is now commonly believed to imply damaging influences (Driscoll et al. 2019b) when this is often not the case. Here we defined the impact of feral horses to be any change that occurs because of feral horses, whether detrimental, neutral or beneficial.

Eldridge et al. (2020) analysed data from 78 studies across five continents and reported that the impacts of feral horses reduced 'the status of ecosystem characteristics that are important for sustaining all living organisms' (Eldridge et al. 2020) by 13% overall. The magnitude of this decline increased with increases in the density or frequency of use of an area by horses. The direct impact from grazing and trampling of vegetation and soil further increased soil erosion and reduced plant biomass and litter cover (Eldridge et al. 2020). The evidence presented by these studies supports an increased effort to control or remove feral horses, where necessary, to achieve some lower level of environmental impact (TWS 2011, Worboys et al. 2018, Driscoll et al. 2019b, Schoenecker et al. 2021). In contrast, feral horses have been positive and instrumental in the management of vegetation in European conservation reserves (Duncan 1992, Menard et al. 2002, Nuñez et al. 2016) by maintaining short lawns within a matrix of tall grass areas, which promotes plant and animal diversity, thereby increasing the abundance of food for endangered rabbits, waterfowl and seed-eating bird species (Fleurance et al. 2011, Fleurance et al. 2012). Grazing by feral horses or other introduced ungulates can facilitate some native ungulate species in North America (Berger 1986) and macropods in Australia (Newsome 1971). Coprophagous insects or predatory beetles benefit from horses in Europe (Fleurance et al. 2012). Increased insect abundance near dung also provides food for lizards or amphibians (Duncan 1992). These ecological benefits of horses may be associated with returning ecosystem structure or function to the period prior to the disappearance of mega-fauna (Freeland 1990). This could include reduced fuel for wildfire (Rule et al. 2012) or increased plant species and structural diversity (Duncan 1992, Wild and Poll 2012). Evidence that horses can have both damaging and beneficial impacts means there is no one simple message to bring community groups together to support feral horse management. Detailed, careful, scientific monitoring conducted in a way that attracts community support is required to guide sensible and effectual feral horse management, such as was demonstrated in central Australia over 30 years ago (Berman 1991, Dobbie et al. 1993, Nuñez et al. 2016, Braysher 2017).

In New Zealand and the United States of America (USA), feral horses were protected by laws enacted due to fear that these culturally valuable horses would be lost (Rogers 1991, National Research Council 2013). Horse numbers have increased since their protection, putting ecosystems at risk and requiring ongoing management to restrict population growth. The intention in the USA is to keep feral horse populations at 'appropriate management levels' to achieve a 'thriving natural ecological balance and prevent rangeland deterioration with minimal management' (National Research Council 2013). Native wildlife benefits in places where feral horse populations have been maintained at or below appropriate management levels, but feral horse numbers have commonly increased above the intended levels (Coates et al. 2021).

In Australia and Argentina, feral horses have generally been considered invasive species managed to minimise adverse environmental impacts (Nuñez et al. 2016, Scorolli 2018), though this policy position is not unanimous. In the New South Wales section of the Australian Alps (Kosciuszko National Park) the cultural heritage value of feral horses has been explicitly acknowledged since the passing of the Kosciuszko Feral Horse Heritage Bill (2018), where feral horses are to be protected without compromising other values of the national park. While there are highly variable impacts of feral horses on ecosystems worldwide (Monsarrat et al. 2020) and management goals vary considerably, one common theme that could unite the various viewpoints is to maintain horse densities at levels where their ecological damage is minimised and their ecological and cultural benefits are maximised. To achieve this balance it is essential to know how vegetation and ecosystem functioning varies with feral horse density (Dobbie et al. 1993, Walters and Hallam 1993, Dawson 2009, National Research Council 2013), but this area of research has received very little attention (Berman and Jarman 1988, Berman 1991, Beever and Herrick 2006).

Successful management requires selecting a threshold level of population control or density of feral horses, below which their impact is acceptable to the agency responsible and the community, based on scientific measurement of beneficial and detrimental impact. This threshold level could be zero horses in some cases, but removing all horses (local eradication) is rarely achieved and it may or may not be the most desirable goal (Bomford and O'Brien 1995). Feral horses often inhabit remote and rugged areas where it is extremely difficult to locate and remove every individual. Community groups often oppose eradication, and agencies responsible rarely have the resources to overcome the logistical and technical difficulties of achieving eradication. Control methods may also be inadequate, and the required perseverance can be lacking for a very long and challenging process (Matthews et al. 2001, Berman 2013). The high cost of eradication, including continued work to maintain the area free of horses, must be weighed against the cost of ongoing management of a sustainable population that retains its heritage, aesthetic and ecological value. A plan to 'do nothing' with feral horse populations is likely to be unsatisfactory since, without management, feral horses tend to increase to high abundances where food becomes limiting (Dobbie et al. 1993, Dawson and Hone 2012, Zabek et al. 2016b, Scorolli 2021). Where this occurs, horse density will almost certainly be higher than the threshold for acceptable environmental impact. There will also be unacceptable horse welfare impacts as horses suffer prolonged deterioration and death due to lack of food or water (Berman and Jarman 1987, Berman 1991). If eradication is not a viable option and doing nothing is unacceptable, then a target density somewhere in between must be determined (Braysher 1993, Dobbie et al. 1993).

Density-impact functions, such as the relationship between the density of feral horses and the reduced viability of populations of threatened species, can help determine the minimum intensity of control required to reach a target level of impact (Braysher 1993, Hone 1994, Choquenot and Parkes 2001, Hone 2007, Norbury et al. 2015, Braysher 2017). When feral horse density coincides with a maximum acceptable conservation impact, a threshold can be set for imposing control. When feral horse density exceeds this threshold, conservation impacts will be unacceptable, and horse control will be required. Where feral horse density is below the threshold, further control will incur some opportunity cost (for example, the loss of the chance to control deer or feral pigs due to resources wasted by continuing to manage horses with no benefit achieved) and is not recommended (Choquenot et al. 1997, Choquenot and Parkes 2001).

Feral horse management is particularly controversial in the Australian Alps, which is considered home for not only horse people and feral horses but also a range of vulnerable or endangered native animals endemic to this uniquely moist and higher-altitude part of Australia (Carr and Turner 1959a, Calaby and Wimbush 1964, Costin et al. 2000, Green and Osborne 2003). Consequently, there is considerable conflict between stakeholder groups, with the extremes holding very inflexible ideologies, and reciprocal mistrust of Government agencies, community groups, and scientists (ITRG 2016, SAP 2020), as is the case in many other parts of the world (Boyce et al. 2021). Science is essential for informing Government policy and the public to help ensure sensible, successful, and acceptable management of feral horses (Berman 2012). Community engagement and involvement of community groups in research and management have been recommended and shown to improve the acceptance of scientific findings and management approaches (Dobbie et al. 1993, Berman 2013, Scasta et al. 2020, Berman 2021).

Feral horses were established in the Australian Alps by the 1830s, but reports of potential environmental issues did not appear until the 1950's (Costin 1954). It was then almost another 40 years before feral horse impacts were first examined scientifically (Dyring 1990). There has been a recent rapid increase in scientific activity in the Australian Alps (Driscoll et al. 2019a). The hard hooves, grazing, and trampling of feral horses are assumed to cause damage because these large, introduced ungulates did not evolve with the Australian vegetation, soil and water (Wimbush and Costin 1979, Dyring 1990, Lawrence 1995, McDougall 2007, Department of the Environment 2015, Cherubin et al. 2019, Foster and Scheele 2019, Robertson et al. 2019, Schulz et al. 2019). Exclusion plots and sampling at sites with or without feral horses demonstrate their direct impacts, such as selective removal of vegetation by grazing or trampling, exposure and compaction of soil, and deposition of dung (Dyring 1990, Wild and Poll 2012, Williams et al. 2014, Robertson et al. 2019). These studies recorded the direct impact of horses, but how this impact varies with horse density has not been reliably quantified. Consequently, existing studies provide very little guidance for managers seeking to define targets for feral horse density or monitor changes resulting from management actions.

In this study, we aimed to 1) measure the direct environmental impact of feral horses in two separate parts of the area occupied by feral horses in the Victorian part of the Australian Alps, one with low and the other with higher feral horse density, 2) compare impact associated with the sign of feral horses, with impact associated with the sign of other potential agents of impact such as deer *Rusa unicolor*, *Dama dama*, feral pigs *Sus scrofa*, fire, and humans 3) and derive density-impact functions using faecal pile density as an index of feral horse density. Our overall goal was to identify any threshold of feral horse density that stakeholders could use to inform more collaborative management of feral horse environmental effects. We further suggest ways to improve the measurement of the environmental effects of feral horses and other potential agents of impact.

Material and methods

Study area – general

The mountainous area known as the Australian Alps includes the southeastern part of New South Wales (predominantly Kosciuszko National Park) and the northeastern part of Victoria (predominantly the Australian Alpine National Park). In March 2020 and again in February–March 2021, we visited sites in the Victorian part of the Australian Alps (sites were within 11 km of latitude −36.91302°, longitude 147.30136° and within 40 km of latitude −36.92271°, longitude 148.09149°; Fig. 1). In the Victorian part of the Australian Alps, two separate areas occupied by feral horses, about 50 km apart, are known locally as the Bogong high plains (BHP) and the Eastern Victorian Alps (EVA). These areas, respectively, have different feral horse densities of 0.8 and 1.72 feral horses km^{-2}, different population sizes of ~

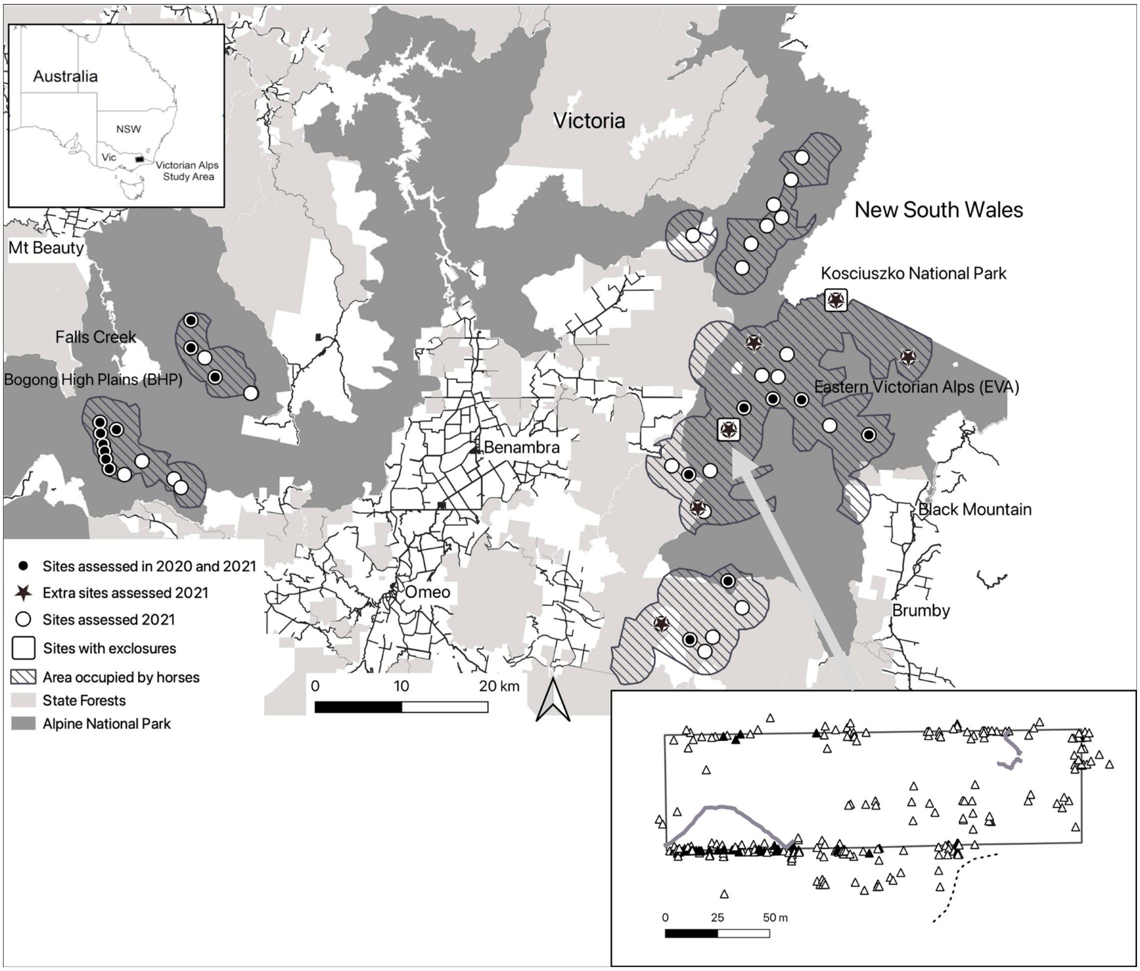


Figure 1. The location of sites visited by us in March 2020 and February–March 2021 in the Victorian part of the Australian Alps (Australian Alpine National Park and State Forests) to measure the environmental impact associated with feral horses. Sites with streams present in alpine treeless drainage line vegetation type, were visited in the two separate parts of the area occupied by feral horses in the Victorian Alps, the Bogong High Plains (BHP) and the Eastern Victorian Alps (EVA). The area occupied by feral horses (hashed) within three kilometres of the entire set of sites established in 2012 (Robinson et al. 2019). Horse occupation confirmation was based on aerial survey (Cairns 2019) in the EVA and ground surveys (Tolsma and Shannon, 2018, Robertson et al. 2019) on the BHP. An example site map shows the 500 m rectangular Site-transect (thin black line), 50 m Streambank-transect (dashed line), animal paths (thick grey lines), horse faecal piles (open triangles) and deer pellet groups (black triangles). Work conducted in 2020 allowed development of methods suitable for the full study in 2021.

100 and ~ 3200 and different rates of increase of < 1.2 and 12.2% per year (Curran 2018, Cairns 2019, Dawson and Miller 2008, Parks Victoria 2017). The management objective on the BHP is eradication, whereas the intention is for feral horses to remain in the EVA, requiring ongoing population control (Parks Victoria 2017, Cairns 2019).

Both areas have a long history of pastoralism involving seasonal grazing by cattle, sheep, and horses (Cabena 1980, Crabb 2003). The EVA was the first area in the Australian Alps where horses were abandoned and allowed to become feral in 1843 (Dyring 1990). Cattle and horses were the main introduced grazing animals in the BHP and the EVA, with horses being the most numerous (Carr and Turner 1959a). For example, Osbourne Young mustered (rounded up) 1500 horses off the BHP in one season in the 1880s for sale as Waler horses (Butler G and Associates 1996), indicative of the relatively large numbers of horses that were there at that time. Breeding and sale of horses for the remount trade was an important economic enterprise during this period, particularly on the BHP. Stocking rates for sheep, cattle and horses were very high, particularly during the droughts of 1884/1885, 1902/1903 and 1914/1915, with sheep numbers reaching at least 40 000 and cattle peaking at over 25 000 in the whole of the Victorian part of the Australian Alps (Cabena 1980). In 1935 there were 6500 cattle on the BHP. However, in the 1950s, sheep and cattle were progressively reduced until all grazing licences were withdrawn from the Australian Alpine National Park in 2005 (Williams et al. 2006). The present population of ~ 3400 feral horses and ~ 3200 cattle (Victorian Department of Energy Environment and Climate Action) represent the few (compared to historical numbers) feral or domestic ungulates remaining in the area. The Australian Alpine National Park is now primarily managed for conservation and recreation. In contrast, the State Forests surrounding it are used for native timber production, with some cattle grazing continuing (Forests Licences and Permits Regulations 2019 Victoria).

Other introduced grazing or browsing species present on the BHP and in the EVA during the study period included sambar *R. unicolor* and fallow *D. dama* deer. There were potentially over 7000 deer in the EVA during our study (Cairns 2019), assuming the reported rate of increase continued. No survey data were available to estimate the number of deer on the BHP. Still, ground survey results suggest the deer density was similar on the BHP to the EVA. Unknown numbers of European rabbits *Oryctolagus cuniculus*, European hares *Lepus europaeus* and feral pigs *S. scrofa* were also present. Native grazing and browsing animals on both BHP and in the EVA include the common wombat *Vombatus ursinus*, red-necked wallaby *Macropus rufogriseus*, swamp wallaby *Wallabia bicolor*, and eastern grey kangaroo *Macropus giganteus*.

The elevations of the study areas ranged from 900 to 1800 m a.s.l. A significant snow cover can last for a few weeks at elevations from 1200 to 1400 m and persist for up to four months of the year (usually June–September) on the highest peaks (Whetton et al. 1996). A mix of open native tussock grassland dominated by *Poa* spp. and heathland *Epacris serpyllifolia* occurred above 1700 m on the BHP. At lower elevations between 1100 and 1700 m, there were patches of open native grassland, heathland and snow gum *Eucalyptus pauciflora* woodland. The snow gum woodland understory was dominated by native grass (*Poa* spp.) or low shrubs *Grevillea australis*, *Hovea montana*, *Leucopogon hookeri*, *Orites lancifolia*, *Bossiaea foliosa* and *Ozothamnus hookeri*. The lower altitude sites of the EVA, between 900 and 1400 m, were predominantly in tall Alpine ash *Eucalyptus delegatensis* open forest with tiny (1–5% of the area) patches of open riparian habitat dominated by *Sphagnum* spp., shrubs *Epacris paludosa*, grass (*Poa* spp.) or *Carex* spp.

The Bogong high plains

The area occupied by feral horses on the BHP was around 132 km^2 (Fig. 1). In this area, there were 90 horses in 2001 (Dawson 2005) and 80–100 horses in 2008 (Dawson and Miller 2008, Parks Victoria 2017). This population appears to have been relatively stable, suppressed by ongoing trapping and removal, with the most recent estimate of 109 horses determined by an aerial survey in 2018 (Curran 2018), shortly before our study commenced. The area occupied by horses is from 1400 to 1800 m a.s.l., with minimum temperatures as low as −9°C in the winter and as high as 30°C in the summer (Falls Creek, Commonwealth of Australia 2021, Bureau of Meteorology). The mean annual rainfall for the year prior to our study for our sites on the BHP was 1499 mm (Interpolated data from the Queensland Government Department of Environment and Science (DES) and the Commonwealth of Australia 2021, Bureau of Meteorology).

The Eastern Victorian alps

The area occupied by feral horses in the EVA is around 1906 km^2. In the EVA, aerial survey indicated that there were approximately 655 horses in 2005 (Dawson et al. 2006) and 3282 horses in 2019 (Cairns 2019), shortly before the commencement of our study. This population had increased within the surveyed area by 15% per year from 2014 to 2019 despite trapping and removal operations (Cairns 2019, Parks Victoria 2017). Our surveys focused on a sub-section of the area occupied by feral horses of around 720 km^2 (Fig. 1) within 3 km of the monitoring sites established by Robertson et al. (2019). This area is between 900 m and 1700 m a.s.l., with minimum temperatures as low as −6°C in the winter and as high as 38°C in the summer (Black Mountain, Commonwealth of Australia 2021, Bureau of Meteorology). At higher altitudes, the temperature was probably more similar to the BHP, but no meteorological stations were closer to these areas. The mean annual rainfall for the year prior to our study for our sites in the EVA was 1018 mm (Interpolated data from the Queensland Government Department of Environment and Science (DES) and the Commonwealth of Australia 2021, Bureau of Meteorology).

Monitoring sites

We established a total of 47 monitoring sites to measure the density and effects of feral horses in the two areas (BHP and EVA) occupied by feral horses in the Victorian portion of the Australian Alps (Fig. 1). All sites were selected within the same vegetation classification (Alpine treeless drainage lines) described by Robertson et al. (2019). No site was all grassland (meadow) or all forest. All sites had a mix of open grassland, heathland or open woodland. A total of 16 sites were on the BHP, and 31 sites were in the EVA. Thirty-three of our 47 sites were a systematically selected (process described below) subsample of a set of study sites established in 2012. These sites were originally randomly selected for the Australian Alps-wide study of the impact of feral horses on treeless alpine drainage lines (Robertson et al. 2019); we refer to these here as Rob-sites (Table 1). There were 8 Rob-sites on the BHP and 25 in the EVA. To increase the number of sites on the BHP, we selected six more existing sites. These sites were originally established in 2017 to study the environmental impact of feral horses (Tolsma and Shannon 2018), and we refer to these here as TS-sites (Table 1). The Rob-sites and TS-sites were selected, independent of the density of feral horses. We monitored the previously studied Rob-sites and TS-sites for comparative purposes. We do not report the outcomes of those comparisons here because they are outside the scope of our objectives, but in summary, we observed little change or difference in horse impact at these sites between 2012/2017 and the time we undertook our surveys in 2020 and 2021. The Rob-sites and TS-sites were used here to compare the two regions (BHP and

Table 1. Description four types of sites visited in February–March 2021 in the Victorian part of the Australian Alps (Australian Alpine National Park and surrounding State Forests) to measure the environmental impact associated with feral horses on the Bogong High Plains (BHP) and in the Eastern Victorian Alps (EVA).

Selection	Site types (number of sites)	Description
Selected independent of feral horse density and used for comparison between BHP and EVA	Rob-sites (33)	Randomly selected in 2012 for the Australian Alps-wide study of the impact of feral horses on alpine treeless drainage lines (Robertson et al. 2019)
	TS-sites (8)	Originally established in 2017 to study the environmental impact of feral horses on the BHP (Tolsma and Shannon 2018).
Selected because of suspected or known high horse activity to populate the upper part of the density-impact curve	Extra-sites (4)	Selected by us in areas with evidence suggesting high horse activity (reports of or observation of horses, an abundance of horse faecal piles and hoof prints)
	Exclusion-sites (2)	At the location of exclosures built in 1999 to demonstrate the impact of feral horses (Theile and Prober 1999)

EVA) because these sites were selected within the area occupied by feral horses independent of observed feral horse activity levels. We selected our remaining six sites by targeting areas with evidence suggesting high horse activity (reports of or observation of horses and abundance of faecal piles and hoof prints etc.) to ensure a wide range of horse densities were available for plotting the relationships between impact and density (as per our objectives, described above). In exploratory analyses not reported here, when we plotted the faecal pile density against impact, there was a clump of 16 sites with very low faecal pile density and only one with substantially higher faecal pile density. A straight line was the best fit for these data and identification of a meaningful density-impact relationship was difficult. With only randomly selected sites like these, we would have required many more such sites to determine the shape of the density-impact function at the upper level of the curve where faecal pile densities are high, but this was not possible in the time available nor necessary for our purpose. Sites with relatively high horse activity or horse density or faecal pile density are easy to identify as you drive or walk through the Australian Alps; horses are seen, and the high densities of hoof prints, faecal piles, and paths are obvious. However, our non-random selection meant that these six sites were not included in analyses comparing the BHP and the EVA because they would have exaggerated the difference between these regions. Four of these remaining six sites are therefore referred to here as Extra-sites (Table 1). The other two sites were at the location of exclosures built in 1999 (Theile and Prober 1999) and are referred to here as Exclusion-sites. Four small (30 × 10 m) horse-proof exclosures were present at each Exclusion site (Table 1). Our surveys were conducted outside the exclosures in areas fully accessible to feral horses. These Exclusion-sites were an excellent reference for the present study showing the extremes of the density-impact relationship. We expected a considerable difference between the randomly selected sites (Rob-sites, TS-sites) and the subjectively selected sites (Extra-sites and Exclusion-sites) for horse faecal density and impact of horses.

No other exclosures established to measure the impact of horses were present in the area. There is one other area near an Extra-site where grazing was excluded in the State Forest. This fence was established to protect rare plants from cattle. Another set of exclosures was established outside of our study area in Kosciuszko National Park. These are mentioned in our discussion. Other long-term exclosures were established on the BHP to measure the impact of cattle (Carr and Turner 1959b), but these are outside the area occupied by feral horses.

To define the geographical extent of our study area, a three-kilometre buffer area from the complete set of Rob-sites was overlayed on the area occupied by feral horses determined by aerial survey (Parks Victoria 2017, Cairns 2019) in the EVA. Aerial and ground surveys on the BHP were used to define the area occupied by feral horses there (Dawson 2005, Tolsma and Shannon 2018, Robertson et al. 2019). We selected the sub-sample from the full set of Rob-sites so that the sites we monitored were 0.5–3.0 km apart and spread as evenly as possible across the full extent of our study area. We had limited access to wilderness areas in the northeastern and eastern parts of the EVA, so a lower density of sites was selected there. Still, our observations indicate that the sites visited adequately represented the feral horse density for alpine treeless drainage lines in these wilderness areas.

Field visits

Pilot study

A subset of 10 sites were visited on the BHP and six in the EVA between 13 March 2020 and 23 March 2020. Bush fires delayed the commencement of fieldwork, and COVID-19 border closures and travel restrictions caused an early end to fieldwork in 2020. Consequently, this field visit was relegated to a pilot study (Fig. 1), and the 2020 data were not included in the results presented here.

Complete survey

All 16 sites were visited (see below for methods) on the BHP and all 31 were visited in the EVA from 19 February 2021 to 24 March 2021.

Monitoring method selection

We based our methods on those developed for feral horse impact assessment in the Australian Alps (Robertson et al. 2019) adapted from the Ephemeral Drainage Line Assessment method (Tongway and Ludwig 2011). Exploratory analyses of data collected during our pilot study indicated that the categorical scoring system used by Robertson could only identify large differences between areas in impact, so we developed a more sensitive method using continuous measures described in the present paper. We used both the Robertson categorical scoring system as well as our continuous scoring system at all sites, but only data from the continuous scoring system are reported here. The Robertson method involves scoring a site by viewing a 20 × 50 m plot and giving scores from 1–5 for various impact variables while for our method we walked along 50 m of stream bank and 500 m through the site recording the number of metres with observed impact and the number of metres with no observed impact. The observer needed to identify vegetation changes caused by horses, such as grass that had been grazed, and the presence of horse faeces or hoof prints. While doing this, the observer recorded signs of other introduced and native wildlife and humans. In other words, we recorded the presence or absence of impact for every metre walked through the site, which is akin to quantifying what an observant bushwalker, park ranger, or horseback rider would see as they move through the site.

Impact of feral horses, deer, feral pigs, rabbits, hare, cattle, fire and humans

Trampling and grazing

Evidence of trampling, and grazing was recorded where there was vegetation broken, bent or trodden into the ground, soil disturbed by hooves or feet or animals rolling/wallowing or vehicles, bitten grass leaves indicated by being square at the top and shorter than pointed leaves of the same grass species, and grass pulled out of the ground.

Potential causes of impact

For each record of grazing or trampling, the potential agent(s) of impact was/were allocated by identifying sign (e.g. foot/hoof prints, faeces, wheel tracks etc.) of the presence of feral or domestic horses, deer, feral pigs, rabbits, hare, cattle or humans observed along the same section of transect. Single or multiple potential causes were assigned to records with no attempt to quantify the relative contribution. Human sign consisted of vehicle paths, wheel tracks or footprints.

Fire impact

Blackened, dead vegetation and bare ground with charcoal and ashes indicated fire occurrence. Nine out of 47 sites surveyed for our study were burnt during the 2019/2020 bushfires.

Streambank-transect

At each site, we surveyed a 50 m section of stream (Fig. 1). The start of each 50 m section was located at the closest point on the nearest stream (drainage line) to the published random point (Robertson et al. 2019), site point (Tolsma et al. 2018) or the additional site waypoints selected by us. The streambank-transect was along both banks parallel to and 1 m from the water or 1 m from the centre of a dry drainage line. The length in metres of transect passing over evidence of trampling or grazing was recorded along with evidence of species (horse, deer, feral pig, rabbit, hare, cattle or human) presence or fire associated with the impact. Since both banks were included, 100 m of stream-bank-transect was surveyed for evidence of grazing or trampling and animal signs at each site.

Site-transects, faecal density and impact

We established and walked along each Site-transect quantifying faeces and evidence of grazing or trampling (Fig. 1). Observations started at each site point, located by GPS, and the observer (Berman) walked along a rectangular transect 200 m east, 50 m north, 200 m west, then 50 m south, ending back at the site point (total 500 m). The number and location of horse, deer, rabbit, hare, cattle or feral pig faecal piles were recorded. For each location where faecal piles were observed, the perpendicular distance from the transect to the pile's centre was recorded to determine density using Distance Sampling (Buckland et al. 1993). The location of both stallion piles (stud piles or dung piles) and individual faecal deposits was recorded as the location of feral horse faecal piles. Stallion piles are multiple faecal deposits on top of each other created by stallions marking their presence. Any faecal pile with more than one faecal deposit was recorded as a stallion pile. The age and size of the pellets distinguished the different faecal deposits on a stallion pile. The number of faecal pellets was counted in stallion piles and in individual faecal deposits to allow estimation of the number of individual faecal deposit equivalents in stallion piles. The mean number of faecal pellets in individual faecal deposits was also counted. We counted the pellets that were not decayed and could still be identified as horse faecal pellets in stallion piles and in individual deposits. If faecal material was too decayed to be identifiable as horse faeces it was not recorded. So we counted the same thing for stallion piles as we did for individual faecal deposits. Stallions eat in the same place with the mares and young they accompany, so their diet should not vary much from other horses. Horse faecal pile density was calculated as the density of faecal pile locations (individual faecal deposits and stallion piles) with no correction for the number of individual defections in stallion piles. Presenting faecal pile density without determining the number of individual faecal deposits in a stallion pile allows others to more easily compare their faecal pile density estimates with ours without the requirement to count individual pellets or weigh faeces. Only if someone wishes to convert faecal pile density to actual horse density is it advisable to follow the methods described to count pellets. Conversions from horse faecal pile density to horse density were based on the number of individual horse faecal deposits, including those in stallion piles. We converted horse faecal pile density to horse density using a 426 day decay rate and eight defecations per day (Linklater et al. 2001, Zabek 2015a, Zabek et al. 2016a).

We also obtained horse faecal pile densities for 20 sites in the EVA surveyed in 2018 by Cherubin et al. (2019) and converted these to horse densities for comparison with our results. Cherubin's sites were in the same area we surveyed in the EVA. They were a sub-sample of the Robertson sites, just like ours were. Some sites were possibly the same as ours, but we could not obtain the actual location of Cherubin's sites. We converted deer faecal pellet group density to deer density using a 71 day decay rate (Davis and Coulson 2016) and 12 defecations per day (Ratcliffe 1987, Nugent et al. 1997, Mayle et al. 1999, Forsyth 2005).

The number of metres of the Site-transect passing over evidence of grazing or trampling and the sign of potential agents of impact was recorded. If a path (large animal trail) was observed crossing the Site-transect, it was mapped using a GPS by walking along its entire length within the area bounded by the 500 m rectangular Site-transect. All paths within the rectangular Site-transect were mapped at each site and were photographed at the points where they crossed the Site-transect, their average width was estimated, and the animal tracks or faeces present on the path were recorded. We recorded that a path had been used by a species based on the presence of their faeces or foot/hoof prints, or wheel tracks. For example, a 50 m section of GPS-mapped path where both horse and deer hoof prints were detected would be assigned to both horses and deer. Assigning a section of path to a species meant that there was evidence that that section of path had been used by that species within the period that animal tracks or faeces remained visible.

Density-impact functions

Feral horse faecal pile density (as a surrogate of horse density) and impact (evidence of grazing or trampling where there was sign of feral horses) data collected at all 47 sites were used to determine density-impact functions for feral horses. Three impact measures: 1) metres of impact along Site-transects, 2) metres of impact along Streambank-transects and 3) the area of path impact at sites – were plotted against horse faecal pile density to determine the strength and characteristics of their relationships. We selected candidate functions after viewing the plots of impact versus faecal pile density. Candidates included the Gompertz (Tjørve and Tjørve 2017), Logistic and Linear functions. Unlike the Logistic function, the Gompertz function is not symmetric around the inflection point.

Statistical analyses

Data were analysed using the R (www.r-project.org) statistical computing environment (ver. 4.1.0). The density of horse and deer faecal piles and the probabilities of detection were determined using the Distance Sampling analysis in the Rdistance package ver. 2.1.3 (www.r-project.org). We conducted t-tests to see if there was a difference between the EVA and BHP in mean feral horse faecal pile density and the mean number of metres with evidence of grazing or trampling associated with the sign of feral horses (Rob-sites and TS-sites data). To see if the mean faecal pile density were different between the four site types, we conducted one-way ANOVAs. Subsequent Tukey multiple comparison post hoc tests were conducted for the significant ANOVA. Density-impact curves were fitted, and the best model was selected based on Akaike's information criterion (AICc) for small samples (Akaike 1998).

Results

Horse faeces and impact

Horse faeces (dung) was detected at 12 out of 16 sites (75%) on the BHP and 24 out of 26 sites (92%) in the EVA. Evidence of feral horse grazing or trampling was detected at all sites where horse faeces was detected (Fig. 2). Horse faecal pile density was lower on the BHP than in the EVA ($t=3.85$, $df=41$, $p < 0.01$; Fig. 2, Table 2), and there were substantial differences between Exclusion-sites/Extra-sites and Rob-sites/TS-sites ($F_{(3,\ 43)}=56.75$, $p < 0.01$; Fig. 3). The probability of detection of a horse faecal pile was 0.30, and the effective strip width was 3.03 m, as determined by Distance Line Transect analysis. The mean length of sections of Site-transects ($t=3.15$, $df=41$, $p < 0.01$) and Streambank transects ($t=3.05$, $df=41$, $p < 0.01$) with evidence of grazing or trampling was lower on the BHP than in the EVA (Table 3).

Horse density

Site horse densities in the BHP and the EVA are shown in Table 3 for Rob-sites and TS-sites. The mean we calculated for 20 sites surveyed in 2018 by Cherubin et al. (2019) in the EVA was 43.20 horses per km^2. This horse density was significantly higher than the mean for our Rob-sites (7.18 horses per km^2) in 2021 ($t=-5.42$, $df=44$, $p < 0.01$) but not statistically different to our Extra-sites (16.52 horses per km^2) or our Exclusion-sites (58.53 horses per km^2). Approximately 72% of locations recorded with horse faeces had individual faecal deposits, and 28% had stallion piles (equivalent to 3.7 individual faecal deposits per stallion pile).

Horse density-impact function

The evidence of grazing and trampling along the Site-transects was very low (< 2% of transects) until density reached around 200–250 feral horse faecal piles per ha. Evidence of grazing and trampling then increased rapidly as feral horse faecal pile density increased (Fig. 4). No BHP site had more than 1% of the Site-transects with evidence of horse grazing/trampling. Only 24% of all sites and 17% of randomly selected sites across the BHP and the EVA had more than 2% of the Site-transects with evidence of feral horse grazing/trampling. Once feral horse faecal pile density was above 250 faecal piles per ha, the grazing and trampling sign became much more obvious. The sites with the highest impact were the Exclusion-sites first established by Theile and Prober in 1999, where we found faecal pile density to be exceptionally high (> 1000 faecal piles per ha).

The relationship between stream bank evidence of feral horse impact and faecal pile density had a greater spread of points, but below 250 faecal piles per ha the sign was still relatively low (Fig. 4b).

The relationships for both impacts, Site-impact and Streambank-impact with horse faecal pile density, fitted best to the Gompertz function (Table 4). Gompertz function is:

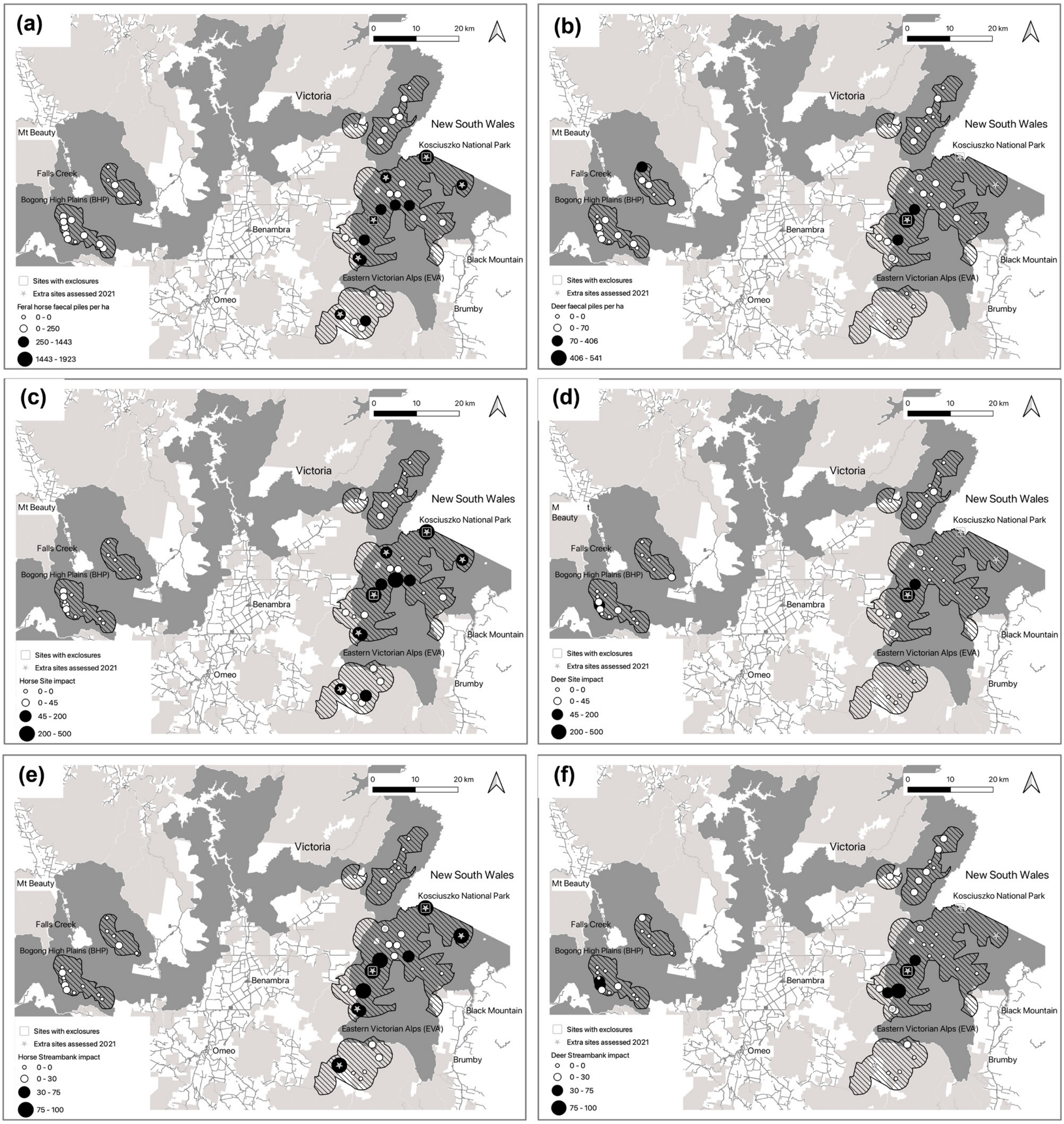


Figure 2. Maps of (a) feral horse and (b) deer faecal pile density, (c) feral horse and (d) deer grazing and trampling evidence along Site-transects, (e) feral horse and (f) deer grazing and trampling evidence along Streambank transects at sites visited by us in February–March 2021 in the Victorian part of the Australian Alps (Australian Alpine National Park – dark grey and surrounding State Forests – light grey). Shown is the area occupied by feral horses (hashed) within three kilometres of the entire set of sites established in 2012 (Robinson et al. 2019) on the Bogong High Plains (BHP) and in the Eastern Victorian Alps (EVA).

$$y = c + (d - c) \times \exp(-\exp[b \times (x - e)])$$

where $b = -0.003$, $c = -12.172$, $d = 100.944$, $e = 266.311$ for Site impact and $b = -0.003$, $c = -2.494$, $d = 527.079$, $e = 576.257$ for Streambank impact.

Deer faecal piles

BHP and the EVA did not differ significantly in deer faecal pile density (Table 2). Deer faecal piles were detected at 11 out of 16 (42%) sites on the BHP and 13 out of 26 (50%) sites in the EVA. We found the probability of detecting deer

Table 2. Horse and deer faecal piles counted along Site-transects, converted to density by Distance sampling. Animal density of horses and deer were calculated using faecal decay rates and defecation rates. Data from Extra-sites and Exclusion-sites were not included because they were selected based on the presence of very high feral horse activity. Includes data from 42 sites visited in February–March 2021 in the Victorian part of the Australian Alps (Australian Alpine National Park and surrounding State Forests) to measure the environmental impact associated with feral horses on the Bogong High Plains (BHP) and in the Eastern Victorian Alps (EVA).

	Region	Faecal piles in 500 m ± SE	Faecal pile density (groups/ha) ± SE	Animal density (animals/km^2) ± SE	Sites (n)
Horse	BHP	4.06 ± 1.19	17.11 ± 5.64	0.64 ± 0.31	16
	EVA	36.54 ± 9.58	191.22 ± 51.44	7.18 ± 1.93	26
Deer	BHP	1.69 ± 0.60	21.24 ± 7.60	3.02 ± 1.08	16
	EVA	3.58 ± 1.71	45.02 ± 21.54	6.39 ± 3.06	26

faecal piles was 0.13, and the effective strip width was 1.30 m according to Distance Line Transect analysis.

Deer density

Site deer densities are shown in Table 2 for Rob-sites and TS-sites on the BHP and in the EVA. Deer density estimates ranged from 0 to 13.96 deer per km^2 on the BHP and 0–60.01 deer per km^2 in the EVA.

Evidence of deer grazing/trampling

The mean lengths of transect sections with evidence of deer grazing or trampling for the BHP and the EVA were not statistically different (Table 3). On the BHP, evidence of deer grazing/trampling was detected at 7 out of 16 randomly selected sites (44%). Evidence of deer grazing/trampling in the EVA was detected at 14 out of 26 randomly selected sites (54%).

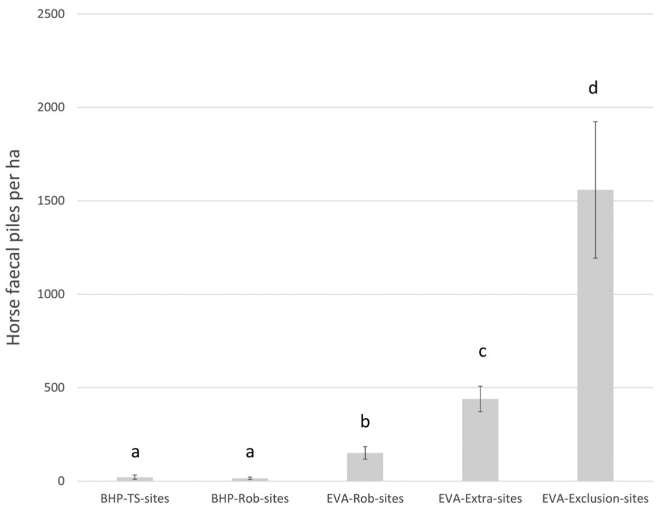


Figure 3. Comparison of feral horse faecal pile density at site types visited by us in February–March 2021 in the Victorian part of the Australian Alps (Australian Alpine National Park and surrounding State Forests) to measure the environmental impact associated with feral horses. Mean ± SE horse faecal pile density for the different site types on the Bogong High Plains (BHP) and in the Eastern Victorian Alps (EVA) are shown with significant differences determine by ANOVA and the Tukey post hoc test indicated by different letters a, b, c and d. TS-Sites were established and surveyed in 2017 by Tolsma and Shannon (2018). Rob-sites were established and surveyed in 2012 by Robertson et al. (2019) on the BHP and in the EVA. Extra sites were established by us in 2021in the EVA to increase the sample size of sites with high levels of horse activity to better determine the shape of the density-impact function at high feral horse density. Exclusion-sites were first established in 1999 to demonstrate the difference between small, fenced areas and areas heavily grazed by horses (Theile and Prober 1999). Our surveys were conducted outside the fenced plots in areas fully accessible to feral horses.

Animal paths

There was no significant relationship between the area of paths with horse tracks and the density of horse faecal piles at a site (Fig. 4c). Table 3 shows the mean area of the site that was path with horse sign and or deer sign. Excluding sites where no paths were detected, 0.19 and 0.21% of the area were impacted by paths on the BHP and in the EVA, respectively. Paths with horse hoof prints or horse faeces were detected at 6 out of 16 (38%) sites on the BHP and 15 out of 26 (58%) sites in the EVA. On the BHP, a total length of 570.98 m of path was recorded at the six sites with paths. The mean width of paths was 19.46 cm, covering a total area of 111.10 m^2 (0.07%) within the total of 160 000 m^2 for 16 sites bounded by the Site-transects (each site 200 × 50 m) on the BHP. In the EVA, a total length of 1992.52 m of path was recorded at the 15 sites with paths. The mean width of paths was 16.46 cm covering a total area of 338.62.10 m^2 (0.13%) within the total of 260 000 m^2 for 26 sites bounded by the Site-transects (each site 200 × 50 m) in the EVA.

Paths with deer hoof prints or faecal piles were detected at 10 out of 47 (21%) sites. There were 0.21 and 0.18% of the area impacted by paths with deer prints or faeces present on the BHP and in the EVA, respectively, at sites with paths detected.

For the total area of all paths where evidence of recent use was recorded, 78% of the area of path had sign of human use, 13% feral horse use, 4% deer use, 3% cattle use and 1% feral pig use. Paths leading to wombat burrows were observed at two sites in the EVA but no sign of wombats was detected on these paths.

Other impact agents

Figure 5 shows the proportion of impact associated with the sign of horses to be small compared to the combined impact of deer, feral pigs, rabbits, hare, cattle, fire and humans along both the Site-transects and Streambank-transects at sites on both the BHP and in the EVA. No feral pig diggings or faeces were seen on the BHP. Out of 31 sites in the EVA, feral pig diggings were detected at seven (23%) sites (3.01/500 m ± 1.54 SE) and pig faeces was recorded at five sites. No rabbit

Table 3. Horse and deer impact along Site-transects, Streambank transects and paths excluding Extra-sites and Exclusion-sites because they were selected based on the presence of very high feral horse activity. Includes data from 42 sites visited in February–March 2021 in the Victorian part of the Australian Alps (Australian Alpine National Park and surrounding State Forests) to measure the environmental impact associated with feral horses on the Bogong High Plains (BHP) and in the Eastern Victorian Alps (EVA).

	Region	Site impact m in 500m ± SE (% of transect)	Stream bank impact m in 100m ± SE (% of transect)	Path impact m^2 in 10 000 m^2 (% of area)	Sites (n)
Horse	BHP	0.75 ± 0.35 (0.15%)	0.69 ± 0.44 (0.69%)	9.10 ± 5.32 (0.09%)	16
	EVA	49.46 ± 20.93 (9.89%)	17.11 ± 6.05 (17.12%)	15.03 ± 3.53 (0.15%)	26
Deer	BHP	5.94 ± 3.30 (1.19%)	6.75 ± 3.52 (6.75%)	2.68± 2.02 (0.03%)	16
	EVA	16.08 ± 11.78 (3.22%)	13.08 ± 5.26 (13.08%)	4.51 ± 2.24 (0.05%)	26

faecal pellets were detected on the BHP. Rabbit faecal pellets were recorded at seven sites (17.94 pellets/500 m ± 12.72 SE) in the EVA. Evidence of rabbit grazing or soil disturbance was detected along a mean of 0.42/500 m ± 0.24 SE at Rob/TS-sites and 9.20/500 m ± 9.20 SE at Exclusion/Extra-sites in the EVA. No evidence of rabbit grazing or soil disturbance was detected in the BHP or along Streambank transects in the EVA. Off-transect, we observed rabbit warrens and the bare ground around warrens presumably created by rabbits. Hare pellets were recorded at six sites on the BHP but at no sites in the EVA. Evidence of hare grazing or soil disturbance was detected along a mean of 2.19/500 m ± 2.19 SE at random sites on the BHP. No evidence of hare grazing or soil disturbance was detected on the EVA or along Streambank transects on the BHP. Cattle sign was detected at four sites out of 31 (13%) sites. All sites with cattle sign were in the State Forest in the southern section of the EVA. Evidence of cattle grazing or trampling was detected along a mean of 6.96/500 m ± 4.67 SE at Rob-sites and 58.33/500 m ± 58.33 SE at Exclusion/Extra-sites in the EVA. No cattle sign was detected on the BHP or in the National Park in the EVA.

Fire impact was present at one of the sixteen (6%) sites surveyed on the BHP and at six of the 31 (19%) sites in the EVA. Two sites out of 16 (13%) had the impact of humans, including vehicle and bush walking tracks on the BHP. Four sites out of 26 (15%) had human impact in the EVA. Evidence of human disturbance was detected along a mean of 0.44/500 m ± 0.33 SE of Site-transect on the BHP and 8.13/500 m ± 6.39 SE in the EVA and a mean of 6.37/100 m ± 6.34 SE of Streambank-transect on the BHP and 0.24/100 m ± 0.24 SE in the EVA Rob-sites and TS-sites. Native macropods probably contributed to grazing and browsing, but we detected no sign along transects associated with these species. We observed wombat burrows off-transect at one site out of 16 on the BHP and six sites out of 31 in the EVA. Evidence of wombat grazing or trampling, or soil disturbance was detected along a mean of 0.29/500 m ± 0.19 SE of Site-transect but not along Streambank-transects in the EVA. There was no evidence of wombat disturbance along transects or off-transects in the BHP.

Discussion

Understanding the role and extent of feral horse impacts is important for ensuring that the best and most agreeable management outcomes are reached. Understanding density impact functions can assist this, but had not been previously attempted for horses in the Australian Alps. We detected and measured the magnitude of difference in direct feral horse impact between two separate parts (the BHP and EVA) of the area occupied by feral horses in the Victorian part of the Australian Alps. As expected, the lowest impact associated with feral horses was on the BHP, where the lowest feral horse density had been determined by aerial survey (Worboys et al. 2018, Cairns 2019) and by our feral horse density derived from faecal pile density (Fig. 2, Table 2). But contrary to assertions made in earlier studies (Tolsma et al. 2018), almost all (>99%) of the area we surveyed by walked transect on the BHP had no detectable evidence of grazing or trampling associated with the presence of horses (Fig. 5, Table 3). Even in the EVA, where feral horse faecal pile density and impact were significantly higher than the BHP, the vast majority of the area surveyed (> 82%) had no evidence of grazing or trampling associated with the presence of feral horses (Fig. 5, Table 3). Previous studies concluded that there was high feral horse impact on the BHP based on the proportion of sites they surveyed with impact detected (i.e. presence of any sign of horse faeces, hoof prints, stream bank, soil or vegetation disturbance), although estimates of the proportion of the area at each site with horse impact were not reported (Tolsma et al. 2018). Instead, they emphasised the high proportion of sites (57%) surveyed with at least some sign of horses having been detected there, meaning that a site with one horse faecal pile or one deer pellet was considered to have an equal impact to sites with many horse faecal piles or deer pellets. Results like those of Tolsma et al. (2018) therefore confirmed horses had been present, but they provided little information about the degree of impact at sites or the extent of impact throughout the BHP, particularly considering that sites were selected because they were ‘known, or suspected, to be utilised by feral horses’ (Tolsma et al. 2018). Like the previous researchers whose sites we resurveyed, we conducted our work within the known distribution of feral horses. Consequently, we also found the proportion of sites with at least some horse impact to be high (75% on the BHP and 92% in the EVA). But a ‘high proportion of sites with horses present’ says nothing about the level of horse impact at those sites, and is also not particularly noteworthy given that sites were initially selected because of horse presence. Thus, although the proportion of sites with sign of horses was understandably high given sampling efforts focused on areas with known horse presence, the proportion of those sampled areas with actual horse impact was extremely low (< 1%) on the BHP and low (< 18%) in the EVA (Table 3, Fig. 5).

Interestingly, the proportion of the area travelled by horses on paths did not differ between the BHP and the EVA

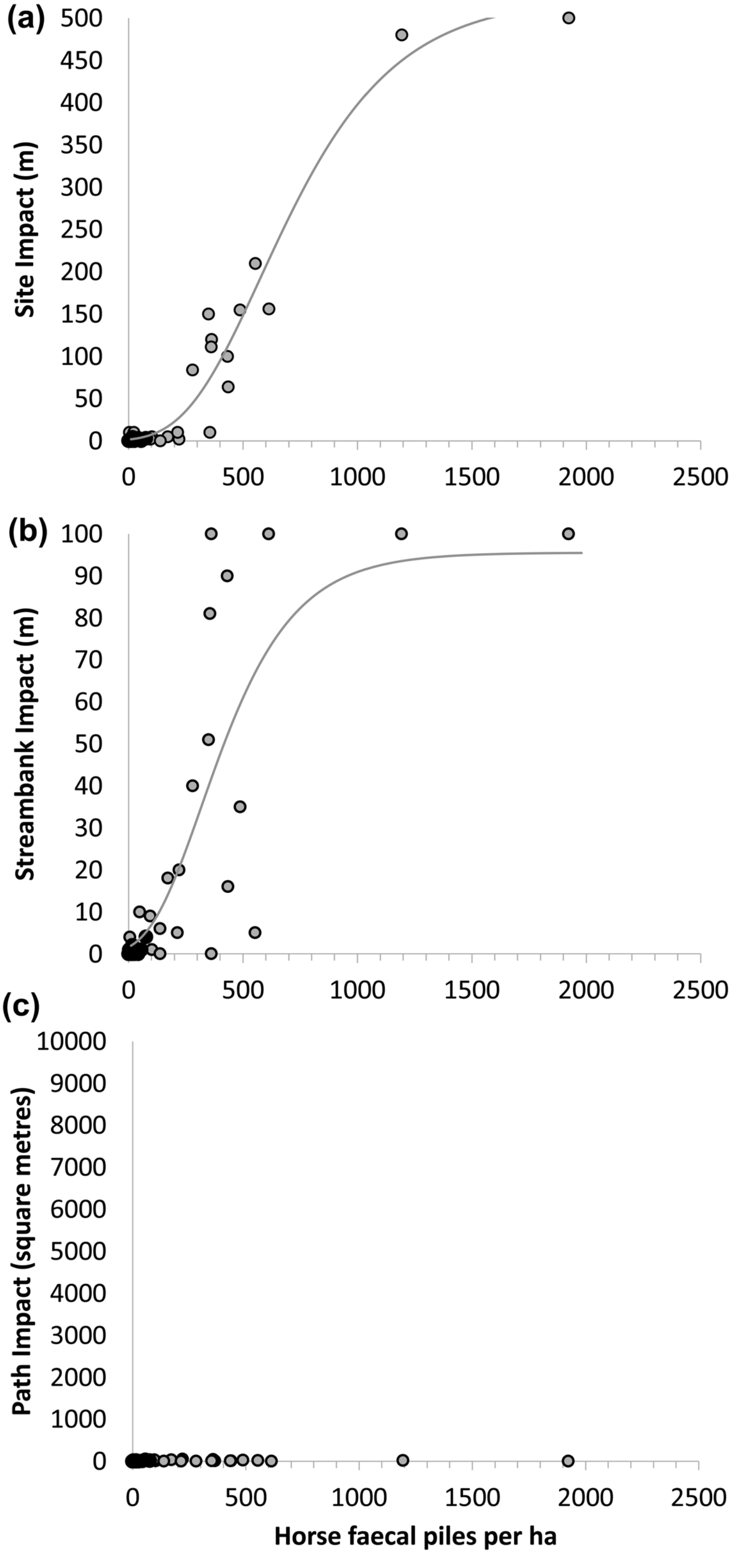


Figure 4. Density/impact relationships for all sites visited by us in February–March 2021 in the Victorian part of the Australian Alps (Australian Alpine National Park and surrounding State Forests) to measure the environmental impact associated with feral horses. (a) Relationship between grazing and trampling impact along Site-transects on the Bogong High Plains (BHP, black dots) and in the Eastern Victorian Alps (EVA, grey dots). (b) Relationship between grazing and trampling impact along Stream bank-transects on the BHP (black dots) and in the EVA (grey dots). (c) Area of path impact within the Site-transect area on the BHP (black dots) and in the EVA (grey dots).

Table 4. Ranking according to AICc (Akaike's information criterion for small samples) of density-impact functions for horse faecal pile density and impact recorded along Site-transects and Streambank transects. Includes data from all 47 sites visited in February–March 2021 in the Victorian part of the Australian Alps (Australian Alpine National Park and surrounding State Forests) to measure the environmental impact associated with feral horses on the Bogong High Plains (BHP) and in the Eastern Victorian Alps (EVA).

	AICc	
Functions	Site impact	Stream bank impact
Gompertz	438.9	416.7
Logistic	442.8	418.4
Linear	468.6	424.2

despite different horse densities (Table 1, 2). These paths are the most visually obvious impact of horses and were a major focus of previous work (Dyring 1990). Paths are created and/or maintained by horses and other animals repeatedly travelling over the same routes, compacting the soil and damaging or removing the vegetation under their feet. Paths running along drainage lines can disturb stream banks and stream flow (Kauffman and Krueger 1984, Hope et al. 2012). While we observed examples of what appears to be severe impact where paths crossed streams (gully erosion and churned-up mud), these were very isolated in the broader landscape given that the actual area of impact on soil and vegetation found in our study was only around 0.2% of the area surveyed at sites where paths were even detected at all. Dyring (1990) also recorded 0.2% of the area of one of her sites subjected to path impact. This small area of impact does not appear to vary with horse density (Fig. 4c), so removing or managing horse populations is unlikely to change the proportion of the area affected by this type of impact. Multiple animal species, including native animals like wombats, or introduced animals like deer, will use and maintain the paths even with reduced feral horse use, so some, if not all, these paths are likely to remain even if horses are removed.

Density-impact functions for Site and Streambank grazing and trampling demonstrate the rarity of sites with very high feral horse density and impact (Fig. 4). The lower part of the curve for feral horse faecal pile density and site impact is relatively flat; then, as faecal pile density increases, the curve rises steeply at around 250 horse faecal piles per hectare before flattening again and approaching an asymptote above 1000–1500 faecal piles per hectare as the impact is detected along 100% of the transect. We found that below the threshold of ~ 250 horse faecal piles per ha, or 9 horses per km^2, increases in faecal pile density resulted in very small increases in evidence of feral horse grazing/trampling (Fig. 4). This is particularly so for Site impact (Fig. 4a), with a relatively rapid increase at low densities for Streambank impact (Fig. 4b). Below this threshold, greater than 97% of the length of site transect and 90% of the stream bank transect had no evidence of feral horse grazing or trampling detected. However, a small increase in faecal pile density above this threshold resulted in a disproportionately large increase in evidence of feral horse grazing/trampling detected. At the sites above the threshold (17% of randomly selected sites), a small population control effort targeting 17% of the area (instead of the entire area occupied by horses) and costing relatively little, may significantly reduce evidence of feral horse grazing/trampling. Below the threshold where horse density is low, and the cost of removal of horses is most likely high (Choquenot et al. 1999), a considerable expense will be incurred in attempts to reduce an already low level of direct impact (Hone 1994, Choquenot and Parkes 2001, Hone 2007, Norbury et al. 2015). Hence, the most efficient way to reach an acceptable level of direct feral horse impact is to target areas with the highest feral horse density for population control, identified here as areas with densities along alpine treeless drainage lines above 250 faecal piles per ha or 9 feral horses per km^2. However, we caution that the threshold determined here may only apply to the current study area and also may not even apply during drier or wetter periods or if the abundance of other species changes. Monitoring needs to be designed to determine site-specific thresholds and to allow re-evaluation of any threshold as conditions change.

We also found the combined impact associated with the sign of deer, feral pigs, humans, and fire to be large compared to the impact associated with the sign of feral horses alone, particularly on the BHP (Fig. 5). Horse sign was associated with no more than 4% of the impact on the BHP and less than 34% of the impact in the EVA (Fig. 5). Comparing impact potential as we did, where the impact is recorded on a transect line, should minimise differences in impact detectability between the various potential agents of impact. Searching rectangular strip transects or sample areas so that impact or sign is recorded at various distances from the observer may also increase the chance of differences in impact detectability between the various potential agents of impact. Detection probability can be determined using Distance Sampling (Buckland et al. 1993), and we found horse faecal piles to be more easily detected than deer faecal pellets. Horse hoof prints are also larger and, therefore, most likely more easily observable at a distance than deer or pig prints. Without searching carefully along a line, if any bias exists, then the proportion of horse impact recorded will be exaggerated compared to those of deer and pigs. Previous authors either excluded the impact of agents other than horses statistically (Robertson et al. 2019) or assumed that the contribution of other agents was insignificant because the sign of horses was much more obvious (Tolsma et al. 2018, Robertson et al. 2019), without considering differences in detectability. Yet these and other similar published studies (Cherubin et al. 2019, Eldridge et al. 2019, Schulz et al. 2019) have lead others to conclude that there is 'unequivocal evidence that feral horses are the single largest cause of widespread environmental degradation throughout their range in the alpine parks' (Driscoll et al. 2019b). This view is clearly inconsistent with data that does account for sampling bias and detectability (Table 2–4, Fig. 2–5) and with the findings of a recent study in Kosciuszko National Park, adjacent to our study site (Hartley et al. 2022). These results show clearly that in spite of deer and feral pig control and management of fire and human activity (Parks Victoria 2016, GSBMPWG

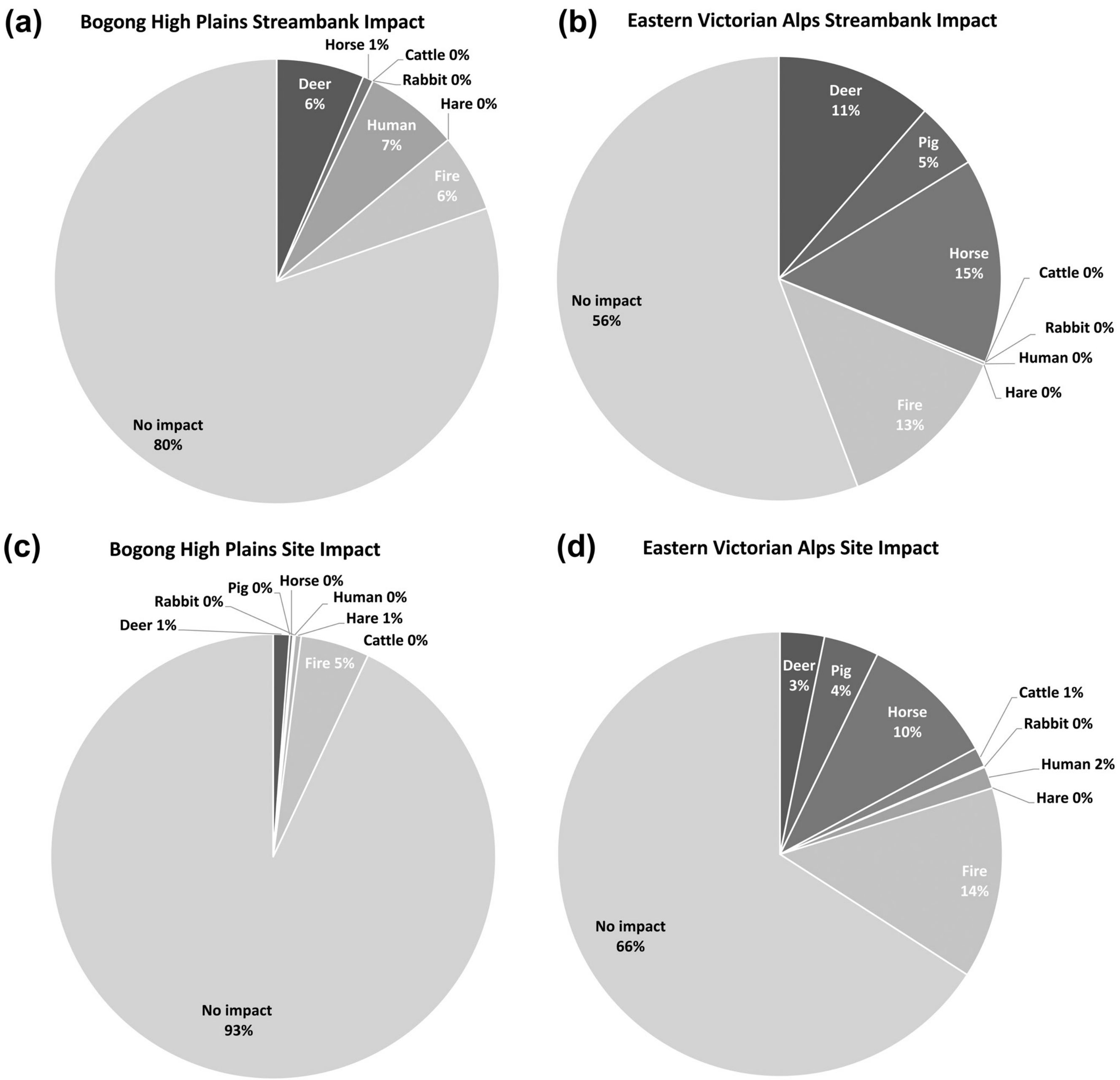


Figure 5. Percentage of impact associated with sign of horses, deer, pig, rabbit, hare, cattle, fire and humans at sites visited by us in February–March 2021 in the Victorian part of the Australian Alps (Australian Alpine National Park and surrounding State Forests) to measure the environmental impact associated with feral horses. Impact and sign of potential causes were recorded along the 500 m Site-transects and 100 m Stream bank-transects on the Bogong High Plains (BHP) and in the Eastern Victorian Alps (EVA). Data from Extra-sites and Exclusion-sites were excluded because these sites were selected because of their known/observed high horse activity. (A) Site impact on the BHP, (B) Site impact in the EVA, (C) Stream bank impact on the BHP and (D) Streambank impact in the EVA .

2020, Comte et al. 2022), these combined impacts far exceed those impacts associated with the sign of feral horses. This does not mean the management of feral horses is unnecessary, but it does mean that 1) feral horses may not be the most important environmental threat and 2) the management of feral horses alone may result in very little reduction in overall environmental impact if these other important factors are not also managed (Braysher 2017).

A study of the impacts of feral horses on threatened native species found horse impacts and faecal piles at 19 out of 20 sites (95%), whereas only seven sites (35%) had evidence of deer (Cherubin et al. 2019). Unlike previous researchers in the Australian Alps, Cherubin et al. (2019) quantified deer faecal piles in the same way as horse faecal piles. Cherubin et al. (2019) also determined decay/disappearance rate for horse faecal piles. But unfortunately, they did not do this for deer faecal pellets and neither was the difference between the detection probability of horse faecal piles and deer faecal piles considered. Hence, they could not compare the level of activity or impact of horses and deer, cautioning that 'we cannot entirely separate the impacts of horses from other species' (Cherubin et al. 2019, Supplementary information). This statement is also true for our study. Another study in Kosciuszko National Park compared deer faecal pile and horse faecal pile densities to determine relative impacts, finding more horse faecal piles than deer faecal piles, whereas camera traps suggested the activity of deer was far greater than that of horses (Ward-Jones et al. 2019). These authors preferred the results of the faecal pile comparison over the camera trap findings because the camera traps were in place for a relatively short period. Short survey periods are a problem for reliably interpreting camera trap data (Meek et al. 2012) but faecal pile indices are also notoriously unreliable when they have not been calibrated to densities with proper measurement of defection rates and decay rates (Allen 2012, Le Pla et al. 2022). Inadequate quantification of the impact or activity of other species, coupled with failure to consider differences in faecal decay/disappearance rate or detection probability, mean that previous assertions about the negative environmental impacts of horses in the Australian Alps are highly likely to be overestimated compared to the impact of other animals, such as deer.

We also found it difficult to disentangle the impact of wild or feral horses from the impacts of domestic horses frequently brought into the study area for recreational purposes, which was likely to have also influenced earlier studies but was not mentioned (Robertson et al. 2019). Likewise, we further found it extremely difficult to disentangle the stream bank impact attributable to deer, horses, or feral pigs given that they all drink, wallow, or cross streams in the same places. In addition to differences in detectability, deer faecal piles can decay more rapidly than horse faecal piles (Linklater et al. 2001, Zabek 2015a, Davis and Coulson 2016). Horse density also varied considerably depending on how and where sites were selected (Fig. 3). Results from the exclusion plot studies that commenced in 1999 (Wild and Poll 2012) are extremely valuable in demonstrating the difference between very high and zero feral horse density. However, such results cannot be extrapolated to all areas of the Australian Alps because those sites only represented extreme horse densities, and not the otherwise normal densities found across almost all other areas of the Victorian Alps (Fig. 2, 3).

The preceding information leads to the observation that 1) unbiased site selection, 2) consideration of detection probability and 3) decay or disappearance rate for horse and deer faecal piles and impact, are each important for reliably comparing density and impacts between species, places, and times. Clearly defining the variable extent and level of impact of feral horses separate to other factors is essential for resolving stakeholder disputes, prioritising management actions, and measuring the success of management. This is not easy and is rarely attempted for horses or deer (Davis et al. 2016), but it is possible.

In our study, to improve comparisons of the relative activity of feral horses and deer, we converted faecal density to animal density using faecal decay/disappearance rate and defecation rate. By doing this, we found that the mean density of horses and deer (number of animals present in a defined area in a 24 h period) along treeless drainage lines in the EVA were very similar (Table 2). This is consistent with aerial survey results indicating feral horse and deer density were similar in 2019 in the EVA (Cairns 2019). The mean horse density for 20 sites surveyed in 2018 by Cherubin et al. (2019) was not significantly different to our Extra-sites or the Exclusion-sites, which were significantly higher than the mean for the randomly selected Rob-sites in our study (Fig. 3). Our Extra-sites and Exclusion-sites were deliberately chosen because they were expected to have high horse activity. Findings from studies where sites are selected in this way (Dyring 1990, Theile and Prober 1999) and subsequent conclusions derived from them (Driscoll et al. 2019b) are not representative of the entire Australian Alps but are relevant only to specific sites with exceptionally high horse activity. In our study, the highest horse density estimate was 60 horses per km^2 at Cowombat Flat (Exclusion-site), and the second highest was 41 horses per km^2 at Native Cat Flat (Exclusion-site). These densities were comparable with densities (46 horses per km^2) determined in the late 1980s at a site studied by Dyring (1990) in Kosciusko National Park, and are exceptionally high compared to the mean determined by us for randomly selected sites in the EVA (Table 2; 8 horses per km^2). In our study at Native Cat Flat where the highest deer faecal pile density was recorded, there were 60 deer per km^2 compared to 41 horses per km^2 based on our faecal pile counts. This is one of the sites where exclusion-fenced plots built in 1999 are commonly used to demonstrate the impact of feral horses (Wild and Poll 2012, Williams et al. 2014) with no consideration of the high relative impact of deer activity at the same site. Our actual horse and deer density estimates from faecal pile density relied on defecation and decay/disappearance rates determined elsewhere. Decay rates can change with the amount of rainfall (Zabek 2015b), so ideally, decay/disappearance rates need to be derived for this location in the period prior to surveys.

With the ultimate goal of threatened fauna conservation in mind, further research is also required to confirm the nature of the relationship between feral horse density or direct impact and threatened fauna behaviour, abundance, or trends. More studies similar to Cherubin et al. (2019) and Schulz et al. (2019) are required over a larger range of feral horse densities, for longer periods and with active manipulation of horse density. The work done on sage-grouse *Centrocercus urophasianus* in the USA which supports the objective of maintaining feral horse density below an 'appropriate management level' (Coates et al. 2021), is a good example of the type of work that is required in the Australian Alps. Nevertheless, this and other previous studies are basically correlative. Finding that the abundance of native species is negatively associated with feral horse activity or density (Schulz et al. 2019, Coates et al. 2021) might well mean that feral horse impact has influenced the native species. Still, such a correlation could also be a result of differences in habitat suitability or preference, independent of feral horse activity or density. Incorporating detailed studies of threatened fauna in planned adaptive management experiments with randomly selected areas for treatment (changing the density of feral horses) would provide the inferential strength of evidence required. More detailed studies of native species at our sites, coupled with an increased number of sites representing other habitats and random manipulation of feral horse density, may allow improved measurement of both direct and indirect horse impact. Identification of ecologically significant, positive, neutral, or negative impacts would then be possible, and with this, the ideal targets for managing feral horse density may be determined.

For the first time in the Australian Alps, our work shows how evidence of the direct impact of feral horses on soil and vegetation varies with feral horse density over the range from very low to very high horse density. If the level of direct impact relates to the level of threat to vulnerable native species, soil, vegetation or water, then our work will guide managers attempting to minimise this threat. Previous studies have reported correlations between horse abundance or impact and vegetation structure (Cherubin et al. 2019), abundance or occurrence of endangered species (Schulz et al. 2019), and abundance of ant nests and soil penetrability (Beever and Herrick 2006), but compared to our study, these earlier reports were based on a limited range of feral horse activity levels and/or were categorised into a small number of classes of horse impact. In our study, grazing and trampling impact correlated with the density of horses, which means feral horse impact should be manageable by changing horse density. However, an exception to this is the impact along paths. This impact appears to reach a maximum at very low horse density and did not increase thereafter with increasing horse density. We also measured the relative extent of evidence of grazing, trampling and vegetation disturbance associated with the sign of deer, feral pigs, fire and humans showing that removal of horse impacts alone could prevent only a small proportion of potentially damaging impact. We are hopeful that we have provided valuable suggestions that will improve the measurement of the environmental impact of feral horses in a way that guides effective and acceptable management.

Management implications

Reduction in direct horse impact can be most efficiently achieved by targeting those few sites where horse impact is highest. Where horse density is low, and the cost of removal of horses is high, a considerable expense will be incurred to reduce an already low level of direct impact. Incorporating experimental monitoring into management will ensure that any unexpected, undesirable consequences of removing horses are detected and the expected benefits of management are measured. Instead of attempting to remove as many horses as possible, a better approach to reducing horse impacts would be to establish an evidence-based target density and then reduce and maintain horse populations at or below that level.

Acknowledgements – We would like to thank V. A. Imhoff for generously providing local knowledge of the Victorian Alps, horses and people and commenting on a draft manuscript. S. G. Radke, J. M. Curatolo, L. Cameron are all thanked for providing views and valuable discussion on various aspects of this work. Parks Victoria staff are thanked for supplying keys for gates, permits and caring for our health and safety. Without the administrative support of the USQ staff, particularly S. Verrall, no field work would have been possible.
Funding – The University of Southern Queensland and the Australian Brumby Alliance jointly funded this research.
Conflict of interest – Jill Pickering is President of the Australian Brumby Alliance, an organisation working to sustainably manage feral horse populations to preserve heritage and environmental benefits. David Berman is Director of a private company that is contracted to manage and sometimes eliminate feral horses from National Parks or State Forests.

Author contributions

David McKenzie Berman: Conceptualization; Methodology; Formal anaylsis; Investigation; Data curation; Funding acquisition (equal); Project administration (equal); Writing original draft; Writing – review and editing; Visualization. **Jill Pickering**: Conceptualization (supporting); Funding acquisition (equal); Project administration (supporting); Resources (supporting); Writing – review and editing (supporting). **Deane Smith**: Conceptualization (supporting); Data curation (supporting); Investigation (supporting); Writing – review and editing (supporting). **Benjamin L. Allen**: Conceptualization (supporting); Writing – review and editing (supporting); Visualization (supporting); Validation.

Transparent peer review

The peer review history for this article is available at https://publons.com/publon/10.1002/wlb3.01107.

Data availability statement

Data are available from the Dryad Digital Repository:https://doi.org/doi:10.5061/dryad.sn02v6x8p (Berman et al. 2023).

References

Akaike, H. 1998. Information theory and an extension of the maximum likelihood principle. – In: Parzen, E., Tanabe, K. and Kitagawa, G. (eds), Selected papers of Hirotugu Akaike. Springer, pp. 199–213.

Allen, B. L. 2012. Scat happens: spatiotemporal fluctuation in dingo scat collection rates. – Aust. J. Zool. 60: 137–140.

Beever, E. A. and Herrick, J. E. 2006. Effects of feral horses in great basin landscapes on soils and ants: direct and indirect mechanisms. – J. Arid Environ. 66: 96–112.

Beever, E. A., Simberloff, D., Crowley, S. L., Al-Chokhachy, R., Jackson, H. A. and Petersen, S. L. 2019. Social–ecological mismatches create conservation challenges in introduced species management. – Front. Ecol. Environ.17: 117–125.

Berger, J. 1986. Wild horses of the great basin: social competition and population size. – Univ. of Chicago Press.

Berman, D. 1991. The ecology of the feral horses in central Australia. – PhD thesis, Univ. of New England, Australia.

Berman, D. 2012. Australian wild horse science and management. In: Annual meeting of the society for range management. – Society for Range Management.

Berman, D. 2013. Feral horse management amongst unexploded bombs and peri-urban people. – In: 'Australasian vertebrate pest conference': Brisbane.

Berman, D. M. 2021. What is required to achieve effective wild horse management?, 25–27 May 2021, Virtual. – Centre for Invasive Species Solutions.

Berman, D. and Jarman, P. 1987. Feral horses in the northern territory, vol. I: the ecology of feral horses in central australia and their interaction with cattle. – Conservation Commission of the Northern Territory.

Berman, D. and Jarman, P. 1988. Feral horses in the northern territory, vol. IV: environmental impact of feral horses in central australia. – Conservation Commission of the Northern Territory.

Berman, D. M., Pickering, J., Smith, D. and Allen, B. L. 2023. Data from: Use of density-impact functions to inform and improve the environmental outcomes of feral horse management. – Dryad Digital Repository, https://doi.org/doi:10.5061/dryad.sn02v6x8p.

Bomford, M. and O'Brien, P. 1995. Eradication or control for vertebrate pests? – Wildl. Society Bull. 1973–2006 23: 249–255.

Boyce, P. N., Hennig, J. D., Brook, R. K. and McLoughlin, P. D. 2021. Causes and consequences of lags in basic and applied research into feral wildlife ecology: the case for feral horses. – Basic Appl. Ecol. 53: 154–163.

Braysher, M. 1993. Managing vertebrate pests: principles and strategies. – Australian Government Publishing Service.

Braysher, M. 2017. Managing australia's pest animals: a guide to strategic planning and effective management. – CSIRO Publishing.

Buckland, S. T., Anderson, D. R., Burnham, K. P. and Laake, J. L. 1993. Distance sampling: estimating abundance of biological populations. – Chapman and Hall.

Butler G and Associates. 1996. Victorian alpine huts heritage survey. – Graeme Butler & associates.

Cabena, P. B. 1980. Grazing the high country: an historical and political geography of high country grazing in victoria, 1835 to 1935. MSc thesis, Department of Geography, The University of Melbourne. Australia.

Cairns, S. 2019. Feral horses in the Australian alps: the analysis of aerial surveys conducted in April–May 2014 and April–May 2019. – Australian Alps Liaison Committee.

Calaby, J. and Wimbush, D. 1964. Observations on the broad-toothed rat, *Mastacomys fuscus* thomas. – CSIRO Wildl. Res. 9: 123–133.

Carr, S. and Turner, J. 1959a. The ecology of the Bogong high plains. I. The environmental factors and the grassland communities. – Aust. J. Bot. 7: 12–33.

Carr, S. and Turner, J. 1959b. The ecology of the Bogong high plains. Ii. Fencing experiments in grassland c. – Aust. J. Bot. 7: 34–63.

Cherubin, R. C., Venn, S. E., Driscoll, D. A., Doherty, T. S. and Ritchie, E. G. 2019. Feral horse impacts on threatened plants and animals in sub-alpine and montane environments in Victoria, Australia. – Ecol. Manag. Restor. 20: 47–56.

Choquenot, D. and Parkes, J. 2001. Setting thresholds for pest control: how does pest density affect resource viability? – Biol. Conserv. 99: 29–46.

Choquenot, D., Lukins, B. and Curran, G. 1997. Assessing lamb predation by feral pigs in australia's semi-arid rangelands. – J. Appl. Ecol. 34: 1445–1454.

Choquenot, D., Hone, J. and Saunders, G. 1999. Using aspects of predator–prey theory to evaluate helicopter shooting for feral pig control. – Wildl. Res. 26: 251–261.

Coates, P. S., O'neil, S. T., Muñoz, D. A., Dwight, I. A. and Tull, J. C. 2021. Sage-grouse population dynamics are adversely affected by overabundant feral horses. – J. Wildl. Manag. 85: 1132–1149.

Comte, S., Thomas, E., Bengsen, A. J., Bennett, A., Davis, N. E., Brown, D., and Forsyth, D. M. 2022. Cost-effectiveness of volunteer and contract ground-based shooting of sambar deer in australia. – Wildl. Res. Online Early Special Edition.

Costin, A. B. 1954. A study of the ecosystems of the Monaro region of New South Wales, with special reference to soil erosion. – Soil Conservation Service of New South Wales.

Costin, A. B., Gray, M., Totterdell, C. and Wimbush, D. 2000. Kosciuszko alpine flora. – CSIRO Publishing.

Crabb, P. 2003. Managing the Australian alps: a history of cooperative management of the Australian alps national parks. – A joint publication of the Australian Alps Liaison Committee and the Centre for Resource and Environmental Studies Australian National University.

Curran, I. 2018. Aerial feral horse survey Bogong high plains and surrounding valleys. – Parks Victoria.

Davis, N. E. and Coulson, G. 2016. Habitat-specific and season-specific faecal pellet decay rates for five mammalian herbivores in south-eastern Australia. – Aust. Mammal. 38: 105–116.

Davis, N. E., Bennett, A., Forsyth, D. M., Bowman, D. M. J. S., Lefroy, E. C., Wood, S. W., Woolnough, A. P., West, P., Hampton, J. O., and Johnson, C. N. 2016. A systematic review of the impacts and management of introduced deer family cervidae) in australia. – Wildl. Res. 43: 515–532.

Dawson, M. J. 2005. The population ecology of feral horses in the Australian Alps management summary. Prepared for the Australian Alps Liaison Committee.

Dawson, M. J. 2009. Scoping study: assessing the impact of feral horses on frost hollow communities (bogs/fens, streams, wet heath and wet grasslands) in the Australian alps national parks. – The Australian Alps National Parks Natural Heritage Working Group.

Dawson, M. J. and Miller, C. 2008. Aerial mark–recapture estimates of wild horses using natural markings. – Wildl. Res. 35: 365–370.

Dawson, M. J. and Hone, J. 2012. Demography and dynamics of three wild horse populations in the Australian alps. – Aust. Ecol. 37: 97–109.

Dawson, M. J., Lane, C., and Saunders, G. National Feral Horse Management Workshop. 2006, Canberra. – Invasive Animals Cooperative Research Centre.

Dobbie, W., Berman, D. and Braysher, M. 1993. Managing vertebrate pests: feral horses. – Australian Government Publishing Service.

Driscoll, D., Scheele, B. and McDonald, T. 2019a. Feral horses in the Australian alps: an introduction to the special issue. – Ecol. Manag. Restor. 20: 3–3.

Driscoll, D. A. et al. 2019b. Impacts of feral horses in the Australian alps and evidence-based solutions. – Ecol. Manag. Restor. 20: 63–72.

Duncan, P. 1992. Horses and grasses: the nutritional ecology of equids and their impact on the camargue. – Springer-Verlag.

Dyring, J. 1990. The impact of feral horses (*Equus caballus*) on sub-alpine and montane environments in Australia. – MSc thesis, University of Canberra, Australia.

Eldridge, D. J., Travers, S. K., Val, J., Zaja, A. and Veblen, K. E. 2019. Horse activity is associated with degraded subalpine grassland structure and reduced habitat for a threatened rodent. – Rangel. Ecol. Manag. 72: 467–473.

Eldridge, D. J., Ding, J. and Travers, S. K. 2020. Feral horse activity reduces environmental quality in ecosystems globally. – Biol. Conserv. 241: 108367.

Department of the Environment. 2015. National recovery plan for the alpine sphagnum bogs and associated fens ecological community. – Department of the Environment, Canberra, Australia.

Fleurance, G., Duncan, P., Farruggia, A., Dumont, B., and Lecomte, T. 2011. Impact of equine pasture land on the diversity of fauna and flora in grazing lands. – Fourrages 207: 10.

Fleurance, G., Edouard, N., Collas, C., Duncan, P., Farruggia, A., Baumont, R., Lecomte, T., and Dumont, B. 2012. How do horses graze pastures and affect the diversity of grassland ecosystems? – In: '6th European workshop on equine nutrition': Lisbon, Portugal.

Forsyth, D. 2005. Protocol for estimating changes in the relative abundance of deer in new zealand forests using the faecal pellet index (FPI). – Landcare Research Contract Report.

Foster, C. N. and Scheele, B. C. 2019. Feral-horse impacts on corroboree frog habitat in the Australian alps. – Wildl. Res. 46: 184–190.

Freeland, W. J. 1990. Large herbivorous mammals: exotic species in northern Australia. – J. Biogeogr. 17: 445–449.

Green, K. and Osborne, W. S. 2003. The distribution and status of the broad-toothed rat *Mastacomys fuscus* (Rodentia: Muridae) in new south wales and the Australian capital territory. – Aust. Zool. 32: 229–237.

GSBMPWG. 2020. Gippsland bushfire management strategy 2020. – Gippsland Strategic Bushfire Management Planning Working Group, The State of Victoria Department of Environment, Land, Water and Planning.

Hartley, R., Blanchard, W., Schroder, M., Lindenmayer, D. B., Sato, C. and Scheele, B. C. 2022. Exotic herbivores dominate Australian high-elevation grasslands. – Conserv. Sci. Prac. 4: e601.

Hone, J. 1994. Analysis of vertebrate pest control. – Cambridge Univ. Press.

Hone, J. 2007. Wildlife damage control. – CSIRO Publishing.

Hope, G. S., Nanson, R. and Jones, P. 2012. Peat-forming bogs and fens of the snowy mountains of new south wales. – Office Environment and Heritage.

ITRG. 2016. Final report of the independent technical reference group: supplementary to the Kosciuszko national park wild horse management plan. – Office of Environment and Heritage NSW.

Kauffman, J. B. and Krueger, W. C. 1984. Livestock impacts on riparian ecosystems and streamside management implications. A review. – Rangel. Ecol. Manag./J. Range Manag. Arch. 37: 430–438.

Lawrence, R. 1995. The effects of grazing activity on the hydrology of the Bogong high plains, Australia. – Rangel. J. 17: 138–153.

Le Pla, M. N., Birnbaum, E. K., Rees, M. W., Hradsky, B. A., Weeks, A. R., Van Rooyen, A., and Pascoe, J. H. 2022. Genetic sampling and an activity index indicate contrasting outcomes of lethal control for an invasive predator. – Aust. Ecol. 47: 1062–1076.

Linklater, W. L., Cameron, E. Z., Stafford, K. J., and Minot, E. O. 2001. Estimating kaimanawa feral horse population size and growth. – Department of Conservation. Do Conservation.

Matthews, D., Bryan, R., and Edwards, G. 2001. Recovery of the black-footed rock-wallaby following horse removal on finke gorge national park, northern territory. – In: 12th Australasian Vertebrate Pest Conference. Department of Natural Resources and Environment, p. 204.

Mayle, B. A., Peace, A. J. and Gill, R. M. A. 1999. How many deer? A field guide to estimating deer population size. – Forestry Commission.

McDougall, K. L. 2007. Grazing and fire in two subalpine peatlands. – Aust. J. Bot. 55: 42–47.

Meek, P., Ballard, G. and Fleming, P. 2012. An introduction to camera trapping for wildlife surveys in Australia. – Invasive Animals Cooperative Research Centre.

Menard, C., Duncan, P., Fleurance, G., Georges, J. Y. and Lila, M. 2002. Comparative foraging and nutrition of horses and cattle in european wetlands. – J. Appl. Ecol. 39: 120–133.

Monsarrat, S., Hansen, O. L. P. and Svenning, J. 2020. Highly variable impacts of feral horses on ecosystems worldwide. – Biol. Conserv. 247: 108616.

National Research Council. 2013. Using science to improve the blm wild horse and burro program: a way forward. – The National Academies Press.

Newsome, A. E. 1971. Competition between wildlife and domestic livestock. – Aust. Vet. J. 47: 577–586.

Norbury, G. L., Pech, R. P., Byrom, A. E. and Innes, J. 2015. Density-impact functions for terrestrial vertebrate pests and indigenous biota: guidelines for conservation managers. – Biol. Conserv. 191: 409–420.

Nugent, G., Fraser, K. W. and Sweetapple, P. J. 1997. Comparison of red deer and possum diets and impacts in podocarp-hardwood forest, Waihaha Catchment, Pureora conservation park. – Sci. Conserv. 50: 64.

Nuñez, C. M. V., Scorolli, A., Lagos, L., Berman, D., and Kane, A. J. 2016. Management of free-roaming horses. – In: Ransom, J. and Kaczensky, P. (eds), Wild equids: ecology, management, and conservation. Johns Hopkins Univ. Press.

Parks Victoria. 2016. Greater alpine national parks management plan. – Parks Victoria.

Parks Victoria. 2017. Protection of the alpine national park: feral horse strategic action plan 2018–2021. – Victoria State Government. P Victoria.

Ratcliffe, P. R. 1987. Red deer population changes and the independent assessment of population size. – Symp. Zool. Soc. Lond. 58: 13.

Robertson, G., Wright, J., Brown, D., Yuen, K. and Tongway, D. 2019. An assessment of feral horse impacts on treeless drainage lines in the Australian alps. – Ecol. Manag. Restor. 20: 21–30.

Rogers, G. M. 1991. Kaimanawa feral horses and their environmental impacts. – N. Z. J. Ecol. 15: 49–64.

Rule, S., Brook, B. W., Haberle, S. G., Turney, C. S. M., Kershaw, A. P. and Johnson, C. N. 2012. The aftermath of megafaunal extinction: ecosystem transformation in pleistocene Australia. – Science 335: 1483–1486.

SAP. 2020. Final report of the Kosciuszko wild horse scientific advisory panel: advice to assist in preparation of the Kosciuszko national park 2020 wild horse management plan. – NSW Department of Planning, Industry and Environment. IaE NSW Department of Planning.

Scasta, J. D., Adams, M., Gibbs, R. and Fleury, B. 2020. Free-ranging horse management in Australia, New Zealand and the United States: socio-ecological dimensions of a protracted environmental conflict. – Rangel. J. 42: 27–43.

Schoenecker, K. A., King, S. R. B. and Messmer, T. A. 2021. The wildlife profession's duty in achieving science-based sustainable management of free-roaming equids. – J. Wildl. Manag. 85: 1057–1061.

Schulz, M., Schroder, M. and Green, K. 2019. The occurrence of the broad-toothed rat *Mastacomys fuscus* in relation to feral horse impacts. – Ecol. Manag. Restor. 20: 31–36.

Scorolli, A. L. 2018. Feral horse management in Parque Provincial Ernesto Tornquist, Argentina. – Hum.–Wildl. Interact. 12: 8.

Scorolli, A. L. 2021. Feral horse population model and body condition: useful management tools in Tornquist park, Argentina? – J. Wildl. Manag. 85: 1162–1168.

Theile, K. R. and Prober, S. M. 1999. Assessment of impacts of feral horses (*Equus caballus*) in the Australian alps. Part 2. Outline of experimental monitoring programs for determining impacts of feral horses on flora and streams in the Cobberas-Tingaringy unit if the alpine national park. – Australian Alps Liaison Committee, Report to the Australian Alps Liaison Committee number. (U report).

Tjørve, K. M. C. and Tjørve, E. 2017. The use of Gompertz models in growth analyses, and new Gompertz-model approach: an addition to the unified-Richards family. – PLoS One 12: e0178691.

Tolsma, A. and Shannon, J. 2018. Assessing the impacts of feral horses on the Bogong high plains, Victoria. – Parks Victoria.

Tolsma, A., Shannon, J., Wright, J., Keatley, M. and Brown, D. 2018. In: Worboys, G., Driscoll, D. and Crabb, P. (eds), Bogong high plains impacts: impact of feral horses on alpine bogs, treeless drainage lines and alpine/sub-alpine vegetation communities on the Bogong high plains, alpine national park, Victoria. – Australian Academy of Science; Fenner School of Environment and Society, The Australian National University; and Deakin University.

Tongway, D. J. and Ludwig, J. A. 2011. Ephemeral drainage-line assessments: indicators of stability. – In: Tongway, D. J. and Ludwig, J. A. (eds.), Restoring disturbed landscapes: putting principles into practice. Island Press/Center for Resource Economics, pp. 151–156.

TWS. 2011. Final position statement: feral horses and burros in North America. – The Wildlife Society.

Walters, B. and Hallam, M. 1993. Feral horses in the alps: report of a workshop, howmans gap, Victoria 12–14 October 1992. – Australian Alps Liason Committee Australian Nature Conservation Agency.

Ward-Jones, J., Pulsford, I., Thackway, R., Bishwokarma, D. and Freudenberger, D. 2019. Impacts of feral horses and deer on an endangered woodland of Kosciuszko national park. – Ecol. Manag. Restor. 20: 37–46.

Whetton, P. H., Haylock, M. R. and Galloway, R. 1996. Climate change and snow-cover duration in the Australian alps. – Clim. Change 32: 447–479.

Wild, A. and Poll, M. 2012. Impacts of feral horses on vegetation and stream morphology in the Australian alps: feral horse exclusion plot monitoring and analysis. – Friends of Cobberas, Parks Victoria and the Australian Alps Liaison Committee.

Williams, R. J., Wahren, C.-H., Bradstock, R. A., and Müller, W. J. 2006. Does alpine grazing reduce blazing? A landscape test of a widely-held hypothesis. – Aust. Ecol. 31: 925–936.

Williams, R., Papst, W., McDougall, K., Mansergh, I., Heinze, D., Camac, J., Nash, M., Morgan, J., and Hoffmann, A. 2014. Alpine ecosystems. – In: Lindenmayer, D., Burns, E., Thurgate, N. and Lowe, A. (eds), Biodiversity and environmental change: monitoring, challenges and direction. CSIRO Publishing.

Wimbush, D. and Costin, A. 1979. Trends in vegetation at Kosciusko. III. Alpine range transects, 1959–1978. – Aust. J. Bot. 27: 833–871.

Worboys, G. L., Driscoll, D. and Crabb, P. 2018. Feral horse impacts: Kosciuszko science conference – Conference abstracts. – Australian Academy of Science, Australian National University and Deakin University.

Zabek, M. 2015a. Understanding population dynamics of feral horses in the tuan and toolara state forest for successful long-term population management. – PhD thesis, The University of Queensland, Australia.

Zabek, M. A. 2015b. Understanding population dynamics of feral horses in the Tuan and Toolara state forest for successful long-term population management. – University of Queensland, Australia.

Zabek, M. A, Berman, D. M., Blomberg, S., and Wright, J. 2016a. Estimating distribution and abundance of feral horses (*Equus caballus*) in a coniferous plantation in Australia, using line-transect surveys of dung. – Wildl. Res. 43: 604–614.

Zabek, M. A., Berman, D. M., Blomberg, S. P., Collins, C. W. and Wright, J. 2016b. Population dynamics of feral horses (*Equus caballus*) in an exotic coniferous plantation in Australia. – Wildl. Res. 43: 358–367.

Peer Review – Report on the 2014 Survey of Feral Horses (*Equus ferus caballus*) in the Australian Alps

Eric Rexstad and Steve Buckland

Centre for Ecological and Environmental Modelling

University of St Andrews

Preface

This document contains a review of the amended report *2014 Survey of Feral Horses (Equus ferus caballus) in the Australian Alps* as recommended by the Independent Technical Reference Group appointed by the New South Wales National Parks and Wildlife Service. The authors of this review, Drs. Eric Rexstad and Steve Buckland from the Centre for Ecological and Environmental Modelling at the University of St Andrews were invited to undertake the review by Dr. Kate Wilson of the New South Wales Office of Environment and Heritage (email dated 26 February 2016).

As defined by the Terms and Conditions, this review has two primary purposes

- Assess scientific rigour and confidence underlying
    - Current (at the time of 2014 survey) horse numbers in the Australian Alps,
    - Whether conclusions regarding population trends can be drawn and
    - Distribution of wild horses in subregions of the Park. This last point was clarified in email dated 16 March 2016 by Terry Koen, in which 'subregions' was defined as the 'block' estimates presented in Tables 5 and 6 of the report.
- Comment on specific issues raised by the ITRG
    - Appropriateness of reanalysis of 2009 survey, using observed group size,
    - Whether ground-based estimates of group size led to overestimate of numbers in 2009 survey,
    - Potential value of analysing alternative transects (in 2014 survey) to reduce risk of double counting for comparative purposes,
    - Potential value of analysing data from each stratum individually (i.e. fitting stratum-specific detection functions) to 2014 survey data for comparative purposes and
    - Inclusion of a strip transect analysis for comparative purposes.

We will discuss each of these points in turn.

1 Assess scientific rigour and confidence

By and large, the report and its conclusions were sound, we comment upon specific matters in the following set of list elements (1.1, 1.2 and 1.3).

1.1 Current (at the time of 2014 survey) horse numbers

Rigour and confidence in survey estimates are dependent upon the rigour and confidence of the elements that went into the study. We will divide our comments about the rigour and confidence of the 2014 estimates on the design, data collection and analysis phases of the 2014 survey.

1.1.1 St Andrews comments on current horse numbers

1.1.1.1 Study design

The use of the 2009 survey results as a basis upon which to base sample size calculations for the purposes of defining survey effort was sound. These calculations, of course, cannot take account of variability in group size, which plays a considerable role in these data. Awkwardly shaped portions of the study area, i.e. the Snowy River Valley, creates a challenge for imposing survey effort. The use of zig-zag transects in that area is acceptable. The only design matter that was not clear was the use of differing sampling intensity by habitat types within blocks. That does not necessarily lead to difficulties, but such an element incorporated into the survey design would expect to be used in the analysis (e.g. habitat-specific estimates of abundance within blocks).

Note also that the achieved overall precision (CV=11%) exceeded the target precision of CV≈17.5%. That might be a consequence of the 2009 survey using one-sided transects and the 2014 survey using two-sided transects.

1.1.1.2 Field methods

No issues regarding field methods. It is not commonly reported, but given the interest in group size, it would be interesting to know how many of the groups detected straddled sighting bands and were therefore broken into “observation” groups rather than “social” groups.

1.1.1.3 Data analysis

The preliminary Poisson regression analysis is not a standard approach in distance sampling analysis. Instead such modelling is done within the Distance software. Consequently we would have advocated including candidate covariates in analysis within Distance; nevertheless, it is unlikely that inclusion or exclusion of covariates in the detection function would have substantial consequences for the density/abundance estimates.

The investigation of the “spike” in detections in the 0-20m band conducted by combining the first two distance bands seems fine. The half-normal detection function is relatively insensitive to spikes close to the transect line.

Not adjusting two of the blocks for the effects of size bias, while adjusting the other two blocks is acceptable. Non-significant size-bias regression (at α=0.15) implies that the average observed group size and the adjusted group size will be almost equivalent. Consequently, there is little difference in the overall estimate of using or not using the size-biased group size estimate.

1.2 Conclusions regarding population trends can be drawn

1.2.1 St Andrews comments on population trends

This is a most challenging topic. That is because the assessment of trend ought to be based upon more than two time points. Furthermore the 2014 survey was designed with objectives that intentionally differed from the 2009 survey (Cairns 2014:1). Hence, there was considerable effort expended in adjusting 2014 estimates to take into account area surveyed in 2009 (Cairns 2014:46). The use of two portions of the surveyed region North Kosciuszko and Byadbo-Victoria (combination of South Kosciuszko, Davies Buenba and Cobberas Buchan from Dawson (2009, Table 1) (Cairns

2014:46) is a proxy for measuring change between 2009 and 2014 because the surveys sampled different overall study areas.

Secondly, assessing change between two surveys requires assessment of the "rigour and confidence" of **both** surveys. There were several issues associated with the 2009 survey that were causes for concern:

- The use of ground-based group size estimates (from 2001) to transform group density to individual density (Dawson 2009:6) is an unusual approach. This assumes that group definition is consistent for both ground-based and aerial observers, that group sizes detected from the ground are representative of sizes of groups seen from the air, and that group sizes did not change over the span of eight years. This topic will be revisited in Section 2.1 of this report.
- The exclusion of detections from the fourth distance band (150-200m) "because of an unexpectedly large number of groups sighted in this category" raises concerns described by Dawson (2009:6) that the aircraft was at the wrong altitude or the survey bar was miscalibrated. It is also possible for observers to include marginal groups near the boundary of the fourth distance band as having a centre within the band rather than outside, and including those groups in the fourth band.

Consequently, the estimation of population trends is difficult. If further investigation of change between 2009 and 2014 incorporating uncertainty in the two estimates is desired, we have some thoughts on that in this report Section 2.1.1.

1.3 Distribution of wild horses between park subregions

Are the block-specific estimates of horses reliable given a pooled (common across all blocks) detection function was used? Presumably the reason pooling was performed was to keep the number of detections used in fitting detection functions >60 (as suggested by Buckland et al. (2001:240). More recently, Buckland et al. (2015:23) note that with more advanced methods (using block as a covariate in the detection function and with 'well behaved' data) this sample size limit can be relaxed. It cannot be relaxed, however, to the point of fitting a detection function to the four sightings in the Snowy Plain block. Note that the Snowy Plain block was estimated to contain <2% of the horses in the total study area. Excluding that block from the study would not have caused a great change in estimated overall abundance.

1.3.1 St Andrews comments on subregion estimates

An alternative to the analysis presented by Cairns (2014) using a pooled detection function, would have been to fit block-specific detection functions (to all but the Snowy Plain block) and used model selection to provide an objective assessment of the relative performance of detection function models that were block-specific or pooled across blocks.

A related topic is whether the uneven coverage by habitats within blocks designed in the 2014 survey would cause the pooling robustness property of distance sampling to fail. Pooling robustness (Burnham et al. 2004:389-392) suggests that distance sampling will produce nearly unbiased estimates of density or abundance even when data collected under variable conditions are combined. Examining coverage in each survey block, using information from Cairns (2014:Table 10), along with a truncation distance of 150m:

Department of Planning, Industry and Environment - Record released under GIPAA - DPIE - 21 - 1345

Table 1. Helicopter effort, area covered, block size and coverage proportion for 4 survey blocks from Cairns (2014) survey.

Block	Effort (km)	Covered area (km^2)	Block area (km^2)	Coverage
NK	671.4	201.4	1366	0.147
SP	32.3	18.7	123	0.152
BM	409.1	122.7	847	0.145
BV	1674.5	502.2	3497	0.144

These calculations of coverage, proportion of block area surveyed, showed coverage to be approximately equal across blocks.

2 Specific issues raised by the ITRG

We turn our attention to the specific outstanding issues from the ITRG report of February 2016.

2.1 Reanalysis of 2009 survey, using observed group size

This was the area where we invested much of our time in this review. To make an assessment of the reanalysis, we needed to understand both the original 2009 survey and analysis, as well as the steps taken in the reanalysis.

As noted in item 4 of the ITRG 2016 comments, there was a puzzle caused by the Dawson (2009:7) estimate of 2.69 individuals km^{-2} and the reanalysis estimates of 1.25 and 0.57 individuals km^{-2} in Cairns (2014:Table A2).

2.1.1 St Andrews comments on reanalysis of 2009 survey

Fortuitously, Dawson (2009:18-20) provide the raw sighting and effort data for the 2009 survey. This allowed us to perform our own reanalysis of these data. We provide the Distance 7.0 project containing this reanalysis as an electronic appendix to this report.

From the 2009 survey data, we were able to duplicate the results presented by Dawson (2009:7) for the fit of the negative exponential detection function, truncating sightings from the fourth distance bin. Our estimates from Distance for group density match Dawson (2009), the AIC value for the fitted model as well as Pr(χ^2 GOF) value.

Following from that success, we attempted to duplicate the estimates of Cairns (2014:Table A2) wherein he used the same data, fitted a half-normal detection function and used the observed group sizes from the 2009 survey (adjusted for size bias) to estimate individual density. We were not able to duplicate the values in Cairns (2014:Table A2), but instead produced individual estimates for the two areas of interest of 2.33 km^{-2} (North Kosciuszko) and 1.13 km^{-2} for South Kosciuszko, Davies Buenba and Cobberas Buchan. Components of analysis leading to these results are presented in Table 2.

Table 2. Components of density estimates derived from reanalysis of Dawson (2009) survey data. Fourth distance band was omitted, half-normal detection function fitted to pooled data, size-bias adjustment was used for both strata. Column labels from left to right: number of detections, kilometres of survey effort, encounter rate, CV(encounter rate), detection probability of group within 150m of transect, CV(probability of detection), observed average group size, size-bias adjusted group size, CV(group size), estimated group density (km^{-2}), CV(group density), estimated individual density (km^{-2}), CV(individual density).

	n	L	n/L	CV	$\widehat{P_a}$	CV	$\bar{s}$	E(s)	CV	$\widehat{D_s}$	CV	$\widehat{D}$	CV
NK	16	387	0.041	28%	0.67	13%	5.9	5.7	26%	0.412	31%	2.33	40%
SK/NV	41	914	0.045	19%	0.67	13%	3.0	2.5	11%	0.446	23%	1.13	25%

We traced this discrepancy to the fact that the Distance software assumes that observers are collecting data from both sides of the transect, such that the estimator of density is

$$\widehat{D} = \frac{n\,\widehat{f(0)}}{2\,L}.$$

However, for the 2009 survey, both observers were seated on the same side of the helicopter. Hence a sampling fraction of ½ needed to be applied as a multiplier. Consequently, the 2009 estimates presented by Cairns (2014:Table A2) were too small by a factor of two. In the case of North Kosciuszko, our estimate is 1.86X that of Cairns; we have been unable to trace the remaining small difference.

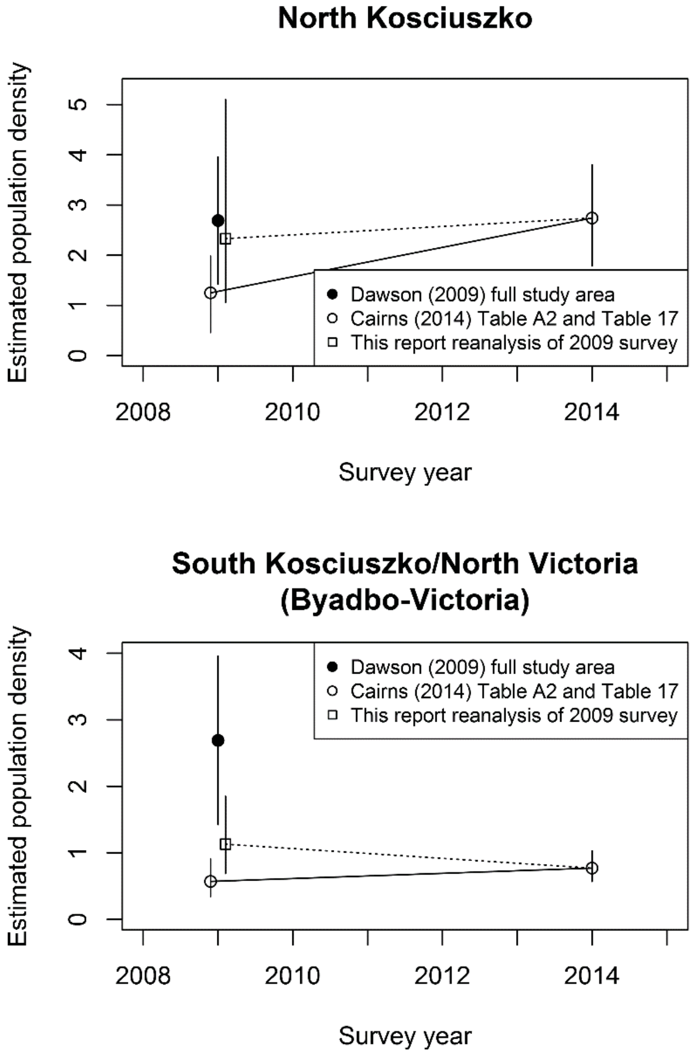


Figure 1. Individual density estimated three ways from 2009 survey and the single estimate of Cairns (2014:Table 17) of the 2014 survey.

Three estimates of individual density in 2009 for two regions (North Kosciuszko and South Kosciuszko/North Victoria) are presented in Figure 1 along with their 95% confidence intervals. The size of the confidence interval for our reanalysis of the 2009 survey from North Kosciuszko is because just 14 sightings contribute to that estimate.

The study-area wide density estimate reported by Dawson (2009:7) of 2.69 individuals km^{-2} was derived from an estimated density of 0.475 groups km^{-2} combined with a ground-based group size of 5.65 individuals per group. The difference in individual density between Dawson's (2009) estimate and the estimate produced in our reanalysis is small for North Kosciuszko because the 2001 group size of 5.65 individuals per group is close to the size bias adjusted group size of 5.66 individuals per group. However, the difference in individual density between the value of 2.69 individuals km^{-2} reported by Dawson (2009) and our estimate is largely a result of the group size estimate applied (5.65 individuals per group for Dawson's estimate and the size bias adjusted group size for South Kosciuszko/North Victoria of 2.52 individuals per group).

Fundamentally it is the adjustment that needed to be done in the reanalysis of the 2009 data for the one-sided transects that likely led to much of the concern regarding the findings of Cairns (2014).

If a measure of the annual rate of change between 2009 and 2014 were desired for these two areas, draws from two log-normal distributions could be made repeatedly. The log-normal distributions would have means equal to the 2009 and 2014 point estimates and standard deviations equal to the standard errors of the 2009 and 2014 estimates of density. These density estimates could be converted to abundance estimates multiplying by appropriate areas, and finite rates of population growth (λ) could be calculated for each draw. This would result in a distribution of $\hat{\lambda}$ incorporating uncertainty in both annual estimates.

2.2 Ground-based estimates of group size led to overestimate of numbers in 2009

Without detailed information regarding data collection methods for the 2001 estimates of group size we are limited in our ability to speak to the appropriateness of that estimate used to convert group density to individual density in the 2009 aerial survey.

2.2.1 St Andrews comments on ground-based group size estimates in 2009 survey analysis

As noted in Section 2.2.1, for the North Kosciuszko region, the 2001 ground-based estimate is almost spot-on with the size bias adjusted estimate of group size from the aerial data. However for the South Kosciuszko/North Victoria region, the ground-based estimate is twice the size of the observed average group size of 3.0 or the bias adjusted estimate of 2.5. We did not consult Walter (2002) to learn the sample size or sampling methodology associated with the ground-based estimate of group size from 2001, therefore we cannot assess whether the 2001 estimate is representative of the entire 2009 study area. The reanalysis suggests not.

2.3 Potential value of analysing alternate transects (in 2014 survey)

The final three points of the ITRG each suggest an alternative analysis of the 2014 data against which to compare the original 2014 analysis. It is certain that each of the alternative analyses proposed will produce estimates different from those presented in Cairns (2014). We suspect that rather than increase confidence in the estimates presented in Cairns (2014), alternative analyses will decrease the confidence in the original findings.

All data analysis undertakings consist of a very large number of decisions; what detection functions to fit, at what perpendicular distance to truncate sightings, how to estimate group size, which transects to include, etc. etc. Each of those decisions influence the outcome. Multiple results stemming from multiple analyses will simply lead to the decision of which result to choose.

2.3.1 St Andrews comments on analysis of alternate transects

Given the line separation, it is possible that animals initially close to one transect are driven towards the next transect. By excluding alternate lines, it may seem that this potential source of bias is

Department of Planning, Industry and Environment - Record released under GIPAA - DPIE - 21 - 1345

removed. However, this is not the case. Because the aircraft traversed all the lines, all lines except the first are potentially affected by this bias. By excluding line 2 say from the analysis, you do not remove the effect on line 3 of animals being driven from line 2. If this was thought to be a serious concern, future surveys could use greater separation between lines. (Overall effort could be retained by traversing lines 1, 3, 5, … in one pass, then lines 2, 4, 6, … in another pass, perhaps on a different day.)

2.4 Potential value of analysing data from each stratum individually (i.e. fitting stratum-specific detection functions) to 2014 survey data for comparative purposes

Brief comments about this alternative analysis follows.

2.4.1 St Andrews comments on stratum-specific detection functions

As noted in Section 1.3.1 of this report, a reanalysis using stratum-specific detection functions could be contrasted with the approach used by Cairns (2014) and arbitrated via AIC. However such refinements to the fitting of detection functions are not likely to produce considerable changes in the overall estimates of density and abundance. As noted in Section 1.3.1 of this report, distance sampling results are robust to detection function heterogeneity when coverage probability is uniform across the study area.

Examining Table 15 of Cairns (2014), note that uncertainty in the density estimates are largely driven by variability in encounter rate (number of groups per km of transect). Small adjustments to the detection function are unlikely to change density and abundance estimates greatly.

2.5 Inclusion of a strip transect analysis for comparative purposes

It is almost certain this analysis will produce a larger estimate of density and abundance than the analysis of Cairns (2014). It will also produce an estimate with greater precision, because one form of uncertainty arising from estimating the parameters of the detection function will be assumed away.

2.5.1 St Andrews comments on strip transect analysis

Performing a strip transect analysis requires the determination of the cutpoint within which detectability is assumed to be 1. Use of different cutpoints will produce different estimates. Detection is assumed to be certain in the first distance bin, however Cairns (2014:36) suggests there might be some heaping of detections in that first interval. That would produce a density and abundance estimate larger than the estimate provided by Cairns (2014). The veracity of a strip transect estimate based upon the first bin will be dependent upon an untestable assessment of the existence of heaping in that interval.

Summary

We consider that the analyses of the 2014 data have been carried out to a high standard. We see little merit in carrying out a range of further analyses. Any analysis involves a range of decisions and different analysts will obtain different results. However, we see nothing in these analyses to suggest that estimates would change appreciably provided the decisions are made to ensure as far as possible that assumptions are met to a good approximation.

The analyses of the 2009 data are more problematic. The two surveys were not designed to be directly comparable, thus compromising the comparisons now being made. The original analyses of the 2009 data adopted a poor model for the detection function (the negative exponential), and estimated group size from ground-based surveys from 2001. Both of these decisions have the potential to generate large bias.

The reanalysis of the 2009 data has been done to a higher standard, apart from the apparent failure to adjust for the fact that the survey was conducted with observers on one side of the aircraft only. This has resulted in abundance estimates for 2009 that are low by a factor of roughly two. If this is corrected, there appears to be little change in the total population size between the two surveys as shown in Figure 1; there is an estimated small decrease in the southern stratum, and a small increase in the northern stratum.

Literature Cited

Buckland, S., D. Anderson, K. Burnham, J. Laake, D. Borchers and L. Thomas. 2001. Introduction to Distance Sampling. Oxford Univ. Press.

Buckland, S., E. Rexstad, T. Marques and C. Oedekoven. 2015. Distance Sampling: Methods and Applications. Springer.

Burnham, K., S. Buckland, J. Laake, D. Borchers, T. Marques, J. Bishop and L. Thomas. 2004. Further topics in distance sampling. Pages 307-392 in Buckland, S., D. Anderson, K. Burnham, J. Laake, D. Borchers and L. Thomas. (eds.) Advanced Distance Sampling. Oxford Univ. Press.

Cairns, S. 2014. Feral horses in the Australian Alps: the design and analysis of surveys conducted in April-May 2014. A report to the Australian Alps Liaison Committee.

Dawson, M. 2009. 2009 aerial survey of feral horses in the Australian Alps. A report prepared for the Australian Alps Liaison Committee.

Walter, M. 2002. The population ecology of wild horses in the Australian Alps. Doctoral thesis. University of Canberra, Canberra

Review – Report on the 2019 Survey of Feral Horses (*Equus ferus caballus*) in the Australian Alps

Eric Rexstad and Steve Buckland

Centre for Ecological and Environmental Modelling

University of St Andrews

Preface

As requested by Clause 3(f) (email correspondence, 04Nov2019), we have been asked to review this report. We treat this review in the same fashion as a peer review of a scientific manuscript.

1 Study design and field protocol

As the emphasis of this report seems to be contrasting abundance estimates from 2014 and 2019, survey design and field protocols are described as equivalent. We have no concerns about design or field methods. The subdivision of groups on the ground into groups defined by distance band boundaries is unusual but appears to be conducted in the same way as was done in 2014 (largely by the same observers). It would have been useful to know the proportion of observed groups that were broken by distance band boundaries.

2 Data analysis

Fitting of detection functions with covariates seems consistent with standard practice. A minor quibble is the use of highly parameterised detection functions in the face of only five detection function bands. With five points of support in the data, the richest model that could be fit while still providing a degree of freedom for a χ^2 goodness-of-fit test would be a three-parameter model. With the suite of covariates in this data set, the only covariate model that fits this criterion is a half normal with the habitat covariate. All other models are too parameter rich to allow assessment of model fit beyond visual inspection. The number of candidate models fitted is more than adequate to explore possible models.

3 The report itself

There are several quibbles about wording noted in the Miscellaneous comments section below. It is not necessary to dwell on differences in detectability between years (line 702) or differences in precision between blocks (line 844); those differences are small.

4 Computation of finite rate of growth

The block-specific finite rates of growth are likely at the centre of the 2019 report. The high rate of growth reported for the North Kosciuszko block are of particular interest as it appears to exceed published maximum growth rates for the species. It is surprising that during the period 2003-2009 there was a 21% increase, while the period 2009-2014 there was very little change and for 2014-2019 again a 22% increase.

The rates of growth and inferences derived therefrom appear to be properly computed. We undertake a brief investigation of the distance sampling estimation components to assess the potential for subjective decisions during data collection or analysis to contribute to the reported change in population size for the North Kosciuszko block.

Using the parameter estimates provided, we decomposed the 5.4X increase in North Kosciuszko open and 3.2X increase in North Kosciuszko medium horse abundance between 2014 and 2019 shown in Table 11 into the respective distance sampling elements.

		2014			**2019**	
		NK Open	NK Medium		NK Open	NK Medium
Encounter rate		0.208	0.075		0.560	0.168
Percent difference					269%	224%
Group size		2.64	3.2		4.98	4.02
Percent difference					189%	126%
Pr(detection)		0.53	0.53		0.47	0.47
Percent difference					113%	113%
Compound multiple					5.73	3.18
Reported multiple (from Table 11 Ns)					5.44	3.16

Starting with the least influential cause of the increase: there was only a 13% change in probability of detection, from 0.53 to 0.47. There was a more considerable increase in group size used in the conversion from group density to individual density: an 89% increase in group size for Open and 26% increase in Medium. The most influential cause in the change in computed abundance is encounter rate that more than doubled for both habitats between 2014 and 2019. Multiplying the three factors approximates the multiplier in abundance calculated from Table 11.

Model selection and fitting the detection function plays little part in explaining the between-year change in abundance. Also (lines 340-341 and 349-350) two of the three observers used in 2019 were the same as the 2014 observers, suggesting similarity in detection processes between years. Assuming the field protocol of splitting herds that straddled distance bands was done identically in 2014 and 2019 (see p39-40 of the 2016 report), there is little investigator influence upon group size; that is more likely a biological phenomenon. The component most responsible for reported increase in population size in both North Kosciuszko habitats is encounter rate; more that doubling in each habitat between years.

At face value, from statistics presented in the report, the reason for the large differences in abundance in North Kosciuszko between years is that the 2019 survey encountered, on average, larger herds, and many more of them per unit distance of search, than in 2014. It is unlikely these events were under investigator control. Consequently there is no reason to doubt the reported abundance estimates and the derived finite rates of population growth.

5 Miscellaneous comments

Line numbers as in the report

- Line 23, because of insufficient detections, not replication transects, I presume
- line 25, was the increase statistically significant?
- line 27, is this level of increase possible?
- line 49, deer rate of growth is roughly as fast as horse rate of growth
- Line 180, apparently an important area to survey but quite small, how much effort could be employed in this stratum?
- Line 286, were the small differences between the survey designs the result of new realizations of the 2014 designs generated
- line 293, were the blocks flown in the same sequence in 2019 as in 2014
- Line 329, regarding the calibrated citing boom what allowance existed for observer head placement, that influences how the sighting boom gradations map onto the Earth's surface
- Line 344, is there any reason for a difference in helicopters in the northern blocks between 2014 and 2019 contributing to the large jump in abundance estimates in NK
- Line 376, one group of animals on the ground could be broken into 2 or more if they straddled distance bands; what proportion of detections were treated this way
- Line 473, hazard rate with three adjustments would be a 5 parameter model; given 5 distance bins degrees of freedom for goodness of fit tests would run out even with a smaller model than this
- Line 513, 3 factor covariates, 2 with 3 levels and one with 2 levels combined in any more than 2 at a time would result in non-identifiable models
- Line 533 actually confidence intervals produced by distance on log based; assuming an underlying lognormal sampling distribution
- Line 557, this is awkward wording, the 73% refers to the portion of the bloc not excluded from consideration in the survey, the proportion of the study area covered by survey efforts is closer to 15%
- Line 589 I think this means a pooled detection function for the two strata was used for NK with stratum specific detection functions. What might have transpired if stratum was used as a covariate? I suppose the model was already over-saturated with the existing covariates.
- Line 609, 16 models seems like too many candidate models for a 5 bin data set
- Line 612, not being able to assess model fit is problematic; might it have been better to restrict the candidate model set to simpler models where fit could be assessed
- Line 702, stating the detection probability was higher in 2014 than 2019 doesn't seem substantiated by the values reported; any differences in probability of detection between years is fairly small and probably not worthy of mention
- line 707, unclear why it needed to be stated that the surveys were not conducted as strip transect surveys, that seems obvious from the outset
- Table 5, If pooled detection functions (NK open and medium) are reported in this table, it is unclear why CV for those pooled estimates would be identical for 2014 but not identical for 2019
- Line 744, for groups spanning multiple bins, were two group sizes recorded by the observers question how prevalent was this?
- line 751, given the splitting of groups that straddle bins, the explanation for large clusters being far from the transect could simply be down to the bins far from the transect being the

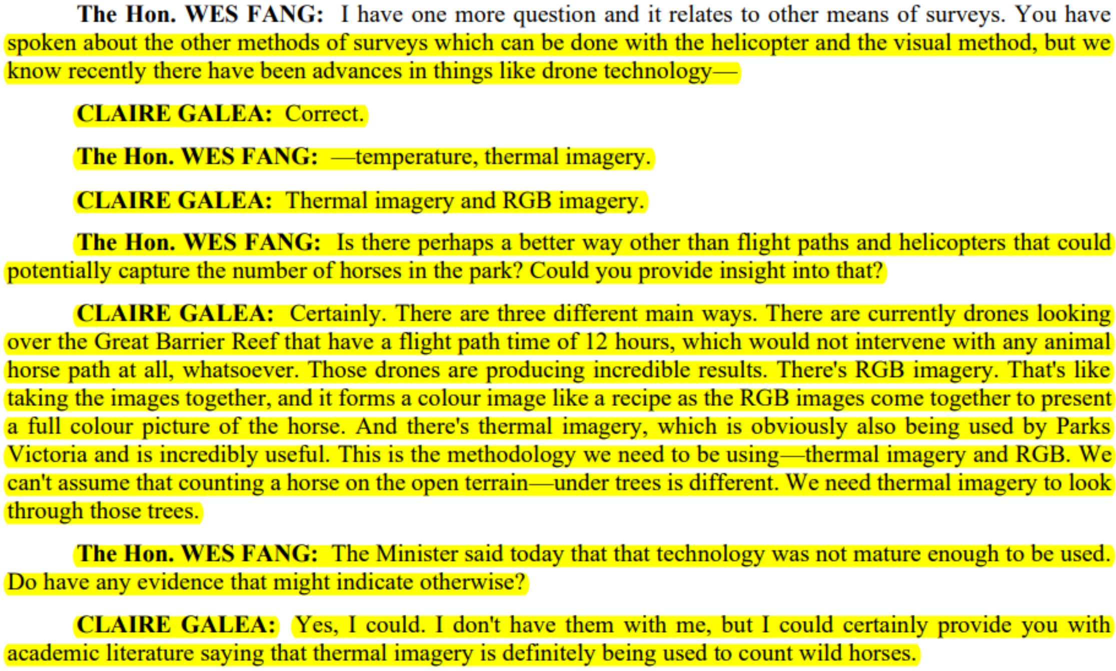
Department of Planning, Industry and Environment - Record released under GIPAA - DPIE - 21 - 1345

widest bins. An alternative approach would have been to ignore clusters and count number of horses in each distance interval and treat each animal as an individual detection

- line 791, not sure I agree that arbitrariness is removed, seems there is another set of decisions observers need to make: where to break the cluster and then to evaluate two cluster sizes rather than one
- Line 818, why are densities of clusters across 3 survey blocks not similar in 2019 as they were in 2014; this seems central to the findings of the report
- Table 7 and 8, given the emphasis of the report is on temporal differences rather than spatial, seems tables could be reorganised to present estimates such that the temporal comparisons are easier to make
- Table 7, Snowy River Valley CV of Ds should be 45.5 rather than 39.7
- Table 8, Caption should indicate results for 2019 not 2014
- I would say differences in precision between years is small; only exception being NK open where precision is much higher in 2019 likely because the point estimate is 3 times larger in 2019
- line 844, this difference in precision is hardly perceptible
- line 851, hard to argue this point about trying to increase precision, given that the target precision was 20% and in most instances the CVs reported here are in the vicinity of 15%
- Line 885-897, it is arguable whether any inference should be made to the areas not surveyed in each block. The assumption in the report is horse density is zero in unsurveyed areas. As noted, the assumption cannot be refuted. However, because the centre of the report seems to be inference about changes in abundance and abundance is based upon size of area over which inference is to be made, the distinction between surveyed area and block area takes on added importance. Was the same extrapolation performed to produce the abundance estimates for 2014 provided in Table 11, i.e. Table 17 from the 2014 report?
- Table 10, Caption should indicate results are for 2019
- Table 11, Australian Alps sum for 2014 should be 9187 to be consistent with Table 9 as well as preceding values in the table
- Table 11, SE values for total could be computed by squaring SEs by area to make variances, then summing the variances and rooting the result to obtain SEs for summed abundance over blocks
- Line 954, the suggestion is that there has been substantial movement into NK since 2014, is there any evidence to support this claim
- Line 1016-1026, this pattern of changes in λ between 5 year periods seems unusual. For the period 2003-2009 there was a 21% increase, while the period 2009-2014 there was very little change and for 2014-2019 again a 22% increase. Any explanation?

esponding to questions on notice transcript

1. Thermal imagery for wild horses



The Hon. WES FANG: I have one more question and it relates to other means of surveys. You have spoken about the other methods of surveys which can be done with the helicopter and the visual method, but we know recently there have been advances in things like drone technology—

CLAIRE GALEA: Correct.

The Hon. WES FANG: —temperature, thermal imagery.

CLAIRE GALEA: Thermal imagery and RGB imagery.

The Hon. WES FANG: Is there perhaps a better way other than flight paths and helicopters that could potentially capture the number of horses in the park? Could you provide insight into that?

CLAIRE GALEA: Certainly. There are three different main ways. There are currently drones looking over the Great Barrier Reef that have a flight path time of 12 hours, which would not intervene with any animal horse path at all, whatsoever. Those drones are producing incredible results. There's RGB imagery. That's like taking the images together, and it forms a colour image like a recipe as the RGB images come together to present a full colour picture of the horse. And there's thermal imagery, which is obviously also being used by Parks Victoria and is incredibly useful. This is the methodology we need to be using—thermal imagery and RGB. We can't assume that counting a horse on the open terrain—under trees is different. We need thermal imagery to look through those trees.

The Hon. WES FANG: The Minister said today that that technology was not mature enough to be used. Do have any evidence that might indicate otherwise?

CLAIRE GALEA: Yes, I could. I don't have them with me, but I could certainly provide you with academic literature saying that thermal imagery is definitely being used to count wild horses.

Below are peer-reviewed publications demonstrating the evidence of imagery and wild horses. The publications have been attached to the email.

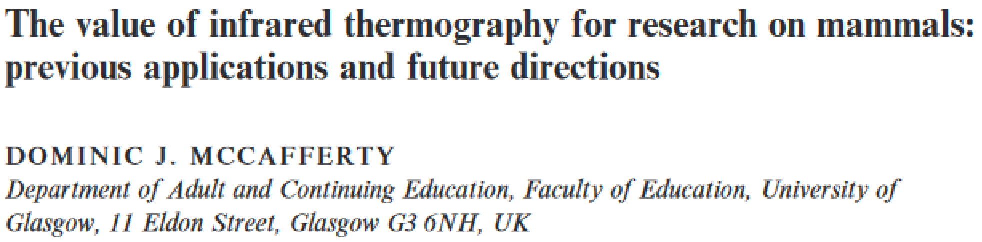
Technical Note

Evaluating Alternative Flight Plans in Thermal Drone Wildlife Surveys—Simulation Study

Julia Witczuk * and Stanisław Pagacz

Fusion of visible and thermal images improves automated detection and classification of animals for drone surveys

B. Santhana Krishnan[1], Landon R. Jones[2], Jared A. Elmore[2,3], Sathishkumar Samiappan[1], Kristine O. Evans[2], Morgan B. Pfeiffer[4], Bradley F. Blackwell[4] & Raymond B. Iglay[2]



The value of infrared thermography for research on mammals: previous applications and future directions

DOMINIC J. MCCAFFERTY

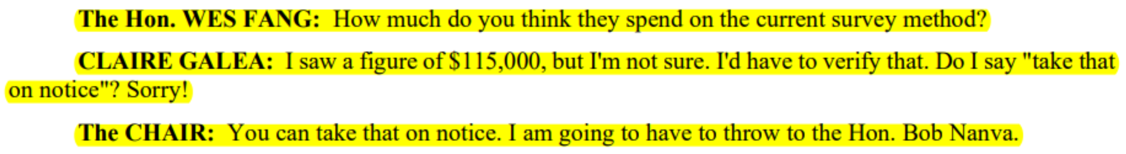
Department of Adult and Continuing Education, Faculty of Education, University of Glasgow, 11 Eldon Street, Glasgow G3 6NH, UK

Feasibility of using high-resolution satellite imagery to assess vertebrate wildlife populations

Michelle A. LaRue,* ¶ Seth Stapleton,† and Morgan Anderson‡

*Department of Earth Sciences, University of Minnesota, 310 Pillsbury Drive SE, Minneapolis, MN 55455, U.S.A.
†Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, St. Paul, MN 55108, U.S.A.
‡Department of Environment, Government of Nunavut, Igloolik, Nunavut X0A 0L0, Canada

2. Cost of Cairns surveys



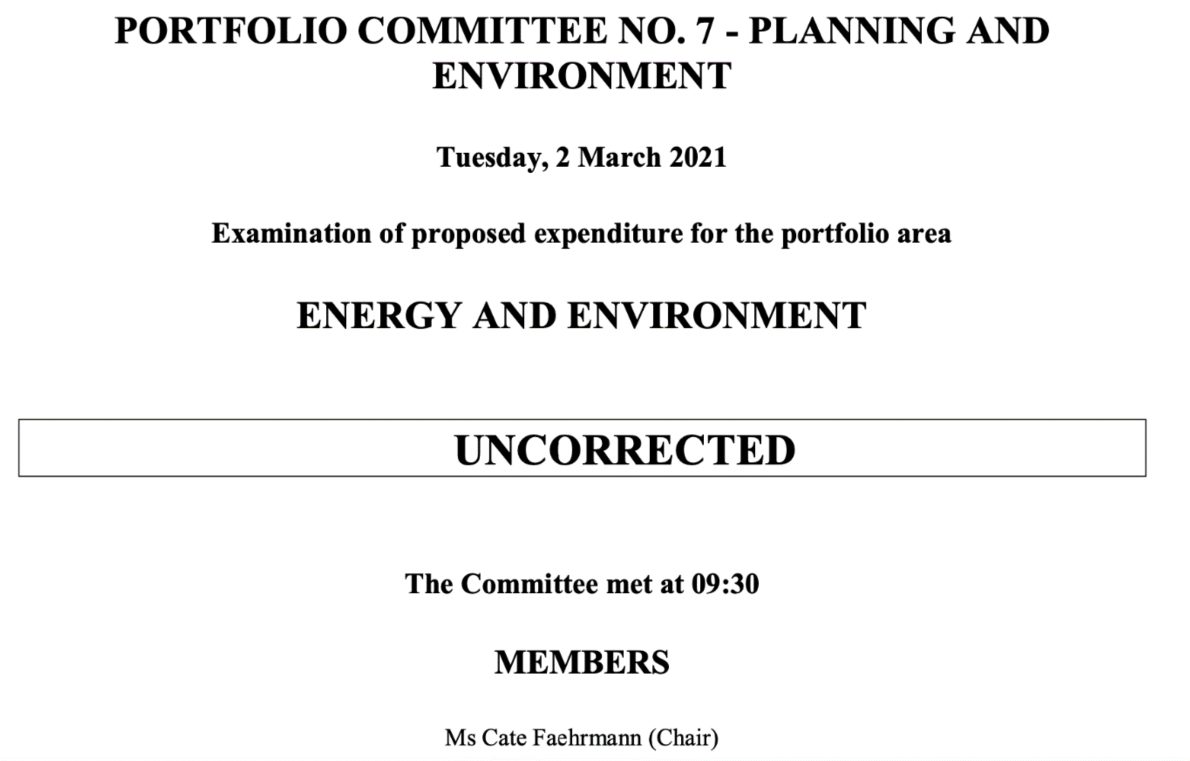
The Hon. WES FANG: How much do you think they spend on the current survey method?

CLAIRE GALEA: I saw a figure of $115,000, but I'm not sure. I'd have to verify that. Do I say "take that on notice"? Sorry!

The CHAIR: You can take that on notice. I am going to have to throw to the Hon. Bob Nanva.

I was informed of the cost of the annual head count surveys undertaken by NPWS to be $116,000 by Snowy Mountains Brumby Sustainability Management Group who obtained this information from the Hon Emma Hurst. MLC.

I am not aware of the cost for Cairns to undertake the wild horses analyses however with respect to the kangaroo analyses which applies the same methodology and is also undertaken by Cairns, this information was obtained through questions on notice in 2021. Please see screen shots below from the documentation.



PORTFOLIO COMMITTEE NO. 7 - PLANNING AND ENVIRONMENT

Tuesday, 2 March 2021

Examination of proposed expenditure for the portfolio area

ENERGY AND ENVIRONMENT

UNCORRECTED

The Committee met at 09:30

MEMBERS

Ms Cate Faehrmann (Chair)

The Hon. MARK PEARSON: It appears that Dr Stuart Cairns commenced working as a paid consultant for your department from at least the late 1990s. Can you advise how much money your department has paid to Dr Stuart Cairns or G E & S C Cairns Consulting Pty Ltd over the past 30 years? I suppose you would want to take that on notice?

Mr MATT KEAN: I might have to take some time to quantify that, Mr Pearson.

Answer

I am advised:

Departmental accounting records show GE and SC Cairns Consulting Pty has been paid $99,800 (excluding GST) for kangaroo management related work. The earliest payment was in 2015.

Payments prior to 2015 for Dr Stuart Cairns' work were made to the University of New England. Payments made for kangaroo work could not be distinguished from University of New England services for other work.

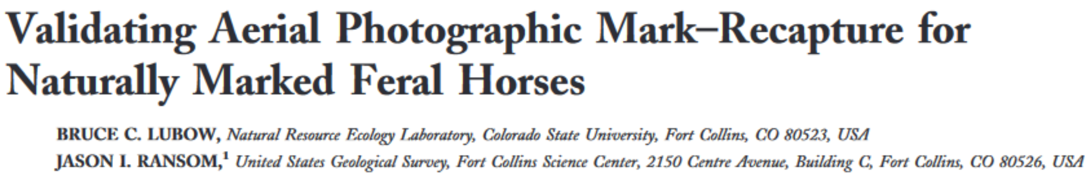
The Department of Planning, Industry and Environment currently has a five-year contract with Dr Cairns to analyse and report on the Tablelands Helicopter Surveys 2017–2021. This contract is for a total of $103,400.

3. Mark resight for wild horses

The Hon. BOB NANVA: In terms of the application of the resight model in Victoria, is that in a small, isolated population?

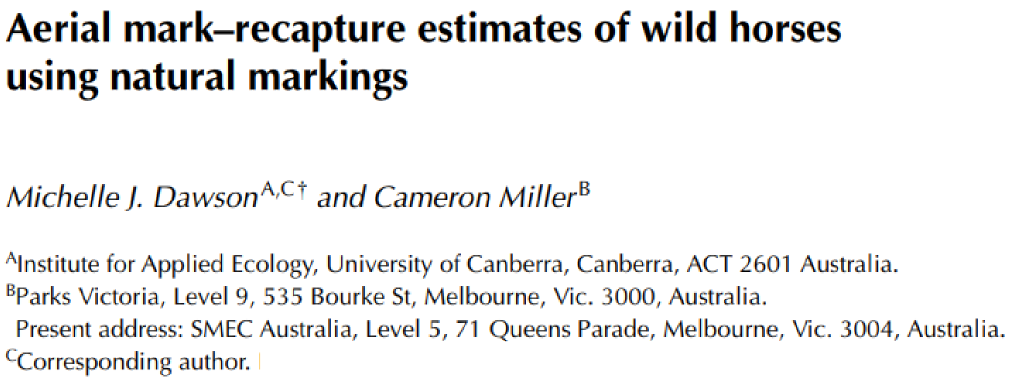
CLAIRE GALEA: It is in a smaller isolated population. But I could provide you so many publications where mark-resight has been used on aquatic animals in much larger areas—many, many examples. It's the gold standard for capturing imagery on wildlife. I can provide you many examples, if that would help.

Below are peer-reviewed publications demonstrating the evidence mark-resight for wild horses. The publications have been attached to the email.



Validating Aerial Photographic Mark–Recapture for Naturally Marked Feral Horses

BRUCE C. LUBOW, *Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO 80523, USA*
JASON I. RANSOM,[1] *United States Geological Survey, Fort Collins Science Center, 2150 Centre Avenue, Building C, Fort Collins, CO 80526, USA*



Aerial mark–recapture estimates of wild horses using natural markings

Michelle J. Dawson[A,C†] *and Cameron Miller*[B]

[A]Institute for Applied Ecology, University of Canberra, Canberra, ACT 2601 Australia.
[B]Parks Victoria, Level 9, 535 Bourke St, Melbourne, Vic. 3000, Australia.
 Present address: SMEC Australia, Level 5, 71 Queens Parade, Melbourne, Vic. 3004, Australia.
[C]Corresponding author.

remote sensing

Technical Note

Evaluating Alternative Flight Plans in Thermal Drone Wildlife Surveys—Simulation Study

Julia Witczuk * and Stanisław Pagacz

Museum and Institute of Zoology Polish Academy of Sciences, Wilcza 64, 00-679 Warszawa, Poland;

* Correspondence:

Abstract: The rapidly developing technology of unmanned aerial vehicles (drones) extends to the availability of aerial surveys for wildlife research and management. However, regulations limiting drone operations to visual line of sight (VLOS) seriously affect the design of surveys, as flight paths must be concentrated within small sampling blocks. Such a design is inferior to spatially unrestricted randomized designs available if operations beyond visual line of sight (BVLOS) are allowed. We used computer simulations to assess whether the VLOS rule affects the accuracy and precision of wildlife density estimates derived from drone collected data. We tested two alternative flight plans (VLOS vs. BVLOS) in simulated surveys of low-, medium- and high-density populations of a hypothetical ungulate species with three levels of effort (one to three repetitions). The population density was estimated using the ratio estimate and distance sampling method. The observed differences in the accuracy and precision of estimates from the VLOS and BVLOS surveys were relatively small and negligible. Only in the case of the low-density population (2 ind./100 ha) surveyed once was the VLOS design inferior to BVLOS, delivering biased and less precise estimates. These results show that while the VLOS regulations complicate survey logistics and interfere with random survey design, the quality of derived estimates does not have to be compromised. We advise testing alternative survey variants with the aid of computer simulations to achieve reliable estimates while minimizing survey costs.

Keywords: bias; BVLOS; population density; survey design; thermal infrared; UAV; ungulates; variability; VLOS; wildlife monitoring

check for updates

Citation: Witczuk, J.; Pagacz, S. Evaluating Alternative Flight Plans in Thermal Drone Wildlife Surveys—Simulation Study. *Remote Sens.* **2021**, *13*, 1102. https://doi.org/10.3390/rs13061102

Academic Editor: Magaly Koch

Received: 8 February 2021
Accepted: 8 March 2021
Published: 14 March 2021

1. Introduction

Estimating population size is one of the fundamental tasks in wildlife management and conservation. Accurate and precise estimates of densities are especially important in the case of wild ungulates, which are not only popular game species, but also cause damage to forestry and agriculture, and are involved in vehicle collisions and transmission of diseases to livestock [1]. Since ungulate populations often occupy large, mostly forested areas, surveying them is logistically difficult and costly. The currently used ground-based survey methods, such as drive counts or snow tracking are acknowledged to be unreliable and inefficient [2]. Alternative methods such as aerial surveys are more effective; however, over the forests, their use remains limited, as camouflage coloration of ungulate coats impedes their detection on the cluttered background of the forest floor. Additionally, in areas where ungulates are active mostly during dusk and dawn [3–5], the applicability of aerial surveys is further reduced; in times of poor visibility, especially at lower altitudes, manned aircraft flights are challenging and dangerous.

The recent development of unmanned aerial vehicles (UAVs, commonly known as "drones") industry, together with the improvement, miniaturization, and increased availability of thermal infrared cameras has brought new possibilities to aerial wildlife surveys. The use of drones and thermography overcomes limitations of manned aircraft, since drones can safely fly even in the dark and animal detection based on thermal contrast is not affected by camouflage coats of animals. Successful tests of such surveys have been reported and interest in the implementation of drones for wildlife research and management is still growing [6–8].

While drones have many advantages over manned aircraft, they also have limitations. The flight range of small or medium drones used by wildlife researchers is constrained by the battery capacity or fuel tank size. More importantly, regulations applied in most countries limit the flight range by restricting drone operation to visual line of sight (VLOS), meaning that during the whole flight the drone must be visible to the operator or assisting visual observer. Some countries (e.g., in the European Union [9]) tend to have more relaxed UAV policies and allow flights beyond the visual line of sight (BVLOS); nevertheless, in most cases, researchers planning to use drones for wildlife monitoring will have to cope with limited drone range under VLOS restrictions. The regulation not only reduces the efficiency but also affects the design of surveys by the need to concentrate flight paths into small sampling blocks. The number and location of these blocks may be additionally limited by the availability of convenient take-off and landing spots. These factors can lead to poor designs and, as a consequence, to biased estimates of population densities. Therefore, it is important to understand whether being constrained by the VLOS rule survey design affects the accuracy and precision of the obtained results. Here, we aimed to answer this question with the aid of computer simulations.

To assess the effect of the survey design on the accuracy (bias) and precision (variability) of density estimates derived from drone collected data, we conducted simulated surveys of three virtual populations of a hypothetical ungulate species (differing in animal density) using two alternative flight plan designs: (1) under VLOS restriction, and (2) assuming no restriction of the flight range (BVLOS). For each design, we additionally tested three levels of survey effort (from one to three survey repetitions).

2. Materials and Methods

The methods of conducting surveys and estimating population density applied in the simulations were the same as those used in the actual field tests of thermal drone surveys conducted in 2017 in several hunting districts in Poland (Table S1). As a virtual study area (Figure 1), we chose one of the surveyed districts (8400 ha). For simulated data collection, we assumed the same conditions as during field tests—flight altitude of 150 m and a 640-pixel camera resolution—resulting in a survey strip width of 65 m.

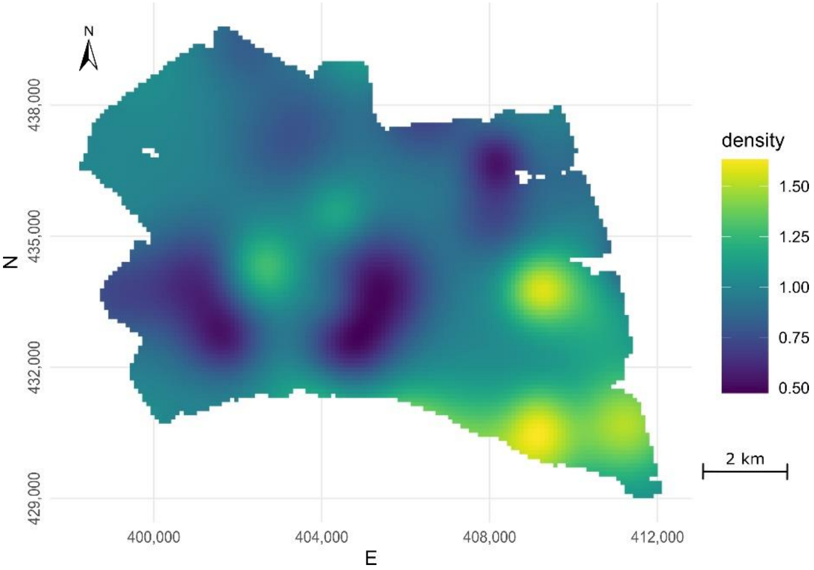


Figure 1. The study area with density surface used to emulate uneven spatial distribution of individuals in the virtual populations.

The simulations were run in three steps: (1) generation of the virtual population, (2) simulation of aerial survey of this population, and (3) estimation of population density based on the data collected during the simulated survey.

2.1. Virtual Populations

Since the population size influences the precision and accuracy of density estimates [10], we conducted simulations for three levels of population densities of a hypothetical ungulate species. Chosen density levels represented the range of local red deer densities in Poland: low (two individuals per 100 ha), medium (seven individuals per 100 ha), and high (12 individuals per 100 ha) [11]. In the study area of 8400 ha, these densities translate to population sizes of 168, 588, and 1008 individuals. In the first step of each simulation, we generated a virtual population of a given size and distributed its individuals throughout the study area. In natural populations, animals often occur in groups of varying sizes; however, in simulations, we assumed that animals occur individually, to reduce the variability in density estimates due to the effect of group size [10]. To emulate uneven spatial distribution of animals across the study area, we used the density surface with randomly placed spots of low and high animal density (Figure 1).

2.2. Surveys

We tested two alternative variants of the flight plan: VLOS design vs. BVLOS design. In the VLOS variant (flights within visual line of sight), the transects were allocated within three sample blocks placed throughout the study area. Based on the results of pilot field tests, we determined the optimal form of sample blocks as the 2 km by 2.5 km rectangle. To place blocks within the study area, we used a quasi-random procedure. The coordinates of the blocks' centroids were randomly selected with two restrictions: the distance between centroids had to be greater than 1.5 km (to avoid blocks overlapping), and the distance of centroids to the boundaries of the study area had to be greater than 1 km (to avoid placing a significant part of the block outside the study area). Then, the blocks were rotated around their centroids for the best fit inside the study area. In case of one block, it was necessary to move its centroid ~200 m from its initially drawn location. Inside each block, we placed 13 parallel transects spaced 200 m apart (Figure 2a). We assumed that such spacing between transects should minimize the chance of double counting moving animals during the surveys [8]. The total length of 39 transects within all three blocks was 104.5 km. Previous field tests revealed that the thermal survey according to such a flight plan using a fixed-wing drone is logistically challenging but can be completed in about three hours.

Figure 2. Sample transects (gray lines) under (**a**) VLOS and (**b**) BVLOS flight variants with example realizations of a simulated survey of low-density population (two individuals per 100 ha, N = 168); open circles represent locations of undetected animals and filled circles represent locations of animals detected with probability $p = 0.6$.

The second variant assumed flights beyond the visual line of sight (BVLOS). In this case, the flight plan had a form of parallel systematic transects running across the entire study area (Figure 2b). Transects were delineated perpendicular to the long axis of the area (to maximize the number of transects). To achieve the same effort as in the VLOS variant (i.e., the same total transect length ~104 km), we set up 16 transects spaced 800 m apart.

Given the thermal swath width of 65 m, the total area surveyed along all transects constituted approximately 8% of the entire area of the hunting district in case of both designs (VLOS and BVLOS).

Over the forest, the probability of detection during aerial surveys with a thermal camera depends mainly on the sensor efficiency as well as the availability of animals for detection (animals under canopy are unavailable for detection). For simplicity, during simulated surveys, we arbitrarily set the overall value of the detection probability to $p = 0.6$, assuming it is equal to availability and independent of the distance to the transect line. This means that every animal in the virtual population, which, during the survey, was located inside the 65 m wide transect strip, had a 60% chance of detection. This rather conservative figure is based on the mean value of the tree cover density (~40%) in the studied area calculated from remotely sensed data [12].

2.3. Population Density Estimation

For the population density estimation based on data from simulated surveys, we used two methods: the ratio estimate [13,14] and distance sampling [15].

In the ratio method, the estimated density per unit area was calculated as

$$\hat{D} = \frac{\sum x_i}{\sum a_i}, \tag{1}$$

where x_i is the number of animals detected on transect i and a_i is the area of transect i. The total population abundance in the study area A was calculated as

$$\hat{N} = \hat{D}A, \tag{2}$$

with the approximated variance calculated as

$$Var(\hat{N}) = \frac{n_t(n_t - n_s)}{n_s(n_s - 1)} \left(\sum x_i^2 + \hat{D}^2 \sum a_i^2 - 2\hat{D} \sum x_i a_i \right), \tag{3}$$

where n_s is the number of surveyed transects, and n_t is the total number of all possible transects. To assess the precision of the estimates, we calculated the coefficient of variation as

$$CV = \frac{\sqrt{Var(\hat{N})}}{\hat{N}} 100. \tag{4}$$

Relative bias was calculated as

$$B = \frac{\hat{D} - D}{D} 100, \tag{5}$$

where D is the true density.

The distance sampling method accounts for the decrease of detectability with the increasing distance from the transect that is often observed in wildlife surveys. The perpendicular distances of detected individuals from the transect centerline recorded during the survey are used to fit the detection function model, which is then used to estimate density. The main assumptions of the method are: detection at distance zero (at the transect line) is certain, distances are measured without error, and animals do not move in response to the observer. Distance sampling is a commonly used method for density estimation in aerial surveys. Here, we used the simplest form of distance sampling—conventional distance sampling (CDS). In each CDS analysis, three alternative models were fit to the distance data generated during survey simulations (half-normal, hazard rate, and uniform with cosine adjustment terms) and the best model was selected based on the minimum Akaike Information Criterion (AIC) value. Distance sampling estimates, along with the corresponding coefficient of variation (CV) were calculated using the Distance package in R [16]. Relative bias was calculated as for the ratio estimate (Equation 5).

For both methods, the ratio estimate and the distance sampling do not account for animals unavailable for detection—present on the transect strip but hidden under the canopy. To correct for bias caused by unavailability, the density values obtained with either of the methods need to be multiplied by the correction factors. In practice, such factors are estimated, but incorporating the correction factor uncertainty in the simulations would introduce additional variability to population density estimates and could obscure possible differences driven by alternative flight plans. Thus, for simplicity, we assumed constant correction factor, equal to the reciprocal of the preconceived availability for detection, i.e., 1/0.6.

2.4. Simulations

For all combinations of the survey design (VLOS, BVLOS), the level of population density (low, medium, high) and level of survey effort (one to three survey repetitions), we ran 100 iterations of simulated surveys; thus, a total of 1800 survey results were generated. For each iteration, the population density was estimated using the ratio estimate and the distance sampling method. For repeated surveys, the density estimates and their variance were calculated from summed detections across repeats. To assess the differences in the accuracy and precision of the results from alternative survey designs, we plotted the distribution of performance metrics (relative bias and coefficient of variation) as a series of boxplots. All procedures were run in R 4.0.0. program [17] using packages DSsim 1.1.5. [18] and Distance 1.0.0 [16].

3. Results

The population density estimates derived from surveys under the VLOS design were in general of a similar quality to those from BVOS surveys (Figure 3). Only in one scenario—the surveys of the low-density population with low effort—did the VLOS design appear to be significantly inferior to BVLOS, showing moderate bias (median underestimation of −14%, Figure 3a) and slightly less precise results (median CVs for VLOS up to 3% higher than for BVLOS, Figure 3b).

Regardless of the survey design, the precision of estimates increased with increasing population density as well as with increasing survey effort (Figure 3). However, the most apparent improvement in precision was observed for two survey repetitions (~30% decrease in median CV values). The third survey repetition resulted in a further increase in precision but to a lesser degree (~16% decrease in median CV values).

In terms of accuracy, both estimation methods—the ratio estimate and distance sampling—provided unbiased estimates of population density in the surveys of medium and high-density populations (median bias from −5 to 2% for the ratio estimate and from 0 to 7% for distance sampling). However, in the case of a low-density population, distance sampling tended to overestimate density (median bias up to 18%). In terms of precision, the methods differed substantially in the quality of the obtained results. For all analyzed survey scenarios distance sampling estimates were less precise, showing CV values from 4 to 15% higher than those for the ratio estimates.

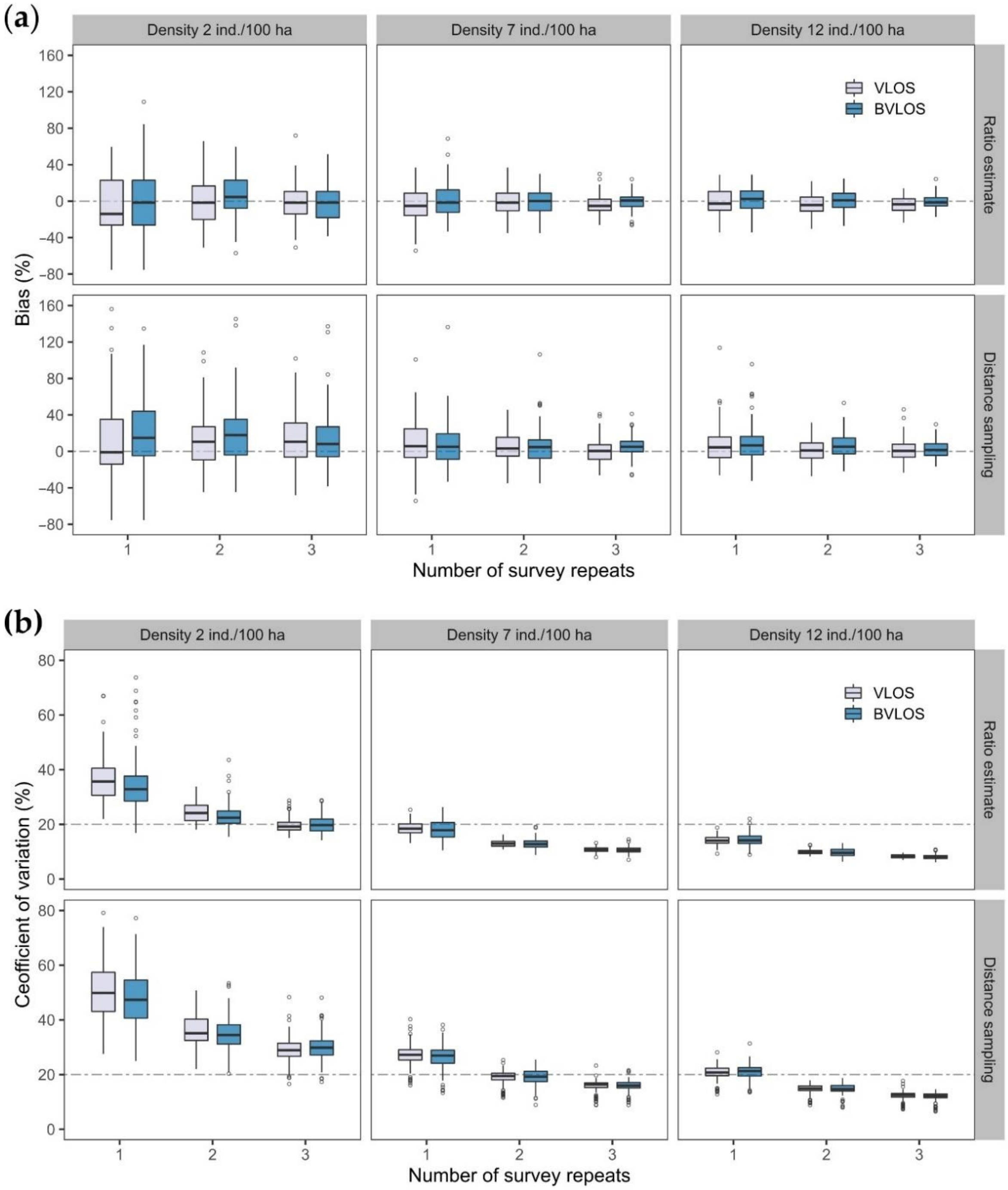


Figure 3. Pairwise comparison of (**a**) bias and (**b**) coefficient of variation distributions for population density estimates from simulated aerial surveys under VLOS and BVLOS design, in relation to different population densities, levels of survey effort, and methods of analysis. Dashed lines denote the desired level of bias (0%) and the acceptable value of the coefficient of variation (20%) for density estimates in wildlife management.

4. Discussion

The main goal of this study was to assess whether VLOS restriction would affect the accuracy and precision of the results of a wildlife aerial survey. We used simulations to compare two alternative flight plans (within VLOS vs. BVLOS) together with several combinations of survey conditions (three density levels of the surveyed population and three levels of survey effort).

The observed differences in the accuracy and precision of the population density estimates derived from the VLOS and BVLOS surveys were relatively small and negligible. Only in the case of a low-density population and low survey effort did the differences between designs become more distinct—the estimates obtained under VLOS design were biased and less precise than those derived under BVLOS (Figure 3). This was a result of the uneven coverage of VLOS design, coupled with the sparseness of the population (2 ind./100 ha) and minimum effort of one survey. However, increasing effort by one additional survey was sufficient to reduce the differences in accuracy and precision between the two designs. These results are encouraging, as the applicability of drones, which are hoped to revolutionize data collection in wildlife research [19], remains to be constrained by the VLOS rule. While the design and logistics of surveys are seriously impacted by the need to maintain visual contact with the aircraft, the results obtained in this study provide evidence that the quality of the derived estimates does not have to be compromised.

Researchers planning surveys under VLOS restriction face several issues and challenges. First, the transects must be grouped in blocks. To make survey logistics efficient, the number of blocks needs to be small, and the size of the blocks as large as possible under the constraints of VLOS in given conditions. At the same time, the set of surveyed transects in the blocks has to be representative of the entire study area. To achieve this, the blocks must be placed randomly. In a situation where the majority of the study area is forested, it may be necessary to adjust block locations by shifting them towards open areas or other suitable places for drone deployment. Such interference with the random placement of blocks may lead to the overrepresentation of open areas on transects and biased population density estimations. Thus, after the adjustment of block locations, it is advisable to compare the proportions of different land cover classes or habitats in the study area with those along transects. It is worth noticing that if locations of blocks significantly depart from random—potentially leading to biased estimates—it is possible to apply more advanced, model-based methods of density estimation [20]. These methods can overcome the issue of non-random transect placement by using spatial covariates (e.g., habitat) to model animal abundance [21].

Another challenge with thermal drone surveys, especially under VLOS restriction, is the relatively small width of the surveyed transect strip. Assuming 10 cm as the minimum ground resolution needed to recognize an animal the size of a deer [8], the surveyed strip using the most common 640-pixel resolution camera is only about 64 m wide. This results in a small area coverage per kilometer of transect. Given the limited size of the sample blocks due to the VLOS rule and the necessity of maintaining sufficient distance between transects within the blocks to avoid double counting, it is difficult to design a survey plan with transects covering more than 10% of the study area. As shown by our results, in the case of low-density populations, this may not be enough to achieve an adequate sample size and thus the desired precision of the estimate. Increasing transect coverage by delineating additional sample blocks will often be impossible because of a lack of space within the study area, and it would also complicate the survey logistics. Therefore, the best solution to increase the sample size in the aerial survey under VLOS conditions is to increase the sampling effort by repeating surveys along the same transects. The results of our simulations show that for the low-density population, three survey repeats are required to obtain a CV of about 20%, a commonly acceptable level of precision for wildlife management purposes [13,22] (Figure 3b). Under BVLOS design, besides survey repeats, higher coverage can be achieved by simply decreasing transect spacing and adding additional transects. The distance between transects in the simulated BVLOS variant was 800 m; changing it to 267 m would triple the area coverage.

Both survey data analysis methods applied in the study—the ratio estimate and distance sampling—in general gave unbiased estimates of population density. Only in the case of the low-density population did we observe a tendency for distance sampling to overestimate density (Figure 3a). This might be caused by a small number of detections (mean 16, range 2 to 42) in simulated surveys of such a sparse population, insufficient to correctly model the detection probability function (uniform in this case). It is assumed that at least 60 observations are required to accurately model the detection function in distance sampling, and with smaller samples, the results should be treated with caution [15]. Comparing the precision of the two methods, we observed that distance sampling gave far less precise results than the ratio estimate (Figure 3b). Density estimates with a CV of about 20% from distance sampling were obtained only for the medium- and high-density populations, but it always required two or three survey repeats (Figure 3b). For the low-density population, acceptable CV values could not be obtained even with three repetitions. The lower precision of distance sampling estimates is caused by uncertainty in modeling the detection probability functions. In the case of the ratio estimates, values of CV equal to or below 20% were achieved for the medium- and high-density populations with minimal effort of one survey, and for the low-density population with three survey repeats (Figure 3b).

It should be pointed out that for the presented survey simulations, the distance analysis was in fact not quite adequate, as the generated detection data were independent of the distance to the transect line. Nevertheless, the purpose of using it was to demonstrate how a non-optimal choice of data analysis method affects the quality of the resulting estimates. The distance sampling method has a long history of use in traditional aerial surveys, where detections are made by a human observer and the detection probability drops with increasing distance from the transect line. The method is robust and has many sophisticated variants suitable for different scenarios (inclusion of covariates in the detection probability function, double observers, etc.) [23,24], and well-developed, free software for survey planning and analysis is available [16,25]. Thus, distance sampling is often a method of first choice for researchers analyzing data from aerial wildlife surveys, including surveys using thermal infrared sensing [26–28]. However, during such surveys, presence of the distance effect in detections underlying the method may depend on factors such as the camera mounting angle (vertical, forward, oblique), type and quality of the camera lenses, or method of image reviewing (i.e., visual or automatic). For example, in research using a vertically looking camera for deer surveys by Kissell and Nimmo [26], uniform detection models gave the best fit to the data, which suggests no or a very limited impact of distance on detection probability. Therefore, before choosing an analysis method, it is important to find convincing support for the distance-dependence of the detection probability in the approach used for data acquisition and treatment. If there is strong support for the presence of the distance effect, then the use of distance sampling is justified; otherwise, an alternative estimation method should be adopted.

In the simulations, we used several simplifications—we ignored the fact that ungulates usually occur in groups of various sizes, and we assumed that the true availability for detection is known. Such simplifications may lead to the over-optimistic evaluation of survey performance; however, they were used here to reduce the variability of density estimates due to factors other than survey design. Since the simplifications affected both of the compared designs in the same way, the differences in accuracy and precision caused solely by the design variant were made more distinct. In practice, the availability of animals for detection is usually unknown, unless surveys are conducted over open areas and at the times of their activity peaks. Not adjusting for imperfect availability for detection is an important source of bias; thus, the reliable assessment of the availability value is crucial to obtain accurate estimates of population density.

Simplifications used here must be avoided when simulations are used to assess the minimum effort—the total length of transects or number of survey repeats—needed to achieve a given level of precision in a survey of a specific population. In such cases, an attempt should be made to simulate the population in focus and survey conditions as closely as possible and account for all sources of possible errors [29]. Data available from previous surveys can be used to set a possible range of population densities and mean group sizes, while the remotely sensed data on land cover and tree-crown density may serve to estimate reasonable values of availability for detection. Simulations using realistic system parameters will ensure the selection of the best survey variants to achieve reliable estimates while minimizing survey costs [30].

5. Conclusions

Wildlife surveys with the use of drones are usually constrained by regulations limiting drone operations to visual line of sight. While the regulations complicate the logistics of surveys and may interfere with random survey design, the accuracy and precision of the derived population density estimates do not have to be compromised. Our results show that, in general, VLOS surveys can provide estimates of similar quality to those from spatially unrestricted randomized designs under BVLOS operations. However, in low-density populations, surveys under VLOS design are prone to bias if the applied survey effort is too small to provide adequate sample sizes. In such cases, increasing effort by repeating surveys along the same flight paths allows the reduction of the sampling variance and achieves unbiased estimates of population density. Comparing the precision of the two estimation methods used in the study, we observed that distance sampling gave far less precise results than the ratio estimate. Thus, unless there is strong evidence that the detection probability in drone surveys depends on distance, other estimation methods than distance sampling should be adopted. To ensure accurate and precise density estimates while minimizing survey costs, we recommend testing alternative survey designs with the aid of computer simulations.

Supplementary Materials: The following are available online at https://www.mdpi.com/2072-4292/13/6/1102/s1, Table S1: The details of ungulate thermal drone surveys conducted in hunting district nr 473 in Krotoszyn Forest District, Poland.

Author Contributions: Conceptualization, J.W. and S.P.; Formal analysis, J.W. and S.P.; Funding acquisition, J.W. and S.P.; Investigation, J.W. and S.P.; Methodology, J.W. and S.P.; Project administration, J.W.; Visualization, S.P.; Writing—original draft, J.W. and S.P.; Writing—review & editing, J.W. and S.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Narodowe Centrum Badań i Rozwoju (National Centre for Research and Development, Poland), grant number PBS3/B8/18/2015.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We are grateful to two anonymous reviewers for their helpful comments on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Putman, R.; Apollonio, M.; Andersen, R. (Eds.) *Ungulate Management in Europe: Problems and Practices*; Cambridge University Press: Cambridge, UK, 2011; p. x + 398.
2. Morellet, N.; Klein, F.; Solberg, E.; Andersen, R. The census and management of populations of ungulates in Europe. In *Ungulate Management in Europe: Problems and Practices*; Apollonio, M., Andersen, R., Putman, R., Eds.; Cambridge University Press: Cambridge, UK, 2011; pp. 106–143.
3. Ikeda, T.; Takahashi, H.; Igota, H.; Matsuura, Y.; Azumaya, M.; Yoshida, T.; Kaji, K. Effects of culling intensity on diel and seasonal activity patterns of sika deer (Cervus nippon). *Sci. Rep.* **2019**, *9*, 17205. [CrossRef] [PubMed]
4. Ogurtsov, S.S.; Zheltukhin, A.S.; Kotlov, I.P. Daily activity patterns of large and medium-sized mammals based on camera traps data in the Central Forest Nature Reserve, Valdai Upland, Russia. *Nat. Conserv. Res.* **2018**, *3*, 68–88. [CrossRef]
5. Stache, A.; Heller, E.; Hothorn, T.; Heurich, M. Activity patterns of European roe deer (Capreolus capreolus) are strongly influenced by individual behaviour. *Folia Zool.* **2013**, *62*, 67–75. [CrossRef]
6. Beaver, J.T.; Baldwin, R.W.; Messinger, M.; Newbolt, C.H.; Ditchkoff, S.S.; Silman, M.R. Evaluating the use of drones equipped with thermal sensors as an effective method for estimating wildlife. *Wildl. Soc. Bull.* **2020**, *44*, 434–443. [CrossRef]
7. Corcoran, E.; Denman, S.; Hanger, J.; Wilson, B.; Hamilton, G. Automated detection of koalas using low-level aerial surveillance and machine learning. *Sci. Rep.* **2019**, *9*, 3208. [CrossRef] [PubMed]
8. Witczuk, J.; Pagacz, S.; Zmarz, A.; Cypel, M. Exploring the feasibility of unmanned aerial vehicles and thermal imaging for ungulate surveys in forests—preliminary results. *Int. J. Remote Sens.* **2018**, *39*, 5504–5521. [CrossRef]

9. European Union Aviation Safety Agency. Easy Access Rules for Unmanned Aircraft Systems (Regulations (EU) 2019/947 and (EU) 2019/945). Available online: https://www.easa.europa.eu/document-library/easy-access-rules/online-publications/easy-access-rules-unmanned-aircraft-systems (accessed on 22 February 2021).
10. Witczuk, J.; Pagacz, S. Wpływ cech populacji i środowiska na dokładność i precyzję wyników symulacji lotniczej inwentaryzacji zwierzyny [Effects of population and habitat characteristics on the accuracy and precision of wildlife aerial surveys results]. *Sylwan* **2020**, *164*, 560–567. [CrossRef]
11. Bobek, B.; Merta, D.; Furtek, J.; Wojciuch-Płoskonka, M.; Kopeć, K.; Maślanka, J.; Ziobrowski, M. Ocena dynamiki liczebności i zagęszczenia populacji dzikich kopytnych przy użyciu różnych metod w czterech regionach Polski [Population dynamics of wild ungulates in various regions of Poland estimated by different methods]. *Studia I Mater. Cepl.* **2013**, *15*, 88–101.
12. European Union. Copernicus Land Monitoring Service, European Environment Agency (EEA). Available online: https://land.copernicus.eu/pan-european/high-resolution-layers/forests/tree-cover-density (accessed on 22 February 2021).
13. Krebs, C.J. *Ecological Methodology*; Benjamin Cummings: Menlo Park, CA, USA, 1999; p. xii + 620.
14. Pierce, B.L.; Lopez, R.R.; Silvy, N.J. Estimating animal abundance. In *The Wildlife Techniques Manual Research*; Silvy, N.J., Ed.; The Johns Hopkins University Press: Baltimore, MD, USA, 2012; Volume 1, pp. 284–310.
15. Buckland, S.T.; Anderson, D.R.; Burnham, K.P.; Laake, J.L.; Borchers, D.L.; Thomas, L. *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*; Oxford University Press: Oxford, UK, 2001; p. xv + 432.
16. Miller, D.L.; Rexstad, E.; Thomas, L.; Marshall, L.; Laake, J.L. Distance sampling in R. *J. Stat. Softw.* **2019**, *89*, 1–28. [CrossRef]
17. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
18. Marshall, L. DSsim: Distance Sampling Simulations. Available online: https://CRAN.R-project.org/package=DSsim (accessed on 16 November 2020).
19. Anderson, K.; Gaston, K.J. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Front. Ecol. Env.* **2013**, *11*, 138–146. [CrossRef]
20. Borchers, D.L.; Buckland, S.T.; Zucchini, W. *Estimating Animal Abundance: Closed Populations*; Springer: London, UK, 2002; p. xiii + 314.
21. Hedley, S.L.; Buckland, S.T. Spatial models for line transect sampling. *J. Agric. Biol. Environ. Stat.* **2004**, *9*, 181. [CrossRef]
22. Skalski, J.R.; Ryding, K.E.; Millspaugh, J.J. Estimating population abundance. In *Wildlife Demography*; Skalski, J.R., Ryding, K.E., Millspaugh, J.J., Eds.; Academic Press: Burlington, NJ, USA, 2005; pp. 435–539.
23. Buckland, S.T.; Anderson, D.R.; Burnham, K.P.; Laake, J.L.; Borchers, D.L.; Thomas, L. *Advanced Distance Sampling. Estimating Abundance of Biological Populations*; Oxford University Press: Oxford, UK, 2004; pp. xvii + 416.
24. Buckland, S.T.; Rexstad, E.A.; Marques, T.A.; Oedekoven, C.S. *Distance Sampling: Methods and Applications*; Springer International Publishing: Cham, Switzerland, 2015; p. 292.
25. Thomas, L.; Buckland, S.T.; Rexstad, E.A.; Laake, J.L.; Strindberg, S.; Hedley, S.L.; Bishop, J.R.B.; Marques, T.A.; Burnham, K.P. Distance software: Design and analysis of distance sampling surveys for estimating population size. *J. Appl. Ecol.* **2010**, *47*, 5–14. [CrossRef]
26. Kissell, R.E.; Nimmo, S.K. A technique to estimate white-tailed deer Odocoileus virginianus density using vertical-looking infrared imagery. *Wildl. Biol.* **2011**, *17*, 85–92. [CrossRef]
27. Lethbridge, M.; Stead, M.; Wells, C. Estimating kangaroo density by aerial survey: A comparison of thermal cameras with human observers. *Wildl. Res.* **2019**, *46*, 639–648. [CrossRef]
28. Schoenecker, K.A.; Doherty, P.F., Jr.; Hourt, J.S.; Romero, J.P. Testing infrared camera surveys and distance analyses to estimate feral horse abundance in a known population. *Wildl. Soc. Bull.* **2018**, *42*, 452–459. [CrossRef]
29. Brack, I.V.; Kindel, A.; Oliveira, L.F.B. Detection errors in wildlife abundance estimates from Unmanned Aerial Systems (UAS) surveys: Synthesis, solutions, and challenges. *Methods Ecol. Evol.* **2018**, *9*, 1864–1873. [CrossRef]
30. Conn, P.B.; Moreland, E.E.; Regehr, E.V.; Richmond, E.L.; Cameron, M.F.; Boveng, P.L. Using simulation to evaluate wildlife survey designs: Polar bears and seals in the Chukchi Sea. *R. Soc. Open Sci.* **2016**, *3*, 150561. [CrossRef] [PubMed]

scientific reports

Fusion of visible and thermal images improves automated detection and classification of animals for drone surveys

B. Santhana Krishnan[1], Landon R. Jones[2], Jared A. Elmore[2,3], Sathishkumar Samiappan[1], Kristine O. Evans[2], Morgan B. Pfeiffer[4], Bradley F. Blackwell[4] & Raymond B. Iglay[2]✉

Visible and thermal images acquired from drones (unoccupied aircraft systems) have substantially improved animal monitoring. Combining complementary information from both image types provides a powerful approach for automating detection and classification of multiple animal species to augment drone surveys. We compared eight image fusion methods using thermal and visible drone images combined with two supervised deep learning models, to evaluate the detection and classification of white-tailed deer (*Odocoileus virginianus*), domestic cow (*Bos taurus*), and domestic horse (*Equus caballus*). We classified visible and thermal images separately and compared them with the results of image fusion. Fused images provided minimal improvement for cows and horses compared to visible images alone, likely because the size, shape, and color of these species made them conspicuous against the background. For white-tailed deer, which were typically cryptic against their backgrounds and often in shadows in visible images, the added information from thermal images improved detection and classification in fusion methods from 15 to 85%. Our results suggest that image fusion is ideal for surveying animals inconspicuous from their backgrounds, and our approach uses few image pairs to train compared to typical machine-learning methods. We discuss computational and field considerations to improve drone surveys using our fusion approach.

Drones (small unoccupied aircraft systems or UAS) are increasingly used for monitoring animals, offering multiple advantages, including time or cost savings, increased safety over occupied aircraft, and more accurate counts than traditional ground-based methods[1–4]. Drones can also quickly collect large amounts of data at fine spatial, spectral, and temporal resolutions. Visible (e.g., red, 650 nm; green, 550 nm; blue, 450 nm) or thermal (7.5–14 µm) cameras, yield image or video data that can be used to detect and classify animals either manually or autonomously by computers[5–8]. Human detection (i.e., finding an animal) or classification (i.e., identifying an animal) can be tedious, costly, and error-prone leading to lower detection rates and misclassification errors[5,9,10]. Some biologists have used crowd sourcing or citizen science efforts to manually detect and classify animals in images[11,12], while others are turning to automated detection and classification through machine learning, specifically deep learning methods like convolutional neural networks (CNN) and computer vision[6,13–15].

Automated detection and classification have been found to be more accurate and time efficient than human detection and classification in aerial images[5,7,9], including citizen science approaches[10,14,16]. Recent work has focused on deep learning methods such as CNN to detect and classify animals in images[13–16]. However, detection and classification can often be difficult, not only because of the absence of prominent distinguishing features[13], but also uncontrollable factors such as obstruction from overhead vegetation or neighboring animals[6,17], confusion between animals and associated ghost images created from the mosaicking process[14], or a lack of contrast between animals of interest and their background (e.g., cryptic in visible imagery or homogenous temperature in thermal imagery; reviewed in[6]).

High success or accuracy of machine learning in computer vision stems from the availability of substantial of labelledimages[18]. Image labelling or annotation is the process of marking areas in an image (usually with

[1]Geosystems Research Institute, Mississippi State University, Mississippi State, Mississippi State, MS 39762, USA. [2]Department of Wildlife, Fisheries, and Aquaculture, Mississippi State University, Box 9690, Mississippi State, MS 39762, USA. [3]Department of Forestry and Environmental Conservation, Clemson University, Clemson, SC 29634, USA. [4]U.S. Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, National Wildlife Research Center, Ohio Field Station, Sandusky, OH 44870, USA. ✉email:

a rectangular box, referred to as a 'bounding box') with class labels such as animal species. However, large, open access databases of annotated animal images from aerial perspectives are lacking to train computer vision algorithms to detect and classify animals in drone images. To our knowledge, primary available databases are those associated with single studies, which often limit the diversity of species, environments, animal poses, and background and color variability surrounding animals captured[14]. In this low-sample learning scenario, typical image augmentation techniques (e.g., rotation, scaling, etc.) often do not account for texture variability in the object and background[19]. Meanwhile, computer vision algorithms are tasked with evaluating entire drone images, not only the cropped regions, which only contain one animal each. Further, unlike camera trap images[20], the background is constantly changing among drone images, which makes learning the animal features among various backgrounds critical for efficient performance of animal detection and classification in drone images, whether manually or with computer vision.

Combining information from multiple sensors (e.g., visible and thermal images) offers another approach to improve the distinguishability of an animal from the background[21]. Image fusion is the process of combining corresponding image information on the scale of each pixel or group or pixels from multiple image modalities (e.g., visible and thermal images) to generate a single image containing more information than either source image alone[22]. Processing the 'fused' image instead of the individual visible or thermal images has shown improved performance among multiple computer vision problems including automated detection and classification in terrestrial imagery[21,23]. Unlike deep learning engines, which use only visible imagery to achieve similar results[24], large quantities of correctly annotated data and ample training resources are often not required for fusion methods.

Fusion of thermal and visible images has been used for a variety of applications, including autonomous driving (especially in low light situations), surveillance, defect identification, electronic testing, medical imaging, and remote sensing[25]. Fortuitously, many newer drone models and associated imaging sensors are equipped with dual thermal/visible cameras capable of collecting both image types simultaneously (e.g., DJI Zenmuse XT2). To date, however, image fusion has only been tested in four studies involving animals, including one study identifying animals posing hazards to autonomously driven vehicles[26], and another to identify livestock from unoccupied ground vehicle[27]. Two additional studies pioneered fusion approaches to identify animal species from drone images based on combining visible and thermal data to detect captive white-tailed deer (*Odocoileus virginianus*[17]) and a few individuals of four species in zoo enclosures[28]. However, large advances in both commercially available drone sensors and computer vision approaches since these studies provide opportunities to improve on their methodology and results.

Fusion of visible and thermal information in drone imagery to automatically detect and classify animals is a promising yet relatively untested avenue for improving the efficiency of drone surveys, particularly when few images are available for training machine learning algorithms[6]. We evaluated the performance of image fusion of thermal and visible information in drone imagery for three animal species: white-tailed deer, domestic cow (*Bos taurus*), and domestic horse (*Equus caballus*). We compared performance metrics of eight image fusion methods in two deep learning classification networks to automatic classification of test species using visible and thermal images alone. Finally, we discuss computational and field considerations in using our fusion approach to maximize the information gained from drone surveys that could be scaled up across a range of animal species and conditions.

Methods

Study area. We collected study images among research facilities located at Mississippi State University, USA in 2021 and 2022 (Supplementary Fig. S1 online). We used deer enclosures on the Forest and Wildlife Research Center (33.439 N, −88.791 W) and paddocks on the H. H. Leveck Animal Research Center (33.436 N, −88.797 W), which is part of the Mississippi Agricultural and Forestry Experiment Station.

Drone data collection. We captured images of white-tailed deer, domestic cattle, and domestic horse (hereafter deer, cow, and horse, respectively), during diurnal hours using a DJI Zenmuse XT2 (8-mm visible and 25-mm thermal lenses) mounted on a DJI Matrice 200 V2 quadcopter (SZ DJI Technology Co., Ltd., Shenzen, China). Flights were conducted by a Part 107 certified remote pilot (FAA 2016) through the DJI Pilot app on a Samsung T500 tablet (Samsung Electronics America, Inc., Ridgefield Park, New Jersey, USA) with the sensor in nadir position (i.e., 90° or straight down). We used either autonomous flights with a lawnmower pattern with > 50% overlap, or conducted manual flights, at 30–121 m altitude above ground level (6.9–28.4 mm Ground Sampling Distance) to simultaneously collect visible and thermal images during missions associated with other UAS efforts. Collected images were stored in the open-source Aerial Wildlife Image Repository-AWIR (https://projectportal.gri.msstate.edu/awir/). Methods were approved by NWRC IACUC Number QA-3267 and MSU IACUC (i.e., methods reviewed but no protocol necessary), and we followed all relevant guidelines and regulations for data collection.

Image processing. Input data totaled 164 images, including 68 images with 265 cows, 53 images with 77 deer, and 43 images with 136 horses (Table 1). From collected images, we first identified image pairs in which one or more animals were present. To maximize variation of animals in our dataset for training fusion methods, we omitted sequential images of the same animal without pose variation in the same series of images on the same day. Second, we annotated images by manually drawing bounding boxes around each animal object and labelling them to species. Finally, because thermal (640 × 512 pixels) and visible (4000 × 3000 pixels) images were of different sizes, we aligned the smaller thermal images by upscaling and translating them with the corresponding region of the larger visible image using image registration procedures[29–31] (see Supplementary Information Sect. 1.1 online) to obtain a final pixel size of 1792 × 1434 pixels for all images.

Category	Cow	Deer	Horse
Training images	51	38	28
Training objects	218	61	88
Validation images	4	5	5
Validation objects	16	5	18
Test images	13	10	10
Test objects	31	11	30

Table 1. Numbers of images (Images) and individual animal objects (Objects) within images used for training, validation, and testing fusion methods for automated classification of domestic cattle (*Bos taurus*), white-tailed deer (*Odocoileus virginianus*), and domestic horses (*Equus caballus*) from images taken by a drone (unoccupied aircraft system or UAS).

Image fusion. After respective pairs of visible and thermal images were acquired and registered, their information was combined through fusion before splitting the dataset for training and classification (Fig. 1a, see Supplementary Information Sect. 1.2 online). The image fusion portion of our process followed three general steps for visible and thermal images in each pair: (1) transform both images to a different feature space, (2) merge the information from both images to create the fused image in the transformed feature space, and (3) reconstruct the fused image by an inverse transform of merged information. For some fusion approaches (optimization-based), transforms during the first step were not applied and only the second step occurred.

To compare the performance of fusion methods to visible or thermal images to detect and classify animals, we tested eight different image fusion methods: four multi-resolution-based approaches, two optimization-based approaches, and two hybrid approaches. We evaluated the following multi-resolution-based approaches: (1) guided filter, (2) Laplacian/Gaussian pyramid (LP), (3) singular-value decomposition (SVD), and (4) sparse representation (sparse; Fig. 2). Multi-resolution approaches transform the original image in multiple scales, where the amount (resolution or number of pixels) and type (approximation, detail) of information differs in each scale (Fig. 1b). Image fusion was then performed in each corresponding scale in the transform space. We also used two optimization-based fusion methods, Gradient and total variation distance (TVM). Optimization-based approaches conduct fusion of visible and thermal images at the pixel level to optimize a chosen criterion without image transformation. Finally, we used two hybrid fusion methods, a wavelet (WL) plus TVM hybrid approach (WL + TVM) and a WL plus swarm hybrid approach (WL + Swarm). Hybrid approaches first transform the image to a multi-resolution representation and then fuse in the transform space based on an optimization criterion, combining some aspects of both multi-resolution and optimization-based approaches.

Object detection and classification. You Only Look Once (YOLO) is a popular deep learning-based object detection architecture. YOLO's key idea is to frame object detection as a regression problem, thus predicting bounding boxes and confidence probabilities in a single pass of the image through the neural network. This one-shot algorithm excels both in accuracy and speed. By considering multiple scales and aspect ratios, YOLO can handle objects of various sizes and shapes efficiently. YOLO has undergone several iterations with latest being YOLOv8. The newer YOLOv7 also provides focal loss, ideal for identifying small objects but also computationally intense compared to YOLOv5. After image fusion, we used YOLOv5[32,33] and YOLOv7[34] to automatically detect and classify objects. Objects evaluated in this study were annotated areas in images. Objects included animals (i.e., animal objects) or non-animals such as annotations by us or incorrect annotations by YOLO architectures (i.e., false positives; see Evaluation criteria for more information). We annotated all animal objects in our image dataset with ground-truth bounding boxes.

We used an approximate 70–10–20% split of images for training–validating-testing classification architectures among species. The same training–validating-testing data were used between classification networks to allow for cross-comparison with our annotated animal objects. However, in the testing procedure, we provided full images without annotations, which often contained multiple animal objects. The trained architectures then created bounding boxes around objects detected as animals and provided the classification of each object in the output. Both architectures were trained and tested on Google Colab Pro using GPU acceleration, using at least 100 Intel Xeon CPUs with a frequency of 2.30 GHz, allocating an average 38 GB of GPU RAM. Because most drone images are larger (our final images were 1792 × 1434 pixels) to cover a large field of view for survey and other applications than typical segmented images processed by these architectures (256 × 256 pixels), our larger image sizes and different network architectures on the same computing resources permitted a maximum batch size of 12 for YOLOv7 and 16 for YOLOv5, using 135 and 100 epochs, respectively. Additional details are available in Supplementary Information Sect. 1.3 online.

Evaluation criteria. We evaluated the performance of fusion methods based on metrics of (1) animal object quality and (2) classification accuracy (additional details available in Supplementary Information Sect. 1.4 online). We used our annotated animal objects in our test image dataset to evaluate animal object quality metrics of entropy, mutual information, and a gradient-based Petrovic metric. Entropy is the average number of bits per pixel needed to represent an image region[35], or the animal object within a bounding box for our purposes. A higher value of entropy implies a larger amount of information in the image region, which typically improves differentiation of animal objects from their respective backgrounds. Mutual information (bits per pixel) repre-

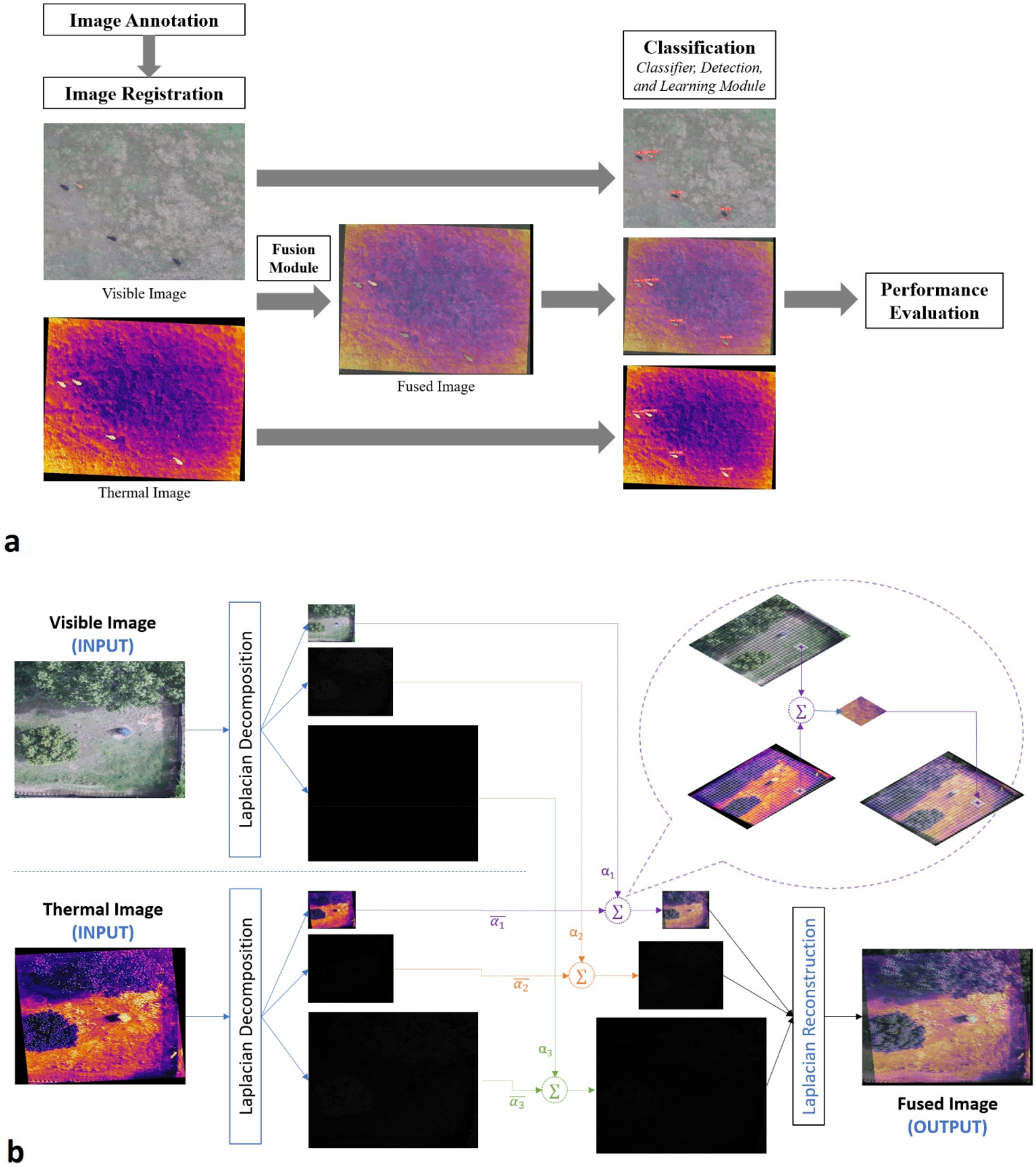


Figure 1. Workflow for fusion of thermal and visible images for learning-based animal object detection and classification from drone (unoccupied aircraft system or UAS) imagery (**a**) and a depiction of the Lalacian fusion algorithm used showing the layer coefficients and block wise coefficients combining in the approximation layer as well as the reconstructed image (**b**).

sents the amount of information transferred from an individual image (visual or thermal) to the fused image. The two values of mutual information from the visual and thermal images were summed; higher values were preferred and indicated that a larger amount of useful information was transferred to the fused image compared to lower values[35]. The gradient-based Petrovic metric is a unitless measure of edge preservation ranging from 0 to 1[36,37]. Values closer to 1 indicated higher preservation of edge information compared to values closer to 0 because the visual perception of an object is first based on identifying its edges. Thus, details in the pixels at the edges of an object contain most of the information comprising its shape compared to middle regions[22], as is the case for animals in our drone images. To visualize patterns and compare the performance of fusion metrics, we created plots of all three metrics of image quality for each animal object (entropy) or object pair (mutual infor-

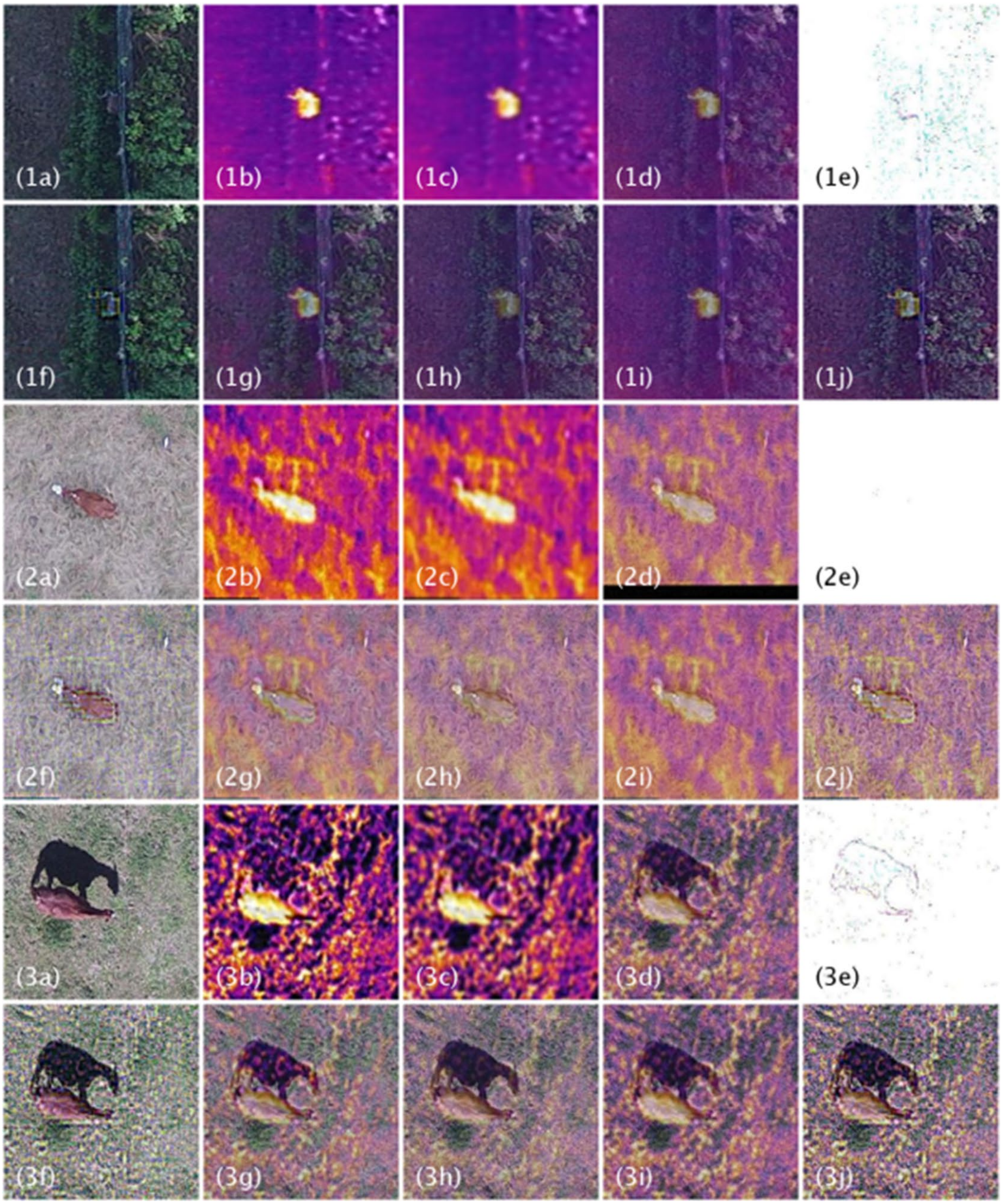


Figure 2. Comparison of aerial imagery captured by drone (unoccupied aircraft system or UAS) containing white-tailed deer (*Odocoileus virginianus*; series 1), cow (*Bos taurus*; series 2), and horse (*Equus caballus*; series 3) among visible (**a**) and thermal (b) images and eight fusion methods: guided filter (**c**), Laplacian (**d**), SVD (**e**), sparse (**f**), gradient (**g**), TVM (**h**), WL + Swarm (**i**), and WL + TVM (**j**).

mation, Petrovic metric) in our dataset. Because these metrics are specific to their respective backgrounds within bounding boxes, we represented them as values for individual animal objects and did not average them across fusion methods for comparison.

Metrics of classification accuracy were computed based on comparing classification of animal objects in images without bounding boxes by architectures after training to original images containing bounding boxes that were manually drawn (ground truth) during image processing. We computed precision, recall, and mean average precision (mAP50; an additional measure of accuracy) as performance metrics of classification. Mean average precision (mAP) measured the correctness of animal detection (i.e., bounding box around animal object) and animal classification (i.e., species identification) for objects annotated by architectures in the test image dataset. Greater mAP values indicate greater model accuracy in animal detection and classification. For mAP50, a 50% threshold was considered for intersection over union (i.e., the overlap or intersection of predicted boundaries and actual animal boundaries; more information available in Supplementary Information Sect. 1.4). These metrics rely on three scenarios of correct or incorrect detection and classification to define[38]. Correct detections (draws a bounding box around the animal object) and classifications of target species (deer, cow, horse) by respective models (combinations of fusion methods and classification architectures) are defined as true positives. False positives occur when the respective model correctly detects (draws a bounding box around an object) but incorrectly classifies that object in an image as the target species, such as a different species (Fig. 3a) or inanimate object (Fig. 3b). False negatives occur when the respective model does not detect an individual of the target species when it occurs in an image (Fig. 3b). Accordingly, precision measures the proportion of true compared to false positives that the model correctly predicted, calculated as:

$$Precision = \frac{true\ positives}{true\ positives\ +\ false\ positives} \tag{1}$$

Recall measures the proportion of true positives compared to false negatives that the model correctly predicted, calculated as:

$$Recall = \frac{true\ positives}{true\ positives\ +\ false\ negatives} \tag{2}$$

Typically, overall accuracy considers false positives and negatives and is often defined as true positives divided by the sum of true positives, false positives, and false negatives (e.g.[14]). Similarly, we evaluated accuracy by comparing (1) human-drawn bounding boxes containing correctly classified animals (ground truth) to (2) objects in a bounding box automatically drawn and classified from respective model output (predicted). However, models did not necessarily draw the bounding box to entirely encompass the animal. Thus, models required a threshold of the number of overlapping pixels to evaluate if the bounding box adequately captured the animal object compared to the manually drawn, correct classifications (e.g., 60%, 80%, overlap with true positives; see Supplementary Information Sect. 1.4 online for additional information regarding intersection over union). Accordingly, we used mAP50 (mean average precision with a threshold overlap of 50%) as an alternative but accepted metric of accuracy, because it accounted for precision and recall while computing an average value for the overlap of predicted and ground truth bounding boxes for a range of values[38]. To further evaluate fusion methods and compare them to visible and thermal results alone, we ranked results for each fusion method with visible and thermal results using mAP50 for each animal species and classification architecture. We then summed the rank scores (1–10) among species and architectures, using the lowest score to determine the best performing fusion methods in context of visible and thermal results.

Results

Between architectures, YOLOv5 (Table 2) outperformed YOLOv7 (Table 3) overall among metrics and animals in visible and thermal images, as well as among fusion methods. Although mAP50 for YOLOv7 was poor overall, we report results for both architectures to demonstrate that some fusion methods provided improvement for animal classification beyond visible and thermal results alone.

Object quality. Metrics of image quality for entropy indicated that Sparse and WL + TVM consistently provided more information than visual (Fig. 4a) and thermal (Fig. 4b) alone, indicating these fusion methods better characterized the animal object compared to the background than unfused images or other fusion methods. For cows and horses, Sparse had the highest entropy values in 54.5–75.0% of animal objects, respectively, compared to 18.8–22.7% of animal objects for WL + TVM. For deer, in contrast, WL + TVM had the highest entropy values (70.0% of deer objects) compared to Sparse (20.0%). Sparse, WL + TVM, and Guided filter had consistently high values for mutual information (Fig. 4c), indicating they transferred more information from visible and thermal

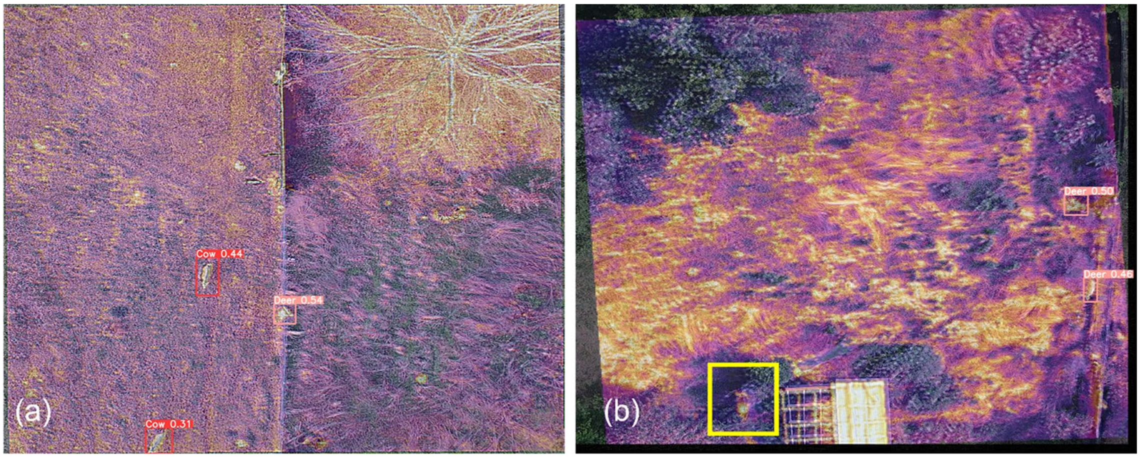


Figure 3. Examples of false positives regarding misclassification (**a**,**b**), and false negatives as non-detection or target animal (**b**). Two of three white-tailed deer (*Odocoileus* virginianus) were misclassified as cows (*Bos taurus*, **a**), and two hay bales were classified as cows (**b**, red boxes). The false negative occurred when the target animal (white-tailed deer, yellow box) was not detected (**b**).

Class	Classification accuracy										
	Metric	Visible	Thermal	GF	LP	SVD	Sparse	Gradient	TVM	WL + Swarm	WL + TVM
Cow	Precision	0.88	0.73	0.79	0.85	0.58	0.87	0.77	0.81	0.84	0.72
	Recall	0.75	0.75	0.75	0.80	0.14	0.80	0.66	0.85	0.84	0.65
	mAP50	0.89	0.77	0.83	0.88	0.16	0.93	0.69	0.84	0.86	0.74
Deer	Precision	0.72	0.81	1.00	0.74	1.00	0.72	0.77	0.61	0.73	0.52
	Recall	0.44	0.56	0.44	0.63	0.00	0.67	0.56	0.86	0.72	0.78
	mAP50	0.63	0.64	0.56	0.66	0.01	0.72	0.69	0.77	0.70	0.62
Horse	Precision	0.95	0.93	0.89	0.93	0.67	1.00	0.92	0.87	0.90	0.93
	Recall	0.93	0.92	1.00	0.93	0.87	0.93	0.77	0.87	0.87	0.88
	mAP50	0.99	0.97	1.00	0.95	0.89	0.99	0.95	0.95	0.95	0.96

Table 2. Classification accuracy metrics of domestic cattle (*Bos taurus*), white-tailed deer (*Odocoileus virginianus*), and domestic horses (*Equus caballus*) for visible, thermal, and eight fusion methods for YOLOv5 learning module from images taken by a drone (unoccupied aircraft system or UAS). *GF* guided filter method, *LP* Laplacian method.

Class	Classification accuracy										
	Metric	Visible	Thermal	GF	LP	SVD	Sparse	Gradient	TVM	WL + Swarm	WL + TVM
Cow	Precision	0.26	0.35	0.53	0.38	0.16	0.54	0.41	0.00	0.32	0.37
	Recall	0.45	0.65	0.62	0.45	0.25	0.75	0.70	0.00	0.54	0.80
	mAP50	0.26	0.48	0.49	0.37	0.10	0.59	0.55	0.00	0.47	0.60
Deer	Precision	1.00	0.55	0.99	0.78	0.00	0.68	0.52	1.00	0.67	0.33
	Recall	0.00	0.55	0.44	0.77	0.00	0.33	0.67	0.00	0.57	0.50
	mAP50	0.00	0.51	0.50	0.85	0.00	0.37	0.62	0.00	0.44	0.46
Horse	Precision	0.26	1.00	0.46	0.50	0.00	0.41	0.54	0.80	1.00	0.70
	Recall	0.87	0.27	0.47	0.40	0.67	0.80	0.47	0.53	0.27	0.47
	mAP50	0.53	0.47	0.44	0.44	0.24	0.64	0.54	0.52	0.50	0.54

Table 3. Classification accuracy metrics of domestic cattle (*Bos taurus*), white-tailed deer (*Odocoileus virginianus*), and domestic horses (*Equus caballus*) for visible, thermal, and eight fusion methods for YOLOv7 learning module from images taken by a drone (unoccupied aircraft system or UAS). *GF* guided filter method, *LP* Laplacian method.

images to fused images. Like entropy results, Sparse had the greatest values of mutual information for cows and horses (54.5–75.0% of animal objects, respectively) versus deer (20.0%) compared to WL + TVM (cows,18.8%; horses, 22.7%; deer, 80.0%). The Guided filter method performed best for mutual information in 6.8% of cow objects, 9.4% of horses, and no deer. Results for the Petrovic metric were approximately the same for all three animal species, and all but the SVD method performed similarly well (Fig. 4d), indicating consistent preservation of edge information of animal objects among fusion methods. The following four fusion methods had the greatest values for the Petrovic metric for all three animal species: TVM (34.4–40.0% of animal objects), Guided filter (25.0–27.3%), WL + TVM (15.0–15.9%) and Sparse (9.1–15.6%).

Object detection and classification. Classification accuracy (mAP50) of cows for YOLOv5 was 16% higher in visible compared to thermal images (Table 2). Only Sparse improved overall accuracy beyond the performance for visible images alone (Table 2). In contrast, for YOLOv7, classification was about 85% better in thermal compared to visible images (Table 3). Classification improved over thermal images with three fusion methods (WL + TVM, 25%; Sparse, 23%; Gradient, 15%), and minimally with Guided filter (2%, Table 3).

For deer, classification accuracy was similar for visible and thermal images with YOLOv5 (i.e., < 2% increase from visible to thermal; Table 2). Four fusion methods provided considerable improvement (TVM, 21%; Sparse, 13%; WL + Swarm, 10%; Gradient, 9%), and one method (LP, 4%) provided minimal improvement compared to visible and thermal results (Table 2). For YOLOv7, visible was never accurate but increased fivefold for thermal (Table 3). Fusion improved classification of deer substantially over thermal results with the LP (67%) and Gradient (13%) methods (Table 3).

Classification accuracy of horses was near 100% for both visible and thermal images for YOLOv5 (Table 2). Among fusion methods, only Guided filter improved results beyond visible results (Table 2), although the gain was minimal (1% for visible, 3% for visible). For YOLOv7, visible improved accuracy 13% compared to thermal (Table 3). Among fusion methods, Sparse provided substantial gains in overall accuracy (21% improvement to visible), whereas Gradient and WL + TVM provided minimal (< 2%) gains compared to visible results (Table 3).

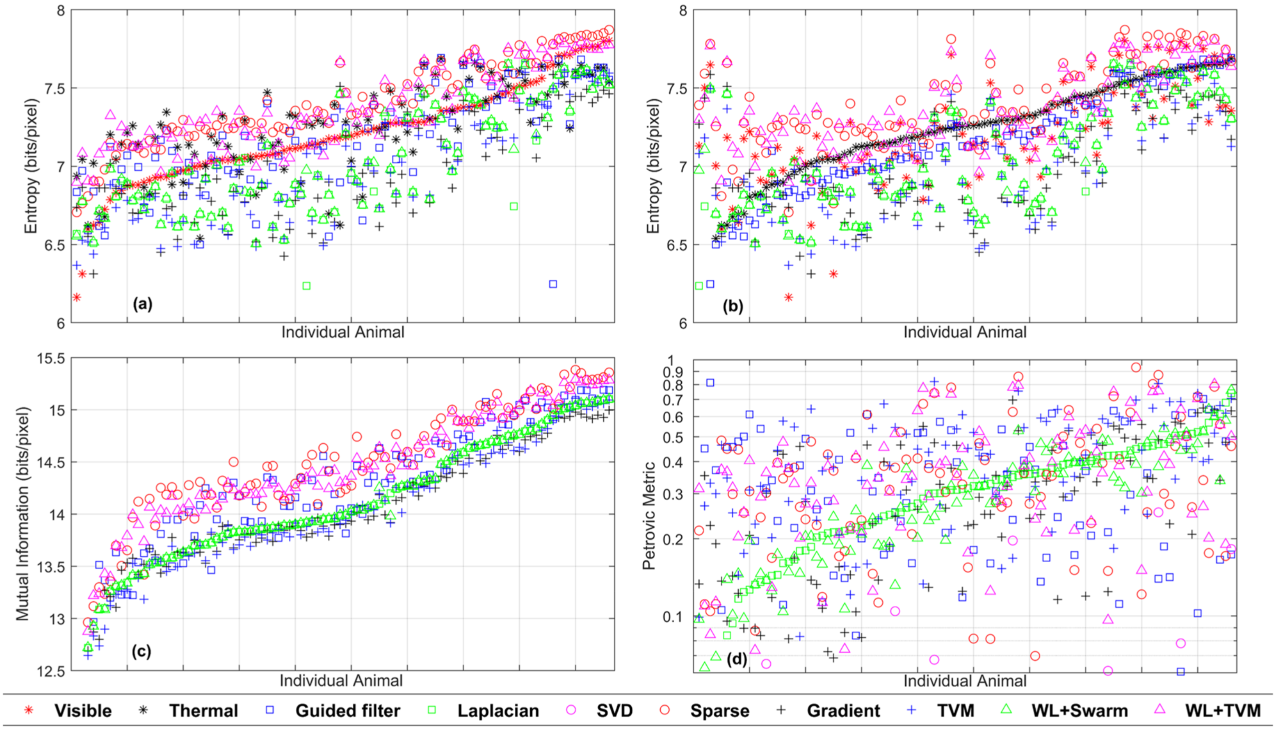


Figure 4. Plots of four metrics of image quality for 96 animal objects of domestic cattle (*Bos taurus*, cow), white-tailed deer (*Odocoileus virginianus*), and domestic horses (*Equus caballus*) in drone (unoccupied aircraft system or UAS) images automatically detected and classified using visible, thermal, and fused images produced by eight fusion methods. Animal objects on the x-axis do not correspond to the same individuals in each plot and are ordered in each plot to aid in interpreting the relative performance of fusion methods. Entropy values (bits/pixel), a measure of object information compared to the background, are ordered from descending to ascending values based on visible values (**a**) and thermal values (**b**). Values for mutual information (bits/pixel), a measure of the amount of information transferred from an individual image (visual or thermal) to the fused image (**c**), and the Petrovic metric (**d**), a measure of edge-preservation from 0 to 1 (unitless, shown on log scale), are ordered from descending to ascending values based on the values of a middle-ranked fusion method for this metric (Laplacian) and do not contain values for visible or thermal images.

	YOLOv5			YOLOv7		
Rank	**Cow**	**Deer**	**Horse**	**Cow**	**Deer**	**Horse**
1	Sparse	TVM	Guided filter	WL + TVM	Laplacian	Sparse
2	Visible	Sparse	Visible (2)	Sparse	Gradient	Gradient (2)
3	Laplacian	WL + Swarm	Sparse (2)	Gradient	Thermal	WL + TVM (2)
4	WL + Swarm	Gradient	Thermal	Guided filter	Guided filter	Visible
5	TVM	Laplacian	WL + TVM	Thermal	WL + TVM	TVM
6	Guided filter	Thermal	Laplacian (5)	WL + Swarm	WL + Swarm	WL + Swarm
7	Thermal	Visible	Gradient (5)	Laplacian	Sparse	Thermal
8	WL + TVM	WL + TVM	TVM (5)	Visible	Visible (8)	Guided filter (7)
9	Gradient	Guided filter	WL + Swarm (5)	SVD	SVD (8)	Laplacian (7)
10	SVD	SVD	SVD	TVM	TVM (8)	SVD

Table 4. Ranking of animal classification results based on overall accuracy (mAP50) for domestic cattle (*Bos taurus*, cow), white-tailed deer (*Odocoileus virginianus*, deer), and domestic horses (*Equus caballus*, horses) for visible, thermal, and eight fusion methods for YOLOv5 and YOLOv7 learning modules from images taken by a drone (unoccupied aircraft system or UAS). Numbers in parentheses indicate ties in rank for corresponding numbers and methods.

Rankings based on overall accuracy (mAP50) for animal classification among animals and architectures for fusion methods indicated that Sparse was the highest-ranking fusion method (Table 4), with the lowest rank score (15), followed by Gradient (25). The LP and WL + TVM methods ranked equally (28), followed by WL + Swarm (29), just above the rank score for visible and thermal results (30). Guided filter (31) and TVM (33) ranked below visible and thermal results (Table 4). SVD (51) was consistently poor and typically ranked last in most tests (Table 4).

Discussion

Our results further promote fused thermal and visible imagery for improved detection and classification of animals in drone imagery as initially explored in two previous studies[17,28]. Broadening past approaches, we found that some fusion methods increased both image quality and classification metrics consistently over thermal and sometimes visual results alone, but these results differed by animal species. For deer, the most accurate fusion methods substantially increased classification accuracy over visible and thermal images alone. However, the most accurate fusion methods provided little improvement over classification of cows and horses from visible images alone. These differences are likely explained by the contrasting search images of the animals we surveyed, suggested in two previous studies[17,28]. Cows and horses were typically conspicuous in visible images compared to deer, which were more cryptic against their respective backgrounds and required additional thermal information for classification. Our results suggest that for cryptic species such as deer, the fusion of information in thermal and visible images improves classification over either image type alone. Understanding the tradeoffs in using fusion compared to visible images alone for automated animal classification can improve the results and efficiency of drone surveys among animal species that differ across a gradient of conspicuous to cryptic against their respective backgrounds.

Intrinsic (e.g., animal size, color, and shape) and extrinsic (e.g., image background and shadows) factors can influence accurate automated classification of animals[7,8,17,28,39]. If size, shape, color, or a combination of these or other features are distinctive, visible images often contain most of the identifying characteristics needed to accurately identify animal species, at least during diurnal periods with adequate lighting[8,17,28,39]. Cows and horses in our study represented large-bodied mammals with distinctive body shapes and colors against open pastures or contrasting color backgrounds. Larger animals in images comprise more pixels than smaller animals at the same ground sampling distance (GSD), providing more information and often better classification performance for automated approaches[28,39]. Similarly, distinct differences in body shape or appendages (e.g. beaks, hooves, antlers, etc.) can also provide information used to automatically classify animals to species[28,39]. Color contrast with background environments, like our black and brown cows and horses against a green pasture has also been shown to improve automatic detection and classification of animals[7,8,17,28,39]. A combination of the above factors likely explains our findings for cows and horses, where fusion methods provided minimal, if any, gains in classification performance compared to visible results alone.

In many natural situations encountered during surveys using visible imagery, animals have little to no contrast with surrounding environments , are partially obstructed, or occur in low light conditions[6,17,28]. In such cases, thermal images provide critically important complementary information needed for detection or classification of species, such as animal heat signatures against typically cooler ambient backgrounds[6,7,17,28]. None of the deer in our images were obstructed, but many (~ 60%) occupied shadowed areas. Thermal images alone provide little classification information for animals of similar sizes absent distinct shape features, such as large mammal species in our study. This lack of information is particularly evident in drone images recorded at higher flight altitudes because animals typically appear as indistinct color clusters against the ambient background and provide few distinguishing features[28]. Hence, our observed poor classification results for thermal images alone compared to visible results alone or fused results. Similarly, automated detection and classification of 5 Gy wolves (*Canis lupus*) and 6 fallow deer (*Dama dama*) in zoo enclosures was more difficult, due to their cryptic pelage against respective backgrounds and similar sizes, compared to 4 American bison (*Bison bison*) and 3 elk (*Cervus canadensis*), which were larger and more conspicuous[28]. Meanwhile, fusion of the two image types helped to improve classification for cryptic or shadowed species over visible or thermal alone in our study, resulting in an increase in performance for fused images from15–85% for deer, far exceeding fusion results for self-driving cars during daytime (5% better than visible alone and 29% better than thermal alone[26]).

Our study highlights some important methodological and computational constraints, strengths, and potential future improvements. Computing resources limited the maximum batch size for YOLOv7, which likely explained poorer results for this classification architecture compared to YOLOv5. Using larger batch sizes can increase performance for YOLOv7[40] and is a likely next step for future research. Similarly, future studies could test fusion methods in other classification architectures such as CNNs[6,13–15] and deep residual networks[13].

Sparse and WL + TVM fusion methods performed consistently best for metrics of image quality, but these results translated to improved classification of animals only for Sparse, the top-ranking method by far. In contrast, WL + TVM ranked slightly better than either visible or thermal results for all three species. In our study, we trained our models to consider all three species simultaneously in non-annotated images to correctly detect target species, classify them, and exclude detecting or confusing them with other non-target objects. For studies of animals automatically classified from drone images, our fusion results represent an improvement compared to classifying a single species[6,7,14,15], or multiple species limited to annotated boxes where the animal is already detected but not classified[13]. Our methodology and results also extend the utility of fusion approaches for drone imagery beyond detection of single species[17] and classification among species with a few individuals present in the image[28]. Among fusion methods in our study, Sparse performed consistently well across two classification architectures, as well as three image quality and three classification metrics for three mammal species (one cryptic, two conspicuous). Future research could test Sparse performance with other image fusion methods[26].

Our results demonstrate that image fusion is a viable option when images are limited (43–68 images of 77–265 animal objects in our study) for automated and accurate animal classification taken from visible and thermal drone sensors. Studies for other computer vision methods of animal classification from drone images used much larger numbers of images including > 900 images for koalas (*Phascolarctus cinereus*)[6], and > 2000 tiles from image mosaics for caribou (*Rangifer tarandus*)[14]. However, increasing the number of training and testing images could also improve fusion results compared to the relatively few images collected for our study, as increasing the number of pre-classified images available to train models typically leads to better performance[38]. One solution is to use open-source repositories of pre-annotated objects, which provide large numbers of images and benchmark datasets for training and standardized comparisons across studies for other fields (e.g. ImageNet[18]). Such open-source, collaborative repositories for drone images of wild animals could advance automated classification for a variety of animal species; however, to our knowledge, only two such repositories are currently available (OUR[14], AWIR—this study), Benchmark datasets for animals in drone images would benefit from high variation in image backgrounds, animal positions, group sizes, species, color, and other features, each of which typically improves performance of classification models, as demonstrated for camera trap studies[41,42].

Our fusion results also are indicative of the benefits of employing drones capable of collecting visible and thermal images simultaneously when conducting animal surveys. Classification with fusion methods will yield the best results when the survey maximizes information provided by both visible and thermal sensors. Accordingly, characteristics of target animal species, environment, and time of day are critical considerations. Surveys that target animals that are distinct in size, shape, color, and background contrast relative to each other, will provide the most information for accurate detection and classification in visible images[13,28,39]. Endothermic compared to ectothermic animals will typically provide the most heat contrast of body compared to ambient temperatures in thermal images, unless ambient temperature is high[6,7,15,17].

For visible images, conducting surveys at midday can minimized potential effects of shadows, which can hide or confuse detection in these images[43,44]. In contrast, in some instances shadows can enhance detection[10]. However, activity for endothermic animals is often greater in crepuscular periods, which could improve detection, but might also cause errors in double-counting animals[14,45]. For thermal surveys, early mornings provide the coolest temperatures compared to other times of day, even in warm environments where the image background approaches or exceeds the surface temperatures of endothermic animals[6,7,15,17]. Thus, conducting surveys in the morning in warm environments will likely maximize the benefit of heat contrast with target endothermic animals for thermal images and detection of these animals in visible images. If shadows do not enhance detection[10], our results indicate that the fusion of both image types will offset potential drawbacks in decreased animal detection in shadows of visible images due to gains in information from fusing visible with thermal images leading to improved classification accuracy. Other image processing possibilities, particularly targeting ectotherms, include incorporation of algorithms utilizing color correlation measurements found in some camera trap systems (e.g.[46]). In other environments or seasons where the contrast between ambient temperatures and animal body temperatures are high, time of day may not be important for surveys or may be dictated by the constraints of animal behavior or other logistics. Additionally, using higher-resolution sensors or flying drones at lower altitudes can improve classification results or permit accurate classification of smaller animals[11,39]. Our results indicate that fusion methods are promising to advance automated detection and classification of animals from drone surveys, particularly for cryptic animals.

Data availability

Imagery collected and analyzed are available as unprocessed image pairs (with EXIF) in the Aerial Wildlife Image Repository (https://projectportal.gri.msstate.edu/awir/). Code developed to generate Fusion modules will be made available in Github. Classification architectures were cloned from https://github.com/ultralytics/yolov5/releases/tag/v6.1 for YOLOv5 and https://github.com/WongKinYiu/yolov7 for YOLOv7.

Received: 18 April 2023; Accepted: 19 June 2023

References

1. Christie, K. S., Gilbert, S. L., Brown, C. L., Hatfield, M. & Hanson, L. Unmanned aircraft systems in wildlife research: Current and future applications of a transformative technology. *Front. Ecol. Environ.* **14**, 241–251 (2016).
2. Linchant, J., Lisein, J., Semeki, J., Lejeune, P. & Vermeulen, C. Are unmanned aircraft systems (UASs) the future of wildlife monitoring? A review of accomplishments and challenges. *Mamm. Rev.* **45**, 239–252 (2015).
3. Elmore, J. A. *et al.* Evidence on the efficacy of small unoccupied aircraft systems (UAS) as a survey tool for North American terrestrial, vertebrate animals: a systematic map. *Environ. Evid.* **10**, 1 (2023).
4. Hodgson, J. C., Baylis, S. M., Mott, R., Herrod, A. & Clarke, R. H. Precision wildlife monitoring using unmanned aerial vehicles. *Sci. Rep.* **6**, 22574 (2016).
5. Hodgson, J. C. *et al.* Drones count wildlife more accurately and precisely than humans. *Methods Ecol. Evol.* **9**, 1160–1167 (2018).
6. Corcoran, E., Denman, S., Hanger, J., Wilson, B. & Hamilton, G. Automated detection of koalas using low-level aerial surveillance and machine learning. *Sci. Rep.* **9**, 1 (2019).
7. Seymour, A. C., Dale, J., Hammill, M., Halpin, P. N. & Johnston, D. W. Automated detection and enumeration of marine wildlife using unmanned aircraft systems (UAS) and thermal imagery. *Sci. Rep.* **7**, 1–10 (2017).
8. Lyons, M. B. *et al.* Monitoring large and complex wildlife aggregations with drones. *Methods Ecol. Evol.* **10**, 1024–1035 (2019).
9. Eikelboom, J. A. J. *et al.* Improving the precision and accuracy of animal population estimates with aerial image object detection. *Methods Ecol. Evol.* **10**, 1875–1887 (2019).
10. Rey, N., Volpi, M., Joost, S. & Tuia, D. Detecting animals in African Savanna with UAVs and the crowds. *Remote Sens. Environ.* **200**, 341–351 (2017).
11. Bowley, C., Mattingly, M., Barnas, A., Ellis-Felege, S. & Desell, T. An analysis of altitude, citizen science and a convolutional neural network feedback loop on object detection in unmanned aerial systems. *J. Comput. Sci.* **34**, 102–116 (2019).

12. Wood, S. A., Robinson, P. W., Costa, D. P. & Beltran, R. S. Accuracy and precision of citizen scientist animal counts from drone imagery. *PLoS ONE* **16**, 1–14 (2021).
13. Zhou, M. *et al.* Improving animal monitoring using small unmanned aircraft systems (sUAS) and deep learning networks. *Sensors* **21**, 5697 (2021).
14. Lenzi, J. *et al.* Artificial intelligence for automated detection of large mammals creates path to upscale drone surveys. *Sci. Rep.* **13**, 1–13 (2023).
15. Chen, A., Jacob, M., Shoshani, G. & Charter, M. Using computer vision, image analysis and UAVs for the automatic recognition and counting of common cranes (*Grus grus*). *J. Environ. Manag.* **328**, 116948 (2023).
16. Torney, C. J. *et al.* A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods Ecol. Evol.* 779–787 (2019). https://doi.org/10.1111/2041-210X.13165.
17. Chrétien, L. P., Théau, J. & Ménard, P. Visible and thermal infrared remote sensing for the detection of white-tailed deer using an unmanned aerial system. *Wildl. Soc. Bull.* **40**, 181–191 (2016).
18. Deng, J. *et al.* Imagenet: A large-scale hierarchical image database. in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255 (IEEE, 2009).
19. Hermann, K. L., Chen, T. & Kornblith, S. The origins and prevalence of texture bias in convolutional neural networks. in *34th Conference on Neural Information Processing System* (2020).
20. Tabak, M. A. *et al.* Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: MLWIC2. *Ecol. Evol.* **10**, 10374–10383 (2020).
21. Liu, Y., Liu, S. & Wang, Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf. Fusion* **24**, 147–164 (2015).
22. Gonzalez, R. C., Eddins, S. L. & Woods, R. E. *Digital Image Publishing Using MATLAB* (Prentice Hall, 2004).
23. Pajares, G. & De La Cruz, J. M. A wavelet-based image fusion tutorial. *Pattern Recognit.* **37**, 1855–1872 (2004).
24. Druzhkov, P. N. & Kustikova, V. D. A survey of deep learning methods and software tools for image classification and object detection. *Pattern Recognit. Image Anal.* **26**, 9–15 (2016).
25. Jin, X. *et al.* A survey of infrared and visual image fusion methods. *Infrared Phys. Technol.* **85**, 478–501 (2017).
26. Yadav, R., Samir, A., Rashed, H., Yogamani, S. & Dahyot, R. *CNN Based Color and Thermal Image Fusion for Object Detection in Automated Driving*. (2020).
27. Meena, S. D. & Agilandeeswari, L. Smart animal detection and counting framework for monitoring livestock in an autonomous unmanned ground vehicle using restricted supervised learning and image fusion. in *Neural Processing Letters*. Vol. 53 (Springer, 2021).
28. Chrétien, L.-P., Théau, J. & Ménard, P. Wildlife multispecies remote sensing using visible and thermal infrared imagery acquired from an unmanned aerial vehicle (UAV). *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **40**, 241–248 (2015).
29. Bulanon, D. M., Burks, T. F. & Alchanatis, V. Image fusion of visible and thermal images for fruit detection. *Biosyst. Eng.* **103**, 12–22 (2009).
30. Liu, F. & Seipel, S. Infrared-visible image registration for augmented reality-based thermographic building diagnostics. *Vis. Eng.* **3**, 1–15 (2015).
31. Maurya, L., Mahapatra, P. & Chawla, D. A registration strategy from scale adjustment to fine for thermal-visible face images. *Infrared Phys. Technol.* **120**, 104001 (2022).
32. Xu, R., Lin, H., Lu, K., Cao, L. & Liu, Y. A forest fire detection system based on ensemble learning. *Forests* **12**, 217 (2021).
33. Zhu, L., Geng, X., Li, Z. & Liu, C. Improving YOLOv5 with attention mechanism for detecting boulders from planetary images. *Remote Sens.* **13**, 3776 (2021).
34. Wang, Z., Cui, Z. & Zhu, Y. Multi-modal medical image fusion by Laplacian pyramid and adaptive sparse representation. *Comput. Biol. Med.* **123**, 103823 (2020).
35. Qu, G., Zhang, D. & Yan, P. Information measure for performance of image fusion. *Electron. Lett.* **38**, 1 (2002).
36. Xydeas, C. S. & Petrovic, V. Objective image fusion performance measure. *Electron. Lett.* **36**, 308–309 (2000).
37. Petrovic, V. S. & Xydeas, C. S. Gradient-based multiresolution image fusion. *IEEE Trans. Image Process.* **13**, 228–237 (2004).
38. Szeliski, R. *Computer Vision: Algorithms and Applications*. (Springer, 2022).
39. Chabot, D. & Francis, C. M. Computer-automated bird detection and counts in high-resolution aerial images: A review. *J. F. Ornithol.* **87**, 343–359 (2016).
40. Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M. & Tang, P. T. P. On large-batch training for deep learning: Generalization gap and sharp minima. in *5th International Conference on Learning Representations* (OpenReview.net, 2017).
41. Swanson, A. *et al.* Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. *Sci. Data* **2**, 1–14 (2015).
42. Tabak, M. A. *et al.* Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods Ecol. Evol.* **10**, 585–590 (2019).
43. Patterson, C., Koski, W., Pace, P., McLuckie, B. & Bird, D. M. Evaluation of an unmanned aircraft system for detecting surrogate caribou targets in labrador1. *J. Unmanned Veh. Syst.* **4**, 53–69 (2015).
44. Jones, L. R. *et al.* Controllable factors affecting accuracy and precision of human identification of animals from drone imagery. *Ecosphere* (2023) **(in review)**.
45. Brack, I. V., Kindel, A. & Oliveira, L. F. B. Detection errors in wildlife abundance estimates from unmanned aerial systems (UAS) surveys: Synthesis, solutions, and challenges. *Methods Ecol. Evol.* **9**, 1864–1873 (2018).
46. Bal, P., Lyons, D. & Shuter, A. A new ectotherm 3D tracking and behavior analytics system using a depth-based approach with color validation, with preliminary data on Kihansi spray toad (*Nectophrynoides asperginis*) activity. *Herpetol. Rev.* **51**, 37 (2020).

Acknowledgements

Our study was funded by the USDA Animal and Plant Health Inspection Service (APHIS), Wildlife Services (WS) via cooperative agreements between WS and Mississippi State University (AP20WSNWRC00C026, AP21WSNWRC00C014), extending from an Interagency Agreement between WS and the Federal Aviation Administration (No. 692M15-19-T-00017). We thank the Forest and Wildlife Research Center and College of Forest Resources at Mississippi State University (MSU) for additional support. This publication is a contribution of the MSU Geosystems Research Institute, Mississippi Agriculture and Forestry Experiment Station, and Center for Resolving Human-wildlife Conflicts, College of Forest Resources. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the USDA or FAA policy decisions regarding drones, UAS, or animals.

Author contributions

B.S.K.: Conceptualization, methodology, formal analysis, validation, writing-original draft, writing-review and editing. L.R.J.: Data acquisition, formal analysis, validation, writing-original draft, writing-review and editing.

J.A.E.: Data acquisition, formal analysis, validation, writing-original draft, writing-review and editing. S.S.: Conceptualization, methodology, funding acquisition, validation, writing-review and editing. K.O.E.: Funding acquisition, supervision, writing-review and editing. M.B.P.: Writing-review and editing. B.F.B.: Project administration, funding acquisition, supervision, writing-review and editing. R.B.I.: Project administration, funding acquisition, supervision, methodology, resources, writing-review and editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-37295-7.

Correspondence and requests for materials should be addressed to R.B.I.

Reprints and permissions information is available at www.nature.com/reprints.

Mammal Rev. 2007, Volume 37, No. 3, 207–223. *Printed in Singapore*.

The value of infrared thermography for research on mammals: previous applications and future directions

DOMINIC J. MCCAFFERTY
Department of Adult and Continuing Education, Faculty of Education, University of Glasgow, 11 Eldon Street, Glasgow G3 6NH, UK

ABSTRACT

1. Infrared thermography (IRT) involves the precise measurement of infrared radiation which allows surface temperature to be determined according to simple physical laws. This review describes previous applications of IRT in studies of thermal physiology, veterinary diagnosis of disease or injury and population surveys on domestic and wild mammals.

2. IRT is a useful technique because it is non-invasive and measurements can be made at distances of <1 m to examine specific sites of heat loss to >1000 m to count large mammals. Detailed measurements of surface temperature variation can be made where large numbers of temperature sensors would otherwise be required and where conventional solid sensors can give false readings on mammal coats. Studies need to take into account sources of error due to variation in emissivity, evaporative cooling and radiative heating of the coat.

3. Recent advances in thermal imaging technology have produced lightweight, portable systems that store digital images with high temperature and spatial resolution. For these reasons, there are many further opportunities for IRT in studies of captive and wild mammals.

Keywords: disease, infrared thermography, injury, population surveys, temperature measurement, thermal physiology

Mammal Review (2007), **37**, 207–223
doi: 10.1111/j.1365-2907.2007.00111.x

INTRODUCTION

Infrared thermography (IRT) involves the precise measurement of infrared radiation emitted by an object, which allows the surface temperature to be determined according to relatively simple physical laws and known properties of the surface (see Speakman & Ward, 1998). Specialized thermographic cameras produce images that show the variation in temperature of a surface by representing different temperatures with a grey or coloured shaded scale (Fig. 1). Although thermal imaging was developed principally for industrial, medical and military applications (Burnay, Williams & Jones, 1988), it has been used to study many animal groups including insects, reptiles, birds and mammals (see McCafferty *et al*., 1998).

Infrared thermography can examine many different aspects of thermal physiology, diagnose injury and disease and is a useful technique for counting animal populations. The great advantage of IRT in animal research is that measurements can be made without touching or disturbing the animal and depending on the instrument type and application, measurements can be made either at close range (<1 m) or at large distances (>1000 m). Detailed measurements of the temperature variation of mammals can be made where large numbers of temperature sensors would otherwise be required. Conventional solid probes can also give

---

Correspondence: D. J. McCafferty. E-mail:

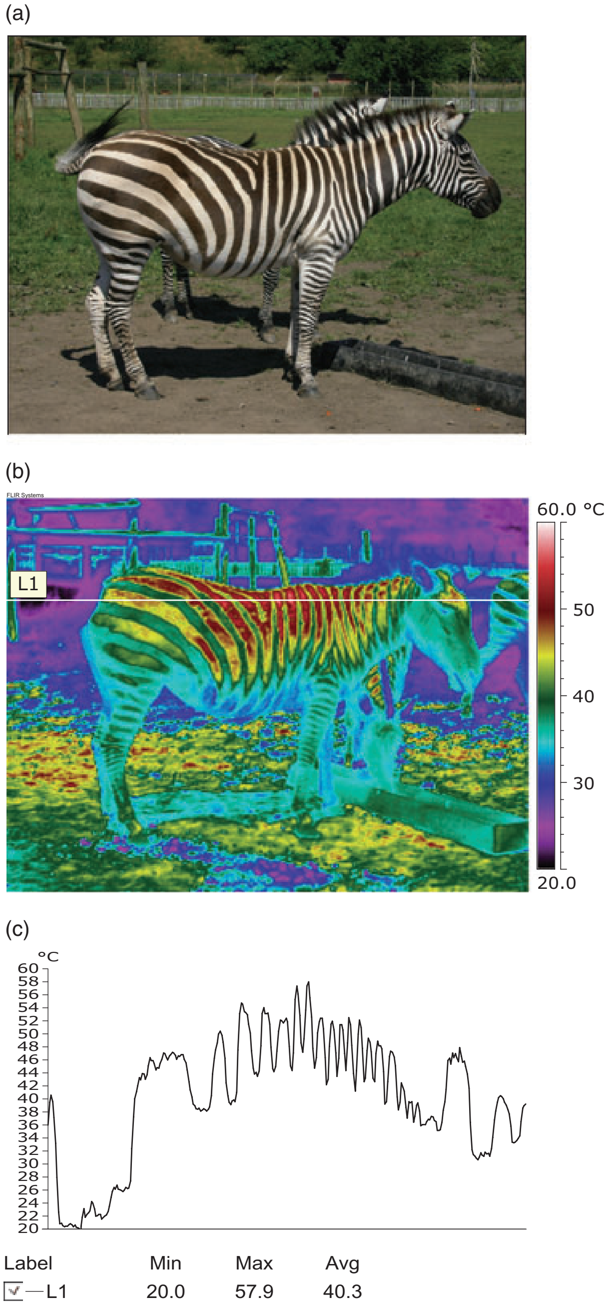


Fig. 1. Photograph (a) of Grant's zebra *Equus burchelli boehmi* with corresponding infrared image (b) in full sun. The temperature profile L1 displayed in the graph below (c) shows the variation in temperature across the body, with black stripes more than 10 °C warmer than white striped areas of the coat. Mean air temperature = 28.3 °C, relative humidity = 44%, solar radiation = 860 $\mathrm{Wm^{-2}}$ and wind speed = 0.3 $\mathrm{ms^{-1}}$.

false readings due to the difference in heat capacity between sensor and coat or through disruption of the hair fibres by sensors (Cena, 1974; Mohler & Heath, 1988). Previously, Cena & Clark (1973) outlined important theoretical aspects of this technique for research on domestic and zoo animals, Yang & Yang (1992) reviewed biomedical and veterinary applications and Speakman & Ward (1998) gave an account of the principles of IRT and demonstrated its usefulness for studying thermoregulation. More recently, Kastberger & Stachl (2003) highlighted several interesting veterinary and physiological applications.

The aim of this review was to examine the value of thermal imaging for research on non-human mammals. In particular, this paper brings together findings from physiological, ecological and veterinary investigations to generate new ideas on how to use IRT to investigate wild mammal populations. This review is timely given recent advances in thermal imaging technology and a reduction in the cost of these devices, both of which will provide future research opportunities.

APPLICATIONS

For this review, a literature search was undertaken using ISI Web of Knowledge^{SM} (http://wok.mimas.ac.uk/). This was followed by compiling a reference list from each of these papers to include older studies that may not have been listed in current electronic databases and supplementing these with other known studies. This is therefore not an exhaustive list as this is a widely used technique, but it is likely to cover a large proportion of the main empirical studies to date. For the purposes of this review, studies on humans and closely related clinical applications were not considered.

Seventy-one empirical studies using IRT on mammals since 1968 (Tables 1–3) were examined. These studies involved domestic and wild mammals from 11 mammalian orders. Two-thirds of the studies involved terrestrial species and a third were on aquatic mammals, mostly marine species. These included 34 studies on thermal physiology (48%), 19 involving veterinary diagnosis of disease and injury (27%) and 18 population surveys (25%). Seventy per cent of studies were on captive mammals.

Thermal physiology

Infrared thermography has been used to examine many different aspects of thermoregulation (Table 1) and much of this work has focused on identifying parts of the body with relatively high temperature which can be related to an animal's anatomy and physiology. This has signaled that the head is a major source of heat loss for most species of mammals and also identified the importance of appendages in controlling heat loss. These studies demonstrate the clear link between surface temperature and underlying blood circulation and brown adipose tissue, as well as the role of fur in reducing heat loss from the skin surface. Many studies have examined the relationship between body surface temperature and air temperature. However, a novel approach with IRT has been to examine the relationship between environmental temperature and the sensitivity of vibrissal follicles in seals and dolphins (Dehnhardt, Mauck & Hyvärinen, 1998; Mauck, Eysel & Dehnhardt, 2000). These studies demonstrated that even in the cold, blood is circulated to these areas to maintain the function of these essential sensory organs.

A major strength of IRT is its ability to relate changes in surface temperature to particular physiological states or associated with certain behaviours such as huddling or vocalization. Recent studies have also shown that IRT is capable of detecting surface temperature changes in response not only to physical activity but also to fear. Particularly significant were the findings of Nakayama *et al.* (2005) which showed that changes in facial surface temperature patterns of Rhesus monkeys *Macaca mulatta* occurred in response to the threat of capture. IRT is particularly suited to examining changes in surface temperature during activities such as running, flying and even swimming. The latter application on marine mammals was an interesting applied study to examine the significance of changes in circulation associated with exercise in dolphins when chased and captured in the Pacific tuna fishery (Pabst *et al.*, 2002). This study found that dolphins increased their rate of heat dissipation from dorsal fins to the environment from the start of the chase. During prolonged chases, animals had higher skin surface temperatures, presumably as a result of greater blood flow to these areas.

Table 1. Thermo-physiology studies using IRT on captive (c) and wild (w) mammals showing measurements taken, distance (m) and imaging system used

Species			Measurement	Distance	Camera type	Author
Harp seal	*P. groenlandicus*	c	Exercise and heat loss	–	–	Ørtisland (1968)
Elephant and zebra	*L. africana, E. chapmani*	c	Colouration	5–15	Agavision 680	Cena & Clark (1973)
Polar bear	*U. maritimus*	c	Sites of heat loss	–	AGA Thermovision 720	Ørtisland *et al.* (1974)
Jackrabbit	*L. californicus*	c	Exercise and heat loss	–	AGA Thermovision 680-120B	Hill *et al.* (1976)
Harp seal	*P. groenlandicus*	w	Vasoconstriction/dilation	–	AGA 750 Thermovision	Blix *et al.* (1979)
Jackrabbit	*L. californicus*	c	Vascular patterns	–	AGA Thermovision 680-120B	Hill *et al.* (1980)
Raccoon dogs and blue foxes	*N. procyonoides, A. lagopus*	c	Thermoregulation and heat loss	–	AGA Thermovision 720	Korhonen & Harri (1986)
Rabbit	*O. cuniculus*	c	Thermoregulation of ear	–	Inframetrics 525	Mohler & Heath (1988)
Coypu	*M. coypus*	c	Insulation	–	Philips Medical	Doncaster *et al.* (1990)
Mongolian gerbil	*M. unguiculatus*	c	Responses to air temperature	0.5–1	Inframetrics 525	Klir *et al.* (1990)
Elephants	*L. africana, E. maximas*	c	Avenues of heat loss	–	AGEMA 728	Williams (1990)
Laboratory rat	*R. norvegicus*	c	BAT tissue and vocalization	–	Thermovision 870	Blumberg *et al.* (1992)
3 species foxes	*V. vulpes, A. lagopus, V. macrotis*	c	Sites of heat loss	0.5–15	Inframetrics 525	Klir & Heath (1992)
Long eared bats	*P. auritus*	c	Body temperature and energy savings	–	EEV P7000	Webb *et al.* (1993)
29 different species		c	Control of heat loss	–	Inframetrics	Phillips & Heath (1995)
Egyptian fruit bats	*R. aegyptiaeus*	c	Temperature changes during flight	<2	AGEMA Thermovision 880	Lancaster *et al.* (1997)
Harbour seal	*P. vitulina*	c	Vibrissae sensitivity	0.8	AGEMA Thermovision THV450D	Dehnhardt *et al.* (1998)
Bottlenose dolphin	*T. truncatus*	?	Circulation system in fluke	–	–	Williams *et al.* (1999)
Harbour seal & River dolphin	*P. vitulina, S. fluviatilis*	c	Sensory physiology	0.07–0.8	AGEMA 870	Mauck *et al.* (2000)
Field vole	*M. agrestis*	c	Non-shivering thermogenesis	0.5	AGEMA Thermovision 880	Jackson *et al.* (2001)
Woodchuck	*M. monax*	c	Temperature responses	–	Inframetrics 525	Phillips & Heath (2001)
Laboratory rats	*R. norvegicus*	c	Huddling	–	FLIR	Sokoloff & Blumberg (2001)
Bottlenose dolphin	*T. truncatus*	?	Superficial veins	–	AGEMA 570	Meagher *et al.* (2002)
Spinner and spotted dolphins	*S. attenuata, S. longirostrus orientalis*	w	Temperature effects of capture	–	AGEMA 570	Pabst *et al.* (2002)
3 seal species	*H. grypus, P. vitulina, P. groenlandicus*	c	Evaporation	0.5–2.5	AGEMA 870	Mauck *et al.* (2003)
Golden mantled ground squirrel	*S. lateralis*	c	Hypoxic metabolic response	–	Inframetrics 522	Tattersall & Milsom (2003)
Zebra	*E. quagga, E. zebra, E. grevyi*	c	Effect of solar radiation	–	Thermovision 695	Benesch & Hilsberg (2003)
Ground squirrel & marmot	*S. tridecimlineatus, M. flaviventris*	c	Hibernation	–	Inframetrics 525	Phillips & Heath (2004)
Greater mouse eared bat	*M. myotis*	w	Roosting behaviour and microclimate	2.5	Thermacam PM595	Sandel *et al.* (2004)
Dairy cattle	*B. taurus*	c	Climate effects on housing conditions	–	Thermotracer 6T62 NEC	Zähner *et al.* (2004)
Steller sea lion	*E. jubatus*	c	Sensor placement	–	Thermacam PM695	Willis *et al.* (2005)
Laboratory rat	*R. norvegicus*	c	Response to fear	0.7	TVS-100, Avio	Vianna & Carrive (2005)
Rhesus monkeys	*M. mulatta*	c	Assessment of emotional state	0.23	TH5100 NEC	Nakayama *et al.* (2005)
Laboratory rat	*R. norvegicus*	c	Inflammation and thermoregulation	–	Thermovision A20M	Almeida *et al.* (2006)
Grey seal	*H. grypus*	c	Effect of instrument attachment	1–3	Thermacam PM595	McCafferty *et al.* (2007)

IRT, infrared thermography.

Table 2. Veterinary studies using IRT on captive (c) and wild (w) mammals showing measurements taken, distance (m) and imaging system used

Species			Measurement	Distance	Camera type	Author
Horse	*E. caballus*	c	Clinical disorders	–	AGA Thermovision 680	Purohit & McCoy (1980)
Horse	*E. caballus*	c	Posture and anaesthesia	0.01	Mikron 25	Palmar (1981)
Horses	*E. caballus*	c	Podotrochlosis	–	Dynarad 209a	Turner *et al*. (1983)
Dairy cattle	*B. taurus*	c	Estrus detection	2.5	AGA Thermovision 750	Hurnik *et al*. (1985)
Bull	*B. taurus*	c	Scrotal disease	–	AGA IR system	Purohit *et al*. (1985)
Cattle	*B. taurus*	c	Effects of transportation	–	AGEMA 782	Schaefer *et al*. (1988)
Horse	*E. caballus*	c	Back pain	–	Insight Visions Systems	Colles *et al*. (1995)
Horse	*E. caballus*	c	Rug and whip damage	–	AGEMA 450	Holah (1995)
Cattle	*B. taurus*	c	Scrotal temperature	1	AGA Thermovision 782	Kastelic *et al*. (1996)
Cattle	*B. taurus*	c	Inflammation and branding	–	Thermovision 470	Schwartzkopf-Genswein & Stookey (1997)
Spanish Ibex	*C. pyrenaica*	w	Sarcoptic mange disease	39–221	Milcam	Arenas *et al*. (2002)
Wapiti	*C. elaphus canadensis*	c	Antler removal	–	Inframetrics 760	Cook & Schaefer (2002)
Horse	*E. caballus*	c	Injections and digital neurectomy	1.2	DTIS-500 Emerge Vision	Van Hoogmoed & Snyder (2002)
Dairy cattle	*B. taurus*	c	Barn management	3	AGA 570 DEMO	Knizkova *et al*. (2002)
Dairy cattle	*B. taurus*	c	Variation in udder temperature	2–2.5	Inframetrics 760	Berry *et al*. (2003)
Cattle	*B. taurus*	c	Infection detection	1.3	FLIR Inframetrics 760	Schaefer *et al*. (2004)
Dairy cattle	*B. taurus*	c	Climate and housing conditions	–	Thermotracer 6T62 NEC	Zahner *et al*. (2004)
Dairy cattle	*B. taurus*	c	Health and condition of hooves	1.5–2.0	FLIR Inframetrerics 760	Nikkhah *et al*. (2005)
Holstein heifers	*B. taurus*	c	Tail docking and pain	–	eMerge, Sebastian FL	Eicher *et al*. (2006)

IRT, infrared thermography.

Table 3. Mammal surveys using IRT on captive (c) and wild (w) mammals showing measurements taken, distance (m) and imaging system used

Species			Measurement	Distance	Camera type	Author
White-tailed deer	*O. virginianus*	w	Population counts	300	Classified	Croon *et al*. (1968)
White-tailed deer	*O. virginianus*	w	Population counts	300	Classified	McCullough *et al*. (1969)
Polar bear	*U. maritimus*	w	Detection & counts	150	Test	Brooks (1972)
White-tailed deer	*O. virginianus*	w	Population census	300	Test equipment	Graves *et al*. (1972)
Ringed seal	*P. hispida*	w	Snow lairs	180	FLIR 1000 A	Kingsley *et al*. (1990)
Walrus	*O. rosmarus divergens*	w	Numbers & group size	400–2400	DFORS	Barber *et al*. (1991)
Whales	5 species	w	Counts and temperature	10–70	AGEMA Thermovision 880	Cuyler *et al*. (1992)
Ringed seal	*P. hispida*	w	Snow lairs	30	Inframetric 600	Sipilä & Kurlin (1992)
White-tailed deer	*O. virginianus*	w, c	Population census	170–450	FLIR 2000G	Wiggers & Beckerman (1993)
Squirrels, hares & mice	4 species	w	Detection of animals, nests and burrows	2–40	Thermovision 210, Inframetrics 5226	Boonstra *et al*. (1994)
Gray bats	*M. grisescens*	w	Population estimates	–	AGEMA 782	Sabol & Hudson (1995)
Grey whale	*E. robustus*	w	Migration counts	2000	AN/KAS-1 A (US Navy)	Perryman *et al*. (1999)
Manatee	*T. manatus latirostris*	c	Detection	–	PalmIR Pro Raytheon	Keith (2002)
Harbour seal	*P. vitulina*	w	Population counts	3000	–	Duck & Thompson (2003)
Polar bear	*U. maritimus*	w	Den Surveys	61–244	FLIR Safire AN/AAQ-22	York *et al*. (2004)
Deer	*O. virginianus*	w	Detection to aid capture	max 72	PalmIR 250 Raytheon	Ditchkoff *et al*. (2005)
Walrus	*O. rosmarus divergens*	w	Group size and counts	800–3200	Airborne Multispectral Scanner	Burn *et al*. (2006)
Deer	*O. virginianus* *O. hemionus*	w	Detection to aid capture	15–50	PalmIR 250 Raytheon	Butler *et al*. (2006)

IRT, infrared thermography.

Thermal imaging is also a useful tool for refining research methods, for example as a guide for the placement of heat flux sensors to study metabolic heat production of Steller sea lions *Eumetopias jubatus* (Willis *et al.*, 2005) and to determine the effects of attaching bio-logging devices to the pelage of grey seals *Halichoerus grypus* (McCafferty, Currie & Sparling, 2007).

Veterinary diagnosis of disease and injury

Infrared thermography has largely been a diagnostic tool in veterinary science in combination with other indicators of disease. A major application of this technique has been to diagnose injury and disease in horses and there have been several useful studies detailing factors influencing normal temperature distributions and outlining appropriate measurement protocols (see review by Eddy *et al.*, 2001). Abnormal or asymmetrical temperature distributions have been used as indicators of underlying problems with blood circulation or inflammatory responses (Table 2).

The non-invasive nature of IRT makes it particularly suited for studying farm animal welfare (see review by Stewart *et al.*, 2005). Studies have examined the extent and duration of inflammation observed on branding sites, effects of antler removal, changes in the thermal status of cattle during transportation, detecting hoof disorders and rises in body temperatures due to infection. An interesting veterinary application has been to detect estrus in cows by examining temperature distribution of the gluteal region. In this case, IRT was more effective than experienced dairy staff in detecting estrus in early stages but was less accurate in later postpartum due to a greater number of false positives (Hurnik, Webster & DeBoer, 1985).

Thermal imaging on captive species other than horses and cattle is less common, although Kouba & Willard (2005) reported anecdotally how IRT was being used to monitor a range of illnesses in zoo species. One of the first attempts to use IRT to detect disease in a wild mammal population was undertaken to diagnose sarcoptic mange in wild Spanish ibex *Capra pyrenaica*. Unfortunately, this was found to be not as affective as visual observation due to the limitations of the thermal imaging system used for distances greater than 100 m (Arenas *et al.*, 2002).

Population surveys

A variety of thermal imaging devices have been used from aircraft or road vehicles to detect and/or count large mammals (Table 3). This application does not require precise temperature measurements but simply detects individuals or dens by a warm signal against a cool background. IRT has been used in this way for counts of deer and pinnipeds. Thermal imaging has also been able to detect the blows of large whales. For example, a remotely operated thermal imaging system from a shore based station was used to count Pacific grey whales *Eschrichtius robustus* over a period of a month and across three years. Numbers of whales were detected from their blows and showed that migration rates were greater during the night than throughout the day (Perryman *et al.*, 1999). Although IRT was also found to be effective in detecting relatively small mammals, transect surveys on foot with handheld infrared cameras have been less commonly used in the past, most probably limited by the relatively large size of imaging systems. More recently, counts of grey bats *Myotis grisescens* using IRT have produced colony estimates similar to those counted visually and have opened up possibilities of using automated systems for monitoring purposes (Sabol & Hudson, 1995).

These studies demonstrate the usefulness of using thermal imaging to survey remote geographical areas. Similar to conventional aerial photography, thermal imaging from aircraft can be hindered by cloud cover since infrared radiation is absorbed by water vapour. The success of the technique relies on a relatively large temperature difference between the study animal and the ground surface. This is dependent on the temperature of the ground surface and the insulation properties of the animal. Surveys using IRT are therefore often undertaken at night when the thermal contrast between animal and background is greatest. Animals living in open habitats such as coastal areas or areas with sparse vegetation are suited to aerial survey methods compared to forest dwelling species. The usefulness for population monitoring relies on being able to ground-truth thermal imaging counts with visual counts and to choose periods of the day or season of the year when animals can be most easily detected.

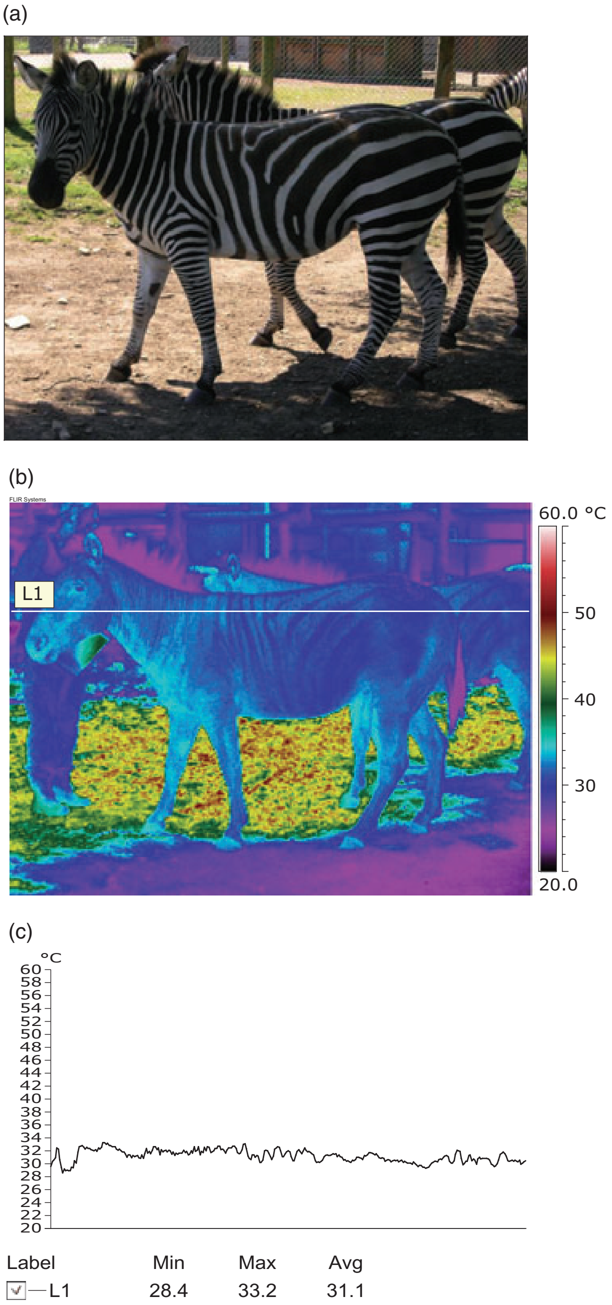
INTERPRETATION OF THERMAL IMAGES

For some applications, such as population counts, accurate temperature measurements of detected animals are not required. However, for the study of thermal physiology and energetics, the infrared radiation detected by the equipment must be converted to an accurate estimate of surface temperature. Infrared radiation emitted by bare-skinned animals is governed by the skin surface temperature but the radiation emitted from most mammals may originate either from the skin, if this is incompletely obscured by hairs, or from the hairs themselves. The radiating surfaces of the hairs are at a range of temperatures determined by the temperature gradient between the skin and the coat surface. The exchange of radiation may be further complicated by external fluxes that contribute to the heat balance of the hairs. For an animal with thick fur, the surface temperature measured by IRT is typically several millimeters beneath the physical surface of the coat. The equilibrium temperature of this surface is determined by the loss of heat from radiation and convection to the surroundings, the conduction of heat through the coat and the exchange of thermal and short wave radiation (Cena, 1974). The radiative environment in which measurements are carried out is therefore important because of its influence on coat temperature. It has been clearly shown that different coloured coats influence solar heating at the surface, with black areas of a coat having greater surface temperatures than white areas in strong sunshine (Cena & Clark, 1973; Benesch & Hilsberg, 2003). This is clearly seen in infrared images of zebras that show black stripes to be more than 10 °C warmer than white strips in full sun (Fig. 1). The temperature pattern does not reflect underlying circulation or large differences in emissivity as the temperature pattern almost disappears after a few minutes in the shade (Fig. 2). Even where solar radiation is excluded care should be taken to use enclosures that have wall temperatures close to air temperature to avoid additional radiative heating and avoid small enclosures that reflect significant amounts of thermal radiation from the animal.

Surprisingly, there have been relatively few comparisons between IRT and solid temperature probes. In a study of a rabbit pinna, Mohler & Heath (1988) showed that although thermocouple measurements gave the same trends in surface temperature, thermocouples consistently recorded higher temperatures when the pinna was vasodilated and recorded lower temperatures when vasoconstricted. The added value of IRT is its ability to measure easily the spatial variation in surface temperature and therefore produce more accurate temperature records of whole body regions.

The surface temperature of a mammal will not only be influenced by its skin temperature but by the thickness, density and quality of hair covering different parts of the body and this may differ between individuals and vary due to seasonal moult. Some veterinary studies on horses have controlled for this by shaving small sections of hair from limbs in order to determine the temperature of the underlying skin surface (Holah, 1995). This is not feasible or indeed desirable for most investigations. Studies should therefore take into account these sources of variation most easily by following the same individual throughout experiments or by sampling a large group of individuals to account for this variation.

Fig. 2. Photograph (a) of Grant's zebra with corresponding infrared image (b) after 5–10 min in the shade of a tree. The temperature profile L1 displayed in the graph below (c) shows the variation in temperature across the body, with black stripes on average less than 2 °C warmer than white striped areas of the coat. Mean air temperature = 27.4 °C, relative humidity = 45%, wind speed = 0.6 ms^{-1} and solar radiation was not recorded.



Label	Min	Max	Avg
☑—L1	28.4	33.2	31.1

In order to obtain accurate surface temperature measurements a surface emissivity value is a required parameter for infrared imaging systems. Bare skin has an emissivity of 0.98 and the emissivity of dry fur is relatively uniform in mammals, in the range 0.98–1.0 (Monteith & Unsworth, 1990). The emissivity of the coat can also be changed by dirt or other materials (e.g. soil = 0.93–0.96 or water = 0.96, Campbell & Norman, 1998). This can be easily addressed with captive animals by brushing or cleaning coats prior to measurement. Since radiative heat transfer scales linearly with emissivity and as surface temperature scales to the

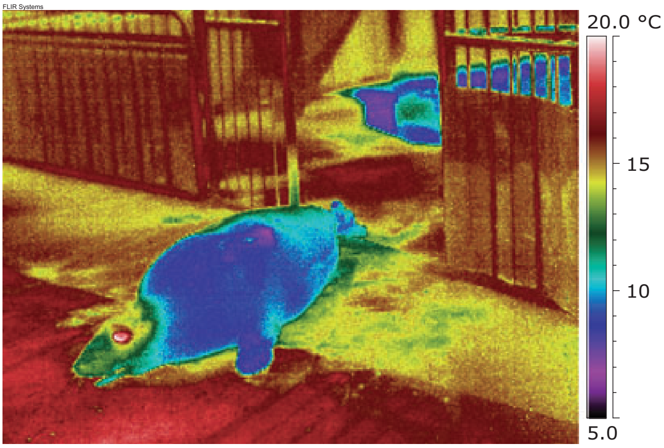


Fig. 3. Infrared image of female adult grey seal recently hauled out from a seawater pool (background) in captivity. Note that most of body is at uniform surface temperature corresponding to the temperature of seawater. The head is warmer than the body trunk as the seal held its head above water prior to leaving the pool. A small temperature logger for recording stomach temperature is also visible on the centre of the back. Air temperature = 16.2 °C.

power of four, these small differences in emissivity can be shown by calculation to account for less than 0.5 °C difference at typical mammalian coat temperatures. In this case, computer software for image analysis can be useful in providing error analysis by simply changing the emissivity of different regions. Alternatively, the temperature of fur with and without dirt/water can be measured to exclude this source of variation.

Temperature errors associated with alterations in the emissivity of a wet coat are small in comparison to changes in coat temperature due to evaporative cooling. This is pertinent for studies on aquatic mammals or animals wet by precipitation in natural conditions. Wetting leads to an apparent uniformity in surface temperature due to the retention of water in the coat. In addition, the greater thermal conductivity of water means that heat may be rapidly conducted from warm parts of the body, particularly as aquatic mammals are seen to leave the water. Both these factors may obscure the variation in underlying skin temperature. This can be seen in an image of an adult grey seal recently hauled out from a seawater pool in captivity, where the temperature of the body corresponds to the temperature of seawater trapped in the fur (Fig. 3). Care should be taken therefore to ensure that animals are kept dry or in the case of aquatic mammals, the period of time out of water is standardized. The influence of wetting may therefore be problematic for studies in the field when accurate temperature measurements are required. One way to correct for this would be to first determine rates of drying from animals in captivity (Mauck *et al*., 2003) or to use heat transfer models in the laboratory to determine the relationship between surface temperature and wetting (e.g. McArthur & Ousey, 1994).

Wet environments are not usually a problem for most IR imaging systems because of the environmental protection/waterproofing of these devices to high industrial standards. However, water on the lens due to rain or spray is a potential difficulty for accurate temperature measurements in the field. Pabst *et al*. (2002) took images from a boat and therefore covered the lens with polyethylene film and recalibrated the temperature measurements. Similarly, Tattersall & Milsom (2003) took images through a polyethylene 'window' to take images of animals in a metabolic chamber. This is possible over a limited range of temperatures, typical in animal studies but it should be remembered that this additional coating will alter the spectral sensitivity of the device.

The detection of radiation by infrared cameras means that curved surfaces are subject to detection errors compared to flat surfaces. This gives rise to a cool edge effect seen on many images of animals. For a surface with emissivity of 0.98, the associated temperature error has been shown to be independent of viewing angle up to about 30° but increased from 0.5 to 3 °C at 30–70° and was greater than 4 °C at angles above 70° (Watmough, Fowler & Oliver, 1970; Clark, 1976). If necessary, this can be overcome using a composite image produced from several images taken from different positions.

FUTURE DIRECTIONS

Developments in technology have meant that infrared imaging devices are now the size of conventional video cameras or smaller and it is relatively easy to capture and store high-resolution thermal images in single image or video format. In the past, IR imaging systems relied on liquid nitrogen cooled detectors that made field studies difficult. Imaging systems nowadays have electronically cooled detectors allowing them to be easily used in remote areas. Custom written software is also available that allows rapid image analysis and summary statistics. Lower cost devices <£10k compared with more advanced systems costing £30–40k with similar temperature resolution (±0.1 °C) are now becoming available and therefore there are future opportunities for using IRT in mammal research.

The non-invasive nature of this technique will continue to provide the basis of future applications and previous studies show that IRT can be used to answer many interesting research questions. Unique opportunities now exist to examine thermoregulation of wild mammals in natural conditions. By combining measurements of surface body temperature with measurements of internal body temperature using implanted temperature loggers or other physiological parameters such as heart rate (e.g. Butler *et al.*, 1995), we will more fully understand thermal responses of animals to a range of environmental conditions. Bakken *et al.* (2005) have shown in birds that by removing a very small section of plumage to reveal the skin temperature, cloacal temperature could be estimated to within 1 °C. Although this is subject to some error, it does provide a method of estimating internal body temperature without the need for internal temperature loggers that has not often been considered in IRT applications with mammals. Preliminary reports also suggest that eye temperature recorded by IRT can be used to determine rectal and vaginal measurements in domestic animals (Sykes *et al.*, 2006; Willard, Vinson & Godfrey, 2006). If this method can be substantiated further, then it may provide opportunities of monitoring internal temperature non-invasively in captive experiments and field studies.

Previous studies have used surface temperature measurements to determine rates of heat loss and thereby estimate metabolic heat production (e.g. Williams, 1990). As yet, there has been no evaluation of how accurate these estimates are likely to be for domestic or free ranging mammals. IRT together with indirect calorimetry could validate heat transfer models that estimate metabolic costs of mammals. This has indeed been successfully carried out on captive birds where the metabolic power of flight determined by heat transfer modelling agreed with measurements using doubly labelled water and mask respirometry (Ward *et al.*, 1999, 2004). Surprisingly, there have been relatively few IRT studies examining changes in surface temperature during exercise in mammals. Surface temperatures could parameterise biophysical models of heat loss that investigate how exercise metabolism compensates for thermoregulatory costs and determine energy costs associated with locomotion or foraging behaviour. However, it should be remembered that although IRT can be used to derive reasonable estimates of heat loss by convection and radiation from the surface of animals, heat losses through respired gases (particularly by latent heat loss) must also be considered in order to estimate total heat loss from the organism.

The absolute accuracy of metabolic rate derived from IRT measurements may be relatively uncertain unless cross-calibration is made with existing metabolic methods as described above for birds. However, IRT is of great value in determining relative estimates of metabolic rate, particularly where natural behaviour does not occur in small metabolic chambers. Ward & Slater (2005) used this approach to estimate the increased metabolic cost of bird song by comparing heat loss between singing and non-singing birds in captivity. This approach could also be used to derive relative energy costs of a wide range of behaviours in the wild.

It is likely that IRT will continue to be a useful tool for the diagnosis of disease and injury in domestic and zoo animals, used in conjunction with existing veterinary procedures (Head & Dyson, 2001; Webbon, 2002). The development of small handheld instruments might soon allow these to be standard pieces of equipment for vets. Given concerns about infectious diseases among farm animals or within wild animal populations, IRT will be useful for early detection of disease, if further clinical trials can be undertaken. This may be achieved by remote monitoring systems such as those outlined by Stewart *et al.* (2005) that are recording the eye temperature of cattle with an automated system. The requirement of studies such as these will be to demonstrate convincingly that surface temperatures strongly correlate with the occurrence of infection. One of the most exciting opportunities in this area will be to extend veterinary applications of thermal imaging to study the health of wild mammal populations. Although an earlier attempt to diagnose disease in wild mammals with IRT was unsuccessful because of the distances involved (Arenas *et al.*, 2002), this may not apply in all cases and more appropriate choice of camera lenses may make distance work feasible.

The use of IRT for population monitoring is likely to be limited as much by the cost of aircraft or ship time as it is by the cost of imaging systems. However, surveys on foot or by vehicle will be easier with the highly portable imaging systems. IR imaging systems are likely to be particularly useful for monitoring nocturnal species. There is already considerable interest in using IRT to monitor large colonies of bats (Sabol & Hudson, 1995; Hristov, Betke & Kunz, 2005; Reichard, Frank & Kunz, 2005). For this purpose, automated image recognition systems provide the opportunity to monitor large colonies, not easily undertaken using traditional methods.

CONCLUSION

Infrared thermography has been successfully used in studies of thermal physiology, disease and population monitoring of captive and wild mammals since the 1960s. Its main advantage is that it is a non-invasive technique for measuring radiative surface temperature and therefore it can be either used to infer underlying circulation that is related to physiology, behaviour and disease or simply to detect a warm body against a cool background. The major limitation of this technique is that radiative surface temperature is also influenced by solar radiation, wetting and evaporation. For accurate temperature measurements in the field, it is therefore best suited for studies at night or in situations where animals experience low solar irradiances. Where environmental conditions prevent accurate temperature measurement comparative studies can still be undertaken provided conditions are equivalent between groups. For studies in captivity, experimental design should also consider the radiative environment of housing where measurements are made and also how underlying physiological responses and disease may influence surface temperature patterns. Nevertheless, if these factors are taken into account, the increased portability and reduced cost of IR imaging systems provide further opportunities for a range of studies that wish to measure surface temperature or detect animals non-invasively.

ACKNOWLEDGEMENTS

I would particularly like to thank John Currie (School of the Built Environment, Napier University, Edinburgh) for use of thermal imaging equipment and for his continued support and enthusiasm for this work. Thanks to Graeme Ruxton for useful comments on an earlier draft of the manuscript. I am grateful to Blair Drummond Safari Park and the Sea Mammal Research Unit, University of St Andrews for permission to take IR images of zebras and grey seals, respectively. Thanks to two reviewers who provided useful comments on an earlier version of the manuscript.

REFERENCES

Almeida, M.C., Steiner, A.A., Branco, L.G.S. & Romanovsky, A.A. (2006) Cold-seeking behavior as a thermoregulatory strategy in systemic inflammation. *European Journal of Neuroscience*, **23**, 3359–3367.

Arenas, A.J., Gómez, F., Salas, R., Carrasco, P., Borge, C., Maldonado, A., O'Brien, D.J. & Martínez-Moreno, F.J. (2002) An evaluation of the application of infrared thermal imaging to the tele-diagnosis of sarcoptic mange in the Spanish ibex (*Capra pyrenaica*). *Veterinary Parasitology*, **109**, 111–117.

Bakken, G.S., Van Sant, M.J., Lynott, A.J. & Banta, M.R. (2005) Predicting small endotherm body temperatures from scalp temperatures. *Journal of Thermal Biology*, **30**, 221–228.

Barber, D.G., Richard, P.R., Hochheim, K.P. & Orr, J. (1991) Calibration of aerial thermal infrared imagery for walrus population assessment. *Arctic*, **44**, 58–65.

Benesch, A.R. & Hilsberg, S. (2003) Infrared thermographic study of surface temperature in zebras. *Zoologische Garten*, **73**, 74–82.

Berry, R.J., Kennedy, A.D., Scott, S.L., Kyle, B.L. & Schaefer, A.L. (2003) Daily variation in the udder surface temperature of dairy cows measured by infrared thermography: potential for mastitis detection. *Canadian Journal of Animal Science*, **83**, 687–693.

Blix, A.S., Grav, H.J. & Ronald, K. (1979) Some aspects of temperature regulation in newborn harp seal pups. *American Journal of Physiology*, **236**, R188–R197.

Blumberg, M.S., Efimova, I.V. & Alberts, J.R. (1992) Thermogenesis during ultrasonic vocalization by rat pups isolated in a warm environment: a thermographic analysis. *Developmental Psychobiology*, **25**, 497–510.

Boonstra, R., Krebs, C.J., Boutin, S. & Eadie, J.M. (1994) Finding mammals using far-infrared thermal imaging. *Journal of Mammalogy*, **75**, 1063–1068.

Brooks, J.W. (1972) Infra-red scanning for polar bear. In: *Bears Their Biology and Management Papers and Proceedings of the International Conference on Bear Research and Management Calgary, Alberta, Canada 6–9 Nov. 1970*. (Ed. by S. Herrero), pp. 138–141. IUCN, Morges, Switzerland.

Burn, D.M., Webber, M.A. & Udevitz, M.S. (2006) Application of airborne thermal imagery to surveys of Pacific walrus. *Wildlife Society Bulletin*, **34**, 51–58.

Burnay, S.G., Williams, T.L. & Jones, C.H. (1988) *Applications of Thermal Imaging*. Adam Hilger, Bristol, UK.

Butler, D.A., Ballard, W.B., Haskell, S.P. & Wallace, M.C. (2006) Limitations of thermal infrared imaging for locating neonatal deer in semiarid shrub communities. *Wildlife Society Bulletin*, **34**, 1458–1462.

Butler, P.J., Bevan, R.M., Woakes, A.J., Croxall, J.P. & Boyd, I.L. (1995) The use of data loggers to determine the energetics and physiology of aquatic birds and mammals. *Brazilian Journal of Medical and Biological Research*, **28**, 1307–1317.

Campbell, G.S. & Norman, J.M. (1998) *An Introduction to Environmental Biophysics*, 2nd edn. Springer Verlag, New York, NY, USA.

Cena, K. (1974) Radiative heat loss from anaimals and man. In: *Heat Loss from Animals and Man* (Ed. by J.L. Monteith & L.E. Mount), pp. 33–58. Butterworths, London, UK.

Cena, K. & Clark, J.A. (1973) Thermographic measurements of the surface temperatures of animals. *Journal of Mammalogy*, **54**, 1003–1007.

Clark, J.A. (1976) Effects of surface emissivity and viewing angle on errors in thermography. *Acta Thermographica*, **1**, 138–141.

Colles, C., Holah, G. & Pusey, A. (1995) Thermal imaging as an aid to the diagnosis of back pain in the horse. In: *The Thermal Image in Medicine and Biology* (Ed. by K. Ammer & E.F.J. Ring), pp. 164–167. Uhlen-Verlag, Wien, Austria.

Cook, N.J. & Schaefer, A.L. (2002) Stress responses of wapiti (*Cervus elaphus canadensis*) to removal of velvet antler. *Canadian Journal of Animal Science*, **82**, 11–17.

Croon, G.W., McCullough, D.R., Olson, C.E. & Queal, L.M. (1968) Infrared scanning techniques for big game censusing. *Journal of Wildlife Management*, **32**, 751–759.

Cuyler, L.C., Wiulsrød, R. & Ørtisland, N.A. (1992) Thermal infrared radiation from free living whales. *Marine Mammal Science*, **8**, 120–134.

Dehnhardt, G., Mauck, B. & Hyvärinen, H. (1998) Ambient temperature does not affect the tactile sensitivity of mystacial vibrissae in harbour seals. *Journal of Experimental Biology*, **201**, 3023–3029.

Ditchkoff, S.S., Raglin, J.B., Smith, J.M. & Collier, B.A. (2005) From the field: capture of white-tailed deer fawns using thermal imaging technology. *Wildlife Society Bulletin*, **33**, 1164–1168.

Doncaster, C.P., Dumonteil, E., Barré, H. & Jouventin, P. (1990) Temperature regulation of young coypus (*Myocastor coypus*) in air and water. *American Journal of Physiology*, **259**, R1220–R1227.

Duck, C.D. & Thompson, D. (2003) The status of British common seal populations. Scientific advice on matters related to the management of seal populations SCOS Briefing Paper 03/7, 47–53.

Eddy, A.L., Van Hoogmoed, L.M. & Snyder, J.R. (2001) The role of thermography in the management of equine lameness. *The Veterinary Journal*, **162**, 172–181.

Eicher, S.D., Cheng, H.W., Sorrells, A.D. & Schutz, M.M. (2006) Short communication: behavioral and physiological idicators of sensitivity or chronic pain following tail docking. *Journal of Dairy Science*, **89**, 3047–3051.

Graves, H.B., Bellis, E.D. & Knuth, W.M. (1972) Censusing white-tailed deer by airborne thermal infrared imagery. *Journal of Wildlife Management*, **36**, 875–884.

Head, M.J. & Dyson, S. (2001) Talking the temperature of equine thermography. *The Veterinary Journal*, **162**, 166–167.

Hill, R.W., Christian, D.P. & Veghte, J.H. (1976) Pinna temperature in exercising jackrabbits, *Lepus californicus*. *Journal of Mammalogy*, **61**, 30–38.

Hill, R.W., Christian, D.P. & Veghte, J.H. (1980) Jackrabbit ears: surface temperatures and vascular responses. *Science*, **194**, 436–438.

Holah, G.D. (1995) Some practical aspects of thermography in competition horses. In: *The Thermal Image in Medicine and Biology* (Ed. by K. Ammer & E.F.J. Ring), pp. 156–163. Uhlen-Verlag, Wien, Austria.

Hristov, N., Betke, M. & Kunz, T.H. (2005) Seasonal monitoring of Brazilian free-tailed bats at Carlsbad Caverns National Park using advanced infrared thermal imaging. *Bat Research News*, **46**, 182.

Hurnik, J.F., Webster, A.B. & DeBoer, S. (1985) An investigation of skin temperature differentials in relation to estrus in dairy cattle using a thermal infrared scanning technique. *Journal of Animal Science*, **61**, 1095–1102.

Jackson, D.M., Hambly, C., Trayhurn, P. & Speakman, J.R. (2001) Can non-shivering thermogenesis in brown adipose tissue following NA injection be quantified by changes in overlying surface temperatures using infrared thermography? *Journal of Thermal Biology*, **26**, 85–93.

Kastberger, G. & Stachl, R. (2003) Infrared imaging technology and biological applications. *Behavior Research Methods, Instruments and Computers*, **35**, 429–439.

Kastelic, J.P., Cook, R.B., Coulter, G.H., Wallins, G.L. & Entz, T. (1996) Environmental factors affecting measurement of bovine scrotal surface temperature with infrared thermography. *Animal Reproduction Science*, **41**, 153–159.

Keith, E.O. (2002) *Boater Manatee Awareness System. Final Report*. Fwc01-02-14. Florida Fish and Wildlife Conservation Commission, St Petersburg, FL, USA.

Kingsley, M.C., Hammill, M.O. & Kelly, B.P. (1990) Infrared sensing of the under-snow lairs of the ringed seal. *Marine Mammal Science*, **6**, 339–347.

Klir, J.J. & Heath, J.E. (1992) An infrared thermographic study of surface temperature in relation to external thermal stress in three species of foxes: the red fox (*Vulpes vulpes*), arctic fox (*Alopex lagopus*), and kit fox (*Vulpes macrotis*). *Physiological Zoology*, **65**, 1011–1021.

Klir, J.J., Heath, J.E. & Bennani, N. (1990) An infrared thermographic study of surface temperature in relation to external thermal stress in the Mongolian gerbil *Meriones unguiculatus*. *Comparative Biochemistry and Physiology A*, **96**, 141–146.

Knizkova, I., Kunc, P., Koubkova, M., Flusser, J. & Dolezal, O. (2002) Evaluation of naturally ventilated dairy barn management by a thermographic method. *Livestock Production Science*, **77**, 349–353.

Korhonen, H. & Harri, M. (1986) Heat loss of farmed raccoon dogs and blue foxes as evaluated by infrared thermography and body cooling. *Comparative Biochemistry and Physiology A*, **84**, 361–364.

Kouba, A. & Willard, S. (2005) What's new and hot in zoo technology: thermography. American Zoo and Aquarium Association Communiqué March 2005, 10–13.

Lancaster, W.C., Thomson, S.C. & Speakman, J.R. (1997) Wing temperature in flying bats measured by infrared thermography. *Journal of Thermal Biology*, **22**, 109–116.

McArthur, A.J. & Ousey, J.C. (1994) Heat-loss from a wet animal – changes with time in the heat-balance of a physical model representing a newborn homeotherm. *Journal of Thermal Biology*, **19**, 81–89.

McCafferty, D.J., Moncrieff, J.B., Taylor, I.R. & Boddie, G.F. (1998) The use of IR thermography to measure the radiative temperature and heat loss of a barn owl (*Tyto alba*). *Journal of Thermal Biology*, **23**, 311–318.

McCafferty, D.J., Currie, J. & Sparling, C.E. (2007) The effect of instrument attachment on the surface temperature of juvenile grey seals (*Halichoerus grypus*) as measured by infrared thermography. *Deep Sea Research II*, **54**(3–4), 424–436.

McCullough, D.R., Olson, C.E. & Queal, L.M. (1969) Progress in large animal census by thermal mapping. In: *Remote Sensing in Ecology* (Ed. by P.L. Johnson), pp. 138–147. University of Georgia Press, Athens, GA, USA.

Mauck, B., Eysel, U. & Dehnhardt, G. (2000) Selective heating of vibrissal follicles in seals (*Phoca vitulina*) and dolphins (*Sotalia fluviatilis guiaensis*). *Journal of Experimental Biology*, **203**, 2125–2131.

Mauck, B., Bilgmann, K., Jones, D.D., Eysel, U. & Dehnhardt, G. (2003) Thermal windows on the trunk of hauled-out seals: hot spots for thermoregulatory evaporation? *Journal of Experimental Biology*, **206**, 1727–1738.

Meagher, E.M., McLellan, W.A., Westgate, A.J., Wells, R.S., Frierson, D. & Pabst, D.A. (2002) The relationship between heat flow and vasculature in the dorsal fin of wild bottlenose dolphins *Tursiops truncatus*. *Journal of Experimental Biology*, **205**, 3475–3486.

Mohler, F.S. & Heath, J.E. (1988) Comparison of IR thermography and thermocouple measurment of heat loss from rabbit pinna. *American Journal of Physiology*, **254**, R389–R395.

Monteith, J.L. & Unsworth, M.H. (1990) *Principles of Environmental Physics*, 2nd edn. Edward Arnold, London, UK.

Nakayama, K., Goto, S., Kuraoka, K. & Nakamura, K. (2005) Decrease in nasal temperature of rhesus monkeys (Macaca mulatta) in negative emotional state. *Physiology and Behavior*, **84**, 783–790.

Nikkhah, A., Plaizier, J.C., Einarson, M.S., Berry, R.J., Scott, S.L. & Kennedy, A.D. (2005) Infrared thermography and visual examination of hooves of dairy cows in two stages of lactation. *Journal of Dairy Science*, **88**, 2749–2753.

Ørtisland, N.A. (1968) Variations in the body surface temperature of the harp seal. *Acta Physiologica Scandinavica*, **73**, 35–36.

Ørtisland, N.A., Lentfer, J.W. & Ronald, K. (1974) Radiatve surface temperatures of the polar bear. *Journal of Mammalogy*, **55**, 459–461.

Pabst, D.A., McLellan, W.A., Meagher, E.M. & Westgate, A.J. (2002) *Measuring Temperatures and Heat Flux from Dolphins in the Eastern Tropical Pacific: is Thermal Stress Associated with Chase and Capture in the ETP-tuna Purse Seine Fishery*. LJ-02-34C, pp. 1–52. National Marine Fisheries Service NOAA, La Jolla, CA, USA.

Palmar, S.E. (1981) Use of the portable infrared thermometer as a means of measuring limb surface temperature in the horse. *American Journal of Veterinary Research*, **42**, 105–108.

Perryman, W.L., Donahue, M.A., Laake, J.L. & Martin, T.E. (1999) Diel variation in migration rates of eastern pacific gray whales measured with thermal imaging sensors. *Marine Mammal Science*, **15**, 426–445.

Phillips, P.K. & Heath, J.E. (1995) Dependency of surface temperature regulation on body size in terrestrial mammals. *Journal of Thermal Biology*, **20**, 281–289.

Phillips, P.K. & Heath, J.E. (2001) An infrared thermographic study of surface temperature in the euthermic woodchuck (*Marmota monax*). *Comparative Biochemistry and Physiology A*, **129**, 557–562.

Phillips, P.K. & Heath, J.E. (2004) Comparison of surface temperature in 13-lined ground squirrel (Spermophilus tridecimlineatus) and yellow-bellied marmot (Marmota flaviventris) during arousal from hibernation. *Comparative Biochemistry and Physiology A*, **138**, 451–457.

Purohit, R.C. & McCoy, M.D. (1980) Thermography in the diagnosis of inflammatory processes in the horse. *American Journal of Veterinary Research*, **41**, 1167–1174.

Purohit, R.C., Hudson, R.S., Riddell, M.G., Carson, R.L., Wolfe, D.F. & Walker, D.F. (1985) Thermography of the bovine scrotum. *American Journal of Veterinary Research*, **46**, 2392.

Reichard, J.D., Frank, A.J. & Kunz, T.H. (2005) Ecological and behavioral energetics of Brazilian free-tailed bats (*Tadarida brasiliensis*) using infrared thermography. *Bat Research News*, **46**, 211–212.

Sabol, B.M. & Hudson, M.K. (1995) Tecnique using thermal infrared-imaging for estimating populations of gray bats. *Journal of Mammalogy*, **76**, 1242–1248.

Sandel, U., Kiefer, A., Prinzinger, R. & Hilsberg, S. (2004) Behavioural thermoregulation in Greater mouse-eared bats, *Myotis myotis*, studied by infrared thermography. *Myotis*, **41–42**, 129–142.

Schaefer, A.L., Jones, S.D.M., Tong, A.K.W. & Vincent, B.C. (1988) The effects of fasting and transportation on beef cattle. 1. Acid-base-electrolyte balance and infrared heat loss of beef cattle. *Livestock Production Science*, **20**, 15–24.

Schaefer, A.L., Cook, N., Tessaro, S.V., Deregt, D., Desroches, G., Dubeski, P.L., Tong, A.K.W. & Godson, D.L. (2004) Early detection and prediction of infection using infrared thermography. *Canadian Journal of Animal Science*, **84**, 73–80.

Schwartzkopf-Genswein, K.S. & Stookey, J.M. (1997) The use of infrared thermography to assess inflammation associated with hot-iron and freeze branding in cattle. *Canadian Journal of Animal Science*, **77**, 577–583.

Sipilä, T. & Kurlin, K. (1992) Infrared videothermography as a method for locating the under-snow lairs of the ringed seal. In: *Wildlife Telemetry: Remote Monitoring and Tracking of Animals* (Ed. by I.M. Priede & S.M. Swift), pp. 35–40. Ellis Horwood, New York, USA.

Sokoloff, G. & Blumberg, M.S. (2001) Competition and cooperation among huddling infant rats. *Developmental Psychobiology*, **39**, 65–75.

Speakman, J.R. & Ward, S. (1998) Infrared thermography: principles and applications. *Zoology: Analysis of Complex Systems*, **101**, 224–232.

Stewart, M., Webster, J.R., Schaefer, A.L., Cook, N.J. & Scott, S.L. (2005) Infrared thermography as a non-invasive tool to study animal welfare. *Animal Welfare*, **14**, 319–325.

Sykes, D., Chromiak, A., Couvillion, S., Gerard, P., Crenshaw, M., Willard, S. & Ryan, P. (2006) Estrus detection in gilts using digital infrared thermal imaging. *Journal of Animal Science*, **84**, 1.

Tattersall, G.J. & Milsom, W.K. (2003) Transient peripheral warming accompanies the hypoxix metabolic response in the golden-mantled ground squirrel. *Journal of Experimental Biology*, **206**, 33–42.

Turner, T.A., Fessler, J.F., Lamp, M., Pearce, J.A. & Geddes, L.A. (1983) Thermographic evaluation of horses with podotrochlosis. *American Journal of Veterinary Research*, **44**, 535–539.

Van Hoogmoed, L.M. & Snyder, J.R. (2002) Use of infrared thermography to detect injections and palmar digital neurectomy in horses. *The Veterinary Journal*, **164**, 129–141.

Vianna, D.M.L. & Carrive, P. (2005) Changes in cutaneous and body temperature during and after conditioned fear to context in the rat. *European Journal of Neuroscience*, **21**, 2505–2512.

Ward, S. & Slater, P.J.B. (2005) Heat transfer and the energetic cost of singing by canaries *Serinus canaria*. *Journal of Comparative Physiology A – Neuroethology Sensory Neural and Behavioral Physiology*, **191**, 953–964.

Ward, S., Rayner, J.M., Moller, U., Jackson, D.M., Nachtigall, W. & Speakman, J.R. (1999) Heat transfer from starlings *Sturnus vulgaris* during flight. *Journal of Experimental Biology*, **202**, 1589–1602.

Ward, S., Moller, U., Rayner, J.M.V., Jackson, D.M., Nachtigall, W. & Speakman, J.R. (2004) Metabolic power of European starlings *Sturnus vulgaris* during flight in a wind tunnel, estimated from heat transfer modelling, doubly labelled water and mask respirometry. *Journal of Experimental Biology*, **207**, 4291–4298.

Watmough, D.J., Fowler, P.W. & Oliver, R. (1970) The thermal scanning of a curved isothermal surface: implications for clinical thermography. *Physics in Medicine and Biology*, **15**, 1–8.

Webb, P.I., Speakman, J.R. & Racey, P.A. (1993) The implication of small reductions in body temperature for radiant and convective heat loss in resting endothermic brown long-eared bats (*Plecotus auritus*). *Journal of Thermal Biology*, **18**, 131–135.

Webbon, P.M. (2002) Prohibited practices in equine sport – how to root out malpractice. *The Veterinary Journal*, **164**, 83–84.

Wiggers, E.P. & Beckerman, S.F. (1993) Use of thermal infrared sensing to survey white-tailed deer populations. *Wildlife Society Bulletin*, **21**, 263–268.

Willard, S.T., Vinson, M.C. & Godfrey, R.W. (2006) Digital infrared thermal imaging of the eye as correlated to rectal and vaginal temperature measurements in the ewe. *Journal of Animal Science*, **84**, 434.

Williams, T.M. (1990) Heat transfer in elephants: thermal partitioning based on skin temperature profiles. *Journal of Zoology (London)*, **222**, 235–245.

Williams, T.M., Noren, D.P., Berry, P., Estes, J.A., Allison, C. & Kirtland, J. (1999) The diving physiology of the bottlenose dolphin (*Tursiops truncatus*). III. Thermoregulation at depth. *Journal of Experimental Biology*, **202**, 2763–2769.

Willis, K., Horning, M., Rosen, D.A.S. & Trites, A.W. (2005) Spatial variation of heat flux in Steller sea lions: evidence for consistent avenues of heat exchange along the body trunk. *Journal of Experimental Marine Biology and Ecology*, **315**, 163–175.

Yang, W.-J. & Yang, P.P.T. (1992) Literature survey on biomedical applications of thermography. *Bio-Medical Materials and Engineering*, **2**, 7–18.

York, G., Amstrup, S. & Simac, K. (2004) *Using Forward Looking Infrared (FLIR) Imagery to Detect Polar Bear Maternal Dens Operations Manual*. OCS Study Report MMS, 2004-062.

Zähner, M., Schrader, L., Hauser, R., Keck, M., Langhans, W. & Wechsler, B. (2004) The influence of climatic conditions on physiological and behavioural parameters in dairy cows kept in open stables. *Animal Science*, **78**, 139–147.

Submitted 29 August 2006; returned for revision 6 February 2007; revision accepted 22 June 2007
Editors: RM & JS

Conservation Biology

Conservation Methods

Feasibility of using high-resolution satellite imagery to assess vertebrate wildlife populations

Michelle A. LaRue,[*] ¶ Seth Stapleton,† and Morgan Anderson‡

[*]Department of Earth Sciences, University of Minnesota, 310 Pillsbury Drive SE, Minneapolis, MN 55455, U.S.A.
†Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, St. Paul, MN 55108, U.S.A.
‡Department of Environment, Government of Nunavut, Igloolik, Nunavut X0A 0L0, Canada

Abstract: *Although remote sensing has been used for >40 years to learn about Earth, use of very high-resolution satellite imagery (VHR) (<1-m resolution) has become more widespread over the past decade for studying wildlife. As image resolution increases, there is a need to understand the capabilities and limitations of this exciting new path in wildlife research. We reviewed studies that used VHR to examine remote populations of wildlife. We then determined characteristics of the landscape and the life history of species that made the studies amenable to use of satellite imagery and developed a list of criteria necessary for appropriate use of VHR in wildlife research. From 14 representative articles, we determined 3 primary criteria that must be met for a system and species to be appropriately studied with VHR: open landscape, target organism's color contrasts with the landscape, and target organism is of detectable size. Habitat association, temporal exclusivity, coloniality, landscape differentiation, and ground truthing increase the utility of VHR for wildlife research. There is an immediate need for VHR imagery in conservation research, particularly in remote areas of developing countries, where research can be difficult. For wildlife researchers interested in but unfamiliar with remote sensing resources and tools, understanding capabilities and current limitations of VHR imagery is critical to its use as a conservation and wildlife research tool.*

Keywords: conservation research, GIS, population monitoring, remote-sensing methods

Viabilidad de la Utilización de Imágenes Satelitales de Alta Resolución para Evaluar Poblaciones Silvestres de Vertebrados

Resumen: *Aunque la telemetría se ha utilizado por más de 40 años para aprender sobre la Tierra, el uso de imágenes satelitales de muy alta resolución (MAR) (<1-m de resolución) se ha vuelto más extendido en la última década como herramienta de estudio de la vida silvestre. Conforme incrementa la resolución de las imágenes, existe una necesidad de entender las capacidades y limitaciones de este nuevo y emocionante camino en la investigación de la vida silvestre. Revisamos estudios que utilizaron MAR para examinar poblaciones remotas de vida silvestre. Después determinamos las características del paisaje y de la historia de vida de las especies que hicieron a los estudios más dispuestos para el uso de imágenes satelitales y desarrollamos una lista de criterios necesarios para el uso correcto de MAR en la investigación de la vida silvestre. De 14 artículos representativos determinamos tres criterios primarios que deben cumplirse para que un sistema y las especies sean estudiadas correctamente con MAR: paisaje abierto, los contrastes de color del organismo estudiado con el paisaje y si el organismo estudiado es de un tamaño detectable. La asociación de hábitat, la exclusividad temporal, la colonización, la diferenciación del paisaje y el incremento en la verificación en el terreno incrementan la utilidad de MAR para la investigación de la vida silvestre. Existe una necesidad inmediata por las imágenes MAR en la investigación de la conservación, particularmente en áreas remotas de los países en desarrollo. Para los investigadores de la vida silvestre que están interesados pero que no están familiarizados con los recursos ni con las herramientas de telemetría, entender las capacidades y las limitaciones actuales de las imágenes MAR es crítico para su uso como una herramienta de conservación e investigación de la vida silvestre.*

¶*email*

Paper submitted June 29, 2016; revised manuscript accepted August 12, 2016.

Conservation Biology, Volume 31, No. 1, 213–220

DOI: 10.1111/cobi.12809

Palabras Clave: investigación de la conservación, métodos de telemetría, monitoreo poblacional, SIG

Introduction

The use of very high-resolution satellite imagery (VHR) (i.e., submeter resolution images) to assess wildlife populations has increased substantially over a relatively short time. Although Schwaller et al. (1984) first determined that 15-m resolution Landsat could be used to index Adélie Penguin (*Pygoscelis adeliae*) populations on Ross Island, Antarctica, Barber-Meyer et al. (2007) pioneered the use of VHR as a resource for research in polar ecology by using a supervised classification technique (i.e., training a computer to differentiate between image pixels) on panchromatic QuickBird-2 images to determine abundance of 7 populations of Emperor Penguins (*Aptenodytes forsteri*) in the Ross Sea. Since Barber-Meyer et al. (2007), >20 studies have been published in which VHR was used to assess populations of wildlife. Such work has also included estimates of animal abundance in the Arctic, African grasslands, and open ocean (Fretwell et al. 2014*a*; Stapleton et al. 2014*a*; Yang et al. 2014). The ability to remotely assess and monitor wildlife populations has the potential to revolutionize such monitoring in remote regions, particularly as technology advances and spatial and spectral resolution of satellite platforms improve.

The use of VHR will likely become more widespread in wildlife ecology and conservation, particularly as costs of the images decrease. To efficiently and effectively implement this resource, however, researchers must understand its capabilities and limitations. Our objective was to develop a feasibility guide for the use of VHR imagery in wildlife research. We conducted a comprehensive literature review of published articles in which VHR was used to assess wildlife populations. We qualitatively described the characteristics of the systems, including the landscape and the life history traits of the target species, and then identified those features that made VHR a suitable method for measuring occupancy and abundance of wildlife. Across 14 articles, we identified 3 primary criteria that must be met but do not guarantee success and 5 additional criteria that enhance the effectiveness of use of VHR in wildlife research.

Primary Criteria for Use of VHR

We identified an open landscape, organism color contrasts with the landscape, and sufficient organism size as the 3 main criteria that must be met for VHR to have utility for assessing wildlife (Table 1). The landscape over which the images would be taken must be open, such that the species of interest cannot be hidden. Polar vertebrates such as emperor penguins, Adélie penguins, Weddell seals (*Leptonychotes weddellii*), southern elephant seals (*Mirounga leonine*), polar bears (*Ursus maritmus*), and walruses (*Odobenus rosmarus*); large-bodied vertebrates on the African plains such as elephants (*Loxodonta africana*); and right whales (*Eubalaena australis*) at the surface of the water in calm conditions are examples of species living in areas where the landscape is accessible for detection via VHR (LaRue et al. 2011; Boltunov et al. 2012; Fretwell et al. 2012; Fretwell et al. 2014*a*; Lynch & LaRue 2014; Stapleton et al. 2014*a*; Yang et al. 2014). The open landscape is the primary reason that many of these studies have been conducted in polar regions, where vegetation cover that could preclude detection does not exist. Even in open landscapes, cloud cover and steep topography render detection difficult and sometimes impossible (Fretwell et al. 2012; Lynch & LaRue 2014).

An individual organism must be large enough for detection on images with resolutions of approximately 0.60 m (panchromatic images). Polar bears, for example, which are approximately 2 m long and 1 m wide, clearly appear on VHR as large white spots on a snow-free landscape (Stapleton et al. 2014*a*; LaRue et al. 2015). If organisms are too small to be individually identified on imagery, proxies may be used to infer presence. Such proxies are called positive indicators and are considered indirect remote sensing (LaRue & Knight 2014). Masked Boobies (*Sula dactylatra*) have a "nest signature"; the ground is cleared away leaving a conspicuous circle of dirt where nesting is occurring (Hughes et al. 2011). Adélie Penguins leave a unique guano stain, the size of which is positively correlated with the number of concurrent breeding pairs, which allows researchers to measure abundance (LaRue et al. 2014*b*).

The color of the target species (or of its positive indicator) must contrast with the surrounding landscape. For example, Weddell seals appear as black spots on white ice (Fig. 1) (LaRue et al. 2011; Ainley et al. 2015), and Masked Booby nests appear as bare ground among the surrounding green grass (Hughes et al. 2011). All 3 criteria are currently required for a vertebrate species to be detected via 0.5-m resolution VHR imagery. However, we suggest that these criteria are an absolute minimum and alone may not be sufficient to guarantee the effective use of imagery for wildlife research.

Secondary Criteria

Additional criteria increase the likelihood of detection and enhance the utility of using remote sensing as a monitoring and assessment tool. The ability to differentiate

Table 1. Criteria of imagery, landscape, and life history of vertebrate species necessary for using very high-resolution satellite imagery (VHR) in wildlife research.[a]

	Minimum necessary				*Enhances utility*				
Species (reference)	*open landscape*	*body size*	*positive indicator*	*contrast*	*habitat association*	*temporal exclusivity*	*coloniality*	*landscape differentiation*	*ground-truthing*
Emperor Penguins (Fretwell et al. 2012)	x	x	x	x	x	x	x	x	x
Weddell seals (LaRue et al. 2011)	x	x		x	x	x	x	x	x
Adélie Penguins (LaRue et al. 2014*b*; Lynch & LaRue 2014)	x		x	x	x	x	x	x	x
Chinstrap penguins (Naveen et al. 2012)	x		x	x	x	x	x	x	x
Southern right whales (Fretwell et al. 2014*a*)	x	x		x	x	x	x	x	
Walrus (Boltunov et al. 2012)	x	x		x	x	x	x	x	
Domestic cattle (Begall et al. 2008)	x	x		x	x	x	x	x	
Masked Boobies (Hughes et al. 2011)	x		x	x		x	x	x	
Gentoo Penguins (Lynch et al. 2012)	x		x	x			x	x	
Macaroni Penguins (Lynch et al. 2012)	x		x	x			x	x	
Southern elephant seals (McMahon et al. 2014)	x	x		x			x	x	
Polar bears (Stapleton et al. 2014*a*; LaRue et al. 2015)	x	x		x	x			x	x
African animals (Yang et al. 2014)	x	x		x			x	x	
Musk ox	x	x		x			x		x

[a] *An open landscape; large body size or a positive indication of species' presence; and organism's color contrasting with the landscape are the minimum necessary (but not sufficient) requirements for using VHR.*

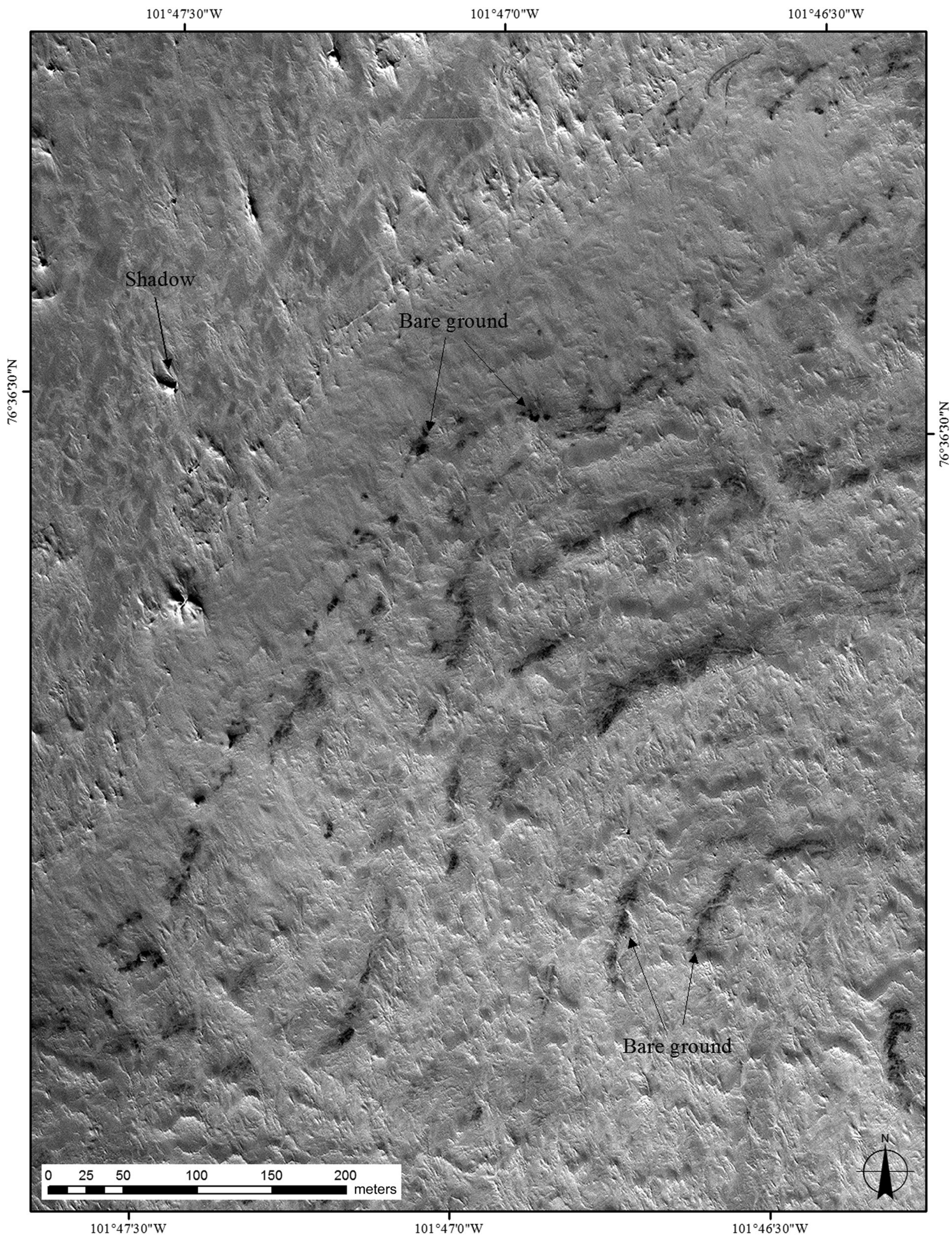


Figure 1. Snow-covered WorldView-1 image (0.6-m resolution, courtesy DigitalGlobe Foundation) of Bathurst Island, Canada, acquired 29 April 2013. Muskoxen were available for detection on Bathurst Island in April 2013 but because of the bare ground, shadows, and rock outcrops on images, we were unable to positively identify muskoxen on the imagery.

the target species from other objects on the landscape improves detection (Table 1). We contend that this is the most important secondary criterion. An individual may be a sufficient size for detection and may contrast with the landscape, but if it is not possible to differentiate the target species from other objects (e.g., rocks or pools of water), then one has less confidence in the accuracy of the estimate derived from the imagery. Polar bears on Rowley Island, Canada, provide a unique example in which combining knowledge of a species' life history with remote-sensing techniques can solve the problem of definitive differentiation (LaRue et al. 2015). In this case, image differencing—where one image is subtracted from another such that only the different pixels remain—allows researchers to differentiate rocks from bears.

Habitat associations of the target species (Table 1), referring to the organism's reliable presence at a specific location, improve detection of vertebrates via VHR. For example, the emperor penguin has a circumpolar distribution in the Antarctic and is typically associated only with land-fast sea ice (also called fast ice) (Kooyman et al. 2000; Fretwell et al. 2012). Detecting the guano signature on the ice is a sure indication of an emperor penguin colony and importantly nothing else because these birds are the only sea birds that breed on Antarctic fast ice (Le Maho 1977). Using VHR, Fretwell et al. (2014*b*) identified that some emperor penguin colonies use glacial or ice shelf in years of poor sea-ice conditions. This guano stain has facilitated the detection of >20 previously unknown colonies via remote sensing (Fretwell & Trathan 2009; Fretwell et al. 2012; LaRue et al. 2014*a*).

Temporal exclusivity—meaning only the target population occupies an area at specific time—is another criterion that improves the utility of VHR. Weddell seals have a circumpolar distribution and are associated with fast ice. Every austral spring, Weddell seals return to the same locations to give birth and raise pups (Stirling 1969; Siniff et al. 1977; Siniff 1981), but there is a specific window of time from which one can determine the size of a breeding population (Banner 2012). Although reproductive female seals are present on the fast ice from early spring through the molt (the following January), their activity is least variable during the first 2 weeks in November. By early November, females have given birth and are spending more time hauled out on the ice raising their pups (Stirling 1969; Cameron & Siniff 2004); nonreproductive females tend to be excluded from these core breeding locations (Stirling 1971). Thus, the likelihood of including nonbreeders in abundance estimates is minimal. Perhaps most importantly, because the pups are very young in early November, they are too small to be detected via satellite imagery (LaRue et al. 2011) and are not included in estimates of the breeding population size.

Detection improves if the target species is colonial or congregates in herds (Table 1). For example, Yang et al. (2014) used VHR and object-based image analysis to detect large-bodied animals on the African savannah. Begall et al. (2008) used Google Earth to demonstrate the magnetic orientation of domestic cattle in pastures across 6 continents. When animals congregate, detection is far easier than if individuals were solitary because detection probability generally increases with greater herd or colony size (Rivest et al. 1998). Conversely, polar bears live in open spaces in the Arctic, are large-bodied, and contrast with ice-free landscapes in some regions. Despite this, reliable detection can still be challenging because of their typically solitary behavior (Stapleton et al. 2014*b*; LaRue et al. 2015).

Finally, ground-truthing is an important secondary criterion that allows for increased confidence in both detection rates and abundance estimates. The first global estimates of both Antarctic penguin species were completed with ground validation built into models (Fretwell et al. 2012; Lynch & LaRue 2014), and confidence in detecting variability in the Weddell seal population in Erebus Bay was only possible because of comparison with concurrent ground counts (LaRue et al. 2011). In fact, it was only through ground validation that LaRue et al. (2011) determined differences in detection probability based on sea-ice conditions and that the increase in counts from images reflected a true increase in the population from 2004 to 2009. Conversely, Fretwell et al. (2014*a*) determined the presence of southern right whales (*E. australis*) off the coast of Argentina without direct ground-truthing because general knowledge of the species' life history was sufficient. Thus, ground validation benefits any pilot project, but detection of large vertebrates via VHR imagery is still possible without it.

Muskoxen

We evaluated our criteria while using VHR to search for muskoxen (*Ovibus moschatus*) on Bathurst Island, Canada. We searched approximately one-third of the 16,000-km^2 island. From an ecological perspective, muskoxen fulfill all primary criteria: their landscape is open and flat, adult muskoxen primarily live in herds and are large bodied, and they contrast with the surrounding landscape (Heard 1992). We hypothesized that they are readily apparent on snow-covered images and anticipated that generating an abundance estimate for comparison with aerial surveys (as in Stapleton et al. 2014*b*) would be straightforward.

Briefly, we obtained WV-01 and WV-02 VHR images of Bathurst Island, Canada (DigitalGlobe Foundation, Westminster, Colorado) acquired during April and May 2013. In an effort to identify muskoxen on VHR, we selected locations that overlapped in space and time with aerial surveys conducted by the Government of Nunavut. Although we knew that >1500 muskoxen were available for detection (Anderson 2014), we were unable to definitively identify them via manual detection on VHR images in ArcGIS 10.2 (for similar methods, see LaRue et al. [2011], Stapleton et al. [2014*b*], and LaRue et al. [2015]). We suspect the reason that we could not detect muskoxen was because of lack of habitat association and, in particular, insufficient differentiation from the landscape (Fig. 1). In other words, despite that muskoxen contrast spectrally with the landscape (black animals on white background), they were not different enough from other items on the landscape, such as rocks and outcrops, to be positively identified. Bathurst Island in April and May is mostly covered in snow, so large, dark muskoxen should appear easily on the images. The preference of muskoxen for well-vegetated lowlands is well known by Inuit and biologists (Thing et al. 1987; Ferguson 1991; Larter & Nagy 1997; Taylor 2005), but although they generally prefer productive lowlands, they are not restricted to these areas and use upland tundra and dwarf shrub habitats. Their habitat associations are not restrictive enough to allow us to rule out large areas of Bathurst Island. Furthermore, there are enough rock outcrops and boulders on the island and windswept ridges and deltas to make differentiating muskoxen from rocks impossible on 60-cm resolution images, even when toggling back and forth between 2 images taken at different times to visually inspect the images, as in Stapleton et al. (2014*a*).

Image differencing presents a possible solution to this problem. However, image differencing with VHR (especially in polar regions with topographic relief) may not have sufficient orthorectification necessary to line up both images exactly. Thus, even when one subtracts one image from another, there can still be a mismatch, such that rocks and outcrops appear to be different objects when they are not. In the case of polar bears on Rowley Island, LaRue et al. (2015) avoided this because the island is flat and relatively uniform across the landscape (little rubble and most rocks are large). Thus, our attempt to detect muskoxen on VHR provided an important lesson in the limitations of VHR for wildlife ecology. Our 3 primary criteria represent the bare minimum necessary for detection. These criteria alone, however, are not necessarily sufficient for accurate detection, and some combination of the secondary criteria may also need to be met.

Discussion

Satellite imagery is an important tool in ecological and conservation research (Turner et al. 2003; Leimgruber et al. 2005). Ecologists historically have been less inclined to use remotely sensed data and geographic information systems (GIS) programs due to lack of technical capacity (Leimgruber et al. 2005), but it appears that this capacity has increased recently along with an urgency in conservation research. Indeed, our review demonstrates the utility of VHR for vertebrates and is an encouraging step forward in better understanding populations of remote wildlife. However, what is most crucial at this point is gaining perspective on the current limitations and appropriate uses of this imagery. As scientists turn to VHR as a valuable monitoring tool, it is important to establish appropriate uses because misjudging the capacity of the method can cost valuable time, effort, and money without achieving desired outcomes. Traditionally, VHR imagery has been cost-prohibitive, and although licensing images can still be expensive, there are now multiple ways of piloting imagery, such as through the DigitalGlobe Foundation (where we obtained imagery used here) or through discounts for educational or nonprofit purposes.

There is an important and time-sensitive need for VHR imagery in ecological research, specifically for conservation efforts in remote areas of developing countries, where research efforts are depressed (Ripple et al. 2015). For example, developing countries contain 88% of populations of large herbivores that are amenable to research via VHR imagery, yet the number of peer-reviewed articles that include these threatened species in developing countries is substantially lower than those that include nonthreatened species in developed countries (Ripple et al. 2015). Presumably much of this disparity is due to funding, feasibility, and accessibility of remote regions in developing countries particularly in Africa and southeast Asia—some of which can be addressed through VHR imagery.

Given our criteria, it is obvious that research on these large herbivore species (e.g., African elephants [*Loxodonta africana*], white rhinoceros [*Ceratotherium simum*], and black rhinoceros [*Diceros bicornis*]) in Africa would benefit from the inclusion of VHR imagery in gathering some relatively basic data on population status, distribution, and trends. Not only do large African herbivores fit the criteria for successful VHR imagery use (open landscape, large-bodied herd animals distinct from the landscape, and specific habitat associations), but their populations are particularly threatened and declining (Ripple et al. 2015; Wasser et al. 2015; Kideghesho 2016). Poaching is one of the primary threats to elephants (Maisels et al. 2013); >100,000 elephants were poached in Africa from 2010 to 2012 (Wittemyer et al. 2014). Poaching of rhinoceros in South Africa increased from 13 in 2007 to >1000 in 2013 (Biggs et al. 2013; South African National Parks 2014). Most studies cited here (Table 1) were conducted in polar regions—which benefit from several of the criteria listed herein, such as open landscapes and contrast with the landscape—and on species that have been the subject of long-term research. Differentiating species (e.g., rhinos from elephants) will certainly take careful ground-truthing to determine differences in size, habitat associations, study areas, and times of year when animals are in identifiable locations. Only then can researchers be more confident in species identification and differentiation. The VHR imagery is increasing in spatial resolution (WV-03 has a spatial resolution of approximately 30 cm), and this will almost certainly help in species identification.

In addition to direct human persecution, conflict with livestock is another threat to African herbivores (Ripple et al. 2015), of which both populations should be detectable via VHR imagery. Pilot projects with VHR imagery used to determine distribution of elephant populations have occurred already (e.g., Satellite Sentinel Project), but we suggest that with our criteria, researchers will be able to more effectively plan for research on these and other species in similar habitats and subject to similarly large conservation threats. Lack of understanding of how to effectively use these resources may detract from important conservation efforts, so knowing the feasibility of use of VHR imagery is critical in the planning phases of research.

Acknowledgments

We thank the Government of Nunavut for funding this research and D. Libby at the DigitalGlobe Foundation for providing high-resolution satellite imagery of Bathurst Island, Canada. The Resolute Bay Hunters and Trappers Association provided support and assistance with fieldwork. T. Arnold provided crucial input on study design and comments on previous drafts of this manuscript.

Literature Cited

Ainley DG, LaRue MA, Stammerjohn SE, Siniff DB, Stirling I. 2015. Apparent decline of Weddell seal numbers along the Northern Victoria Land Coast. Marine Mammal Science **31:**1338–1361.

Anderson M. 2014. Distribution and abundance of Peary caribou *(Rangier tarandus pearyii)* and muskoxen *(Ovibus moschatus)* on the Bathurst Island group, May 2013. Status report, Government of Nunavut, Igloolik.

Banner KM. 2012. Correcting for time-of-day to compare satellite images and ground counts of Weddell seals in Big Razorback haul-out Erebus Bay, Antarctica. MS thesis. Montana State University, Bozeman.

Barber-Meyer SM, Kooyman GL, Ponganis PJ. 2007. Estimating the relative abundance of emperor penguin at inaccessible colonies using satellite imagery. Polar Biology **30:**1565–1570.

Begall S, Cerevny J, Neef J, Oldrich V, Burda H. 2008. Magnetic alignment in grazing and resting cattle and deer. Proceedings on the National Academy of Sciences of the United States of America **105:**13451–1355.

Biggs D, Courchamp F, Martin R, Possingham HP. 2013. Legal trade of Africa's rhino horns. Science **339:**1038–1039.

Boltunov A, Evtushenko H, Knijnikov A, Puhova M, Semenova V. 2012. Space technology for the marine mammal research and conservation in the Arctic: results of the pilot project to develop methods of finding walruses on satellite images. WWF, Moscow, Russia.

Cameron MF, Siniff DB. 2004. Age-specific survival, abundance, and immigration rates of a Weddell seal (*Leptonychotes weddellii*) population in McMurdo Sound, Antarctica. Canadian Journal of Zoology **82:**601–615.

Ferguson RS. 1991. Detection and classification of muskox habitat on Banks Island, Northwest Territories, Canada, using Landsat Thematic Mapper data. Arctic **44:**66–74.

Fretwell PT, LaRue MA, Morin P, Kooyman GL, Wienecke B, Ratcliffe N, Fox AJ, Fleming AH, Porter CC, Trathan PN. 2012. The first global, synoptic survey of a species from space: emperor penguins. PLoS ONE **7**(4):(e33751). doi:10.1371/journal.pone.0033751.

Fretwell PT, Staniland IJ, Forcada J. 2014*a*. Whales from space: counting southern right whales by satellite. PLoS ONE **9**(2):(e88655). doi:10.1371/journal.pone.0088655.

Fretwell PT, Trathan PN. 2009. Penguins from space: faecal stains reveal the location of emperor penguin colonies. Global Ecology and Biogeography **18:**543–552.

Fretwell PT, Trathan PN, Wienecke B, Kooyman GL. 2014*b*. Emperor penguins breeding on ice shelves. PLoS ONE **9**(1):(e85285). doi:10.1371/journal.pone.0085285.

Heard DC. 1992. The effect of wolf predation and snow cover on musk-ox group size. The American Naturalist **139:**190–204.

Hughes BJ, Martin GR, Reynolds SJ. 2011. The use of GoogleEarth satellite imagery to detect the nests of masked boobies *Sula dactylatra*. Wildlife Biology **17:**210–216.

Kideghesho J. 2016. The elephant poaching crisis in Tanzania: a need to reverse the trend and the way forward. Tropical Conservation Science **9:**369–388.

Kooyman GL, Hunke EC, Ackley SF, van Dam RP, Robertson G. 2000. Moult of the emperor penguin: travel, location and habitat selection. Marine Ecological Progress Series **204:**269–277.

Larter NC, Nagy JA. 1997. Peary caribou, muskoxen and Banks Island forage: assessing seasonal diet similarities. Rangifer **17:**9–16.

LaRue MA, Knight J. 2014. Applications of very high resolution imagery in the study and conservation of predators in the Southern Ocean. Conservation Biology **28:**1731–1735.

LaRue MA, Kooyman GL, Lynch HJ, Fretwell PT. 2014*a*. Emigration in emperor penguins: implications for interpretation of long-term studies. Ecography **38:**114–120.

LaRue MA, Lynch HJ, Lyver POB, Barton K, Ainley DG, Pollard AM, Ballard G. 2014*b*. A method for estimating colony sizes of Adélie penguins using remote sensing imagery. Polar Biology **37:**507–517.

LaRue MA, Rotella JJ, Garrott RA, Siniff DB, Ainley DG, Stauffer GE, Porter CC, Morin PJ. 2011. Satellite imagery can be used to detect variation in abundance of Weddell seals (*Leptonychotes weddellii*) in Erebus Bay, Antarctica. Polar Biology **34:**1727–1737.

LaRue MA, Stapleton SS, Porter CC, Atwood T, Lecomte N, Atkinson S. 2015. Testing methods for using high-resolution satellite imagery to monitor polar bear abundance and distribution. Wildlife Society Bulletin **39:**772–779.

Leimgruber P, Christen CA, Laborderie A. 2005. The impact of Landsat satellite monitoring on conservation biology. Environmental Monitoring and Assessment **106:**81–101.

Le Maho Y. 1977. The emperor penguin: a strategy to live and breed in the cold: morphology, physiology, ecology and behavior distinguish the polar emperor penguin from other penguin species particularly from its close relative, the king penguin. American Scientist **65:**680–693.

Lynch HJ, LaRue MA. 2014. The first global census of the Adélie penguin. The Auk **31:**457–466.

Lynch HJ, White R, Black AD, Naveen R. 2012. Detection, differentiation, and abundance estimation of penguin species by high-resolution satellite imagery. Polar Biology **35:**963–968.

Maisels F, et al. 2013. Devastating decline of forest elephants in central Africa. PLoS ONE **8**(3):(e59469). doi:10.1371/journal.pone.0059469.

McMahon CR, Howe H, van den Hoff J, Alderman R, Brolsma H, Hindell MA. 2014. Satellites, the all-seeing eyes in the sky: counting elephant seals from space. PLoS ONE **9**(3):(e92613). doi:10.1371/journal.pone.0092613.

Naveen R, Lynch HJ, Forrest S, Mueller T, Polito M. 2012. First direct, site-wide penguin survey at Deception Island, Antarctica, suggests significant declines in breeding chinstrap populations. Polar Biology **35:**1879–1888.

Ripple WJ, et al. 2015. Collapse of the world's largest herbivores. Science Advances **1:**e1400103. doi:10.1126/sciadv.1400103.

Rivest LP, Couturier S, Crepeau H. 1998. Statistical methods for estimating caribou abundance using postcalving aggregations detected by radio telemetry. Biometrics **54:**65–76.

Schwaller MR, Benninghoff WS, Olson Jr CE. 1984. Prospects for satellite remote-sensing of Adélie Penguin rookeries. International Journal of Remote Sensing **5:**849–853.

Siniff DB. 1981. Seal population dynamics and ecology. Journal of the Royal Society of New Zealand **11:**317–327.

Siniff DB, DeMaster DP, Hofman RJ, Eberhardt LL. 1977. An analysis of the dynamics of a Weddell seal population. Ecological Monographs **47:**319–335.

South African National Parks. 2014. Update on rhino poaching statistics. South African National Parks, Pretoria, South Africa. Available from http://sanparks.org/about/news/default.php?id=55976 (accessed December 2015).

Stapleton S, LaRue MA, Lecomte N, Atkinson S, Garshelis D, Porter CC, Atwood T. 2014*a*. Polar bears from space: assessing satellite imagery as a tool to monitor *Ursus maritimus*. PLoS ONE **9**(7)**:**(e101513) . doi:10.1371/journal.pone.0101513.

Stapleton SS, Atkinson S, Hedman D, Garshelis D. 2014*b*. Revisiting Western Hudson Bay: using aerial surveys to update polar bear abundance in a sentinel population. Biological Conservation **170:**38–47.

Stirling I. 1969. The ecology of the Weddell seal in McMurdo Sound, Antarctica. Ecology **4:**573–586.

Stirling I. 1971. Population dynamics of the Weddell seal (*Leptonychotes weddellii*) in McMurdo Sound, Antarctica, 1966–1968. Antarctic Research Series **18:**141–160.

Taylor ADM. 2005. Inuit Qaujimajatuqangit about population changes and ecology of Peary caribou and muskoxen on the High Arctic Islands of Nunavut (Master's thesis). Queen's University, Kingston, Ontario.

Thing H, Klein DR, Jingfors K, Holt S. 1987. Ecology of muskoxen in Jameson Land, northeast Greenland. Holarctic Ecology **10:**95–103.

Turner W, Spector S, Gardiner N, Fladeland M, Sterling E, Steininger M. 2003. Remote sensing for biodiversity science and conservation. Trends in Ecology & Evolution **18:**306–314.

Wasser SK, Brown L, Mailand C, Mondol S, Clark W, Laurie C, Weird BS. 2015. Genetic assignment of large seizures of elephant ivory reveals Africa's major poaching hotspots. Science **349:**84–87.

Wittemyer G, Northrup JM, Blanc J, Douglas-Hamilton I, Omondi P, Burnham KP. 2014. Illegal killing for ivory drives global decline in African elephants. Proceedings on the National Academy of Sciences of the United States of America **111:**13117–13121.

Yang Z, Wang T, Skidmore AK, de Leeuw J, Said MY, Freer J. 2014. Spotting east African mammals in open savannah from space. PLoS ONE **9**(12)**:**(e115989). doi:10.1371/journal.pone.0115989.

Tools and Technology Article

Validating Aerial Photographic Mark–Recapture for Naturally Marked Feral Horses

BRUCE C. LUBOW, *Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO 80523, USA*
JASON I. RANSOM,[1] *United States Geological Survey, Fort Collins Science Center, 2150 Centre Avenue, Building C, Fort Collins, CO 80526, USA*

ABSTRACT Accurately estimating large mammal populations is a difficult challenge because species of interest often occupy vast areas and exhibit low and heterogeneous visibility. Population estimation techniques using aerial surveys and statistical design and analysis methods provide a means for meeting this challenge, yet they have only rarely been validated because wild populations of known size suitable for field tests are rare. Our study presents field validations of a photographic aerial mark–recapture technique that takes advantage of the recognizable natural markings on free-roaming feral horses (*Equus caballus*) to accurately identify individual animals and groups of animals sighted on multiple occasions. The 3 small populations of feral horses (<400 animals each) in the western United States used in the study were all closely monitored on a weekly basis by local researchers, thus providing test populations of known size. We were able to accurately estimate these population sizes with aerial surveys, despite rugged terrain and dense vegetation that created substantial heterogeneity of sighting probability among horse groups. Our best estimates at the 3 sites were within −6.7%, 2.6%, and −8.6% of known truth (−4.2% mean error, 6.0% mean absolute error). In contrast, we found undercount bias as large as 32% before any statistical corrections. The necessary corrections varied both temporally and spatially, in response to previous sighting history (behavioral response), and by the number of horses in a group. Despite modeling some of the differences in horse-group visibility with sighting covariates, we found substantial residual unmodeled heterogeneity that contributed to underestimation of the true population by as much as 22.7% when we used models that did not fully account for these unmeasured sources. We also found that the cost of the accurate and validated methods presented here is comparable to that of raw count (so called, census) methods commonly employed across feral horse ranges in 10 western states. We believe this technique can assist managers in accurately estimating many feral horse populations and could be applied to other species with sufficiently diverse and distinguishable visible markings. (JOURNAL OF WILDLIFE MANAGEMENT 73(8):1420–1429; 2009)

DOI: 10.2193/2008-538

KEY WORDS aerial survey, *Equus caballus*, feral horse, heterogeneity, mark–recapture, population estimation, sighting probability, validation.

Accurately estimating population sizes of large free-roaming animals is a challenging and critical task for successful wildlife management (Williams et al. 2002), yet up to one third of ungulates in the western United States are missed by standard visual aerial surveys (Samuel et al. 1987, Ackerman 1988, Singer and Garton 1994, Bodie et al. 1995, Bowden and Kufeld 1995). Visibility of ungulates can vary tremendously among survey sites and occasions, depending upon transect spacing and sighting factors such as snow cover, group size, activity of the animals, tree cover, and experience of the observers (Pollock and Kendall 1987, Samuel et al. 1987, Unsworth et al. 1994, Bodie et al. 1995, Lubow and Ransom 2007). Despite these well-known biases that result in variable and unknown degrees of undercounting, the use of so-called census methods that make the unjustified assumption of 100% sighting probability remains commonplace (Rabe et al. 2002), while published evaluations of population estimation methods for feral horses (*Equus caballus*) are rare. Modern survey methods based on statistical models have been applied in Australia (Bayliss and Yeomans 1989, Graham and Bell 1989, Walter and Hone 2003, Laake et al. 2008), and recently Lubow and Ransom (2007) applied a technique to a North American feral horse population. All of these studies employed the simultaneous double-count method of mark–recapture and were limited by incorporating only 2 occasions (mark and recapture sightings), making testing for and correcting of biases due to unmodeled heterogeneity impossible. Furthermore, none of these prior studies was able to validate the methods in a population of accurately known size.

Our study focuses on a form of mark–recapture sampling technique that adjusts for sightability bias similar to a sightability bias correction model. Mark–recapture methods do not necessarily require physically capturing animals, only that individual animals or coherent groups can be reliably identified by natural or artificial marks or other unique characteristics on ≥ 2 occasions; individual capture histories can thus be recorded and used to estimate the number of unobserved animals (Seber 1973; Huggins 1989, 1991; Neal et al. 1993; Pledger 2000). Repeated observations can be made simultaneously by multiple observers or at different times. Sighting heterogeneity among individual animals or groups is common; some groups, due to their size, distance, coloration, location within cover, or other factors, are easier or more difficult for observers to see (Pollock and Kendall 1987). Unless heterogeneity is measured and modeled to correct for these differences, the unmodeled heterogeneity will result in underestimating population size (Borchers et al. 2006, Laake et al. 2008). There are 2 fundamental approaches to modeling heterogeneity, which we refer to as explicit and implicit. The more intuitive explicit method is to record a set of covariates that can be used to explicitly model the differing sighting probabilities of different animals or groups (Samuel et al. 1987). The alternative method requires more sighting occasions (typically >4) and estimates heterogeneity implicitly from the distribution of

---

[1] *E-mail:*

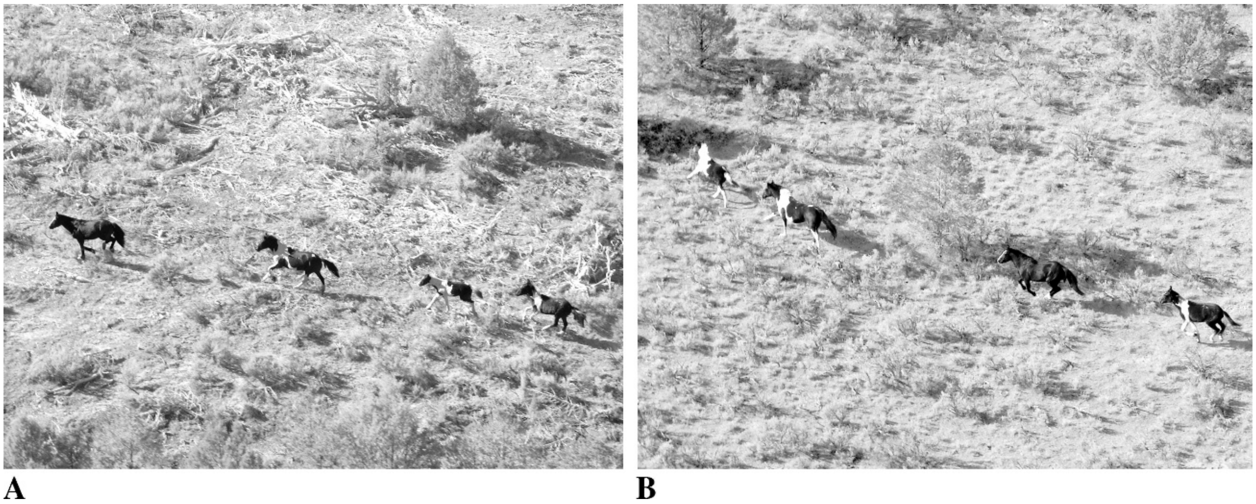


Figure 1. Initial mark photo (A) of a feral horse (*Equus caballus*) group at Little Book Cliffs Wild Horse Range (Mesa County, CO, USA, 2006) and a resight photo (B) of the same group on a subsequent flight, showing the unique individual markings that enable identification of groups without artificial marking. Individuals with uniform pelage coloration, such as the lead horse in photo A, were identified by unique combinations of pelage color and leg and face markings, whereas individuals exhibiting pinto coloration have unique pelage patterns. In this group, the uniformly brown horse exhibits a wide white blaze on its face and 2 white socks on its rear legs, and the pinto horses all have distinct patterns on their left side.

sighting frequencies rather than attempting to explicitly explain most of the sighting variation through covariates. This method is based on fitting mixture models in which multiple subpopulations are assumed to exist, each with its own sighting probability (Pledger 2000).

Feral horses present an opportunity to employ a photographic mark–recapture methodology because, unlike most North American terrestrial mammals, they often have sufficient distinctive natural markings to be uniquely identifiable; therefore, artificial marking is unnecessary for use with a mark–recapture estimation technique (Fig. 1). Natural markings for feral horses include unique pelage colors, and face and leg patterns (Gower 2000, Green 2001). A similar identification strategy has been used for whale (*Megaptera novaeangliae*) body pigmentation patterns (Hammond 1986, da Silva et al. 2000, Schweder 2003), manatee (*Trichechus manatus*) scar patterns (Langtimm et al. 2004), and zebra (*Equus grevyi*) stripe patterns (Rubenstein 1986, Williams 1998). Additionally, the composition of identifiable individuals in family groups can be used to help identify the group. Another advantage of using natural markings is that the cost of marking (sighting once) is no higher than for recapture (sighting again). Animals become marked the first time they are observed during a survey, which can be on any survey occasion. Therefore, precision continues to improve as larger portions of the population become marked when they are first observed during the course of the survey.

Our objective was to field-test a mark–recapture method for feral horses using natural markings for identification. We also sought to explore the importance of sighting heterogeneity and the ability to correct for it with appropriate statistical models. We were able to fully evaluate the bias, precision, and cost of this method under some of the most difficult sighting conditions (dense vegetation and complex terrain) often encountered in North American feral horse habitats.

STUDY AREA

We conducted 4 tests of the photographic mark–recapture technique with sightability bias correction covariates on 3 known populations of feral horses: Little Book Cliffs Wild Horse Range (WHR), McCullough Peaks Herd Management Area (HMA), and Pryor Mountain WHR. Survey 1 was conducted at the McCullough Peaks HMA, located in Park County, Wyoming, USA, 32 km east of the town of Cody. The area consisted of 44,440 ha of predominantly flat, open sagebrush park, with rugged badlands along the western edge of the range. Vegetation was characterized by small shrubs, grasses, and forbs with sparse stands of cottonwood (*Populus* sp.) and willow (*Salix* spp.) along the ephemeral streambeds. Elevations ranged from 1,200 m in the lowlands to 1,964 m at the summit of McCullough Peaks. Horses were distributed across the entire area, although the more rugged western portion of the area was known to have lower densities of horses than the eastern portion.

Surveys 2 and 3 were conducted at the Little Book Cliffs WHR, located in Mesa County, Colorado, USA, 13 km northeast of Grand Junction. This area consisted of 14,614 ha of rugged terrain, with sloping plateaus, sagebrush parks, and 4 major canyon systems. Dense stands of Colorado piñon (*Pinus edulis*) and Utah juniper (*Juniperus osteosperma*) occurred across the area and were intermixed with small sagebrush parks. Elevations ranged from 1,500 m to 2,250 m, with some impassable steep-walled shale and sandstone cliffs. Horses were distributed across the entire area.

Survey 4 was conducted at the Pryor Mountain WHR, located 21 km north of Lovell, Wyoming, within northern Bighorn County, Wyoming, and southeastern Carbon County, Montana (USA). The study area ranged in elevation from 1,190 m to 2,625 m and consisted of 16,046 ha of lowland desert, foothill slopes, forested montane slopes, steep canyons, and isolated grassy plateaus. Vegetation at lower elevations included small desert shrubs and grasses, with mid-elevations dominated by curl-leaf mountain mahogany (*Cercocarpus ledifolius*), Utah juniper, and various grasses, and high elevations dominated by limber pine (*Pinus flexilus*), subalpine fir (*Abies lasiocarpa*), and alpine bluegrass (*Poa alpinum*). Horses were located mostly between mid- and high elevations in moderate to sparse tree cover.

METHODS

Aerial Surveys

We conducted the first 2 surveys with 2 flights each and the final 2 surveys with 6 flights each; we conducted each survey over a 2-day period. We changed the number of flights and amount of time allocated to the latter 2 surveys to better address heterogeneity in group sighting variables and provide greater accuracy and precision to the estimate, which proved necessary for the populations living in relatively dense tree cover and complex topography. We spaced transects closer together and flew more slowly in 2-flight surveys to obtain greater sighting probability per occasion. We designed the time allocated for each survey to be comparable to the cost of one intensive raw-count census flight based on the flight times used by local managers prior to this study. In other words, we held the survey cost constant and examined the difference in precision and accuracy among various methods.

We conducted all surveys with helicopters to allow for mobility in navigating terrain and stability for photographing groups. For all analyses, we defined a group as ≥ 1 horses in association with each other. We conducted both Little Book Cliffs WHR surveys in a B4763B1 (Soloy Aviation Solutions, Olympia, WA) helicopter, and the McCullough Peaks HMA and Pryor Mountain WHR surveys in a 206B-III Jet Ranger (Bell Helicopter, Hurst, TX) helicopter. Survey altitudes ranged from 60 m to 150 m above ground level. We photographed every group using high-resolution digital cameras (Minolta Dimage GT™ [Ramsey, NJ] with 50–200-mm lens or Canon EOS 350D™ [Lake Success, NY] with 70–300-mm image-stabilized lens), and recorded additional covariate data for each group at the time of sighting. Covariates collected on all surveys included locomotion of group (stationary or moving), time of day, sun direction, topography type (rugged or level), vegetation type (trees, shrubs, or grasses), percent vegetation cover (to the nearest 10%), and group size, as well as Global Positioning System (GPS) coordinates. The pilot served as a secondary observer and was accompanied by one primary observer on all surveys. An additional crewmember for data recording was present on 2 flights, but we later deemed this unnecessary. During survey 1, an additional observer took part in the survey to compare film with digital photographs. We found film photography to be unnecessary for adequate resolution and more costly and cumbersome than digital photography.

Immediately following each entire survey, we matched photographs of groups to determine mark–recapture data with a collective effort between multiple members of our team and, in some cases, with participation of the herd manager. Horses at Little Book Cliffs WHR and McCullough Peaks HMA exhibited predominately black, bay, and pinto pelage, whereas Pryor Mountain WHR did not contain any pinto horses and was dominated by dun- and grullo-colored animals. Nevertheless, the diversity of phenotypic expression among individual horses (face and leg markings, in addition to pelage color) made this a straightforward comparison exercise and consensus was reached relatively quickly among participants analyzing each photo.

These 3 herds are part of a larger research project for which each individual horse in each population had previously been identified, catalogued, and monitored weekly on the ground (Ransom 2009); thus, the true population size, as well as individual band composition, was known prior to each aerial survey (however, this information was not shared with the aerial observers or the statistical analyst prior to producing estimates of the aerial survey). We tested our ability to identify groups in the photographs by comparing our decisions with these independently generated population composition data. We confirmed pelage color and markings for each horse, and in all cases our aerial photo decisions matched the individuals and groups of known animals on the lists, making it highly unlikely that we misidentified a photo. A small number of horses were not observed on the ground within a short time prior to each aerial survey and were, thus, considered to be of unknown status (possibly deceased or emigrated). Therefore, the known population sizes for each herd used for validation in this study are presented as a range from the minimum number known present to the maximum number representing all animals that could have been present.

Transects for all surveys were linear, predetermined, and followed by GPS route navigation while in the air. We chose spacing of transects by relying on the prior judgment of field personnel for the distance that would provide at least a moderate (30–50%) probability of sighting each horse group anywhere within the study area during the course of the entire survey. The density and extent of tree cover, together with the terrain type, which determined the ease or difficulty of detecting groups, were the primary factors influencing these judgments.

For survey 1, we stratified transect spacing such that a portion of the range known to have a greater density of horses was more intensely surveyed than areas with a lower density. We spaced the high-density transects 1.6 km apart and spaced the low-density transects 3.2 km apart. Sighting conditions across both areas were homogeneous. We performed the second flight 2 days later, with transects oriented the same direction but offset by 0.8 km in the high-density area and by 1.6 km in the low-density area. Mean transect spacing per flight based on actual GPS tracks was 2.6 km and mean flight speed on transects was 43 km/hour.

We spaced survey 2 transects 1.6 km apart uniformly across the range, and then repeated them the following day

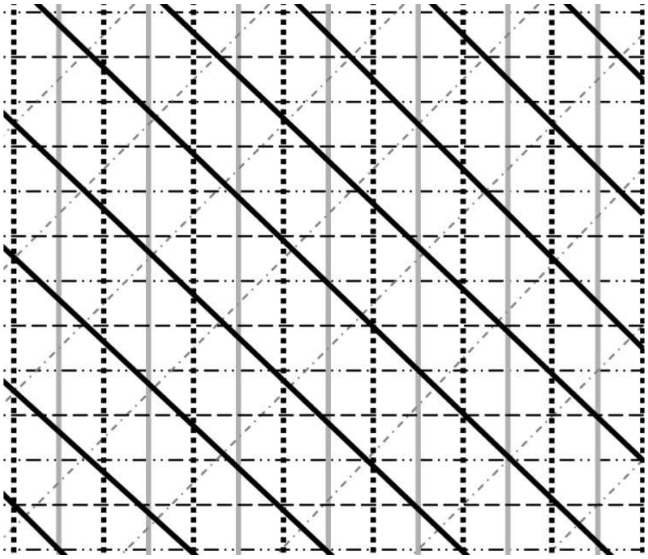


Figure 2. Schematic layout of 6 independent sets of transects used for aerial surveys 3 and 4 at Little Book Cliffs Wild Horse Range, 2006 (Mesa County, CO, USA) and Pryor Mountain WHR, 2007 (SE Carbon County, MT, USA), depicting the flight direction and spacing relative to other sets of transects for each of the 6 flights per survey.

with the same spacing, but perpendicular to the transects of the previous day. On both days, we flew additional transects to follow elevation contours within canyons and other areas where visibility was poor from the original parallel transects. Based on GPS tracks, the average transect spacing we achieved was 0.5 km. Average survey speed was 42 km/hour.

We conducted surveys 3 and 4 over 2 days each, with 6 total flights for each survey (3/day). We restructured these final 2 surveys to better address problems of the sighting probability heterogeneity encountered in survey 2. We collected data in these latter surveys on 6 independent sighting occasions, which provided the required data to apply the implicit method of heterogeneity correction. We spaced transects farther apart and flew at somewhat higher speeds to complete the 6 separate passes in roughly the same time (and therefore at the same cost) as a 2-flight survey required. Flight 1 of each survey was a parallel set of transects, with flight 2 transects rotated by 90°, and flight 3 transects rotated by 45°. The following day, we conducted the 3 flights again, with transects offset between those from the first day (Fig. 2). We spaced survey 3 transects 1.6 km apart and spaced survey 4 transects 3.2 km apart. Average transect spacing was 1.1 km in survey 3 and 1.3 km in survey 4. Average flight speed was 49 km/hour in survey 3 and 75 km/hour in survey 4.

Data Analysis

We used Program MARK (version 5.1, http://welcome.warnercnr.colostate.edu/~gwhite/mark/mark.htm, accessed 20 Mar 2008) to perform analyses with likelihood-based closed-capture population models, as described by Pledger (2000) and Huggins (1989, 1991). We fit a set of models for each analysis that included effects of time (*t*), behavioral response (*b*), heterogeneity (*h*), and a covariate for group size (*N*). We fit interactive and additive versions of models where they were supported by the data for that analysis. The models we considered (in the notation of Pollock et al. [1990] and Pledger [2000]) were M(0), M(*t*), M(*b*), M(*h*), M(*th*), M(*t*+*h*), M(*t*+*b*), M(*bh*), and M(*b*+*h*), each of which also considered the group-size covariate, *N*, as an additive effect. We excluded additional models with covariates for animal movement, sun position, tree cover, and terrain type after preliminary analyses proved them to have minimal predictive value as assessed using corrected Akaike's Information Criterion (AIC_c) weights (Burnham and Anderson 2002). We created 2-sighting occasion data from surveys with 6 actual sighting occasions by combining data from the first 3 occasions to create a single synthetic first occasion and from the last 3 occasions to create a single synthetic second occasion.

The independent observable unit was a horse group, not an individual horse. Therefore, mark–recapture analysis estimated the number of groups. We also estimated a mean group size from the data and the population estimate was the product of these 2 estimates. We computed the estimated number of groups for each analysis by averaging the individual estimates from each model across all of the models considered, weighting each by its AIC_c model weight (Burnham and Anderson 2002). We computed average group size by weighting the observed group size by the inverse of the probability that groups of that size would be observed on ≥ 1 flight during the survey. This procedure adjusted for the bias that would otherwise result from the average of observed groups being larger than the true average of group sizes in the population due to lower sighting probability for smaller groups.

We based standard errors for the number of groups on AIC_c model-weighted estimates, which account for model-selection uncertainty (Burnham and Anderson 2002). We corrected average group size for sighting bias. We computed the error for population using the Delta method (Seber 1973) and computed confidence intervals based on a log-normal distribution for the estimated number of horses not observed as follows:

$$c = \exp\left\{1.96 \times \sqrt{\ln\left[1 + \left(\frac{\hat{s}}{\hat{N} - N_{obs}}\right)^2\right]}\right\}$$

$$95\%\mathrm{CI} = \left[N_{obs} + \frac{\hat{N} - N_{obs}}{c}, N_{obs} + c\left(\hat{N} - N_{obs}\right)\right]$$

Where $\hat{N}$ was the population estimate, $\hat{N}_{obs}$ was the number observed, $\hat{s}$ was the standard error of the population estimate, and the constant 1.96 was the value at which the cumulative standard normal distribution had 97.5% [1 − (0.05/2)] probability. This method was simply assuming a 95% confidence interval on the log-transformed estimate of missed horses and back-transforming that to the linear scale. This was a logical and plausible distribution of errors on a quantity with a range of [0, ∞] and excluded lower

Table 1. Strength of evidence for effects on sighting probability of time, group size, behavioral response to prior sighting, and unmeasured heterogeneity among groups based on the sum of corrected Akaike's Information Criterion model weights across all models that include indicated effect (models may have multiple effects included, so evidence for the several effects will not sum to 100%, but evidence for any single effect will always be 0–100%) for aerial surveys of feral horses at McCullough Peaks Herd Management Area (HMA; Park County, WY, USA), Little Book Cliffs Wild Horse Range (WHR; Mesa County, CO, USA), and Pryor Mountain WHR (SE Carbon County, MT, USA), 2003–2007.

	Survey 1 (McCullough Peaks HMA)	Survey 2 (Little Book Cliffs WHR)	Survey 3 (Little Book Cliffs WHR)		Survey 4 (Pryor Mountain WHR)	
Effects	2 Occ[a] (%)	2 Occ[a] (%)	2 Occ[a] (%)	6 Occ[a] (%)	2 Occ[a] (%)	6 Occ[a] (%)
Time	21.0	36.4	99.0	43.3	23.7	91.8
Group size	44.8	49.7	34.7	51.0	25.5	63.9
Behavior				53.8		14.2
Heterogeneity				46.2		69.4

[a] Occasions (Occ) are separate flights within each survey.

confidence intervals that were less than the minimum known alive (i.e., actual no. observed).

We computed true error rates using the known population sizes for each population at the time of the survey. Because of some ambiguity about the possible death or emigration of a few horses, we knew the true numbers only within a narrow range. Our comparisons of estimates to truth were based on the mean of the minimum and maximum number of horses that could have been present during the survey.

RESULTS

The strength of evidence (AIC_c model wt) supporting the various effects of covariates on sighting probability differed considerably among surveys (Table 1). Support for variation over occasion (time) ranged from as low as 21% in survey 1 to a high of 99% for survey 2. Group-size effect on sighting probability had moderate support on all occasions, ranging from AIC_c weight of 25.5% to 63.9%. Evidence for behavioral response of horses to a prior close approach by the helicopter prior to sighting was moderate in survey 3 (33.5% of groups were running), but nearly absent (only 5.4% of groups running) in survey 4 (the 2 surveys with a 6-flight design for which this effect could be examined). Finally, heterogeneity of sighting probability among different groups had moderate support in both surveys in which it was estimated by the implicit method (surveys 3 and 4).

The effect of group size on sighting probability differed substantially among surveys (Fig. 3). Small groups were most easily seen at McCullough Peaks HMA (survey 1), where trees were absent and visibility was excellent. In stark contrast, sites with tree cover (surveys 2–4) had sighting probabilities for the smallest groups (i.e., 1 horse) well under 50% (Fig. 3). Sighting probability increased with group size at different rates for different flights, but was >90% for groups ≥27 horses in all flights.

Behavioral response to prior observation differed between the 2 surveys for which this effect could be measured (the 6-flight surveys), not only in magnitude, but also in direction. In survey 3, groups had a 20.8% (±7.9% SE) probability of being sighted the first time and 59.2% (±4.5% SE) probability of being sighted again if they had already been sighted. The effect was smaller and in the opposite direction in the Pryor Mountain WHR survey. Initial sighting probability was 40.4% (±7.2% SE) and recapture probability was 29.2% (±3.9% SE). All of these probabilities were based on the model that holds all other factors constant and includes only a behavioral effect. Notice that the trap-shy response (reduced sighting probability following first sighting) is not strongly supported for the Pryor Mountains (Table 1) and that the confidence intervals overlap, but the trap-happy (increased sighting probability following first sighting) effect is moderately supported for Little Book Cliffs WHR with nonoverlapping confidence intervals.

Model-weighted averaging across all models produced estimates of the total number of groups (Table 2). We calculated mean group sizes with bias-corrected weights (Table 2) and multiplied these by the estimated number of groups to obtain the estimated size of the herd (Fig. 4). We compared these estimates to the minimum and maximum possible true number of horses that we knew to be present at the time of the survey (Fig. 4).

Raw (uncorrected) counts of horses actually seen during aerial surveys were consistently less than the known population by between −7.5% to −32.0% ($\bar{x} = -18.2\%$, ±5.1% SE), whereas estimated population size without implicitly accounting for heterogeneity was closer to the true

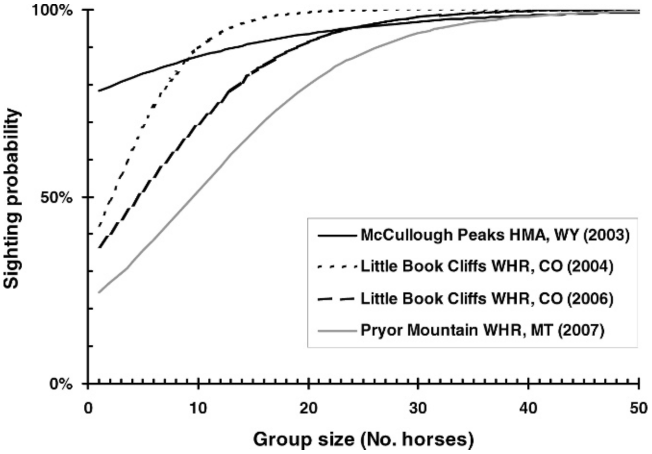


Figure 3. Estimated effect of group size on sighting probability of feral horses from aerial surveys at McCullough Peaks Herd Management Area (HMA; Park County, WY, USA), Little Book Cliffs Wild Horse Range (WHR; Mesa County, CO, USA) and Pryor Mountain WHR (SE Carbon County, MT, USA), 2003–2007. Sighting probabilities are for a single helicopter flight. Surveys are not comparable because transect spacing differed among surveys. Group-size effect is based on a model with all other sighting factors held constant [model $M(0 + N)$].

Table 2. Data and estimates from aerial surveys of feral horse populations at McCullough Peaks Herd Management Area (HMA; Park County, WY, USA), Little Book Cliffs Wild Horse Range (WHR; Mesa County, CO, USA), and Pryor Mountain WHR (SE Carbon County, MT, USA), 2003–2007. Estimates are based on corrected Akaike's Information Criterion–weighted model average of maximum-likelihood closed-capture models in Program MARK. Models with time, group size, and behavioral response effects were included for all surveys, and models of heterogeneity in capture probabilities were included for surveys done with 6 sighting occasions (as indicated).

	Survey 1 (McCullough Peaks HMA)	Survey 2 (Little Book Cliffs WHR)	Survey 3 (Little Book Cliffs WHR)		Survey 4 (Pryor Mountain WHR)	
Population parameter	**2 Occ[a]**	**2 Occ[a]**	**2 Occ[a,b]**	**6 Occ[a]**	**2 Occ[a,b]**	**6 Occ**
Groups (no.)						
Seen	48	42	38	38	36	36
Estimate	49.3	52.0	39.5	46.6	39.4	41.4
SE[c]	0.95	2.53	0.34	3.64	0.84	1.41
Group size (horses/group)						
Bias corrected $\bar{x}$[d]	7.1	3.0	3.6	3.6	3.4	3.4
SE[c,d]	0.16	0.26	0.27	0.27	0.34	0.34

[a] Occasions (Occ) are separate flights within each survey.

[b] Calculated using data collected on 6 occasions by combining the first 3 occasions and the last 3.

[c] Finite population correction factor applied.

[d] Group size corrected for sized-biased sighting probability.

size, but remained negatively biased ($\bar{x} = -13.8\%$, ±3.3% SE; Table 3). Error was smallest at McCullough Peaks HMA (survey 1), where sightability was excellent due to very open terrain and low vegetation. The standard deviation of errors in our surveys provided a means of estimating the range of potential errors in future surveys. For raw counts, the 90% prediction interval was (−34.9%, −1.4%), indicating that about 5% of raw count (so called census) surveys would result in undercounts of >35% and most (>90%) would result in undercounts of some magnitude.

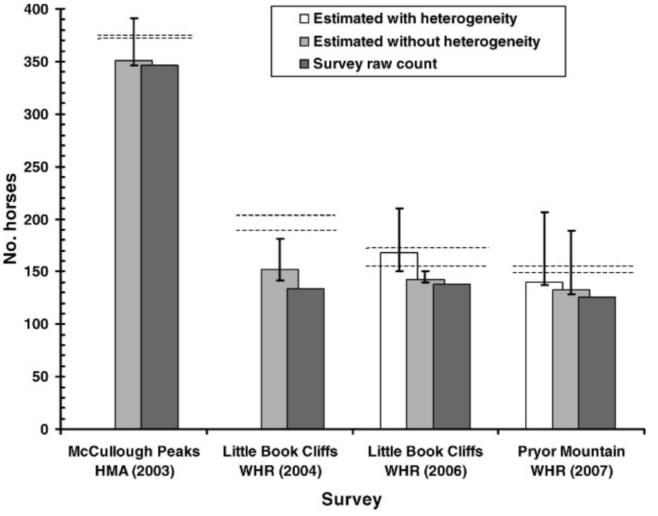


Figure 4. Population estimates (bars), 95% lognormal confidence intervals (black error bars), and minimum and maximum possible known true population size at time of survey (pairs of dashed lines) for aerial surveys of feral horse populations at McCullough Peaks Herd Management Area (HMA; Park County, WY, USA), Little Book Cliffs Wild Horse Range (WHR; Mesa County, CO, USA), and Pryor Mountain WHR (SE Carbon County, MT, USA), 2003–2007. Estimates are based on corrected Akaike's Information Criterion–weighted model average of maximum-likelihood closed-capture models in Program MARK. Models with time and group-size effects were included for all surveys, and models of behavioral response and heterogeneity in capture probabilities were either included (white bars) or not (light gray bars) when the number of occasions supported estimation of these models. Uncorrected raw counts (dark gray bars) from each survey are shown for comparison.

Estimates dramatically improved when sets of models included some that modeled heterogeneity implicitly. Errors on the 2 surveys with implicit heterogeneity information were positive once and negative once and averaged only 5.6% (±3.0% SE) in absolute magnitude (Table 3). The 90% prediction interval for the 6-occasion surveys using the implicit heterogeneity correction was (−16.0%, 10.0%). Fitting a simple exponential model to the errors as a function of the number of sampling occasions per survey provided a projection of the effort (no. of sampling occasions/flights) required for any desired level of precision (Fig. 5).

Statistically estimated precision was uniformly high (<10.5%); however, some of these estimated rates were overly optimistic due to negative bias in methods without implicit heterogeneity correction, which failed to produce confidence intervals that covered the true population size at Little Book Cliffs WHR during either survey 2 or 3. In contrast, the 95% confidence intervals for the estimates easily covered the true population sizes in both cases when implicit heterogeneity was modeled (i.e., in both 6-occasion surveys) and also in surveys 1 and 4, even without implicit heterogeneity correction (Fig. 4).

Survey cost was dominated by the cost of helicopter time. For comparison, we computed the cost per hectare based on actual survey flight time (excluding ferrying to and from the survey site), average helicopter cost of US$650/hour, and actual area covered by the transects. Cost was lowest (US$0.07/ha) at McCullough Peaks HMA (survey 1), where visibility was excellent and enabled wide transect spacing. At the 2 sites dominated by tree cover, cost was US$0.36/ha, which is 5 times the cost at McCullough Peaks HMA. The mean cost (US$0.36/ha) of the 2 surveys with 6 sampling occasions (survey 3 and 4) was nearly identical to the cost (US$0.37/ha) of the comparable 2-occasion survey in rugged, treed terrain (survey 2). This was the result of our

Table 3. Errors in estimates compared to known population size (using mean of min. and max. known animals present during each survey) from aerial surveys of feral horses at McCullough Peaks Herd Management Area (HMA; Park County, WY, USA), Little Book Cliffs Wild Horse Range (WHR; Mesa County, CO, USA), and Pryor Mountain WHR (SE Carbon County, MT, USA), 2003–2007. Errors are the differences between the known population size and the 3 types of estimates: raw counts, statistical estimates without heterogeneity correct, statistical estimates with heterogeneity correct (negative value indicates estimate was lower than actual). Error as a percent of the actual population size is shown. Mean errors and the mean of the absolute values (magnitude) of the errors are presented in the last columns on the right. Statistically estimated (not actual) errors (SE and CV) are also shown for comparison.

	Survey 1 (McCullough Peaks HMA)	Survey 2 (Little Book Cliffs WHR)	Survey 3 (Little Book Cliffs WHR)		Survey 4 (Pryor Mountain WHR)			
Error	2 Occ[a]	2 Occ[a]	2 Occ[a,b]	6 Occ[a]	2 Occ[a,b]	6 Occ[a]	$\bar{x}$	$\bar{x}$ absolute error
Raw-count errors								
No. horses	−28	−63	−26		−27			
Error (%)	−7.5	−32.0	−15.6		−17.6		−18.2	18.2
Error without heterogeneity correction								
Actual errors								
No. horses	−23	−45	−21		−20			
Error (%)	−6.7	−22.7	−13.0		−13.0		−13.8	13.8
Statistically estimated errors								
SE	7.9	9.4	2.4		13.6			
CV	2.2	6.1	1.7		10.2			5.1
Error with heterogeneity correction								
Actual errors								
No. horses				4.2		−13.2		
Error (%)				2.6		−8.6	−3.0	5.6
Statistically estimated errors								
SE				14.2		14.7		
CV (%)				8.4		10.5		9.5

[a] Occasions are separate flights within each survey.

[b] Calculated using data collected on 6 occasions by combining the first 3 occasions and the last 3.

attempt to design surveys of equal approximate cost, leaving differences only in performance to be measured and compared.

During surveys 3 and 4, the patterns of detecting horse groups on each of the 6 successive flights during each of these 2 separate surveys were similar (Fig. 6). Fewer than 50% of horse groups were seen on all but one flight at Little Book Cliffs (Fig 6A). We sighted new, unique groups on every flight except the last one at Pryor Mountain WHR (Fig. 6B). The cumulative total number of unique groups seen over the course of the 6 flights did not exceed 85% and was not approaching a clear asymptote.

DISCUSSION

The photographic mark–recapture with implicit heterogeneity correction worked well in the feral horse populations we tested, despite the difficult sighting conditions created by topography and vegetation cover. Estimated errors were <10.5% coefficient of variation, confidence intervals covered true population size well, and actual errors were substantially lower (5.6% mean absolute error). However, we observed numerous factors that affect sighting probability, which must be accounted for to obtain accurate estimates with realistic confidence intervals. Specifically, raw counts missed as many as one-third of the horses present, despite our observer's high level of aerial survey expertise. Using prediction interval estimates to project error rates on similar surveys led us to predict that undercounts of >35% of the true population would occur in 5% of surveys. This was consistent with the previously observed undercounts of as

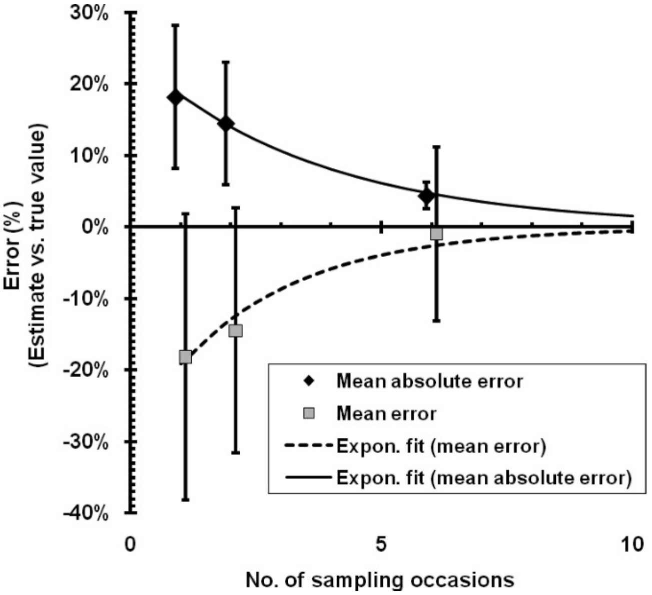


Figure 5. Errors in estimates compared to known true population size (using mean of min. and max. known animals present during survey) from aerial surveys of feral horse populations at McCullough Peaks Herd Management Area (HMA; Park County, WY, USA), Little Book Cliffs Wild Horse Range (WHR; Mesa County, CO, USA), and Pryor Mountain WHR (SE Carbon County, MT, USA), 2003–2007. Mean absolute error is the mean of the absolute values of the errors for each survey. Values for one occasion are the raw counts without correction. Values for 2 occasions are estimates corrected for time, group size, and behavioral response effects on sighting probability, but without modeling unmeasured heterogeneity. Values for 6 occasions are based on a set of models that include some for unmeasured heterogeneity. Error bars for mean absolute error are 95% log-normal confidence interval for estimation of the mean absolute error. Error bars for the mean error (bias) represent the 90% range of predicted outcomes based on a normal distribution (not CI on the mean error).

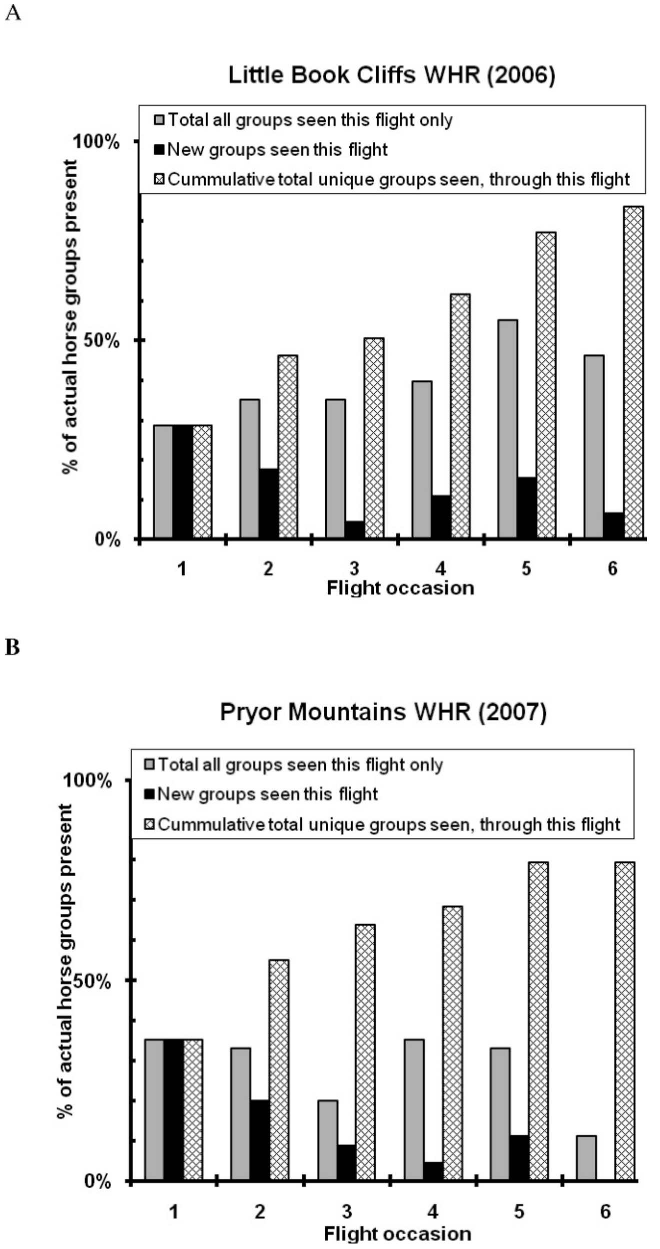


Figure 6. Percent of known feral horse-groups observed in each of 6 independent sighting occasions and cumulative across all sighting occasions, during aerial surveys of Little Book Cliffs Wild Horse Range (WHR), Mesa County, Colorado, USA, 2006 (A), and Pryor Mountain WHR, southeastern Carbon County, Montana, USA, 2007 (B).

much as 39% for feral horses on parts of Assateague Island (east coast of United States) by aerial observation compared to more accurate ground counts (Bashore et al. 1990). Without statistical models and corrections, we do not believe an observer could reliably count all or even most horses within a population unless sighting conditions were extremely good and homogeneous. More importantly, without statistical models and proper sampling methods, there would be no confidence intervals or other evidence to support claims of accuracy.

The corrections required to obtain accurate estimates varied considerably among herds in differing terrain and vegetation types, among flights for the same herd and location, in relation to horse-group size, and due to behavioral response of the horses themselves. No constant correction factor or single sightability-model calibration could account for this extensive spatial and temporal variation in sighting probability. Similar variation in sighting probability with local conditions has previously been observed in aerial surveys (Bayliss and Yeomans 1989, Graham and Bell 1989, Bashore et al. 1990, Walter and Hone 2003, Lubow and Ransom 2007).

Linklater and Cameron (2002) also reported that avoidance behavior of horses to helicopters during aerial surveys led to considerable double-counting of groups, or missing of groups entirely. Using natural markings to identify individuals allowed us to avoid the problems of double-counting and groups merging due to helicopter avoidance response, and we were able to validate this with our extensive record of known individuals in each population. We cannot assess whether helicopter avoidance response was responsible for our not observing specific individuals that were known to be present, but our overall results suggest that whatever heterogeneity in sighting probability this may have caused, our methods and models were able to adequately correct for it.

We found that heterogeneity among groups of horses (some groups are easier to see than others) was among the most important factors affecting sighting probability, particularly under difficult sighting conditions (areas with low sighting probabilities). Despite attempting to correct for differences in sighting probability through the use of covariates for tree cover, terrain, and occasion (time), unmeasured heterogeneity remained a dominant effect in some surveys. This finding casts doubt on the feasibility of applying simple sightability bias correction models (Samuel et al. 1987) that remain constant over time, location, and observers, and that require the measured covariates to explain most or all of the variation in sighting probability, to produce unbiased estimates. The greater the residual, unmodeled, component of heterogeneity, the more severe the negative bias becomes (Laake 1999). Surveys in very open habitat, such as the McCullough Peaks in survey 1, might be sufficiently accurate with only explicit corrections based on a few measured covariates affecting heterogeneity; but in difficult sighting conditions, methods that can implicitly account for heterogeneity are essential. We believe that 6 sighting occasions were needed for sufficient accuracy (precision and bias correction; Fig. 5). Future applications of this method might benefit from experimenting with >6 occasions, although we doubt that >8 would be of any additional value.

The photographic mark–recapture method does not necessarily increase survey time, cost, risk to observers, or disturbance of horses. It is important to note that the poor results we reported here for raw-count (census) surveys were based on the same data from the flights used to produce much more reliable results by applying appropriate statistical models; in other words, the time, cost, effort, risk, and disturbance were identical, and only the methods of data collection and analysis were different.

Obtaining reliable results with the photographic mark–recapture method requires that surveys be properly designed with 6–8 independent sighting occasions, minimal error in identifying groups in photographs, and adequate sighting probabilities for all groups. It is not important that most horse groups are sighted on each occasion, only that each group has a moderate probability (>30%) of being sighted at least once during the survey. Consequently, transects can be designed to merely sample a fraction of the population on each flight. They need not attempt to see all, or nearly all, of the horses on each pass or even once during the course of the survey.

It is common among experienced field biologists to assume that nearly all animals are observed during the course of an aerial survey. Often this belief is held even about the results of a single-pass survey. As our data clearly demonstrate, horses were missed even after 6 independent sighting occasions over the study area, resulting in 16–20% of the known groups never being observed on any occasion. We observed new groups on all sighting occasions except for one. Even by extrapolating beyond the data, there was no clear asymptote to the cumulative raw count that could reliably predict the total population size.

Despite the excellent results we obtained in these trials, it is unlikely that extending this method to all feral horse populations would prove successful. Application to larger herds is constrained by the ability to accurately match (recognize) groups from photographs of the same groups on multiple occasions. Greater uniformity of phenotype among individuals would also complicate the identification of groups. Very dense herds or herds with an atypical social structure that resulted in frequent merging or splitting of bands would also interfere with identification of unique groups. Managers designing surveys using this method must carefully evaluate whether the ability to consistently identify groups correctly can be assured.

We caution users of this method that the survey can only measure horses present in the study area at the time of the survey. In our experience, this often does not correspond to the political boundaries established for managing the herd. Horses often cross fences and travel great distances in search of water, forage, or mates. Whenever possible, it would be important to design surveys so that they cover the entire range of all herds that might intermingle or exchange individuals during a single survey, even if this requires multiple aircraft or multiple days of flights and extended survey boundaries to encompass potential habitat areas adjacent to the public land intended for feral horse use. One distinct advantage of this method is that horse movements do not complicate the survey and there is no requirement that every horse group be observed or that multiple sightings of the same group be avoided (as long as they can be identified accurately in the photographs).

Estimated precision of our estimates was always better (smaller estimated error) when heterogeneity was not modeled implicitly than when it was; this is probably a systematic negative bias in the error estimate that compounds the already negatively biased population estimate in the presence of heterogeneity. The actual error in survey 3 was almost 9 times larger than the statistically estimated standard error due to the combined effect of a negatively biased population estimate and a low error estimate. Thus, heterogeneity leads to estimates that appear to have high precision but are, in fact, neither as precise as they seem nor unbiased. Such estimates can be particularly problematic for managers because they could be misled to believe that they have precise and accurate information when, in fact, they may have neither. Larger confidence intervals at least would alert managers to the weakness of the estimate; seemingly precise but wrong estimates may be worse than none at all, leading to unwarranted overconfidence. Consequently, managers seeking accurate estimates would need to measure and model implicitly the heterogeneity in every population where this method is applied, unless and until multiple repeated surveys demonstrate that a particular population, due to uniform and excellent sighting conditions, such as the one at McCullough Peaks in our study, is not sufficiently heterogeneous to require such corrections. However, given the nearly equal cost of 6-occasion surveys to 2-occasion surveys with similar estimated precision, there should be little incentive to abandon the higher occasion alternative.

Estimation using mark–recapture over multiple sighting occasions provides several important advantages over rival survey methods, including 1) correction for multiple and (temporally or spatially) varying sighting bias; 2) correction for sighting heterogeneity among horse groups using the implicit method; 3) increase in effective sample size in small populations by simply adding additional sighting occasions to the survey; and 4) elimination of potential violation of the assumption that animals not be counted multiple times as they move during the survey. Thus, mark–recapture surveys conducted and analyzed properly can overcome many of the shortcomings that threaten accuracy in other methods such as simple uncorrected censuses and sightability bias correction models. Furthermore, the ability to uniquely and reliably identify individuals or groups over multiple sighting occasions using natural markings in feral horses avoids the prohibitive cost and risk (to both field workers and animals) of capturing and handling large free-ranging animals that is traditionally required to implement mark–recapture with unique artificial markings such as radiocollars or numbered neck-bands. Thus, we believe that the method evaluated here offers one of the best options available for estimating feral horse population size, both for accuracy and cost-effectiveness.

MANAGEMENT IMPLICATIONS

Surveys not employing appropriate sampling designs and statistical models to correct for the pervasive sources of bias and variation present during many aerial surveys may lead to inaccurate population estimates and are, thus, vulnerable to criticism from public stakeholders and unsuitable as a basis for sound management decisions. We have demonstrated that, with the proper field and analysis methods, accurate and precise estimates of feral horse populations are possible and no more costly than many traditional methods employing simple raw counts. The photographic mark–

recapture method is particularly useful for smaller herds and has the distinct advantage of working well under difficult sighting conditions or when horses are moving rapidly and double-counting cannot be reliably avoided; however, it is not applicable to all herds. Our ability to verify this method in multiple populations of known size gives us confidence that results produced by this method will withstand the scrutiny of various public stakeholders, politicians, and advocacy groups that frequently take an interest in the scientific basis for feral horse management decisions.

ACKNOWLEDGMENTS

We thank the Grand Junction, Cody, and Billings Field Offices of the United States Bureau of Land Management for cooperation and assistance with special-use aviation needs over the federal lands they manage. T. R. Stanley, J. E. Roelle, A. B. Shepherd, and 2 anonymous referees provided much appreciated expertise and comments on drafts of this manuscript. This research was funded by the United States Geological Survey Wildlife Program in cooperation with the United States Bureau of Land Management and Colorado State University. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the United States Government.

LITERATURE CITED

Ackerman, B. B. 1988. Visibility bias of mule deer aerial census procedures in southwest Idaho. Dissertation, University of Idaho, Moscow, USA.

Bashore, T. L., R. Keiper, J. W. Turner, Jr., and J. F. Kirkpatrick. 1990. The accuracy of fixed-wing aerial surveys of feral horses on a coastal barrier island. Journal of Coastal Research 6:53–56.

Bayliss, P., and K. M. Yeomans. 1989. Correcting bias in aerial survey population estimates of feral livestock in northern Australia using the double-count technique. Journal of Applied Ecology 26:925–933.

Bodie, W. L., E. O. Garton, E. R. Taylor, and M. McCoy. 1995. A sightability model for bighorn sheep in canyon habitats. Journal of Wildlife Management 59:832–840.

Borchers, D. L., J. L. Laake, C. Southwell, and C. G. M. Paxton. 2006. Accommodating unmodeled heterogeneity in double-observer distance sampling surveys. Biometrics 54:1221–1237.

Bowden, D. L., and R. C. Kufeld. 1995. Generalized mark–resight population size estimation applied to Colorado moose. Journal of Wildlife Management 59:840–851.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Second edition. Springer-Verlag, New York, New York, USA.

da Silva, C. Q., J. Zeh, D. Madigan, J. Laake, D. Rugh, L. Baraff, W. Koski, and G. Miller. 2000. Capture–recapture estimation of bowhead whale population size using photo-identification data. Journal of Cetacean Research and Management 2:45–61.

Gower, J. 2000. Horse color explained. Trafalgar Square, North Pomfret, Vermont, USA.

Graham, A., and R. Bell. 1989. Investigating observer bias in aerial surveys by simultaneous double-counts. Journal of Wildlife Management 53:1009–1016.

Green, B. K. 2001. The color of horses: the scientific and authoritative identification of the color of the horse. Montana Press, Missoula, USA.

Hammond, P. S. 1986. Estimating the size of naturally marked whale populations using capture–recapture techniques. International Whaling Commission Special Issue 8:253–282.

Huggins, R. M. 1989. On the statistical analysis of capture–recapture experiments. Biometrika 76:133–140.

Huggins, R. M. 1991. Some practical aspects of a conditional likelihood approach to capture experiments. Biometrics 47:725–732.

Laake, J. L. 1999. Distance sampling with independent observers: reducing bias from heterogeneity by weakening the conditional independence assumption. Pages 137–148 *in* G. W. Garner, S. C. Amstrup, J. L. Laake, B. F. J. Manly, L. L. McDonald, and D. G. Robertson, editors. Marine mammal survey and assessment methods. A.A. Balkema, Rotterdam, Netherlands.

Laake, J., M. J. Dawson, and J. Hone. 2008. Visibility bias in aerial survey: mark–recapture, line-transect or both? Wildlife Research 35:299–309.

Langtimm, C. A., C. A. Beck, H. H. Edwards, K. J. Fick-Child, B. B. Ackerman, S. I. Barton, and W. C. Hartley. 2004. Survival estimates for Florida manatees from the photo-identification of individuals. Marine Mammal Science 20:438–463.

Linklater, W. L., and E. Z. Cameron. 2002. Escape behaviour of feral horses during a helicopter count. Wildlife Research 29:221–224.

Lubow, B. C., and J. I. Ransom. 2007. Aerial population estimates of feral horses (*Equus caballus*) in Adobe Town and Salt Wells Creek Herd Management Areas using an integrated simultaneous double-count and sightability bias correction technique. U.S. Geological Survey Open File Report 2007-1274, Fort Collins, Colorado, USA.

Neal, A. K., G. C. White, R. B. Gill, D. F. Reed, and J. H. Olterman. 1993. Evaluation of mark–resight model assumptions for estimating mountain sheep numbers. Journal of Wildlife Management 57: 436–450.

Pledger, S. 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. Biometrics 56:434–442.

Pollock, K. H., and W. L. Kendall. 1987. Visibility bias in aerial surveys: a review of estimation procedures. Journal of Wildlife Management 51:502–510.

Pollock, K. H., J. D. Nichols, C. Brownie, and J. E. Hines. 1990. Statistical inference for capture–recapture experiments. Wildlife Monographs 107.

Rabe, M. J., S. S. Rosenstock, and J. C. deVos, Jr. 2002. Review of big-game survey methods used by wildlife agencies of the western United States. Wildlife Society Bulletin 30:46–52.

Ransom, J. I. 2009. Understanding the influences of immunocontraception on equids through competing models of behavior. Thesis, Colorado State University, Fort Collins, USA.

Rubenstein, D. I. 1986. Ecology and sociality in horses and zebras. Pages 282–303 *in* D. I. Rubenstein and R. W. Wrangham, editors. Ecological aspects of social evolution. Princeton University Press, Princeton, New Jersey, USA.

Samuel, M. D., E. O. Garton, M. W. Schlegal, and R. G. Carson. 1987. Visibility bias during aerial surveys of elk in north-central Idaho. Journal of Wildlife Management 51:622–630.

Schweder, T. 2003. Abundance estimation from multiple photo surveys: confidence distributions and reduced likelihoods for bowhead whales off Alaska. Biometrics Journal 59:974–983.

Seber, G. A. F. 1973. The estimation of animal abundance and related parameters. Griffin, London, United Kingdom.

Singer, F. J., and E. O. Garton. 1994. Elk sightability model for the Super Cub. Pages 47–49 *in* J. W. Unsworth, F. A. Leban, D. J. Leptich, E. O. Garton, and P. Zager, editors. Aerial survey manual. Second edition. Idaho Department of Fish and Game, Boise, USA.

Unsworth, J. W., F. A. Leban, D. J. Leptich, E. O. Garton, and P. Zager. 1994. Aerial survey: user's manual. Second edition. Idaho Department of Fish and Game, Boise, USA.

Walter, M. J., and J. Hone. 2003. A comparison of 3 aerial survey techniques to estimate feral horse abundance in the Australian Alps. Wildlife Society Bulletin 31:1138–1149.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations—modeling, estimation, and decision making. Academic Press, San Diego, California, USA.

Williams, S. D. 1998. Grevy's zebra: ecology in a heterogeneous environment. Dissertation, University College, London, United Kingdom.

Associate Editor: Forsyth.

CSIRO PUBLISHING

www.publish.csiro.au/journals/wr *Wildlife Research*, 2008, **35**, 365–370

Aerial mark–recapture estimates of wild horses using natural markings

Michelle J. Dawson[A,C†] *and Cameron Miller*[B]

[A]Institute for Applied Ecology, University of Canberra, Canberra, ACT 2601 Australia.
[B]Parks Victoria, Level 9, 535 Bourke St, Melbourne, Vic. 3000, Australia.
Present address: SMEC Australia, Level 5, 71 Queens Parade, Melbourne, Vic. 3004, Australia.
[C]Corresponding author. Email:

Abstract. Aerial mark–recapture population estimates utilising the natural markings of wild horses to identify individuals was applied in the Bogong High Plains, Alpine National Park, Victoria. A discrete population of wild horses occupying an area of 180 km^2 was sampled over two days in 2005. This study explored the feasibility of a technique that aimed to enable managers to estimate the size of the horse population and monitor it over time. Four observers (including the pilot) searched for horses from a helicopter. Once horses were sighted, photographic and written observations were used to 'mark' each animal. The survey was repeated the following day with observations 'recapturing' individuals. Data were analysed using several mark–recapture estimators, and the derived population estimates ranged from 89 (±5.3, s.e.) horses to 94.7 (±7.9, s.e.) horses. We found that the method gave a level of precision relevant to management, but needs refinement. The technique and its assumptions should be tested further by increasing the number of samples and video should be used to improve identification of individuals. We believe that this is a novel application for aerial surveys, which are typically unsuitable for estimating the size of small populations. This technique was developed for horses but may be used on other conspicuous species with unique natural markings.

Introduction

Knowledge of the size or density of a wildlife population is usually a prerequisite for effective management (Caughley and Sinclair 1994). While there are a range of survey methods available for estimating population size (see Krebs 1999), the most suitable method will be influenced by the species, the environment, and the level of precision and accuracy required. Wild horses (*Equus caballus*) are large, non-native, hard-hoofed animals considered by managers to pose a risk to the delicate ecosystem values of the Bogong High Plains in the Alpine National Park, Victoria, Australia. This wild horse population has the potential to spread in future and may increase in density following recent (2006) termination of cattle grazing leases in the area, and/or if winter snowfalls are less frequent or severe as a result of climate change. As such, Parks Victoria, the government agency that manages the Alpine National Park, required a method for estimating the size and trend in wild horse abundance over time, with reasonable accuracy and precision (coefficient of variation below 35%), utilising parks staff once techniques were established.

Previous studies have investigated suitable methods for surveying wild horses in the Australian Alps. Broad-scale aerial surveys were used to estimate the abundance of wild horses across the wider Australian Alps national parks (Walter and Hone 2003; Laake *et al*. 2008). In that study, a combination of line transect and mark–recapture (i.e. double-count) techniques provided the least-biased population estimates (Laake *et al*. 2008). Unfortunately these techniques are unsuitable for small survey areas such as the Bogong High Plains because the horse population is too small to acquire an adequate sample size. In 2002, a 'best guess' estimate of the wild horse population in the Bogong High Plains was 200 horses, based on anecdotal information from Parks Victoria staff and other sources (Walter 2002). The mean group size of wild horses in the Australian Alps has previously been estimated at 5.65 horses (Walter 2002), which would equate to ~34 groups of horses in the Bogong High Plains study area. In addition, horses have an average linear flush distance of 1 km in response to a helicopter (Linklater and Cameron 2002). Line transect surveys were therefore unsuitable because they usually require a sample size of at least 60 groups to be effective (Buckland *et al*. 1993). The double-count mark–recapture model (Caughley and Grice 1982) often used in aerial surveys (e.g. Walter and Hone 2003) was also unsuitable because transects would need to be closely spaced (less than 2 km apart) to obtain an adequate sample size, which would lead to horses being counted more than once.

Ground-based mark–recapture surveys have also been used to estimate size of wild horse populations in the Australian Alps (Walter 2002). Surveys were undertaken between 1999 and 2002 over three areas ranging from 13 to 41 km^2,and yielded population estimations between 53 and 109 horses. Natural markings were used to identify individual horses in this study, circumventing the need to trap and mark them with conventional methods (e.g. tagging or branding). Identification from natural markings has been used to study a range of mammals including bottlenose dolphins (*Tursiops truncatus*) (Wilson

---

[†]MJD previously published as M. J. Walter.

10.1071/WR07075 1035-3712/08/040365

et al. 1999), horses (Berger 1986; Turner *et al*. 1992) and elephants (*Loxodonta africana*) (Morley and van Aarde 2007). Some of these studies used mark–recapture methods to estimate population size from the ground or boat. Ground-based population estimation was ruled out for the Bogong High Plains horse population following a survey in 2004 because observation rates were very low per unit of sampling effort (Miller, unpubl. data).

The work presented here was an exploratory study that examined the feasibility of estimating the size of a wild horse population utilising the natural markings of individuals for their identification in an aerial mark–recapture survey. It differs from typical aerial mark–recapture surveys that use double-count methodology (Caughley and Grice 1982). Double-count involves two observers in the same aircraft *simultaneously* counting animals in the same strip independently and without collusion. The front observer's observations are the marking event and the rear observer's observations are the recapture event. In our exploratory study, the population was surveyed from the air over two days (sampling periods). On the first day, all observed horses were identified from their natural markings. The population was then resampled the next day as the recapture event. The population estimate was calculated using mark–recapture models.

Methods

Survey Area

The Bogong High Plains are in north-east Victoria in the Alpine National Park near the towns of Omeo and Mount Beauty, and between Mount Hotham and Falls Creek ski resorts (Fig. 1). The survey area was selected to include all areas where horses had been observed in the Bogong High Plains by Parks Victoria staff and also included additional surrounding areas that may not have had horses. The additional areas were included with the aim of detecting any increase in the distribution of the horses over time. The survey area covers an altitudinal range of 1000 to 1800 m above sea level and comprises two main landscape types. Approximately half of the survey area is a gently undulating high plateau (above 1400 m), with vegetation characterised by a mosaic of alpine and subalpine grassland, heathland and woodland. The remainder of the survey area is made up of deeply dissected river valleys that contain montane forest and woodland communities. Areas that were included in the survey to detect an expansion of horse distribution (but not known to have resident horses) were predominantly in the dissected river valleys.

Survey

The survey involved two sampling events from a helicopter, with the same area sampled on two consecutive days. We

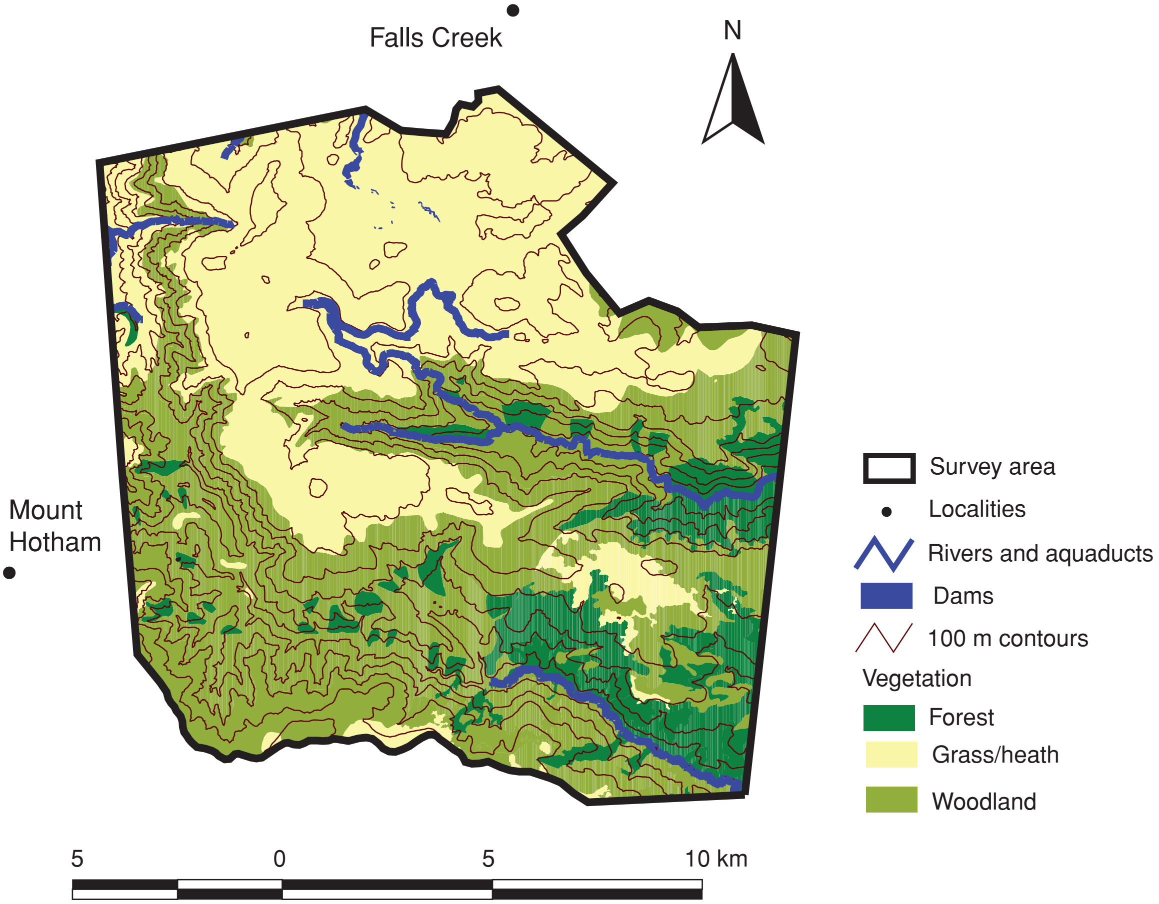


Fig. 1. Bogong High Plains wild horse survey area illustrating vegetation and topography.

limited the survey to two samples because it was a trial survey, for financial reasons and to minimise disturbance to the horses. The first day's survey was a 'marking' event with all individual horses sighted being identified by their natural markings. Individual horse marking details (such as colour, facial markings, socks and age class) were recorded using written descriptions and a digital high-resolution camera (Nikon Coolpix 8700). Fig. 2 shows a photograph taken during the survey showing a group of seven horses, including two juveniles. Note the different white facial markings (stars, snips, stripes and blazes) and variation in the white socks and stockings on their legs. Information on group size, location and vegetation type was also recorded. The second sampling event was conducted the next day and constituted the 'recapture' event.

The survey area was stratified into two zones, plateau and dissected river valleys, with higher sampling effort (h km^{-2}) on the plateau. This strategy aimed to optimise the detection of horses based on their preference for habitats that include grassland (predominantly on the plateau), and minimise the cost of sampling. The plateau survey used a stratified random sampling regime with east–west transects spaced 1 km apart following the map grid at a survey height of 100 m above the ground. The transects on the second day were offset by 500 m to reduce potential bias in capture probabilities. Surveys in the dissected valleys were included primarily to monitor possible range expansion of horse distribution, and to give an indication of the efficacy of surveying horses from the air in this environment. We followed concentric flight patterns (after Choquenot *et al*. 2000) because straight line transects were not practical in such dissected terrain as the helicopter cannot climb and drop fast enough to keep a set survey height. That is, the survey followed the contour of the valleys at mid-slope and also flew along the bottom of the valleys.

On 6 April a trial flight was run and MJD trained two Parks Victoria staff and the pilot on survey protocols. The mark–recapture survey was carried out on 7 and 8 April 2005 in a Bell Jet Ranger helicopter with doors removed except on the right rear. The weather was fine and mild with smoke haze. All aircrew (3) and the pilot searched for horses and when a group was sighted the pilot manoeuvred the helicopter to keep the horses on the left side of the aircraft to enable the left rear observer to describe the group. The observer in the front left (MJD) took photographs (Fig. 2) and also assisted in verbal descriptions during sightings. The right rear observer recorded all data on data sheets. Once a sighting was complete we resumed surveying along the transect. All sighting records were reviewed immediately after the flights and each group was assigned a label (A, B, C etc.).

Confidence in the accuracy of each sighting of individual horses (confident or uncertain) was noted. We were confident of stable groups that had animals with clear distinguishing markings captured with clear photographs. For uncertain identifications, first we assessed the composition of the group, including group size, age of individuals and individuals with distinguishing marks, then looked at their location relative to similar

Fig. 2. Photograph of a group of wild horses taken during the aerial survey for identification purposes. Examples of information for identifying horses include the following. Horse A is an adult, black in colour, with a small star and a left hind sock to below the fetlock. Horse B is an adult red bay with a large star that extends down on the left-hand side and two hind socks to above the fetlocks. Horse C is an adult brown bay with a broad blaze extending over the muzzle and with a left hind sock to above the fetlock.

sightings. If a group with the same attributes had been sighted in the area then they were assumed to be the same group. On two occasions we assigned horses 'resighted' on the second day even though the group composition had changed; however, our confidence was uncertain. These individuals appeared to be bachelors, which often form unstable groups.

The total flight time for the first sample was 3.8 h (including refuelling and transit) and survey distance was ~220 km, while the second sample took 3.9 h and was ~260 km. The actual flight paths were not a direct match with the proposed flight paths because the helicopter had insufficient mapping capabilities. However, straight line transects were flown 1 km apart and MJD assisted the pilot with navigation on the basis of landscape features to ensure the entire survey area was covered on both days.

Analysis

Population size was estimated using four mark–recapture models for closed populations: Chao's moment estimator (Chao 1988), the jacknife estimator (Otis *et al*. 1978), the modified Petersen–Lincoln estimator (Chao and Huggins 2005) and the Huggins estimator (Huggins 1991). The Chao and jacknife estimators were run in Program CAPTURE, a subprogram in Program MARK (White and Burnham 1999), the Huggins estimator was run in Program MARK (with group size as a covariate), and the Petersen–Lincoln estimator was calculated manually. For the Petersen–Lincoln estimator, number of groups was estimated first and the population was estimated as the product of estimated groups and mean group size. These estimators were used because they were able to account for heterogeneity in detection probabilities to varying degrees. This was important because not all horses are equally detectable as horses are more difficult to see if they are in small groups or hidden by vegetation (Walter and Hone 2003).

Results and discussion

Population estimation

We made a total of 128 observations of horses and identified 72 individual horses in 19 groups (Table 1). Mean observed group size was 3.79 (±0.52, s.e.). The four models used returned similar results: population estimates ranged from 89.0 to 94.7 horses and precision ranging from 5.3 to 8.5 (s.e.) (Table 2) with no significant difference in the estimates (as evident by the overlap of the 95% confidence intervals). It is not possible to determine which of the estimates is the most accurate because we do not know actual population size.

This survey method overcame issues of bias that occur in other aerial survey techniques. Our concern over using the double-count mark–recapture technique (Caughley and Grice 1982) often used in aerial surveys was justified because individual horses were sighted more than once in one sampling period (Table 1). A double-count survey would have given an inflated estimate as each sighting would have been assumed to be of a new group/horse. Our method also circumvented the limitation of the double-count technique where both observers (which represent a sampling event) have the same vantage point and that animals that are difficult for one observer to detect are difficult for the other, leading to unequal detection probabilities. Furthermore, some aerial surveys of horses have assumed that one sampling period is a total count (e.g. Garrott *et al*. 1991; see Linklater and Cameron 2002). If either of the two samples from this survey had been assumed to be a total count they would have given estimates 55% and 63% respectively below the Chao estimate.

Key assumptions of the mark–recapture estimators were not met in our study, which could lead to bias in the population estimate. Assumptions that were violated were: (1) that animals are correctly identified and recorded during marking and recapture, and (2) animals act independently. Violation of the first assumption may be largely overcome by modifying the survey technique in the future (see below). It is important to test the effect of violating the second assumption, but this is not possible with only two sampling occasions. Animals clearly do not act independently as they are in groups, and their capture probabilities depend on that of the group. We attempted to overcome this issue by selecting estimators that allowed for heterogeneity in capture probabilities. However, we strongly recommend that five sampling periods be used in the future to enable the testing of this assumption. Alternatively, accuracy could be assessed if change-in-ratio methods were employed through removal of a significant and known number of horses from the population as part of a control program (Krebs 1999). An additional assumption that was not met for the Chao (1988) and jacknife (Otis *et al*. 1978) estimators was that the number of trapping samples should be at least four or five.

High precision was achieved through high sampling intensity and resighting rates on the plateau. Sampling intensity in treeless areas of the plateau was equivalent to 100% because groups of horses are clearly visible over 500 m, and any horses that were in the flight path of the helicopter were seen by the pilot or the front observer. This is contrary to line transect and double-count mark–recapture surveys where animals are missed directly under the aircraft (particularly fixed-wing) as

Table 1. Horses observed during the Bogong High Plains aerial survey, 2005

Note that groups B and C changed composition between days so were split into two groups.

Group	Group size	Times sighted Day 1	Times sighted Day 2
A	2	1	0
B1	2	1	1
B2	2	1	0
C1	3	1	1
C2	2	0	1
D	4	1	1
E	3	1	1
F	3	1	0
G	1	1	0
H	6	1	0
I	9	1	2
J	8	1	2
K	6	1	1
L	1	1	0
M	6	0	1
N	2	0	1
P	4	0	2
Q	4	0	1
R	4	0	1

Table 2. Summary of estimates of the size of the Bogong High Plains horse population
Refer to text for information on estimators.

Estimator	$\hat{N}_{\text{groups}}$ (s.e.)	$\hat{N}_{\text{horses}}$ (s.e.)	Notes
Chao	–	92 (8.5)	Using Chao's notation, $f_1 = 37$, $f_2 = 35$, $S = 72$, $t = 2$
Jacknife	–	89 (5.3)	
Petersen–Lincoln	23.5 (3.5)	89.1 (7.9)	
Huggins	25.0 (3.5)	94.7 (7.9)	p(groupsize)c(groupsize), p(.)c(groupsize), p(groupsize)c(.) and p(.)c(.) models gave estimates within 0.1 groups

observers are looking out to the side and the pilot does not participate. Sampling intensity was lower in areas with tree cover because tree cover reduces the sightability of horses with increasing distance from the helicopter (Walter and Hone 2003). This logically varies depending on the height and density of the trees. Vegetation was also found to be an important covariate in mark–recapture analysis for broadscale aerial survey of wild horses in the Australian Alps (Laake *et al.* 2008).

Survey technique

Not all of the horses were correctly identified and recorded during marking and recapture. We were confident of the identification (or marking) of only 103 of the 128 sightings of horses (80%). This needs to be improved if the technique is to be successfully employed because any misidentification of horses affects population estimates. For example, if five of the horses recorded as being sighted twice in the study were actually being sighted for the first time, the population estimate would have been 105 instead of 92 (using the Chao estimator).

Identification of horses could be improved through better recording and observation techniques. The reasons horses were difficult to correctly identify were limitations in the identification and recording techniques (particularly through vegetation) and/or lack of distinguishing features on horses. Observer inexperience and data recording methods limited the quality of data collected. These effects could be mitigated in the future. The survey required inexperienced parks staff to be observers, and it is well established that observer experience can affect the accuracy of aerial survey estimates (Caughley *et al.* 1976).This was evident in the current survey; however, when the same parks staff conduct the survey in the future their skills will improve. Furthermore, the use of a digital video camera will improve horse identification. The digital stills camera was more useful than verbal descriptions for identifying horses in all cases except where there was a tree canopy. A digital video would improve sightings though tree canopies as it is possible to search frame by frame to obtain a clear image of a horse. Digital video and stills cameras should also resolve issues over identification of horses with few distinguishing features as attributes such as height and body condition of horses can be compared.

All horses were sighted on the undulating plateau areas and none were sighted in the dissected valleys. The lack of sightings in the dissected river valleys was most likely attributable to very low densities (or absence) of horses in this habitat, as well as low sampling effort and low detection through the tree canopy. It would be more efficient to exclude the dissected river valleys from the aerial survey and focus on the plateau for population estimates in the future. This would reduce flight-time and costs by one-third.

Implications for management

The population estimates of wild horses in the Bogong High Plains derived from this study are lower than the previous 'best guess' estimate for the area of 200 horses (Walter 2002). Using the Chao model estimate of 92 (±8.5, s.e.) horses, this equates to a density of 0.51 horses km^{-2} over the entire survey area, or 1.02 horses km^{-2} on the plateau section of the survey area where all of the horses were sighted. This is lower than the average density estimated in a broadscale survey of the Australian Alps in 2001 (2.43 horses km^{-2}) (Laake *et al.* 2008). The level of precision (Table 2) was higher than that required by Parks Victoria for management purposes (coefficient of variation less than 35%). These estimates should be used with caution until the survey technique is refined and model assumptions are tested.

This exploratory study would suggest that aerial survey is most suitable for estimating population size of the core horse population on the Bogong High Plain plateau. Expansion in distribution of horses into the dissected river valleys may be better detected by monitored for the presence of horses through on-ground observations by parks staff and visitors. If horses are frequently detected in a new area then they may be included in future aerial surveys.

Conclusion

This novel population survey method has proven to be viable in this context, but requires refinement. Once practical issues of identifying horses are resolved, the survey should be run with five sampling periods to ensure that model assumptions are not violated. The technique has proven useful for identifying large, conspicuous animals with unique natural markings. Two other studies conducted independently and at the same time as ours used similar techniques and also found the method useful (Freeman 2005; Ransom *et al.* 2005). Neither sampled more than twice, so did not address the underlying assumptions of the models used, but the survey of Ransom *et al.* (2005) was of a population of known size and the estimate was not significantly different.

Acknowledgements

We thank Parks Victoria for funding the survey, Colleen Nagle and Enzo Brotto from Parks Victoria for participating in the aerial survey and the pilot David Empey for keeping us safe. We also thank James Dawson for reviewing a draft of this manuscript, and three anonymous referees who provided valuable input to the final draft.

References

Berger, J. (1986). 'Wild Horses of the Great Basin. Social Competition and Population Size.' (University of Chicago Press: Chicago, IL.)

Buckland, S. T., Anderson, D. R., Burnham, K. P., and Laake, J. L. (1993). 'Distance Sampling: Estimating Abundance of Biological Populations.' (Chapman and Hall: London.)

Caughley, G., and Grice, D. (1982). A correction factor for counting emus from the air, and its application to counts in Western Australia. *Australian Wildlife Research* **9**, 253–259. doi:10.1071/WR9820253

Caughley, G., and Sinclair, A. R. E. (1994). 'Wildlife Ecology and Management.' (Blackwell: Cambridge.)

Caughley, G., Sinclair, R., and Scott-Kemmis, D. (1976). Experiments in aerial survey. *Journal of Wildlife Management* **40**, 290–300. doi:10.2307/3800428

Chao, A. (1988). Estimating animal abundance with capture frequency data. *Journal of Wildlife Management* **52**, 295–300. doi:10.2307/3801237

Chao, A., and Huggins, R. M. (2005). Classical closed population models. In 'The Handbook of Capture Recapture Methods'. (Eds B. Manly, T. McDonald and S. Amstrup.) pp. 22–35. (Princeton University Press: Princeton, NJ.)

Choquenot, D., Bolton, N., and Woods, D. (2000). Evaluating the application of aerial survey to estimate Himalayan thar densities. Landcare Research Contract Report: LC0001/29. Landcare Research, Wellington.

Freeman, M. (2005). Density estimation of feral horses in Guy Fawkes River National Park New South Wales, using an aerial mark–recapture method. B.Nat.Res. Thesis, University of New England, Armidale.

Garrott, R. A., Siniff, D. B., and Eberhardt, L. L. (1991). Growth rate of feral horse populations. *Journal of Wildlife Management* **55**, 641–648. doi:10.2307/3809513

Huggins, R. M. (1991). Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics* **47**, 725–732. doi:10.2307/2532158

Krebs, C. (1999). 'Ecological Methodology.' (Benjamin/Cummings: San Francisco, CA.)

Laake, J., Dawson, M. J., and Hone, J. (2008). Visibility bias: mark–recapture, line-transect or both? *Wildlife Research* **35**, 299–309. doi:10.1071/WR07034

Linklater, W. L., and Cameron, E. Z. (2002). Escape behaviour of feral horses during a helicopter count. *Wildlife Research* **29**, 221–224. doi:10.1071/WR01063

Morley, R. C., and van Aarde, R. J. (2007). Estimating abundance for a savanna elephant population using mark–resight methods: a case study for the Tembe Elephant Park *South African Journal of Zoology* **271**, 418–427.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs* **62**, 1–135.

Ransom, J., Singer, F. J., and Zeigenfuss, L. (2005). Annual report for 2004 wild horse research and field activities. Open-File Report 2005–1316. US Geological Survey, Fort Collins, CO.

Turner, J. W. J., Wolfe, M. L., and Kirkpatrick, J. F. (1992). Seasonal mountain lion predation on a feral horse population. *Canadian Journal of Zoology* **70**, 929–934.

Walter, M. J. (2002). Population ecology of wild horses in the Australian Alps. Ph.D. Thesis, University of Canberra.

Walter, M., and Hone, J. (2003). A comparison of 3 aerial survey techniques to estimate wild horse abundance in the Australian Alps. *Wildlife Society Bulletin* **31**, 1138–1149.

White, G. C., and Burnham, K. P. (1999). Program MARK: survival estimation from populations of marked animals. *Bird Study* **46**, 120–139.

Wilson, B., Hammond, P. S., and Thompson, P. M. (1999). Estimating size and assessing trends in a coastal bottlenose dolphin population. *Ecological Applications* **9**, 288–300. doi:10.1890/1051-0761(1999)009[0288:ESAATI]2.0.CO;2

Manuscript received 27 June 2007, accepted 23 January 2008

COMMENT ON *'Independent biostatistical report on the Brumby population in the Kosciuszko National Park'*

Document tendered by
Mr Jack Gough
Received by
Emily Whittingstall
Date: 18 / 12 / 2023
Resolved to publish Yes / No

Don Fletcher PhD, PSM
4 October 2023

Intentionally blank

Contents

INTRODUCTION AND GENERAL REMARKS

What report is referred to?

The report I comment on *'Independent biostatistical report on the Brumby population in the Kosciuszko National Park'* dated 20 May 2023, is available at https://meetourhorsemeat.com/wp-content/uploads/2023/05/Independent-Biostatistical-report-into-the-counting-of-wild-horses-Claire-Galea.pdf. I refer to the report hereafter as Galea (2023).

Not a scientific process

It is important to recognise that neither Galea (2023), nor this commentary, have been subjected to the normal quality control processes that apply in science, i.e. they have not been published in a journal which deals with wildlife counting methods, which would have subjected them to serious editorial inspection, and the opinions of two or three anonymous peer-reviewers. Also there is the potential for subsequent criticism to be published later in the same journal.

However, any publicly available report such as Galea (2023), which is concerned with scientific matters such as the estimation of animal abundance, should rightly be open to fair criticism or comment on factual and scientific grounds. The justification for such evaluation is increased in this case because the report recommends a major change to the current management of Kosciuszko National Park. Also because the report has not only been placed on the internet, but also sent to the offices of most of the relevant politicians who are involved with management of feral horses in Kosciuszko National Park (KNP).

Not personal

Nothing in the following comments about the content of the report should be read as personal criticism of Mrs Claire Galea herself. I presume she is a fine professional in her own field, as evidenced by the impressive number of scientific papers which she has co-authored, and her acceptance by a university as a PhD student.

Similarity of wildlife controversies

Current controversies over management of feral horses in Australian conservation areas share some features with most other wildlife management controversies. It is particularly common for scientists' population estimates to be disputed. For example, many critics of wildlife management programs have little respect for ecological expertise, and believe there are no more animals present than they have observed themselves.

Importance of this response to Galea 2023

Galea (2023) has been claimed to be both 'independent' and to be based upon greater expertise than that of the professional ecologists and public servants responsible for the official counts of horses in KNP. However I show below that there are numerous deficiencies of science, logic and statistics in the report. On the basis of these deficiencies and the legal, ecological and ethical imperatives outlined below for feral horse population reduction, I challenge the only recommendation in the report, which is for *'Immediate moratorium on the killing of all wild horses in the Kosciuszko National Park and an independent investigation into all population trends and subsequent control needs to be urgently undertaken'.*

Independence

The report title says it is 'independent'. Presumably the reader is meant to infer that the report is unbiased because its author has no links to either side of the horse controversy in KNP. However perusal of the report uncovers instances where a more detached statistician would probably have made a more logical conclusion (see below) or avoided making an error of scholarship (identified below). This made me wary of the claim for independence.

The report displays no understanding of the legal, ecological and ethical contexts of the horse counts

The NSW *Kosciuszko Wild Horse Heritage Act* (2018) requires a horse management plan which recognises and protects wild horse heritage values in Kosciuszko National Park and enables active management of the wild horse population to reduce its impact on the park's fragile environment. Consequently, the Kosciuszko National Park Wild Horse Heritage Management Plan (the Horse Plan) was adopted by the Minister for Energy and Environment on 24 November 2021 following a massive exercise in democracy, including extensive consultation with expert committees, thousands of public submissions and the approval of several key Cabinet Ministers in the NSW Parliament, including the Deputy Premier, John Barilaro who was responsible for introducing the bill that required the plan to be prepared. The Horse Plan, a binding legal instrument, requires the current horse population (estimated in 2022 to be ~19,000) to be reduced to 3,000 by 30 June 2027.

A further requirement for horse population management has been established by the listing of *'habitat degradation and loss by feral horses'* as a Key Threatening Process in Schedule 4 of the NSW *Biodiversity Conservation Act* (2016). The NSW government is thereby required to ameliorate the biodiversity threat where possible.

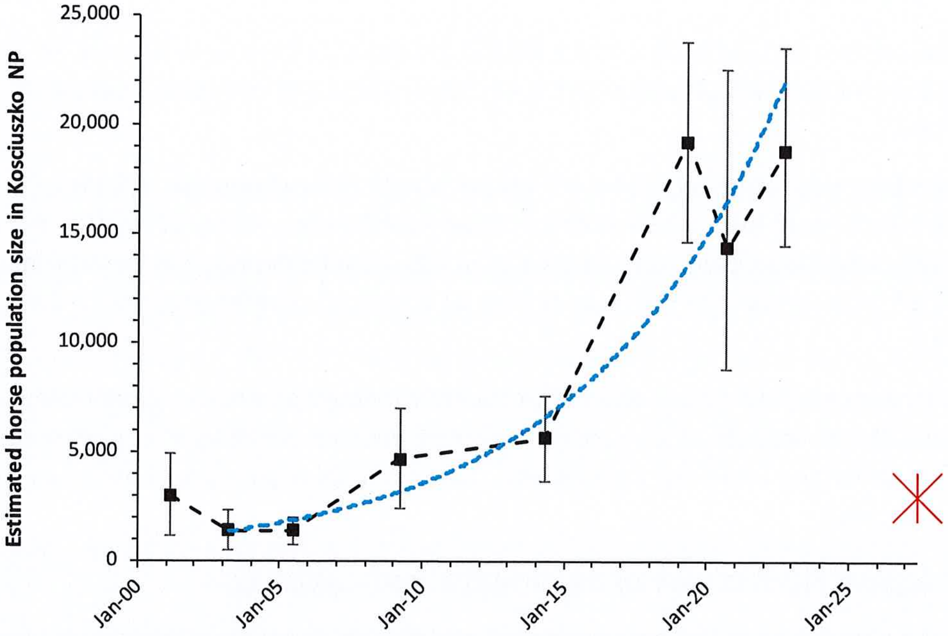
Scientists have estimated the size of the horse population in KNP nine times. The first was in the late 1980s by Dyring (1990). Eight estimates since 2001 all used Helicopter Line Transect Distance Sampling (HLTDS), a method explained below. For years, the exponential increase of the horse population has been plainly evident to anyone who has been interested in the counts (Figure 1 and Appendix table).

It is the nature of exponential population growth to seem slow for a long time while the population is relatively small, then to seem to increase rapidly when the population is larger. Thus the risk to native Australian species of animals and plants has become more acute in the last few years, now that the horse population is large and increasing rapidly.

As a result, although the first horses entered the area more than 100 years ago, in 2023 many plants and animals in the park will be encountering horses for the first time. Others will be experiencing high levels of horse impact for the first time.

Scientists have been pointing to the increasing threat to biodiversity associated with the increasing horse abundance and distribution for 70 years, since 1953. But now it has become worse than a threat, with actual loss of populations having occurred due to horse activity; e.g. populations of tooarrana *(Mastacomys fuscus)* have been lost (Driscoll *et al.* 2019; Eldridge *et al.* 2019; Schulz *et al.* 2019) most likely through reduction of ground layer vegetation which is essential for this native rodent to persist (Cherubin *et al.* 2019). Thus, even if there was not the legal imperative requiring management of the horse population, there are clear ecological and ethical imperatives to act.

Figure 1: Eight estimates of the horse population in KNP (squares) were accompanied by wide 95% confidence intervals (error bars). But overall there has been a consistent trend of 15% annual increase (dashed blue curve) since the 2003 bushfire, except around the time of the 2020 bushfire. The red star marks the commencement of the requirement for there to be only 3,000 horses from 30/6/27. See the Appendix for details.



In this legal, ecological and ethical context, there is no option to recommend that the horse population should not be reduced. Yet that is the only recommendation of this report (Galea 2023, p. 4). On that basis, the report can be regarded as being somewhat adrift from the current legal, ethical and ecological circumstances. In addition there are numerous flaws in the case it presents for doubting the horse counts, as detailed below.

The report does not suggest a better survey method, or provide any way forward

The Horse Plan indicates that feral horses can be found in 53% of the 6,900 sq km KNP, i.e. in an area of 3657 sq km. This area probably includes all places where rangers have seen groups of horses, but counting is limited to a smaller area where horses are more likely to be encountered, i.e. 2,745 sq km. No experienced ecologist would imagine that a population of wild horses spread over even this 2,745 sq km part of KNP could be reduced accurately to 3,000, from more than 10,000 individuals, without further counting. And the estimated population size in 2022 was not 10,000 but was almost twice as large, i.e. 19,000 approximately (Cairns 2022).

As the horse population is reduced closer to the target population size, counts of good precision and accuracy will be needed more frequently than ever before, to guide the culling program so that it achieves a result of acceptable precision and accuracy. The legal, ecological and ethical imperatives referred to above make horse control, and therefore horse counting, essential between now and June 2027. The most suitable counting method available for the terrain and population size at KNP is Helicopter Line Transect Distance Sampling (HLTDS) (Walter and Hone 2003), which is the method used for all eight surveys carried out in the last 22 years.

Galea (2023) identifies many causes of claimed inaccuracies. However, the report does not state whether the claimed problems would make the counts too low, or too high. The important question of whether the claimed inaccuracy is positive or negative appears not even to have been thought of. The author has simply assumed that errors would result in overestimation of horse abundance.

No counting method is perfect. Internationally, distance sampling has more often underestimated than overestimated, and in particular, distance sampling has been shown to be underestimating the count of horses in KNP (Laake *et al.* 2008). If all the methodological deficiencies claimed in Galea (2023) were real, there may be far more horses than suspected and an even greater problem than experts currently appreciate.

Another conceptual omission from the thinking behind the report is the lack of any way forward; which is an extremely important gap, considering the legal, ethical and ecological imperatives mentioned above.

No wildlife population estimation method is perfect but unless some alternative or improvement can be identified, it is fruitless to focus much attention on any deficiencies. Galea (2023) does not outline how an alternative population estimation method to HLTDS could be deployed in the terrain and vegetation of KNP. (In some parts of KNP even helicopter counting is challenging, due to the terrain and vegetation.)

In fact the report does not refer to any alternative scientific method of measuring population size. It simply states (p. 12) that because of (claimed) deficiencies, distance sampling is *'not appropriate methodology for estimating wild horse populations'*. It does not name any other survey method which might be superior.

The report suffers from a lack of experience with ecology

The author of the report is an experienced medical statistician whose name is included in the author-list of a large number (~50) of papers published in peer-reviewed medical journals. Possibly she provided the statistical services needed for data analysis in those projects. Galea is also listed on the Linked-In website (https://www.linkedin.com) as a PhD candidate in the Faculty of Medicine at Macquarie University, Sydney. An on-line article from the university says Galea is *'leading the world's first global evaluation of Dolly Parton's Imagination Library including a research focus on children in the NSW town of Tamworth'*.

I found no evidence of ecological training or experience in Galea's background and there is nothing like this report (Galea 2023) published in any ecology journal. Measuring the abundance of wildlife is a notoriously challenging area of ecology which requires not only skill with the particular statistical challenges of ecology (e.g. see 'B2 Transformation of the data – p. 12') but also a good deal of field experience. The report reads as if it rests on limited field experience in ecology.

In estimating the numbers of horses in national parks like Kosciuszko, the challenges include the impossibility of uniquely identifying all the individual horses in large populations* and the impossibility of seeing all the horses on any one occasion. In the medical studies co-authored by Galea, each person involved would have had unique identifiers (name, date of birth etc). And there is no evidence that Galea's research has involved patients who could not be seen by the researchers. Thus Galea's extensive experience in medical statistics appears unlikely to include experience in two of the main challenges of counting wild animals.

* Individual identification of horses <u>can</u> be used to estimate population size by Mark-Resight analysis, if the population is small enough. Dawson and Miller (2007) observed 50 horses in a 180 sq km area of the Bogong High Plains when they searched it by helicopter one day. The next day they searched it again and recorded 78 individuals. Some of these were seen more than once on the same day and the total number of individual horses seen on either day was 72. From this, the population was estimated to be 89 to 95 horses (so 23% to 32% of horses were never seen on either day). The statistical analysis used the well known Mark-Recapture method. The authors speculated that the method may have practical application for aerial surveys of small populations, subject to methodological improvements such as a change to video rather than still imagery to recognise individual animals (Dawson and Miller 2007). The method of recognising individual horses was not considered suitable for estimating the number of horses to be counted in KNP, which is hundreds of times more. Nor does the counting budget allow time for horse photography and the careful observation required with this method.

Too many simple mistakes

There is a distinct impression that the report (Galea 2023) was written in haste because it has so many simple errors. One example is the mistaken claim that the Kosciuszko HLTDS surveys do not count foals separately (Galea 2023, p17). Yet the foal counts are obvious in the 2020 and 2022 reports. In Cairns (2022), the foal counts are mentioned in the Summary, in the Methods, and in the Results and Discussion. And they occupy more than two entire pages. (see more details in my response to criticism C3). Similar mistakes are mentioned in my comments on criticisms 'B1 – Statistical modelling', 'B2 Transformation of the data', 'B3 - Use of covariates' and 'B4 - Assumptions'. Also there is repetition of the same criticisms under different headings.

The prevalence of such easily identified errors (no statistical knowledge needed) might lead a reader to lose confidence in the main claims of the report.

It is the latest survey (Cairns 2022) which is the most important so it is odd that the report gives no reason for its focus on the 2014 and 2019 surveys with only occasional mention of the latest (2022) survey, and no mentioned at all of the 2020 survey which was completed between those surveys that are mentioned. The omission of the 2020 count report from Galea's criticism was probably just an oversight, and not because Galea approved of the 2020 count.

About the horse counts to date

Over the past 22 years, the feral horses in Kosciuszko have been counted eight times by a number of different people from independent universities and the NPWS. Until recently, most of the counts were funded by the Australian Alps Liaison Committee (AALC), and most provided a single combined estimate of the horse population of the alpine national parks in Victoria and NSW.

In spite of limitations, collectively the set of eight horse counts has provided a consistent and plausible result (Figure 1). The counts are remarkably consistent in showing an average annual increase of 15% except when the horse population was reduced by the bushfires of 2003 and 2020 (Figure 1).

As previously mentioned HLTDS was the method used for all surveys. In the name 'Helicopter Line Transect Distance Sampling', the words 'line transect' refer to the straight lines from which the horses are observed. These transects are parallel east-west lines (in one or two cases over the 20 years a different direction was used in small steep areas) and within a survey block the lines are equidistant (*e.g.* Cairns 2022) so this design is referred to as a 'systematic' layout. Each set of transects has a randomly chosen start point, hence the design is sometimes referred to as 'systematic random' (*e.g.* Cairns 2022). 'Distance Sampling' refers to the analytic method used for statistical analysis of the data, typically using the free program *'Distance'* or an equivalent package in statistical program R. (There are other ecological methods based on transect lines which are not distance sampling and there are other examples of distance sampling which are not from lines or which are not done from helicopters).

In brief, Distance Sampling exploits the fact that fewer animals are detected at greater distance from an observer, in order to enable an estimate of how many animals are not seen, in addition to those which are detected and recorded. The Distance Sampling method (Thomas *et al.* 2010 and https://distancesampling.org/whatisds.html#online-bibliography) is one of the most widely used methods in the world for estimating abundance of wildlife populations. Its mathematical and statistical foundation is comprehensively explained in two books, particularly Buckland *et al.* (2001), and a second book covering more advanced applications (Buckland *et al.* 2004). The results have been evaluated against known populations on numerous occasions and found to be accurate (*e.g.* Hone 1988; Hounsome *et al.* 2005; Glass *et al.* 2015). Thousands of published, peer-reviewed scientific papers exemplify its use. More than 1,400 of them can be found in the bibliography at https://distancesampling.org/dbib.html. There is a wide range of species whose populations have been counted using Distance Sampling, including insects, crabs, fish, reptiles, antelopes, deer, kangaroos, feral pigs, fruit bats, primates, polar bears, whales, dolphins and mice, as well as inanimate objects such as birds' nests, mammal burrows and carcasses (Buckland *et al.* 2001 p11). There is no reason to doubt the Distance Sampling method itself.

In the Australian Alps, HLTDS has been compared to two other counting methods to estimate abundance of horse populations. The study authors recommended HLTDS for future use (Walter and Hone 2003). Since then, the design, analysis and reporting of recent surveys by Cairns (2019, 2020, 2022), and the results obtained, have been reviewed and found acceptable on several occasions by different groups of truly independent scientists who are themselves experienced with statistical analysis of wildlife counts. Overall, there has been a considerable body of scientists who approve of the use of HLTDS, and how it is being applied in KNP to estimate horse abundance.

Aside from the three methods compared by Walter and Hone (2003), only one other method of counting feral horses in the high country has been peer reviewed and published (Dawson and Miller 2007). It relies on recognition of individual horses in small sub-populations, as explained above under the heading 'The report suffers from a lack of experience with ecology'. But it is not the earlier surveys which matter. The most recent survey (Cairns 2022) is the most important for horse management.

The length of transect traversed in the latest survey (survey effort) was 1,496 km. To traverse this distance in a much slower platform than a helicopter (for example on horseback) would contravene statistical requirements (see below) as well as being impractical. To survey a shorter length of transects by helicopter would result in a higher coefficient of variation (i.e. greater uncertainty around the population estimate; see below). There seems almost no chance that a ground based survey method could be found which would enable the horse populations in KNP to be estimated reliably across their full extent.

Accuracy and precision are less important now than later

The upward trend in the horse population is obvious (Figure 1) and corresponds with the experience of long-term observers who have seen at first-hand the horse distribution expanding and the abundance increasing. The expanding distribution of horses has also been reported by several observers including Dawson (2009).

What is apparent from Figure 1 is that previous and current population control measures have not reduced the size of the horse population, nor its rate of increase. While the horse population is growing exponentially, each year the number of horses moves further from the target of 3,000 and the distance it moves away from the target is greater every year.

Therefore at the present time, there is a limit to the importance of accuracy and precision of the count. Count accuracy and precision will become more important after the commencement of a horse control method which has the capability to reverse the current trend and cause the population to move toward the target of 3,000 individuals. Rather than wasting funds and human effort on greater precision and accuracy when the horse population is in the range of 10,000–30,000 individuals, it would be better to focus on the attainment of a program which can actually reduce the horse population as required, and save the counting precision for when it will be needed.

Therefore a detailed discussion of claimed statistical imperfections in the horse count (Galea 2023) is of limited use or importance at this time.

COMMENT ON PARTICULAR CRITICISMS IN GALEA 2023 –

1 Cluster size – page 6

In spite of the title, this criticism is actually about the number of clusters. My response is in three parts.

Ecologically naive statistical approach which is aimed at the wrong survey

The main criticism under this heading is that in some sub-populations of horses counted in the 2014 and 2019 surveys, the number of clusters was below the minimum number recommended for Distance Sampling, i.e. 60. It is the latest survey, conducted in 2022, that should be our main focus but for an unknown reason, this particular criticism is applied only to the 2014 and 2019 counts.

The aim of the surveys was to estimate the number of horses in the Australian Alps (Cairns2014, 2019) or in KNP (Cairns 2022). Therefore both surveys necessarily and rightly included some areas of high horse abundance and some areas of low horse abundance. For all three surveys the total number of clusters was well beyond the minimum of 60 required, i.e. 301, 458 and 419, respectively. The number of clusters was fewer than 60 only in some sub-component areas. Surveys across the range of horses within KNP (or of any species anywhere) inevitably must (and should) include areas where the population is advancing into new areas, or for other reasons is at low density. In these sub-components of the surveyed area it is inevitable that fewer clusters will be recorded than elsewhere. In the 2022 survey fewer than 60 clusters were observed only in the Snowy Plain sub-population, i.e. 47 clusters.

More funds would have enabled more counting over the same ground until at least 60 clusters were seen at Snowy Plain, in order to obtain a better estimate of the number of horses in that site and keep at bay any criticism such as that in this report. (The extra survey effort in that small area would not bias the result, contrary to Galea's Criticism 4, as explained under that heading.) But it would be wrong to apply extra effort to Snowy Plain for a different reason, i.e. because Snowy Plain contained only 4% of the total horse population, so a small improvement in the estimate there would have made negligible difference to the total count. If extra counting could have been done, the effort could have been used more efficiently in the open plains of northern KNP which contained 68% of the estimated population.

The number of clusters counted in the Bago Maragle block in 2014 and 2019 is also criticised on p. 7. The Bago-Maragle block (on the western edge of KNP near Cabramurra) contained 11% of the horses seen in 2014 and only 4% of the horses seen in the 2019 survey. The same comments apply to this criticism.

The *Distance* computer program allows for stratification across different surveys, and after the 2019 survey, Cairns (2019) combined the data for this block from both surveys to obtain a better population estimate. Galea (2023) raises theoretical concerns with the practice of combining surveys, and concludes 'no reliable population estimates can be determined'. Galea (2023) does not indicate what alternative action Cairns (2019) could or should have taken, but the obvious alternative would be a foolish one, to reduce the survey effort where horses were abundant in order to spend more survey effort counting such places as Bago–Maragle where they were uncommon, until more than 60 clusters were seen in each and every sub-population. Instead, by combining results from both surveys, Cairns (2019) has prudently responded to the reality that some survey blocks have few animals.

The report goes on to compound this misunderstanding by complaining that the number of clusters seen in the individual 'medium terrain' stratum within the north Kosciuszko block are too few. The point is that the numbers of clusters seen in the whole block are sufficient, i.e. 84 + 20 =104 in 2014 and 226 + 43 = 269 in 2019. Note that the numbers in the table on p. 6 of Galea (2023) are correct, but mistakes were made when they were repeated in the text on p. 7.

Criticism of kangaroo counting exemplifies ecological naivety

On page 6 of Galea (2023) there are two tables labelled 'Table 3'. The second one presents data from a kangaroo count in New England. The table shows the numbers of Eastern Grey Kangaroos, Wallaroos, Red Necked Wallabies and Swamp Wallabies counted by HLTDS. The district count was subdivided into six sub-component areas. Several of the counts in sub-components recorded fewer than 60 clusters, but never for the main species of interest, the Eastern Grey Kangaroo, only for Wallaroos or the two Wallabies, which are evidently much less common in these areas than Eastern Grey Kangaroos. In these circumstances it is absurd to complain that the uncommon species did not record 60 clusters. I will explain using an example - in semi-arid western NSW where Red Kangaroos and Western Grey Kangaroos dominate, the occasional Eastern Grey Kangaroo is found in some years, but they are less than 0.1% as numerous as the primary species. It would be absurd, and perhaps impossible, to try to record 60 clusters of Eastern Grey Kangaroos in such places. All species are not common in all habitats! However there is an interest in any change over time in the relative abundance of them all, and in their conservation, so it makes sense to record them.

Misunderstood meaning of 'cluster'

The report has misunderstood the meaning of 'cluster' in Distance Sampling. It states that '*a cluster is considered to be more than one animal*'. In fact, the recording and analysis of data in Distance Sampling is based on the reality that animals occur in all sizes of groups or clusters from 1 animal upward (Buckland *et al.* 2001). And, this is a central feature of the method. The importance of clusters arises from the Distance Sampling developers' insistence on rigorous statistical standards, as explained in the next paragraph. If Galea's definition of clusters were adopted, all the animals seen as singletons would be omitted from the count.

To understand clusters in Distance Sampling, we begin with the statisticians' awareness that when an observer has seen one of the animals in a group, there is an increased likelihood that other animals in the group will also be seen. So if animals were recorded separately, as if they had been detected individually, the variance of the population estimate would be underestimated. To achieve impeccable statistical rigour, the observer in Distance Sampling is required to record details of animal clusters (i.e. groups) and the data analysis is done in two parts. The density of animal groups or clusters is estimated (i.e. average number of clusters per unit area) and its variance. Separately, the average size of clusters is estimated, and its variance. Then the two are combined. (In reality however, the analysis is more complex, for example the estimate of group size is usually done as a regression of distance from the observer because observers tend to perceive larger groups at greater distance -Buckland *et al.* 2001).

2 Cluster observation – page 7

The report says *'given insufficient clusters of wild horses .... were seen .... No reliable population estimate can be determined'.*

First, it is not true that insufficient clusters were seen, as explained above. Second, Galea's criticism that mean cluster size is prone to be affected by outliers, reveals a misunderstanding of the mechanics of cluster size analysis in the *Distance* computer program. The estimation of mean cluster size is based on the regression of size over distance (not a simple mean) and has been considered valid by numerous statisticians of international repute. On this point Galea (2023) is incorrect, statistically speaking.

3 Lack of precision – page 8

The criticism here on p. 8 is repeated on p. 15 where it is headed 'B8 Width of the confidence intervals'. One small difference is that under this heading the report focusses on Cairns (2019) but under heading B8 it is Cairns (2022) which is criticised.

To avoid repetition, and because the latest count is the one which matters most, I have responded to this point under B8.

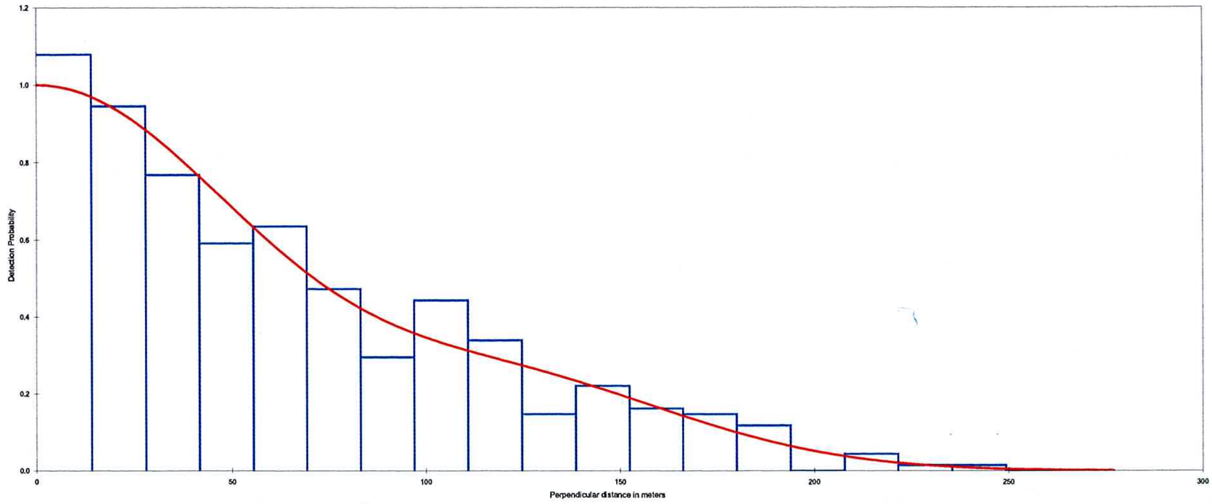
4 Bias sample location – page 11

The report is concerned that counting effort should be applied evenly across the distribution of horses to avoid bias. This comment is plain wrong statistically. It is perfectly acceptable, and in some cases desirable, to apply different levels of survey effort in different parts of an area being surveyed, in order to improve accuracy or precision of the result of the survey (Buckland *et al.* 2001). In Distance Sampling, additional effort can be applied within a stratum, potentially to improve precision and the estimation of the shape of the detection function (explained below). This would improve the accuracy of the estimate of density (horses/sq km) for that stratum. The estimates of the number of horses in the individual strata are subsequently combined to obtain the total population in the counted areas.

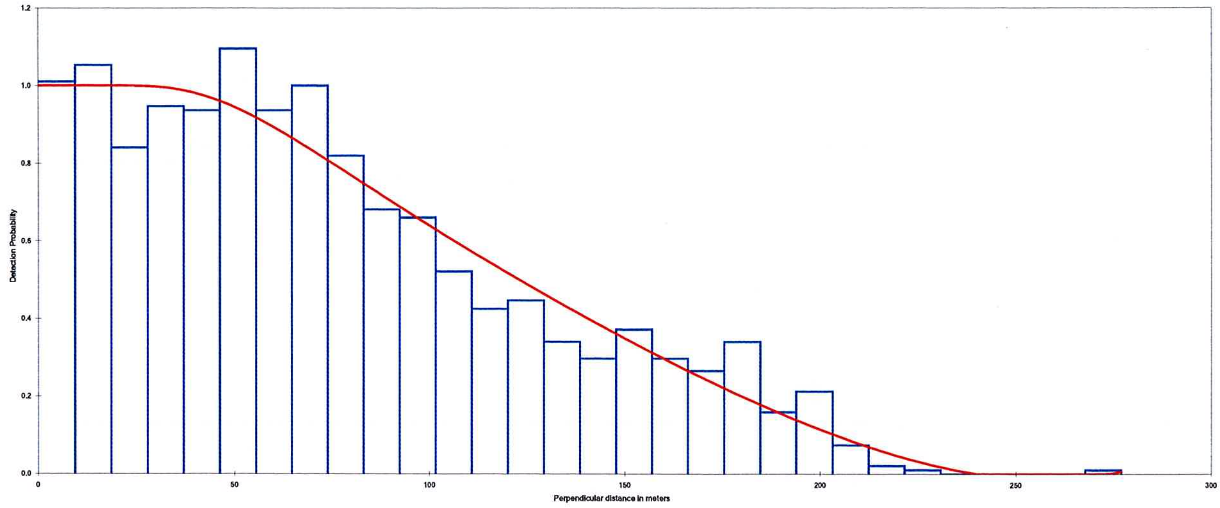
The detection function is the mathematical equation that best describes the proportion of animals present that were actually seen and recorded, as a function of distance from the transect, as illustrated below for two kangaroo surveys. It is a model representing the probability of detecting an object in relation to the covariate of the perpendicular distance that the object might be from the transect centreline, and any other relevant covariates that are measured and recorded. In relation to this is the expectation that objects on the transect centreline (x = 0) will be detected with certainty (g(0) = 1).

Figure 2: Different detection functions in Distance Sampling: bars indicate the proportion of clusters recorded at particular distances from the transects; and red lines are the equations that best describe the shapes of these relationships, (a) a half normal equation with cosine adjustment and (b) a hazard rate equation. Because the transect lines are on a random systematic pattern, in a properly conducted survey it can be safely assumed that on average there will be, in reality, an equal number of clusters at all distances from the transect line. The red lines show the proportion that were seen by the observers.

(a)



(b)



5 Use of line transects with respect to speed of wild horses – p. 12

The report asserts that HLTDS (referred to as '*use of line transects*') is '*not appropriate methodology for estimating wild horse populations*'. Strong claims such as this require strong evidence but the assertion is not even lightly supported with evidence. Also, one might ask what survey platform should be used to count horses if helicopters are too slow to do the job? A large body of highly qualified statisticians with impeccable reputations disagrees with Galea's (2023) assertion. Galea's claim is based on three things:

(i) Advice given by Buckland *et al.* (2001, p. 31) (which was repeated in notes for university students by Owusu 2019, which is the reference cited by Galea) that in order for mobile animals flushed by the survey platform to be recorded at their original location, the animals should travel at no more than half the speed of the survey platform.
(ii) The helicopter speed when counting horses is 93 km/h (Cairns 2022).
(iii) Reference to a web page about racehorses in North America to state that horses can run at 64 km/h.

During counting operations, the feral horses in Kosciuszko do not behave like trained thoroughbreds being ridden hard along the mown track at a racecourse, possibly urged along with whips and spurs. When flushed, the Kosciuszko horses do not move half as fast as the aircraft travelling at 93 kph, at least not for any appreciable distance. In practice it is almost always possible to record their original location, which is all that is required by the Distance Sampling method. Even wild horses which had recently had a bad experience of helicopters, responded to a counting helicopter by moving only 1 km on average from their starting point (Linklater and Cameron 2002).

In any case, it is likely that the theoretically based advice given by Buckland *et al.* (2001) is of a precautionary nature and can be carefully disregarded without biasing the population estimate. My former staff and I compared three counting methods to estimate kangaroo population size in five nature reserves, including Walked Line Transect Distance Sampling (WLTDS) from 220 km of transect. Other methods were Total Counts and Faecal Pellet Counting. For the WLTDS, the observers were required to travel at only 1 km/h. Flushed kangaroos usually travelled much faster than this, yet the population estimates by the three methods were the same. That is, the population estimates were neither significantly different statistically ($p< 0.05$), nor was there any consistency in which method produced the highest or lowest estimate in a reserve (Snape and Fletcher, unpublished data). So the advice about speed of the survey platform in relation to the speed of the animals is not an absolute requirement.

Horses in different countries are comparable to a reasonable extent so Galea's reference to a North American source for the maximum speed of a racehorse is perfectly legitimate but it is inconsistent with her complaint on p. 16 that the examples of high population growth of feral horse populations (Cairns 2019) were observed overseas, as if that somehow makes them less valid.

B1 Statistical modelling ... requires ... three time points – p. 12

The report (Galea 2023) says the 2019 count report '*applied complex statistical modelling techniques*' to data '*comprising only two points*'. Without a more specific reference, it is hard to be sure what this refers to. Most of the modelling in the report is based on the clusters, of which there were 458 in total. So I assume this comment refers to the bootstrap calculation of variance for the population growth rate between the 2014 and 2019 population estimates on p. 40. That is not statistical modelling in the normal sense that most readers would assume. And it is done to calculate the variance of the population growth rate, not the population growth rate itself.

It is perfectly legitimate, and commonplace, to calculate a population growth rate (PGR) between two counts. For example if the population of Melbourne was 2.5 million people in 2010 and 3 million a year later, we might say it had grown 20% that year. And use of the bootstrap method to calculate an estimate of variance around the PGR is also legitimate, and is far better than having no indication of variance, which is the alternative. If indeed this criticism refers to p. 40, it is a surprising one because the method is widely adopted by statisticians.

B2 Transformation of the data – p. 12

Galea (2023) states:

> *'When applying statistical modelling techniques there are various assumptions that the data need to meet in order to apply the techniques. The main one used is for the data to be what is called "normal", that is the raw data follows a normal distribution. When the data does not adhere to this then it is common practice to apply a transformation to the data depending on the shape of the original data. Cairns (2019) states that the* "estimates of cluster density and population density were slightly positively skewed, indicating that the data were not normally distributed*"'.*

Galea (2023) then goes on to speculate that Cairns (2019, 2022) may have applied a log transformation to the data in an incorrect way (by not accounting for zero values) and concludes:

> *'**CONCERN:** If log-transformations are being applied to the raw counts, then all 0 counts will need to be increased and could significantly overestimate the population. Appropriate transformations should be applied that take into consideration 0 counts'.*

This section displays limited experience working with ecological data. First, the report makes the mistake of thinking of data that are not normally distributed as a problem to be fixed, but this is the wrong outlook. Ecological data are almost never normally distributed and therefore require the use of appropriate statistical methods and distributions in their analysis.

Log transformations, or any other transformations, were not needed in this case (not that they are problematic anyway, if applied correctly). Cairns (2019, pp. 36-37) states "*Bootstrap coefficients of variation and confidence intervals were calculated for all estimates, with the bootstrap confidence intervals being given in preference to standard normal-theory confidence intervals (Tables 7 and 8). This approach is becoming more common and is recommended because it relaxes the constraint of assuming that data are normally distributed and [that] confidence intervals are therefore symmetrical (Crawley 2005). The confidence intervals for both the estimates of cluster density and population density were slightly positively skewed, indicating that the data were not normally distributed.*"

So, in short, no transformation was used.

Galea (2023) is incorrect that data need to be normally distributed. It depends what procedure is being used. The comment about log-transformation is based on a total lack of evidence and in spite of the description by the author which clarify that no transformations were involved (Cairns 2019, pp 36-37). This is most likely another example of a criticism written in too much haste but the result has the unfortunate appearance of an attempt to make a target where there is none, or in other words to 'create a straw man' for the purpose of having something easy to criticise.

B3 Use of covariates – p. 13

Galea (2023) says that the count reports (i.e. Cairns 2019, 2022) do not make clear which covariates were used in the detection function modelling and that the reports are 'confusing' in this regard. Galea (2023) concludes '*Determining what covariates were included* *and what impact they had on the accuracy of the models cannot be determined from the reports given the conflicting information provided and therefore the generalisability of the results across the entire four blocks should be interpreted with caution'.*

In fact, the opposite is true. Cairns (2019, 2022) is quite clear about what was done and why. It appears likely that Galea simply overlooked the important words on this topic in both reports, i.e. 'CDS' and 'MCDS', which stand for Conventional Distance Sampling and Multiple Covariate Distance Sampling. As the name suggests, covariates are used with MCDS. Covariates cannot be used with CDS.

Unfortunately, Galea's quotations from Cairns (2019, 2022) omitted the crucial information which answers her criticism. I provide the full text from Cairns (2019) in the two following paragraphs, with underlined text indicating the words that were quoted by Galea (2023) as evidence of *'confusion'*.

On page 19: '*DISTANCE 7.3 has three different analysis engines that can be used to model the detection function (Thomas et al. 2010). Two of these, the conventional distance sampling (CDS) analysis engine and the multiple-covariate distance sampling (MCDS) analysis engine were used here. In analysing survey results using the CDS analysis engine, there is no capacity to include any covariates other than the perpendicular distance of a cluster of horses from the transect centreline in the modelling process. .... If the MCDS analysis engine is used, additional covariates can be included in the analysis.*

On page 21: '*The MCDS analysis engine allows for the inclusion in the detection function model of covariates other than the perpendicular distance from the transect centreline (Thomas et al. 2010). ......* The covariates used in these analyses were related to individual detections of clusters of horses and were identified as observer, cloud cover score and habitat cover at point-of-detection. All these covariates were categorical. There were three observers (DS, MS and SS), three grades of cloud cover (1 = clear to light, 2 = medium, 3 = overcast to dull) and two categories of habitat cover at point-of-detection (1 = open, 2 = timbered), indicating that horses were either sighted in the open or in timbered habitat. The three covariates were included in the analysis either singly or in pairs.'

In regard to the quoted text Galea (2023) says:
'It is unclear throughout the report from Cairns (2019) as to what covariates were included and when. On page 19 it states that *"there is no capacity to include any covariates other than the perpendicular distance of a cluster of horses from the transect centreline in the modelling process"* yet on page 21 it states that *"The covariates used in these analyses were related to individual detections of clusters of horses and were identified as observer, cloud cover score and habitat cover at ....".*

The evidence shows that it was Galea, not Cairns, who was confused. And it was Galea who was responsible for creating confusion by making an incorrect criticism and misquoting (by omission).

The report (Galea 2023) goes on to apply the same level of scholarship to the corresponding sections of Cairns (2022).

Then there is a third mistake under this heading. Referring to a kangaroo counting report (Cairns 2016) Galea (2023) conflates the listing of putative co-variates for statistical evaluation (such as would be found in the Methods section of a scientific report), with the reporting of which of these putative co-variates had come through the statistical evaluation to prove worthy of retention in the model (typically found in the Results section). This is a surprising mistake for an experienced statistician to make. It is yet another instance in Galea (2023) where ordinary normal text by Cairns (2019, 2022) has been misunderstood or misquoted.

In summary, the criticism under this heading is wrong. The horse counting reports of Cairns (2019, 2022) are not confusing in regard to co-variates.

B 4 Assumptions – p. 14

The report quotes Cairns (2019) who, in full, stated:

'There were parts of each block that were not surveyed either because of the steepness of the terrain or because the land was under private ownership and were therefore not included in the calculations of population abundance (see Section 2.1). ……

Given along with the population abundances in Tables 9 and 10 are a second set of population densities. These are densities derived in relation to the total areas of the survey blocks. Implicit in their estimation are the assumptions that the horse population in a block would be aggregated in its distribution and that the density of horses in the very steep country within the survey blocks would be at trace levels; i.e. *near to zero. This assumption could be open to challenge, but could only be refuted with comparable survey results.'*

In the quoted paragraphs, Cairns is explaining his assumption that there were no horses in the areas he did not count. If that assumption is wrong, then the population size will have been UNDERestimated. Yet Galea comments that *'without comparable survey results there is no way of knowing if this assumption had a significant impact on the … population estimates'*. (Is this an 'own goal' by Galea?) Again, no practical alternative is suggested. To apply counting effort where horses are so sparse that they probably can't be detected from the helicopter, would not be reasonable.

B 5 Grouping of Zones for modelling – p. 14

Galea (2023) states that *'In both the [Cairns] 2019 and 2022 reports the populations across the blocks are merged with a global detection function model applied and a single estimate determined. However, it is clearly evident from the report that the blocks provide significant differences in the wild horse counts along with the sizes and expected detection being different.*
…..
CONCERN: *Independent modelling of the four blocks should be undertaken and no overall population estimate reported.'*

This too is wrong. Cairns (2022, p. 2) states *'The survey of the wild horse population in KNP was conducted in four survey blocks that were identified by NSW NPWS as being in areas known to support wild horses'*. A different level of precision was specified for each block as presented in Table 1 on p. 2 of Cairns (2022), ranging from 20% to 40%. (Precision refers to the uncertainty or variance of each estimate, or in other words, the relative length of the error bars shown with each estimate).

Because of the requirements for different levels of precision the blocks could not be combined into a single analysis, nor were they. Tables 4, 5, 6, and 7 (Cairns 2023) present results separately for each block.

B6 *No increase in the population over 2 surveys – p. 15*

The point is made by Cairns (2022) that the higher population estimate in 2022 is not significantly different statistically (at the conventional 0.05 level) from the count in 2020. It may seem paradoxical to those without ecological experience, but that does not mean the population is not increasing.

The reason is that because all estimates of population size have a confidence interval, estimates that are close enough in time can not differ significantly (Figure 1). An example may communicate this better, so imagine a population of animals that is increasing constantly. If the population size is estimated often enough, e.g. every day or every week or every month, consecutive estimates would not differ from each other by very much even if the confidence intervals around the estimates of population size were unrealistically small (say 5%). Yet we know for certain that this hypothetical population is increasing because we made it that way. The answer is not to spread out the time between counts so that the differences between consecutive counts will be significant. The opposite is true, the more often we count the population, the better we will know it.

The answer is that testing whether consecutive counts are significantly different is a poor way to determine whether a population is increasing, stable or decreasing. There are better ways described in many ecology textbooks (but I do not need to go into them here). Suffice it to say that the horse population is probably still increasing, but even if it were stable, Galea's assertion that this would mean culling should stop is illogical.

The need to reduce the horse population is determined by the impact of the number of horses and by whether the number is higher or lower than the statutory requirement to achieve a population of 3,000 horses. The requirement for culling is unaffected by whether the number is increasing, decreasing or stable. To explain in an easier way, using an example, Galeas' assertion is like claiming that no parachute is needed when jumping from aeroplanes flying level, only for aeroplanes that are climbing.

B7 *Implausible population estimates – p. 15*

The 2019 survey actually saw and counted 1,374 horses, from which the population was estimated to be 15,687, which is 11.6 times as many. Galea evidently thinks this ratio is too high. But it is often unavoidable in ecology, e.g. the leading edge application of thermal drone technology (Brack *et al.* 2023) which estimated the population of swamp deer in Sesc Pantanal Reserve, recorded 66 deer from which they estimated the population in the reserve was 1,856 (95% CI 951–3710), which is 28 times as many. Yet Galea (2023) reports the result with horses to be *'implausible'* without quoting any other evidence or any kind of analysis. This is not scientific.

It is common in human medical research for there to be an interaction with every human subject, something virtually unknown in ecology. A pasture ecologist, for example, cuts and dries vegetation samples from quadrats whose area might represents one millionth of the vegetation patch about which the ecologist will make descriptive statements and predictions. An ecologist studying insect species may never see 99.999% of the individuals of the population under study. Thus the design and practice of sampling is a core activity in ecology, but less central to medical research. Perhaps that is why a medical statistician made this mistake.

B8 *Width of the confidence intervals – p. 15*

The report points to the width of the 95% Confidence Interval for the 2022 survey i.e. 14,501 to 23,535 as a reason the population estimates from horse surveys are *'unreliable'*.

This result was better than the owners of the survey had intended. Their requirements for precision were stated in advance (Table 1 in Cairns 2022) and were bettered in all survey blocks (Table 6 in Cairns 2022).

How can it be that a level of precision that is better than what the experienced people responsible for the survey had paid for, is regarded by Galea as unacceptable?

Wildlife surveys only rarely provide tight precision and to improve precision is often impossible or at best inordinately expensive. Many component measurements are being combined, including the height of the aircraft, its position, the distance from the transect to the horses, and the distribution of horses between transects. In addition, the animals move in and out of view and during the course of the survey there may be births, deaths, immigration and emigration. These potential sources of errors magnify, resulting in wider confidence intervals than would be achievable in some other circumstances, such as lab-based research.

The reference given to support Galea's opinion that the confidence intervals are too wide (Bonham 1989) is actually a web page about the beauty of racehorses and does not contain any such statistical comment.

Please note that precision is expressed in the survey report (Cairns 2022) in terms of the Coefficient of Variation (CV%). This is the normal metric for the purpose and is widely used. Galea's use of the 95%CI as a percentage of the mean is one I have not seen before. It results in a higher percentage than the CV.

C1 Implausible population increases – p. 16

Among brumby activists there is a persistent idea that some of the rates of population increase observed in the wild are impossible. Doubt about the population growth leads to the thought that the population estimates must be wrong. Such observers sometimes point out that their well fed mares only produce a foal every second year. Even if the population comprised an equal number of male and female breeding adults, that would corresponds to 25% increase per year so real populations would increase at lower rates.

Early research on the Kosciuszko feral horses produced an estimate of 22% for the intrinsic rate of increase of horses (Dawson 2002, p. 70). This has been variously misquoted as a maximum population growth rate (PGR) for horses, but the two are different. (Confusion is not helped by a commonly used representation of the intrinsic rate of increase as r_{max} or r_m). A population grows at the intrinsic rate of increase for that species if it has a balanced sex and age distribution, and is not limited by food or other resources, or externally imposed mortality such as harvesting or predation (Caughley 1977). The details of this definition are important.

Most populations, most of the time, have PGR close to zero (Caughley 1977; Hone 1999, 2012; Sibly and Hone 2002), much lower than the intrinsic rate of increase for the species, but population growth rates higher than the intrinsic rate of increase are possible in unusual cases, for example a female biased population can grow faster than the intrinsic rate of increase and so can a population that is biased toward the maximum breeding age. Farmed animal populations normally exhibit both features. So PGRs greater than Dawson's (2002) estimate of 22% for horse r_m are possible and have been observed. Wild horse population growth rates up to 39% have been observed by researchers (Scorolli and Lopez-Cazorla 2010).

In addition, horses move into and out of the areas counted, for example between KNP and parks in Victoria. Changes in the estimated population size that are greater than 22% have been partly attributed to such movement (Cairns 2022). The extent and frequency of such movement has not been researched, so unfortunately there is no way to determine how much of the large increases in estimated population size are due to breeding, how much to immigration, and how much to counting error. In the absence of that research, Galea's (2023) criticism is unfounded.

C2 Movement of horses – p. 17

The report says that because horses are capable of crossing the line bounding the counted area, *'the possibility of double counting cannot be eliminated'*

This is illogical. Observers were not counting the same area more than once so there is no question of so called *'double counting'* due to local movement of horses. If animals move into the counted area before the day of counting they will be included. If they move out before the count, they will not be included. Animals whose home ranges straddle the boundary of the counted area contribute to the inherent variability between successive wildlife counts by being inside the counted area in some years and outside it in others.

The report also says about the Kosciuszko horses that *'a true count cannot be determined'*. Here Galea may have begun unwittingly to uncover an important concept. The important idea is to accept that estimating the size of a wild animal population is not analogous to normal counting (of human research subjects, for example, or farm animals in a yard) where the exact number of counted objects can be determined.

A wild population is more like a river flowing constantly. While the count is underway and after it has been completed, animals are dying, and being born, and moving into and out of the area, and these changes are unseen by the person doing the counting. The exact number of animals in the population at any instant is unknowable. Even the geographic extent of a population at a particular time is usually impossible to know exactly, except perhaps on an island or in artificial settings.

Ecologists almost never determine an exact number for wildlife population size but an exact number is not required. What is needed is an estimate of population size that is good enough for the purpose for which the count was done.

C3 Foals and joeys – p. 17

The report claims that the surveys provide no counts of foals. First, this is a gratuitous claim because separate counts of foals have never been shown to be necessary. There are numerous examples of vertebrate species being managed successfully without separate counts of juveniles. But in any case, foal counts are plainly evident in the reports, for example in Cairns (2022) they are mentioned in the Summary, the Methods on p. 16, in Table 7 that occupies all of p. 27, and their discussion occupies almost an entire page in the Results and Discussion (p. 26). The count of foals goes above and beyond what is required for the management of the Kosciusko horses and is so difficult to miss that it leaves the reader scratching their head how this claim came to be made.

C3 is yet another mistaken claim by Galea (2023).

CONCLUSION

Galea (2023), the report named in the title of this commentary, is not a reliable scientific document because many of its criticisms prove on closer inspection to be mistaken, based on a misreading of the reports being criticized, or based on a misunderstanding of ecological methods. In some cases, criticisms are repeated under a different title, creating a false impression of the number of problems found.

The criticisms of Helicopter Line Transect Distance Sampling to estimate the population abundance of feral horses in KNP are not supported by either evidence, such as references to scientific literature comparing superior alternative methods, or by published results of alternative counts in KNP using well understood methods of abundance estimation that are recognised in the scientific literature. No data are provided and there are very few references to the vast ecological literature on wildlife counting.

There is an established body of scientifically credible material available on the counting of the feral horse population in Kosciuszko National Park (Walter 2003; Walter and Hone 2003; Montague-Drake 2005; Laake et al 2008; Dawson 2009; Dawson and Hone 2012; Cairns 2019, 2020, 2022). Galea (2023) adds nothing either credible, or valuable, to this subject.

ACKNOWLEDGMENTS

Helpful comments were provided on a draft version of this document by Isobel Crawford and three other people.

REFERENCES

Buckland, S.T., Anderson, D.R., Burnham, K.P.., Laake, J.L., Borchers, D.L. and Thomas, L. (2001). Introduction to Distance Sampling: Estimating Abundance of Biological Populations. (Oxford University Press: London.)

Buckland, S.T., Anderson, D.R., Burnham, K.P.., Laake, J.L., Borchers, D.L. and Thomas, L. (2004). Advanced Distance Sampling: Estimating Abundance of Biological Populations. (Oxford University Press: London.)

Cairns, S. (2019) Feral Horses in the Australian Alps: the Analysis of Aerial Surveys Conducted in April-May, 2014 and April-May 2019. A report to the Australian Alps Liaison Committee. Available from https://theaustralianalps.files.wordpress.com

Cairns, S. (2020). The results of a survey of the wild horse populations in the Kosciuszko National Park, October-November. Available from https://theaustralianalps.files.wordpress.com

Cairns, S. (2022). A survey of the wild horse population in Kosciuszko National Park, November 2022. Report to the National Parks and Wildlife Service, NSW. Available from https://www.environment.nsw.gov.au/research-and-publications/publications-search/a-survey-of-the-wild-horse-population-in-kosciuszko-national-park

Cherubin, RC., Venn, SE., Driscoll, DA., Doherty, TS., and Ritchie, E G. 2019. Feral horse impacts on threatened plants and animals in sub-alpine and montane environments in Victoria, Australia. *Ecological Management and Restoration* **20**(1) 47–56. https://doi-org.ezproxy.canberra.edu.au/10.1111/emr.12352

Commonwealth of Australia (2019). Australia's Strategy for Nature 2019–2030. Commonwealth of Australia. 38pp. Available from www.australiasnaturehub.gov.au Accessed 8/6/23.

Dawson, MJ. (2005). The Population Ecology of Feral Horses in the Australian Alps: Management Summary. Australian Alps Liaison Committee Archive. Available at: The population ecology of feral horses in the Australian Alps management summary | Australian Alps National Parks (wpcomstaging.com) accessed 7/3/21.

Dawson, MJ. (2009). 2009 aerial survey of feral horses in the Australian Alps. A report to the Australian Alps Liaison Committee.

Dawson, M.J. and Hone, J. (2012). Demography and dynamics of three wild horse populations in the Australian Alps. Austral *Ecology* **37**(1), 1–13.

Driscoll DA., Worboys GL., Allan H., Banks SC., Beeton NJ., Cherubin RC., Doherty TS., Finlayson CM., Green K., Hartley R., Hope G., Johnson CN., Lintermans M., Mackey B., Paull DJ., Pittock J., Porfirio LL., Ritchie E.G.., Sato CF., Scheele BC., Slattery DA., Venn S., Watson D., Watson M., Williams RM. 2019. Overview: Impacts of feral horses in the Australian Alps and evidence-based solutions. *Ecological Management and Restoration* **20**(1) 63–72. https://doi-org.ezproxy.canberra.edu.au/10.1111/emr.12357

Dyring, J. 1990. The impact of feral horses (Equus caballus) on sub-alpine and montane environments in Australia. MSc thesis, University of Canberra, Australia.

Eldridge DJ., Travers SK., Val J., Zaja A., and Veblen KE. (2019). Horse Activity is Associated with Degraded Subalpine Grassland Structure and Reduced Habitat for a Threatened Rodent. *Rangeland Ecology & Management* **72**(3), 467-473.

Glass, R., Forsyth, DM., Coulson, G. and Festa-Bianchet, M. (2015) Precision, accuracy and bias of walked line-transect distance sampling to estimate eastern grey kangaroo population size. *Wildlife Research* **42**, 633–641.

Galea C., (2023). Independent biostatistical report on the Brumby population in the Kosciuszko National Park. Unpublished report (18pp) circulated on social media by brumby activists and provided to state and federal politicians responsible for decisions about management of feral horses. Available at https://meetourhorsemeat.com/wp-content/uploads/2023/05/Independent-Biostatistical-report-into-the-counting-of-wild-horses-Claire-Galea.pdf Accessed 8/6/23.

Hone, J. (1988). A test of the accuracy of line and strip transect estimators in aerial survey. *Australian Wildlife Research* **15**, 493–497.

Hone, J. (1999). On rate of increase (r) patterns of variation in Australian mammals and the implications for wildlife management. *Journal of Applied Ecology* **36**, 709-718.

Hounsome, TD., Young, RP., Davison, J., Yarnell, RW., Trewby, ID., Garnett, BT., Delahay, RJ., and Wilson, GJ. (2005). An evaluation of distance sampling to estimate badger (*Meles meles*) abundance. *Journal of Zoology* **266** (1) 81–87.

Laake, J., Dawson, MJ. and Hone, J. (2008). Visibility Bias in aerial survey: mark-recapture, line-transect or both? *Wildlife Research* **35**: 299–309.

Linklater W. and Cameron E. (2002). Escape behaviour of feral horses during a helicopter count. *Wildlife Research* **29**: 221–224.

Montague-Drake, R. (2005) Results of Aerial surveys to determine wild horse densities and abundance in northern and southern Kosciuszko National Park: a report by the Reserve Conservation Unit, Parks and Wildlife Division.

Schulz M., Schroder M. and Green K. (2019). The occurrence of the Broad-toothed Rat *Mastacomys fuscus* in relation to feral Horse impacts. *Ecological Management and Restoration* **20**(1) 31–36.

Scorolli, AL. and Lopez-Cazorla, AC. (2010). Demography of feral horses (*Equus caballus*): a long-term study in Tornquist Park, Argentina. *Wildlife Research* **37:** 207–214.

Sibly, RM. and Hone, J. (2002). Population growth rate and its determinants: an overview. *Philosophical Transactions of the Royal Society London B*. **357**,1153–1170.

Thomas, L., S.T. Buckland, E.A. Rexstad, J. L. Laake, S. Strindberg, S. L. Hedley, J. R.B. Bishop, T. A. Marques, and K. P.. Burnham. 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology* **47**: 5–14. DOI: 10.1111/j.1365-2664.2009.01737.x

Walter, M. (2002). 'The Population Ecology of Wild Horses in the Australian Alps.' PhD Thesis (University of Canberra: Canberra.)

Walter, MJ. (2003). The effect of fire on wild horses in the Australian Alps National Parks. A report prepared for the Australian Alps Liaison Committee, unpublished.

Walter, MJ. and Hone, J. (2003). A comparison of 3 aerial survey techniques to estimate wild horse abundance in the Australian alps. *Wildlife Society Bulletin* **31**, 1138-1149

APPENDIX: Horse population estimates in KNP and how calculated

Date	Area surveyed	Report	Size of area surveyed (sq km)	Horse Popn estimate	Lower 95% Conf Limit	Upper 95%Conf Limit	Horse Popn in KNP (calc from AANP survey as needed)	Upper 95%CI for KNP	Lower 95% CI for KNP	How KNP portion calculated
Mar-01	AANP	Walter and Hone 2003	2,789	5,200	1,979	8,421	3,000	1,858	1,858	Kosciuszko NP component estimated by Walter (2005)
Apr-03	AANP	Walter 2003	2,717	2,369	3,937	3,937	1,367	905	905	Same proportional adjustment as above (0.58)
Jun-05	KNP	Montague-Drake	1,052	1,357	759	1,955	1,357	598	598	No change necessary
Apr-09	AANP	Dawson 2009	2,860	7,679	CV 25.4%		4,684	2,332	2,332	Adjusted by the proportion of horse groups counted in Kosciuszko NP, given in Dawson (2009 Table 1), ie 0.61.
May-14	AANP	Cairns 2019	7,443	9,187	7,484	11,595	5,604	1,984	1,984	Adjusted by the proportion of horse groups counted in Kosciuszko NP, given in Cairns (2019 Table 3), ie 0.61 excluding the Bago Maragle block and half of the Byadbo-Victoria count.
May-19	AANP	Cairns 2019	7,443	25,318	CV 12.3%		19,242	4,581	4,581	Adjusted by the proportion of horse groups counted in Kosciuszko NP, given in Cairns (2019 Table 3), ie 0.76 excluding the Bago Maragle block and half of the Byadbo-Victoria count. Confidence Limits estimated from CV.
Oct-20	KNP	Cairns 2020	2,673	14,380	8,798	22,555	14,380	8,175	5,582	No change necessary
Oct-22	KNP	Cairns 2022	2,675	18,814	14,501	23,535	18,814	4,721	4,313	No change necessary

AANP = Australian Alps National Parks; KNP = Kosciuszko National Park

Original Reports of horse counts in KNP

Cairns, S. (2019). Feral Horses in the Australian Alps: the Analysis of Aerial Surveys Conducted in April-May, 2014 and April-May 2019. A report to the Australian Alps Liaison Committee. Available from https://theaustralianalps.files.wordpress.com

Cairns, S. (2020). The results of a survey of the wild horse populations in the Kosciuszko National Park, October-November. Available from https://theaustralianalps.files.wordpress.com

Cairns, S. (2022). A survey of the wild horse population in Kosciuszko National Park, November 2022. Report to the National Parks and Wildlife Service, NSW. Available from https://www.environment.nsw.gov.au/research-and-publications/publications-search/a-survey-of-the-wild-horse-population-in-kosciuszko-national-park

Dawson, M.J. (2009). 2009 aerial survey of feral horses in the Australian Alps. Report prepared for the Australian Alps Liaison Committee. Available from https://theaustralianalps.wordpress.com

Dawson, M.J. and Hone, J. (2012). Demography and dynamics of three wild horse populations in the Australian Alps. Austral Ecology 37(1), 1–13.

Dyring, J. 1990. The impact of feral horses (*Equus caballus*) on sub-alpine and montane environments in Australia. MSc thesis, University of Canberra, Australia.

Montague-Drake, R. (2005) Results of Aerial surveys to determine wild horse densities and abundance in northern and southern Kosciuszko National Park: a report by the reserve conservation unit, Parks and Wildlife Division.

Walter, M.J. (2003). The effect of fire on wild horses in the Australian Alps National Parks. A report prepared for the Australian Alps Liaison Committee. Available from https://theaustralianalps.files.wordpress.com.

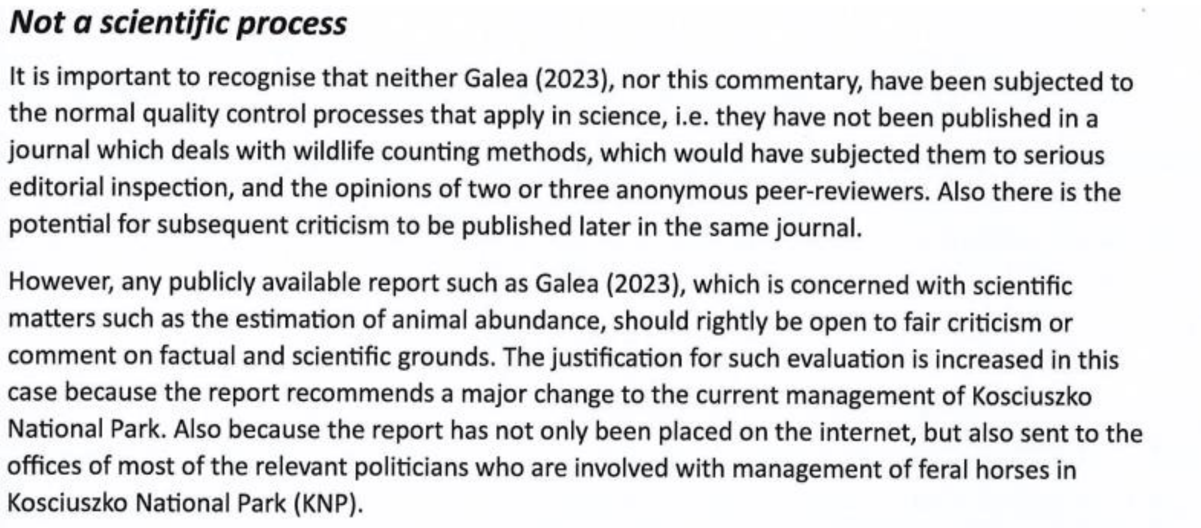
Walter, M. J. and Hone, J. (2003). A comparison of 3 aerial survey techniques to estimate wild horse abundance in the Australian alps. Wildlife Society Bulletin 31, 1138-1149

Claire Galea – responses to Submission 297 (Supplementary submission) Don Fletcher 4 October 2023.

Please note that I did not have access to the 2023 report at the time of writing the initial report therefore any comments pertaining to the 2023 Cairns report are invalid.

Below are "screen shots" from the document written by Dr Flecther and then in blue are my responses.

1.



Not a scientific process

It is important to recognise that neither Galea (2023), nor this commentary, have been subjected to the normal quality control processes that apply in science, i.e. they have not been published in a journal which deals with wildlife counting methods, which would have subjected them to serious editorial inspection, and the opinions of two or three anonymous peer-reviewers. Also there is the potential for subsequent criticism to be published later in the same journal.

However, any publicly available report such as Galea (2023), which is concerned with scientific matters such as the estimation of animal abundance, should rightly be open to fair criticism or comment on factual and scientific grounds. The justification for such evaluation is increased in this case because the report recommends a major change to the current management of Kosciuszko National Park. Also because the report has not only been placed on the internet, but also sent to the offices of most of the relevant politicians who are involved with management of feral horses in Kosciuszko National Park (KNP).

No, my report was not subject to academic peer-review as it was a summary of concerns surrounding secondary data collected and analysed from NPWS and Cairns, not myself, however this data collection and subsequent reports by Cairns were not subject to the peer review process as outlined by Submission 297. This was a concern I have raised in 2 parliamentary inquiries. Submission 297 states that "opinions of two or three anonymous peer-reviewers" would be needed – and again raises the questions why was this not undertaken for any of the Cairns' documents for either kangaroos or wild horses?

Below are three screen clippings:

1. The title page from the transcript of the NSW Parliamentary Inquiry into the Health and Wellbeing of Macropods
2. The quote from Dr Cairns stating that the documents had been peer-reviewed.
3. Concerns and criticism from the University of St Andrews who developed the software and reviewed the methodology Cairns uses for the wild horses.
    a. *Note – this was not an independent peer review according to the requirements of Submission 297, but rather a review by the company who had a conflict of interest as they owned and developed the software Cairns used.*

REPORT ON PROCEEDINGS BEFORE

PORTFOLIO COMMITTEE NO. 7 - PLANNING AND ENVIRONMENT

INQUIRY INTO HEALTH AND WELLBEING OF KANGAROOS AND OTHER MACROPODS IN NEW SOUTH WALES

UNCORRECTED

Virtual hearing via videoconference on Thursday 19 August 2021

The Committee met at 11:00.

PRESENT

The Hon. Cate Faehrmann (Chair)

The Hon. Catherine Cusack
The Hon. Ben Franklin
The Hon. Rose Jackson
The Hon. Shayne Mallard
The Hon. Mark Pearson (Deputy Chair)
The Hon. Penny Sharpe

The CHAIR: We will go to the Hon. Penny Sharpe now if you are ready to recommence. Sorry, you are mute.

The Hon. PENNY SHARPE: Sorry. I am happy for the Hon. Mark Pearson to jump in, if he wants to, at this point.

The CHAIR: Okay. We will go to the Hon. Mark Pearson. I also have a couple of questions as well.

The Hon. MARK PEARSON: Dr Cairns, just while you are here because I know you are leaving shortly, in your professional opinion, should the Department of Planning, Industry and Environment reports be subject to the rigour of the peer review process?

Dr CAIRNS: I would have thought not because they have been peer reviewed in the past. I did some work on horses for DPIE and those reports were peer reviewed. In fact, they were sent overseas to be peer reviewed and the methodologies are similar. They got—

The Hon. MARK PEARSON: But the methodologies are changing, and constantly changing.

Dr CAIRNS: The methodologies at the moment are similar. The method that I am involved in which is used to count kangaroos is the same one that is used to count horses in the alpine national parks.

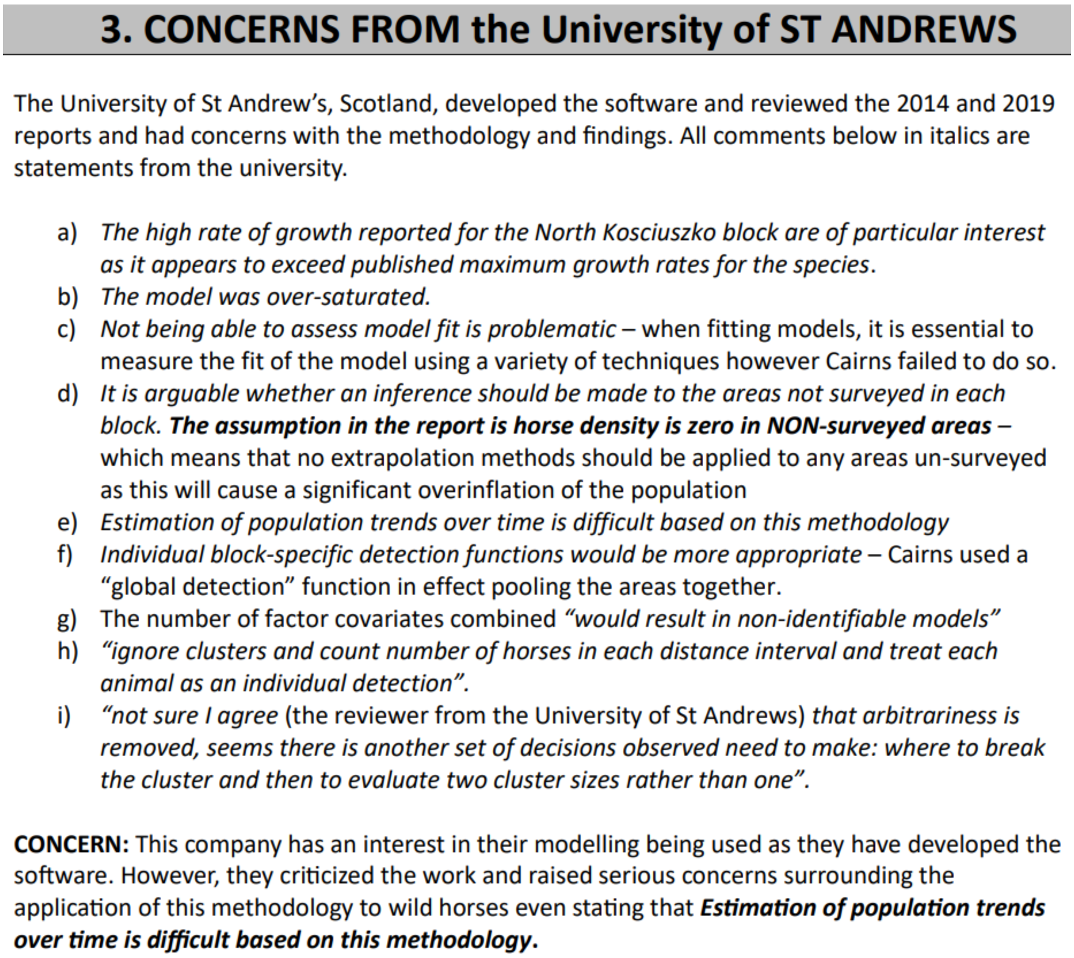
The Hon. MARK PEARSON: But there has been no internal or external audit of your work since the horses?

Dr CAIRNS: No. Sorry, yes. Two reports were peer reviewed in probably 2009 or 2010 and were found to have no problems. I presume that DPIE, or the Office of Heritage and Environment as it was then, felt that if the methods were not changing and they considered I was an honest trader in this area, they would not bother with peer reviews again.

The Hon. MARK PEARSON: But the methodology has changed, Dr Cairns.

Dr CAIRNS: Not from what I do for the tableland surveys.

The Hon. MARK PEARSON: But you just explained that sometimes there is a 100-metre view and then a 200-metre view et cetera, et cetera. So these are changes in the department.



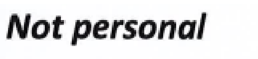
3. CONCERNS FROM the University of ST ANDREWS

The University of St Andrew's, Scotland, developed the software and reviewed the 2014 and 2019 reports and had concerns with the methodology and findings. All comments below in italics are statements from the university.

a) *The high rate of growth reported for the North Kosciuszko block are of particular interest as it appears to exceed published maximum growth rates for the species.*
b) *The model was over-saturated.*
c) *Not being able to assess model fit is problematic* – when fitting models, it is essential to measure the fit of the model using a variety of techniques however Cairns failed to do so.
d) *It is arguable whether an inference should be made to the areas not surveyed in each block.* ***The assumption in the report is horse density is zero in NON-surveyed areas*** – which means that no extrapolation methods should be applied to any areas un-surveyed as this will cause a significant overinflation of the population
e) *Estimation of population trends over time is difficult based on this methodology*
f) *Individual block-specific detection functions would be more appropriate* – Cairns used a "global detection" function in effect pooling the areas together.
g) The number of factor covariates combined *"would result in non-identifiable models"*
h) *"ignore clusters and count number of horses in each distance interval and treat each animal as an individual detection".*
i) *"not sure I agree* (the reviewer from the University of St Andrews) *that arbitrariness is removed, seems there is another set of decisions observed need to make: where to break the cluster and then to evaluate two cluster sizes rather than one".*

CONCERN: This company has an interest in their modelling being used as they have developed the software. However, they criticized the work and raised serious concerns surrounding the application of this methodology to wild horses even stating that ***Estimation of population trends over time is difficult based on this methodology.***

2.



Not personal

Submission 297 states that "nothing in the following comments about the content of the report should be read as personal criticism of Mrs Claire Galea".

Thank you for the opportunity to respond to this document. I found this document to contain a tone that was condescending especially when it was written by someone who was not a statistician. My concerns with the methodology of counting kangaroos and wild horses have always been with just that, the methodology, and not the person undertaking it. Comments made throughout this submission show a level of disrespect to not only myself but my profession as a statistician.

3.

Similarity of wildlife controversies

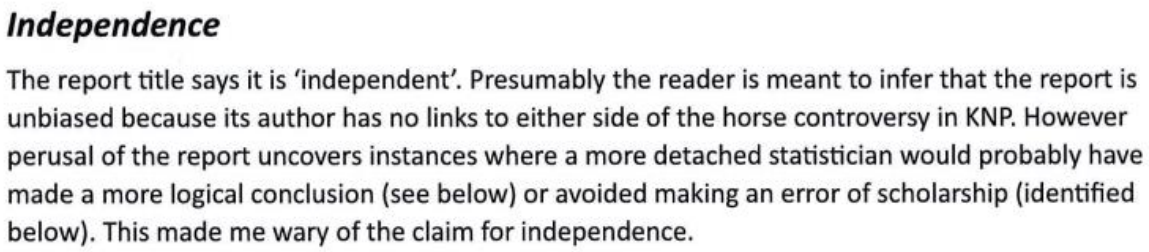
No comment.

4.

Importance of this response to Galea 2023

As this comment is a summary my response is addressed throughout the document below.

5.



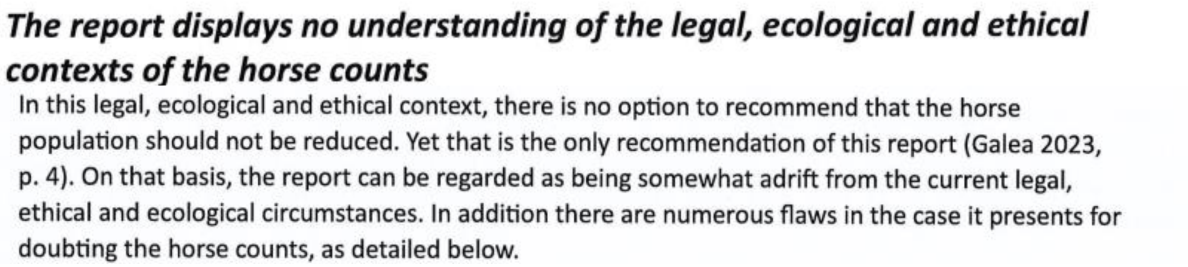
Independence

The report title says it is 'independent'. Presumably the reader is meant to infer that the report is unbiased because its author has no links to either side of the horse controversy in KNP. However perusal of the report uncovers instances where a more detached statistician would probably have made a more logical conclusion (see below) or avoided making an error of scholarship (identified below). This made me wary of the claim for independence.

I am an independent biostatistician, no one has paid me to do any of the kangaroo and wild horse reporting I have undertaken. At the time of writing the report in May 2023 I had no connections to any wild horse or kangaroo groups.

As Submission 297 draws attention to the report released in October 2023 by Cairns I have made comments within the document below pertaining to it.

6.

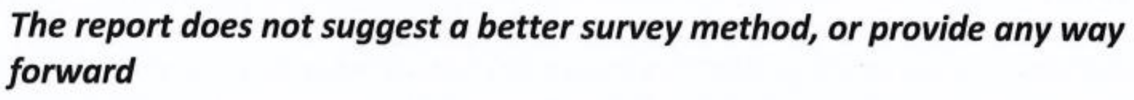


The report displays no understanding of the legal, ecological and ethical contexts of the horse counts

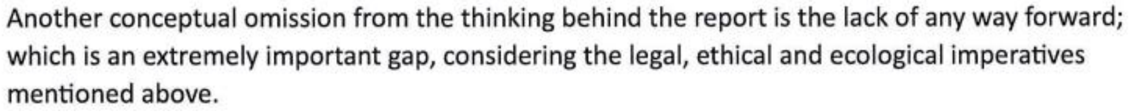
In this legal, ecological and ethical context, there is no option to recommend that the horse population should not be reduced. Yet that is the only recommendation of this report (Galea 2023, p. 4). On that basis, the report can be regarded as being somewhat adrift from the current legal, ethical and ecological circumstances. In addition there are numerous flaws in the case it presents for doubting the horse counts, as detailed below.

My report written in May 2023 was undertaken to draw similarities on the concerns with the methodology used to count kangaroos and wild horses. I am a biostatistician not an ecologist, lawyer or ethicist and therefore my statement calling for a moratorium on the killing of wild horses, which I also made for the kangaroos, was based on the statistical concerns surrounding the methodology used to estimate populations. Given the criticisms also provided by the University of St Andrews who state that "Estimation of population trends over time is difficult based on this methodology" we cannot be sure of how many wild horses are in the park and therefore until an independent recount can be undertaken an immediate moratorium should be put into place. There is currently no imagery evidence to support any of the population estimates for the wild horses in KNP and therefore validation of any of these values cannot be undertaken.

7.



The report does not suggest a better survey method, or provide any way forward



Another conceptual omission from the thinking behind the report is the lack of any way forward; which is an extremely important gap, considering the legal, ethical and ecological imperatives mentioned above.

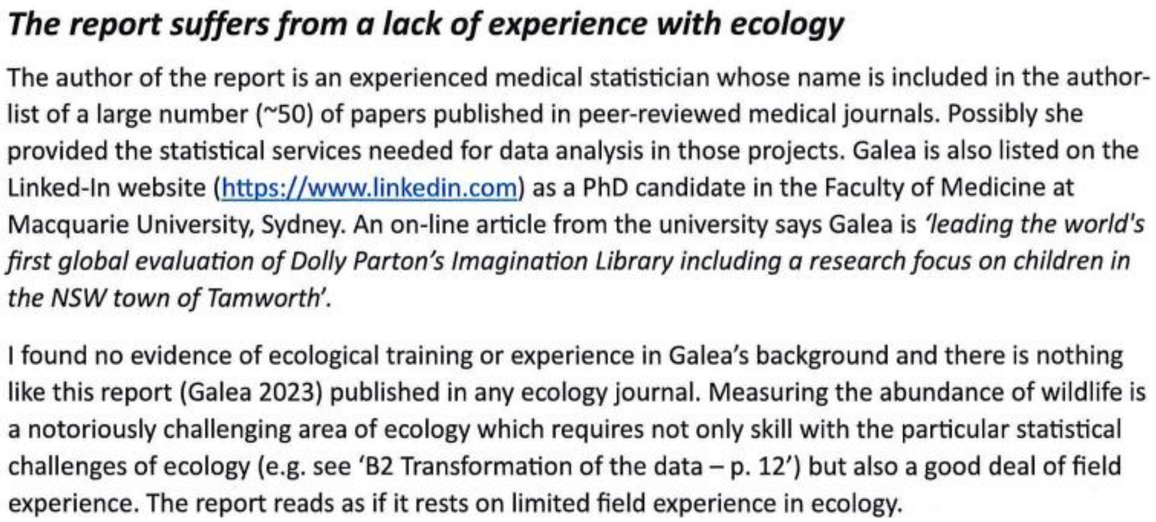
As previously stated, my report written in May 2023 was undertaken to draw similarities on the concerns with the methodology used to count kangaroos and wild horses not to project a way forward for a better survey method.

However, in the last few months I have been contacted by numerous groups, individuals and companies to be involved in undertaking an independent count of the wild horse population in KNP using state of the art imagery.

A company has been contracted to undertake this work completely independent of myself. I will not be doing any of the data collection, analysis or reporting and it will be peer-reviewed with all findings made available for publications.

Note: Documentation surrounding this count has been included in this email to be tabled titled "Summary of Independent Count January 2024"

8.



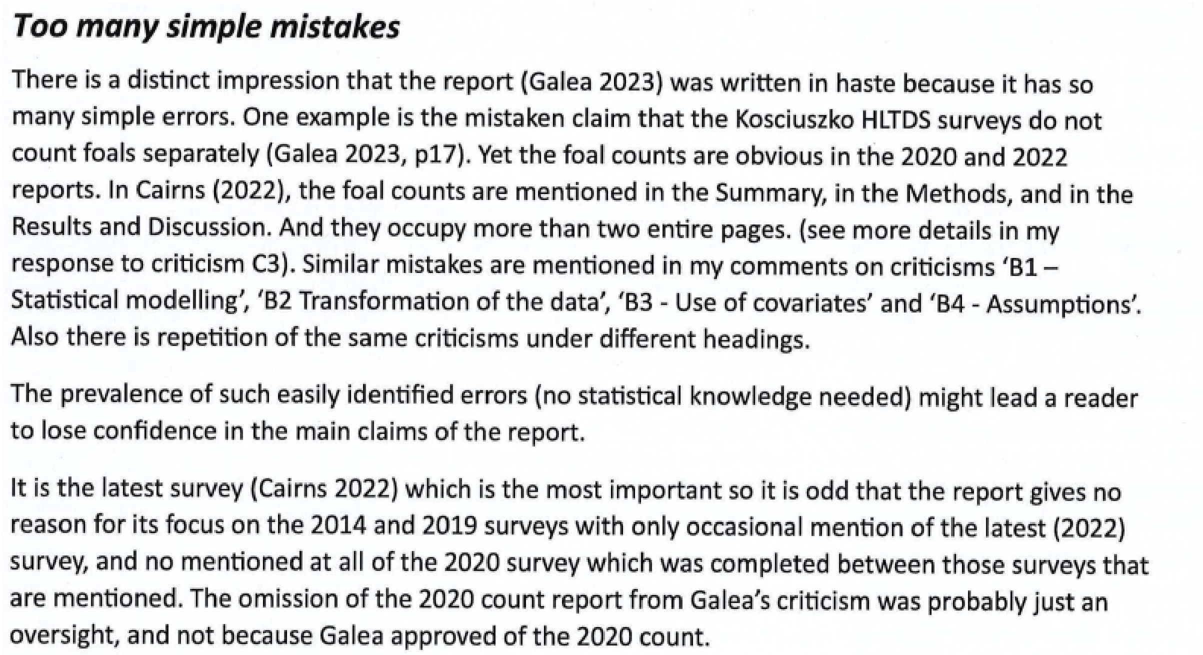
The report suffers from a lack of experience with ecology

The author of the report is an experienced medical statistician whose name is included in the author-list of a large number (~50) of papers published in peer-reviewed medical journals. Possibly she provided the statistical services needed for data analysis in those projects. Galea is also listed on the Linked-In website (https://www.linkedin.com) as a PhD candidate in the Faculty of Medicine at Macquarie University, Sydney. An on-line article from the university says Galea is *'leading the world's first global evaluation of Dolly Parton's Imagination Library including a research focus on children in the NSW town of Tamworth'*.

I found no evidence of ecological training or experience in Galea's background and there is nothing like this report (Galea 2023) published in any ecology journal. Measuring the abundance of wildlife is a notoriously challenging area of ecology which requires not only skill with the particular statistical challenges of ecology (e.g. see 'B2 Transformation of the data – p. 12') but also a good deal of field experience. The report reads as if it rests on limited field experience in ecology.

Over the last 25 years of analysing data and as a statistical consultant I have provided statistical advice, analyses and reported on data which has included a wide range from military to biological, educational and medical. I have not claimed to be an ecologist at any stage, but I specialised in population trends over time and have used this statistical modelling extensively. This modelling has been used for both human and animal studies in many peer-review publications and therefore my expertise in this space can cross both fields. I currently have an ecology paper under peer-review and have reviewed ecology papers from a statistical perspective for a world-renowned wildlife journal.

9.



Too many simple mistakes

There is a distinct impression that the report (Galea 2023) was written in haste because it has so many simple errors. One example is the mistaken claim that the Kosciuszko HLTDS surveys do not count foals separately (Galea 2023, p17). Yet the foal counts are obvious in the 2020 and 2022 reports. In Cairns (2022), the foal counts are mentioned in the Summary, in the Methods, and in the Results and Discussion. And they occupy more than two entire pages. (see more details in my response to criticism C3). Similar mistakes are mentioned in my comments on criticisms 'B1 – Statistical modelling', 'B2 Transformation of the data', 'B3 - Use of covariates' and 'B4 - Assumptions'. Also there is repetition of the same criticisms under different headings.

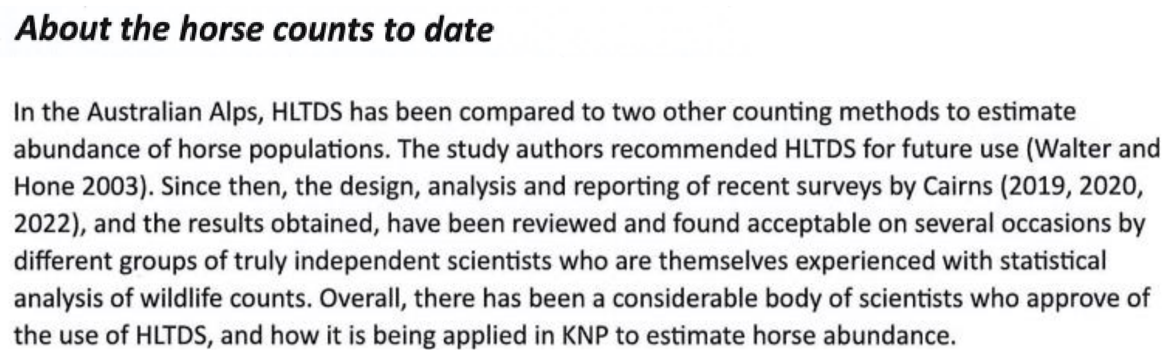
The prevalence of such easily identified errors (no statistical knowledge needed) might lead a reader to lose confidence in the main claims of the report.

It is the latest survey (Cairns 2022) which is the most important so it is odd that the report gives no reason for its focus on the 2014 and 2019 surveys with only occasional mention of the latest (2022) survey, and no mentioned at all of the 2020 survey which was completed between those surveys that are mentioned. The omission of the 2020 count report from Galea's criticism was probably just an oversight, and not because Galea approved of the 2020 count.

As previously stated, my report written in May 2023 was undertaken to draw similarities on the concerns with the methodology used to count kangaroos and wild horses. When Cairns was asked in 2021 in the NSW Parliamentary Inquiry into the Health and Wellbeing of Macropods if his work on counting kangaroos had been peer-reviewed he stated that his work on counting feral horses had been. This started a journey for me to eventually find the reports written by the University of St Andrews critiquing the 2014 and 2019 report and therefore I focussed my report on the counts done on these two occasions. As stated previously the 2014 and 2019 were not independently peer reviewed as per the requirements stated by Submission 297.

With respect to my comment regarding foals and joeys on page 17 – the 2014 and 2019 counts by Cairns do not count foals and therefore my statement is correct. I acknowledge that this was not made clear in my report.

10.



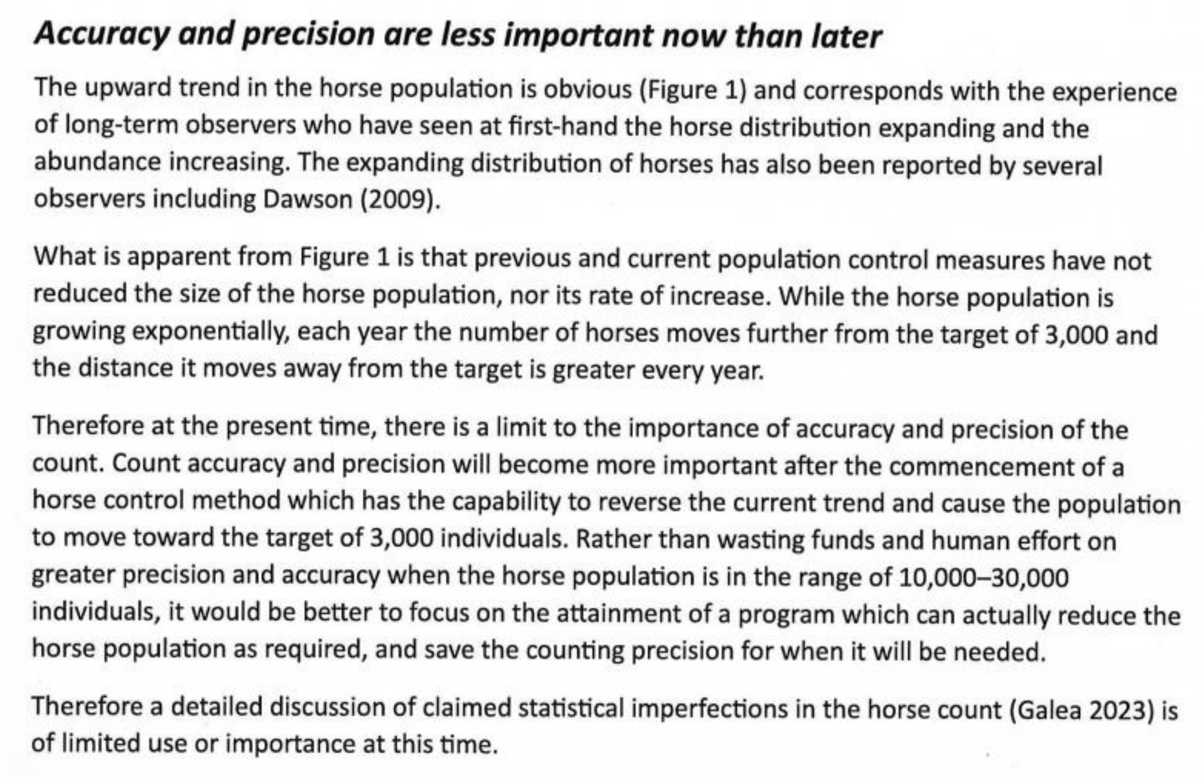
About the horse counts to date

In the Australian Alps, HLTDS has been compared to two other counting methods to estimate abundance of horse populations. The study authors recommended HLTDS for future use (Walter and Hone 2003). Since then, the design, analysis and reporting of recent surveys by Cairns (2019, 2020, 2022), and the results obtained, have been reviewed and found acceptable on several occasions by different groups of truly independent scientists who are themselves experienced with statistical analysis of wildlife counts. Overall, there has been a considerable body of scientists who approve of the use of HLTDS, and how it is being applied in KNP to estimate horse abundance.

Distance sampling has the ability to provide accurate population estimates when all modelling assumptions are met.

Submission 297 states that “a body of scientists”, the question remains are these academically independent and have they followed the protocol for the peer-review process according to Submission 297’s statement on page 1 section 1 of this document?

11.



Accuracy and precision are less important now than later

The upward trend in the horse population is obvious (Figure 1) and corresponds with the experience of long-term observers who have seen at first-hand the horse distribution expanding and the abundance increasing. The expanding distribution of horses has also been reported by several observers including Dawson (2009).

What is apparent from Figure 1 is that previous and current population control measures have not reduced the size of the horse population, nor its rate of increase. While the horse population is growing exponentially, each year the number of horses moves further from the target of 3,000 and the distance it moves away from the target is greater every year.

Therefore at the present time, there is a limit to the importance of accuracy and precision of the count. Count accuracy and precision will become more important after the commencement of a horse control method which has the capability to reverse the current trend and cause the population to move toward the target of 3,000 individuals. Rather than wasting funds and human effort on greater precision and accuracy when the horse population is in the range of 10,000–30,000 individuals, it would be better to focus on the attainment of a program which can actually reduce the horse population as required, and save the counting precision for when it will be needed.

Therefore a detailed discussion of claimed statistical imperfections in the horse count (Galea 2023) is of limited use or importance at this time.

As a biostatistician this section is very concerning. Accuracy and precision are always important especially when working with trends over time. There has been no validation of any of the population estimates from 2014 onwards using imagery or provision of the raw data (where each horse was seen using GPS location and time setting) therefore there can be no certainty on the actual number of wild horses in KNP until an independent, transparent count is undertaken with precision and accuracy.

Any contractual work for the NSW government that does not aim for precision and accuracy every time should be questioned.

I draw the reader to the attention of the annual head count data below conducted by NPWS. These figures were obtained from Equine Scientist Joanne Canning and through GIPA applications. With the estimated population in June 2023 to be 759 horses.

Year	**NPWS annual head count**
2014	1637
2015	Not done
2016	2199
2017	2144

2019	2791
2019	3110
2020	2468
2021	3699
	1649 trapped 890 slaughtered
June 2023	**759 remaining**

12.

COMMENT ON PARTICULAR CRITICISMS IN GALEA 2023 –

1 Cluster size – page 6

As stated earlier, my report focussed on the 2014 and 2019 as they had been critiqued by the University of St Andrews and the reflections of the University were very important in the kangaroo counting given the same methodology was used.

The populations trends which have been established on the wild horses in KNP using the methodology applied by Cairns started in 2014 and 2019 and have progressed from there. My concerns of the application of clusters and the size come from this original work and is supported by the University of St Andrews who stated that ”ignore clusters and count number of horses in each distance interval and treat each animal as an individual detection”.

Submission 297 has misunderstood the use of a global detection function which was applied in the 2014 and 2019 report. When this type of function is used it is necessary that all the individual components which contribute to the global detection also meet the minimum requirement for sample size i.e. a minimum of 60-80 clusters of wild horses. This was not the case in either 2014 and 2019 data. Therefore, referring to the overall total number of clusters does not meet the statistical assumptions and should not be undertaken. My concerns of the application of the global detection is supported by the University of St Andrews who stated that “Individual block-specific detection functions would be more appropriate”. I addressed this concern in the inquiry on the 18th of December.

With respect to not providing a way forward please see my response to 4 above.

Error noted on page 7 of my report.

As Submission has drawn attention to the Cairns document of 2023 it is noted that when counting clusters Cairns makes no distinction between open and treed terrain. These should have been counted separately and had a different probability detection and therefore more areas may not have met the minimum size requirement for the modelling. Cairns states that 86% of cluster sightings were in open terrain in the Northern Block hence 38 clusters (rounded up) were in treed terrain – again less than 60.

13.

Criticism of kangaroo counting exemplifies ecological naivety

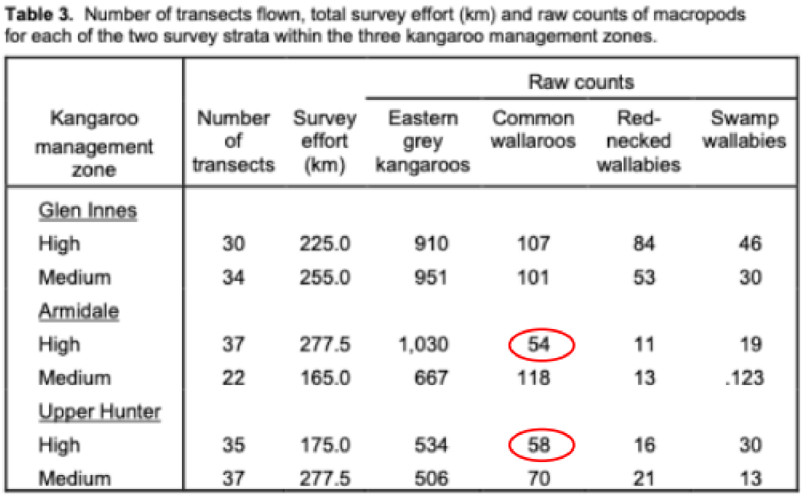


Table 3. Number of transects flown, total survey effort (km) and raw counts of macropods for each of the two survey strata within the three kangaroo management zones.

Kangaroo management zone	Number of transects	Survey effort (km)	Raw counts			
			Eastern grey kangaroos	Common wallaroos	Red-necked wallabies	Swamp wallabies
Glen Innes						
High	30	225.0	910	107	84	46
Medium	34	255.0	951	101	53	30
Armidale						
High	37	277.5	1,030	54	11	19
Medium	22	165.0	667	118	13	.123
Upper Hunter						
High	35	175.0	534	58	16	30
Medium	37	277.5	506	70	21	13

The table being referred to by Submission 297 is above (taken from page 6 of my original report May 2023). My discussion in my report was focussed on the Common Wallaroos not the Eastern grey kangaroos, so these criticisms are misguided. However, I will respond outlining how there were an insufficient number of wallaroos seen. Cairns cites Buckland et al. 2011 and states that "the recommended number of observations, of clusters of horses in this instance, should be 60-80 for reliable modelling of the detection function". From the above table it can clearly be seen that only 54 individual common wallaroos (not clusters) were seen in the Armidale High zone and again only 58 individual animals in the Upper Hunter High Zone. These values were then used to get an overall population estimate of nearly 300,000 common wallaroos in the Northern Tablelands.

Therefore, the comment from Submission 297 "it is absurd to complain that the uncommon species did not record 60 clusters" is related to Eastern Grey kangaroos which was not what my report was on. Therefore, what Submission 297 has stated is confusing as it is clear that 54 individual wallaroos cannot equate to 60 clusters nor can 58 individual wallaroos therefore an insufficient number of animals were seen to undertake any reliable modelling based on Cairns' statement.

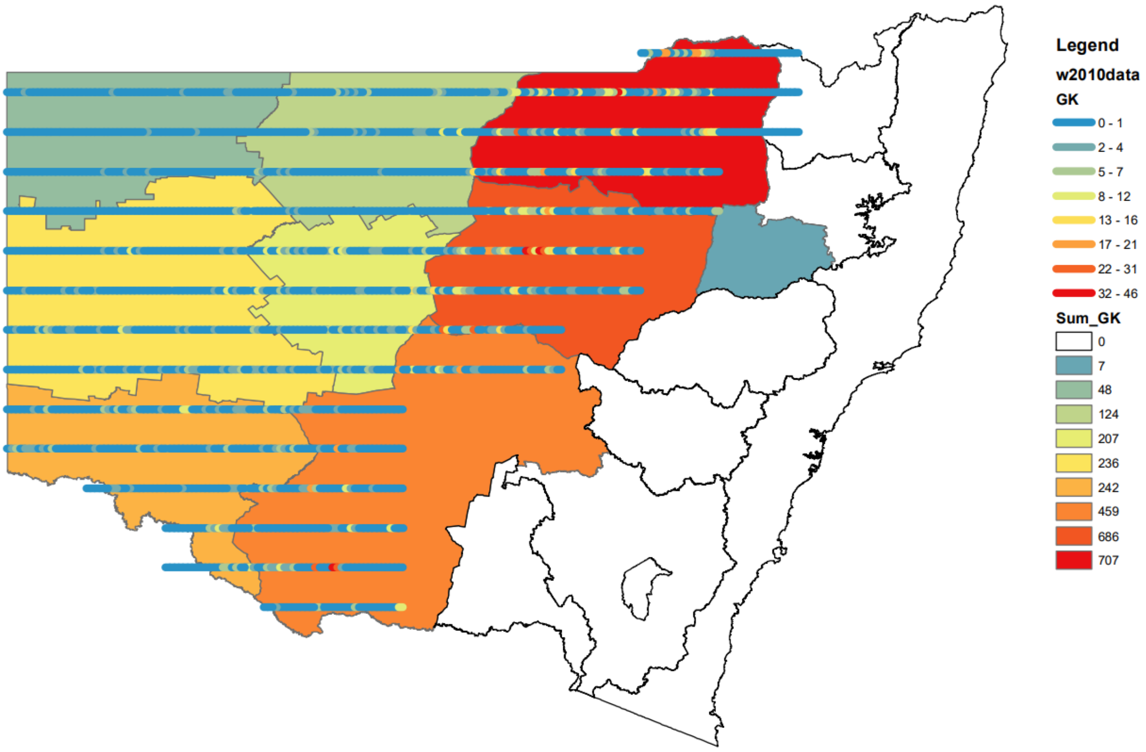
14.

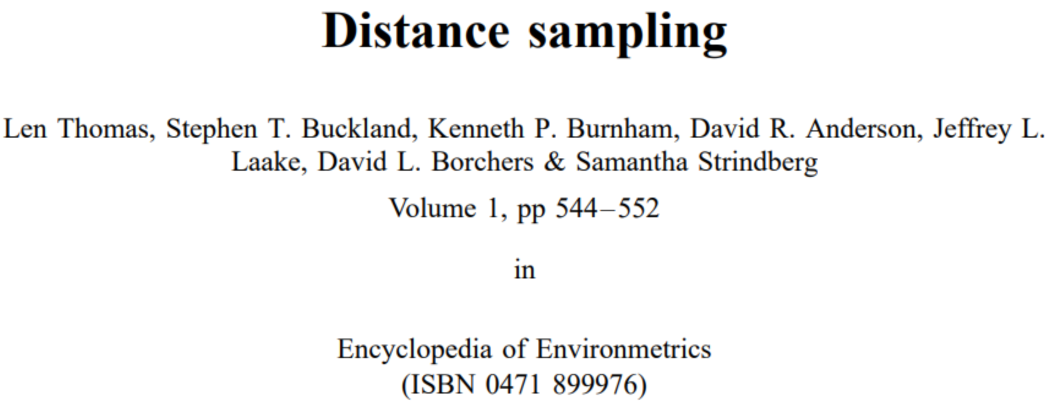
Misunderstood meaning of 'cluster'

Submission 297 raised concern around the statement that "a cluster is considered to be more than one animal' and refers to Buckland et al. (2001). If we refer to the paper titled Distance Sampling (screen shot below) which also includes Buckland ST as an author at no point does any analysis take into consideration if only one animal is seen, it is assumed that if one animal is seen then a cluster is seen and an estimated cluster size is applied – this would produce an overestimation in the population estimate. Hence a cluster, although stated as 1 or more is ONLY ever analysed as "MORE THAN 1" thus my statement.

My concern around this especially with kangaroos was evident when raw data was obtained demonstrating the counts from the observers (please see the map below), there were many instances where 0-1 single kangaroos were seen and modelled as a cluster. This would grossly overestimate the population estimates.

Submission 297 states that "if Galea's definition of clusters were adopted, all the animals seen as singletons would be omitted from the count". From the evidence below it can be seen that this is not the case as they indeed are counted.



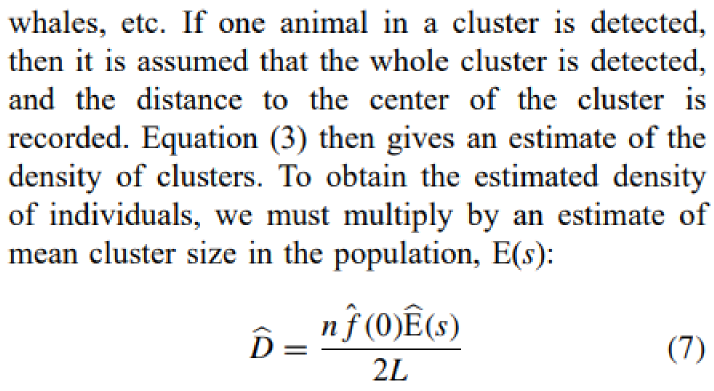

Distance sampling

Len Thomas, Stephen T. Buckland, Kenneth P. Burnham, David R. Anderson, Jeffrey L. Laake, David L. Borchers & Samantha Strindberg

Volume 1, pp 544–552

in

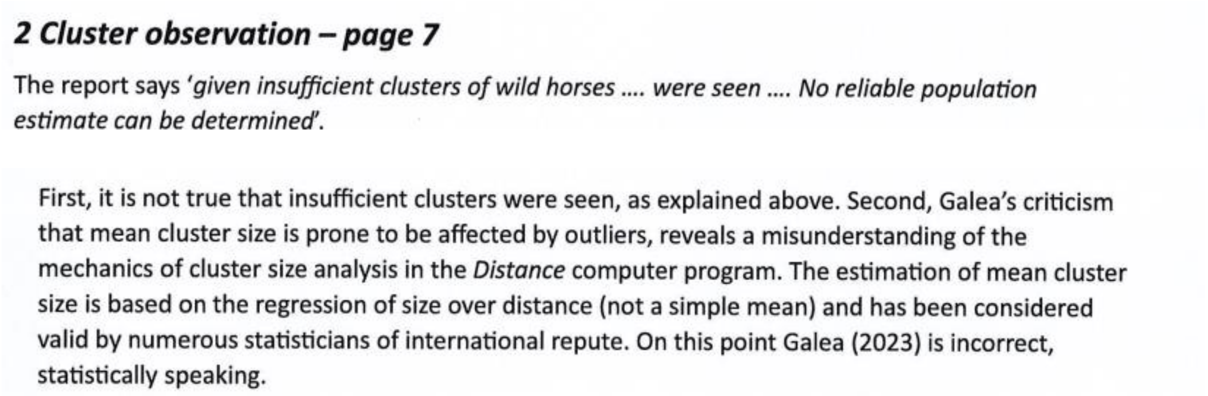
Encyclopedia of Environmetrics
(ISBN 0471 899976)



whales, etc. If one animal in a cluster is detected, then it is assumed that the whole cluster is detected, and the distance to the center of the cluster is recorded. Equation (3) then gives an estimate of the density of clusters. To obtain the estimated density of individuals, we must multiply by an estimate of mean cluster size in the population, E(*s*):

$$\widehat{D} = \frac{n\hat{f}(0)\widehat{\mathrm{E}}(s)}{2L} \qquad (7)$$

15.



2 Cluster observation – page 7

The report says '*given insufficient clusters of wild horses …. were seen …. No reliable population estimate can be determined*'.

First, it is not true that insufficient clusters were seen, as explained above. Second, Galea's criticism that mean cluster size is prone to be affected by outliers, reveals a misunderstanding of the mechanics of cluster size analysis in the *Distance* computer program. The estimation of mean cluster size is based on the regression of size over distance (not a simple mean) and has been considered valid by numerous statisticians of international repute. On this point Galea (2023) is incorrect, statistically speaking.

With respect to cluster size please see my response to numbers 10 and 11 above where I have outlined the serious concerns around using clusters. Please also note the previous comment from the University of Andrews recommending not to use clusters but rather individual horses.

The statement by Submission 297 around the mean cluster size raises the concern which is heavily support in literature that using the mean is affected by outliers. Please see the peer-reviewed manuscript by Nowak-Brsesinska and Gaibei (2022) outlining how outliers, large numbers, influence clustering.

With respect to the method of regression of size over distance this is heavily influenced by the probability of detection which ranged from 0.23 to 0.53 over 2014 and 2019 with no consistency.

16.

4 Bias sample location – page 11

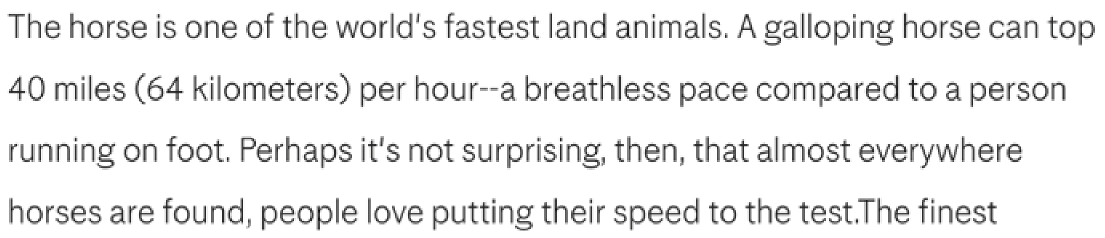
With respect to my reference to the size and survey effort comes under the heading – *Grouping of the zones together* - refers to the fact that when using a global detection function each zone should meet the criteria for the model assumptions which was not the case, seeing a minimum of 60-80 clusters. Please note the earlier statement from the University of St Andrews “Individual block-specific detection functions would be more appropriate”. Survey size and effort can differ within an overall block survey approach however each must meet the criteria set by the assumptions of the model. If they don’t then independent analysis should be undertaken, which is in line with the Universities recommendations.

17.

5 Use of line transects with respect to speed of wild horses – p. 12

In section E of my report, I refer to the statement made by Owusu “that the use of line transects is not appropriate if the object is moving at half the speed of the observer or faster”. This is a peer-reviewed reference and the statement I used for the speed of a running horse did not refer to a racehorse but rather to a galloping horse (screen clipped below). Therefore, both these statements are valid with respect to moving horses.

Also, one of the key assumptions behind distance sampling is that: “objects are detected at their initial location”, that is that they don’t move. Therefore, this choice of modelling violates that assumption.



The horse is one of the world's fastest land animals. A galloping horse can top 40 miles (64 kilometers) per hour--a breathless pace compared to a person running on foot. Perhaps it's not surprising, then, that almost everywhere horses are found, people love putting their speed to the test.The finest

https://www.amnh.org/exhibitions/horse/how-we-shaped-horses-how-horses-shaped-us/sport/bred-for-speed

Please also note the reference here is unpublished therefore not subject to peer-review.

18.

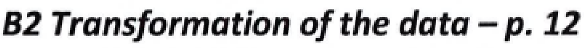
B1 Statistical modelling … requires … three time points – p. 12

Submission 297 states that it is perfectly legitimate to estimate a population growth rate using two specific time points on a basic calculation. However, my argument was that it was not appropriate when applying statistical modelling techniques. Which leads to another concern, which is the lack of model fit provided by Cairns. This was also noted by the University of St Andrews who stated that “Not being able to assess model fit is problematic” without this information we have no way of knowing how well the data fit the model and if any time trends established from the model were reliable.

Submission 297 refers again to the overall total number of clusters which the University of St Andrews have recommended against using and that each block should be analysed and reported separately. Along with this Submission 297 also refers to the bootstrapping which I made no comment about and have myself used on regular occasions for time trend data when the data have met the assumptions of the model.

However, of major concern, which is not noted in Submission 297, Cairns actually only used ONE time point by combing the values from 2014 and 2019 together – ultimately leading to a mean value at approximately the year of 2016.

19.



B2 Transformation of the data – p. 12

Data that is not normally distributed is perfectly "normal" for want of a better word. Normal data is a luxury in statistics and my report did not criticise this, therefore I think there has been a misunderstanding within this section.

Cairns does not provide the evidence of how positively skewed the wild horse data was and therefore it is not possible to determine if the use of the choice of method was appropriate. However, with respect to the kangaroo data it is made very clear that the transformation was used and therefore my concern remains.

20.

B3 Use of covariates – p. 13

I note that the criticism in this section and the tone used is unprofessional. Submission 297 should refer to the kangaroo reports as well as the wild horse reports for 2014 and 2019 and the University of St Andrews criticisms when referring to the confusion surrounding the use of covariates.

For the wild horses' different covariates were used for each block, but then for the overall global detection function only one used. So which covariate contributed to which population estimate within each block and by how much cannot be determined (note page 24 of Cairns 2014 and 2019 states "goodness of fit could not be considered in relation to models produced using the MCDS analysis engine because of a lack of degrees of freedom". Without the model output and fit it is not possible to determine the covariates and their contribution.

With respect to the kangaroo report of 2016 as referred to by Submission 297, there was no consistency in the modelling when each of the zones were grouped together to produce an overall population estimate. One zone used one covariate, another didn't use a covariate but as for the overall model, without the output and model fit it is not possible to determine the covariates and their contribution.

21.

B 4 Assumptions – p. 14

One of the fundamental assumptions on this modelling is that, as stated by the University of St Andrews "It is arguable whether an inference should be made to the areas not surveyed in each block. **The assumption in the report is horse density is zero in NON-surveyed areas**". The report by Cairns also opens the concern of steep non-surveyed areas to be open to challenge therefore, given this along with the statement from the University I am uncertain how this concern is not warranted. There should be no inference made on un-surveyed areas as the assumption is 0 population as stated by the developers of the software.

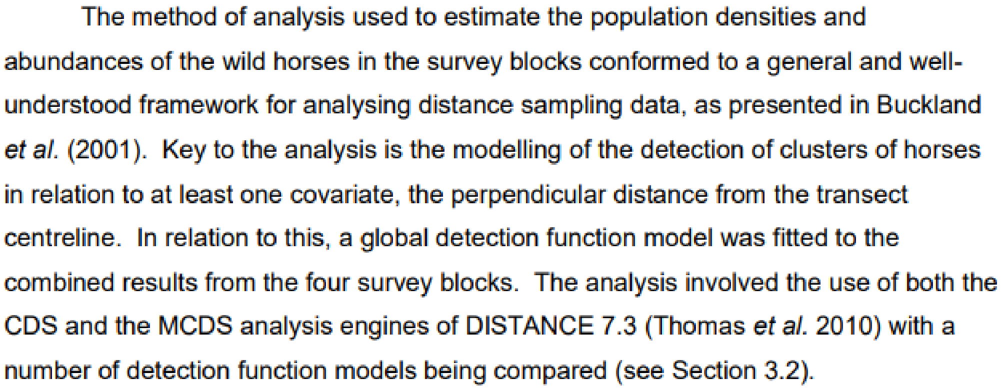
However, the argument remains that without any imagery and provision of the raw data nothing can be verified.

22.

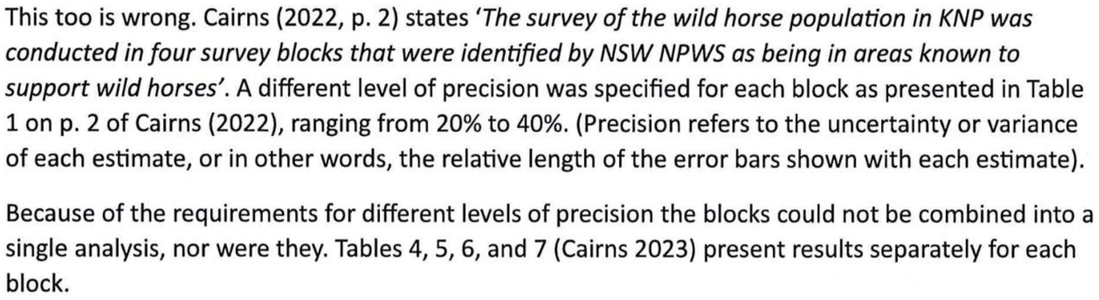
B 5 Grouping of Zones for modelling – p. 14

As previously stated, the University of St Andrews stated that "Individual block-specific detection functions would be more appropriate". Please see my responses to 9, 13 and 17.

The statement from Submission 297 that Cairns 2022 didn't use a global detection function is incorrect. Please see the screen clip below which was taken from page 17 of Cairns 2022 and note that a global detection function was also used in 2020 and 2023.



The method of analysis used to estimate the population densities and abundances of the wild horses in the survey blocks conformed to a general and well-understood framework for analysing distance sampling data, as presented in Buckland *et al.* (2001). Key to the analysis is the modelling of the detection of clusters of horses in relation to at least one covariate, the perpendicular distance from the transect centreline. In relation to this, a global detection function model was fitted to the combined results from the four survey blocks. The analysis involved the use of both the CDS and the MCDS analysis engines of DISTANCE 7.3 (Thomas *et al.* 2010) with a number of detection function models being compared (see Section 3.2).



This too is wrong. Cairns (2022, p. 2) states '*The survey of the wild horse population in KNP was conducted in four survey blocks that were identified by NSW NPWS as being in areas known to support wild horses*'. A different level of precision was specified for each block as presented in Table 1 on p. 2 of Cairns (2022), ranging from 20% to 40%. (Precision refers to the uncertainty or variance of each estimate, or in other words, the relative length of the error bars shown with each estimate).

Because of the requirements for different levels of precision the blocks could not be combined into a single analysis, nor were they. Tables 4, 5, 6, and 7 (Cairns 2023) present results separately for each block.

Note: With respect to the statement referring to Cairns 2023, as my report was published in May I did not have access to this information and therefore cannot respond to that comment.

23.

B6 No increase in the population over 2 surveys – p. 15

Submission 297 makes the statement that “The reason is that because all estimates of population size have a confidence interval, estimates that are close enough in time cannot differ significantly’. As the same methodology is used for counting wild horses and kangaroos how then does the population of grey kangaroos increase by 76.4% in one year or 154.2% in one year? These are just two examples where this has happened - is one year considered close enough in time for population increases for Submission 297? However, if we then look at the increase in the population of wallaroos which was surveyed with a 3 year gap this increased by 269%? (Please see screen shots below).

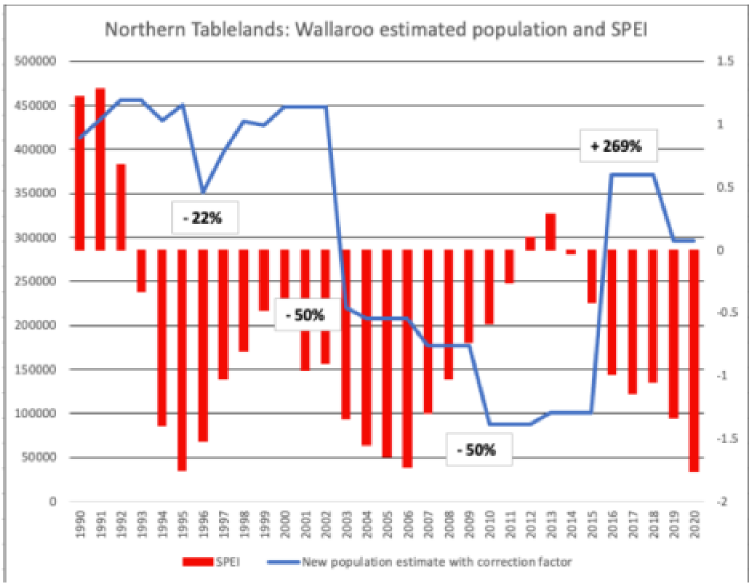
Submission 297 also states “testing whether consecutive counts are significantly different is a poor way to determine whether a population is increasing, stable or decreasing”. However, Cairns does this statistical comparison not me, so Submission 297 has criticised Cairns in this statement.

Table 27 **Grey kangaroo temporal variation – Narrabri**

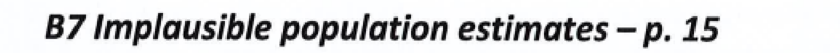
Average density (kangaroos/km²)		14.58	Standard deviation		7.13
Area in km²		65,755			
Year	**Population**	**Density**	**% Change**	**Quota**	**% Population**
1990	1,005,900	15.3	24.5	142,700	17.7
1991	987,000	15.0	-1.9	186,000	18.5
1992	898,500	13.7	-9.0	187,500	19.0
1993	1,964,801	29.9	118.7	188,685	21.0
1994	1,168,552	17.8	-40.5	412,608	21.0
1995	835,633	12.7	-28.5	103,530	8.9
1996	369,992	5.6	-55.7	61,964	7.4
1997	671,027	10.2	81.4	34,931	9.4
1998	1,214,523	18.5	81.0	63,543	9.5
1999	867,516	13.2	-28.6	175,310	14.4
2000	1,491,090	22.7	71.9	119,500	13.8
2001	1,523,954	23.2	2.2	182,500	12.2
2002	1,927,959	29.3	26.5	191,200	12.5
2003	874,080	13.3	-54.7	247,300	12.8
2004	367,179	5.6	-58.0	131,112	15.0
2005	399,672	6.1	8.8	55,077	15.0
2006	398,589	6.1	-0.3	59,853	15.0
2007	697,531	10.6	75.0	59,788	15.0
2008	513,617	7.8	-26.4	104,630	15.0
2009	447,330	6.8	-12.9	77,043	15.0
2010	752,771	11.4	68.3	67,002	15.0
2011	1,229,345	18.7	63.3	112,851	15.0
2012	1,246,675	19.0	1.4	184,304	15.0
2013	1,874,886	28.5	50.4	186,514	15.0
2014	1,321,410	20.1	–29.5	280,485	15.0
2015	813,425	12.4	-38.4	197,692	15.0
2016	1,434,755	21.8	76.4	121,072	15.0

Table 31 Grey kangaroo temporal variation – Griffith

Average density (kangaroos/km^2)		10.07	Standard deviation		3.68
Area in km^2		98,171			
Year	Population	Density	% Change	Quota	% Population
1990	566,000	5.8	41.1	62,800	15.7
1991	704,600	7.2	24.5	105,100	18.6
1992	669,100	6.8	-5.0	190,200	27.0
1993	1,395,898	14.2	108.6	160,584	24.0
1994	1,105,248	11.3	-20.8	108,744	7.8
1995	1,093,657	11.1	-1.0	128,016	11.6
1996	1,288,316	13.1	17.8	113,564	10.4
1997	1,121,800	11.4	-12.9	128,047	9.9
1998	1,349,050	13.7	20.3	74,650	6.7
1999	1,244,734	12.7	-7.7	154,860	11.5
2000	1,157,073	11.8	-7.0	131,050	10.5
2001	1,022,526	10.4	-11.6	147,600	12.8
2002	1,437,265	14.6	40.6	140,100	13.7
2003	874,589	8.9	-39.1	194,450	13.5
2004	722,872	7.4	-17.3	131,188	15.0
2005	701,493	7.1	-3.0	108,431	15.0
2006	677,124	6.9	-3.5	105,224	15.0
2007	571,999	5.8	-15.5	101,569	15.0
2008	638,262	6.5	11.6	85,800	15.0
2009	321,138	3.3	-49.7	95,739	15.0
2010	562,931	5.7	75.3	16,673	5.2
2011	541,306	5.5	-3.8	50,019	8.9
2012	700,388	7.1	29.4	78,012	14.4
2013	1,780,269	18.3	154.2	105,058	15.0



24.



B7 Implausible population estimates – p. 15

Please see my response number 18 where I have outlined in more detail what is below.

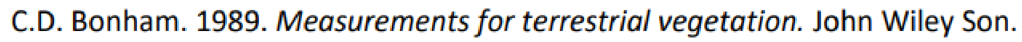
One of the fundamental assumptions on this modelling is that, as stated by the University of St Andrews "It is arguable whether an inference should be made to the areas not surveyed in each block. ***The assumption in the report is horse density is zero in NON-surveyed areas****".*

Please see the attached document from Equine Scientist Joanne Canning outlining the implausible population estimates.

25.

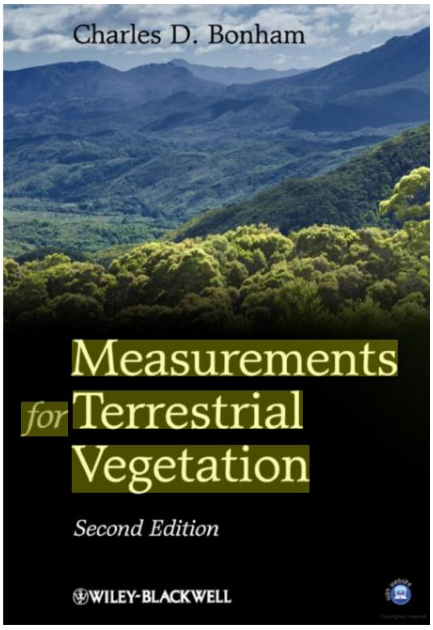
B8 Width of the confidence intervals – p. 15

Submission 297 has stated that the reference supplied in this section is a website that is incorrect. The reference states Bonham (1989) as can be seen on page 18.



C.D. Bonham. 1989. *Measurements for terrestrial vegetation.* John Wiley Son.

Below is a screen clip of the cover of the book.



Charles D. Bonham

Measurements *for* Terrestrial Vegetation

Second Edition

WILEY-BLACKWELL

26.

B7 Implausible population estimates – p. 15

The University of St Andrews stated that "the high rate of growth reported for the North Kosciuszko block are of particular interest as it appears to exceed published maximum growth rates for the species". Please see the report from the University of St Andrews for scientific evidence to support the implausible population rate of growth.

Also, along with my previous response to number 20, the implausible population growths far exceed the biological growth rate.

Please see the attached document from Equine Scientist Joanne Canning outlining the implausible population increases.

Submission 297 states that "there is no way to determine how much of the large increases in estimated population size are due to breeding, how much to immigration, and how much to counting error. In the absence of that research Galea's criticism is unfounded". This statement itself validates the criticism, Cairns does not have the evidence to support any of the population estimate claims – no imagery in any form. This evidence is vital!

27.

C2 Movement of horses – p. 17

The comments made by Submission 297 cannot be verified as there is no imagery to support any of the population estimates obtained. Double counting can occur during one session of an observer counting a cluster of horses, it does not have to be on two sperate occasions. But without images there is no evidence to demonstrate that double counting did not take place.

With respect to the important concept that a true count cannot be obtained – the same argument as in the previous paragraph stands. Without imagery covering the whole survey area we are left with the observers views of what was seen.

Note: As stated earlier an independent count that provides imagery of the entire survey area is needed to determine the population of wild horses in KNP. Please see documentation submitted with this email of an independent count being undertaken to do just that.

28.

C3 Foals and joeys – p. 17

Foals and Joeys

Please see my response to number 6 which outlines from the 2014 and 2019 report there is no mention of foals.

29.

CONCLUSION

No comment.

INDEPENDENT COUNT OF THE WILD HORSES IN KOSCIUSZKO NATIONAL PARK

An independent count has been organised by Rocky Harvey and Claire Galea with a contract being issued with AirborneLogic and funded on GoFundMe. This count will be undertaken by the company and the report obtained will be peer-reviewed. Both Rocky and Claire will not be involved in any counting, analysis or reporting of the data.

Below is a summary of the project.

QUOTATION AND CONTRACT

We have negotiated a fixed price contract with Airborne Logic for $76,000 including GST for combined imagery capture, data processing, reporting and peer review. We have raised sufficient funds (less funding platform fees) for this sum. The survey area is per previous updates and the contract has provision to increase the size of the survey area, including data processing, should more funds be made available prior to flight day.

Should additional funds become available prior to the imagery flight, the contract allows for additional area to be added to the survey.

TECHNOLOGY

Airborne Logic will be using the latest technology for the survey and data analysis including:

- Fixed wing plane flying at about 3500 feet (1050m) above ground with twin 5cm per pixel cameras. The elevation of the flight means the horses will not be disturbed or animated by the presence of the plane, but also that no approval from NPWS is required.
- Imagery will be sewn into a complete mosaic of the survey area – like a photographic map of the entire survey area
- A.I. software will be trained in horse recognition and count horses in the images
- We are advised that horse detection will be incredibly high with 5cm RGB imagery
- We were keen on surveying both RGB (red, green blue – normal colour) imagery combined with thermal but were advised that this presented a number of issues:
    - Cost was a big one with imagery capture cost about double for the same area
    - Inclusion of thermal resulted in the narrowest swarth width (distance between flight paths) of all options
    - Thermal results are best in the morning when temp variation between animal and ground is highest, but this is the worst time for RGB imagery when light levels are low and shadows long
    - The machine learning programs will be different for thermal and RGB data and significantly increase the cost and timeframe of data processing
    - Stand-alone thermal imagery will not distinguish between a horse and any other large animal such as sambar deer
- Use of dual cameras increase width of image with each pass meaning the survey area is covered more quickly

- Flight paths have been designed with relatively short flight runs before turning and flying a parallel path. This limits the opportunity for horse movement as the time between passes is minimised.
- Covering the survey area quickly, from a high elevation in short parallel runs will all minimise potential for double counting or mis counting horses.
- The scope of machine training for this project excludes detection of double counting (we are advised this would be extremely complex and costly), however to mitigate this risk, the pilot will make a log (including aircraft GPS position) of any significant horse sightings and movement particularly on the extremities of the imagery.
- This automation of the data collection and analyses results in almost complete removal of human error and potential bias – the polar opposite of techniques currently used by NPWS

STEPS INVOLVED

Airborne Logic's progressive steps for the project are as follows:

1. Procure imagery
2. Train photogrammetry software which will create a mosaic of the survey area to tag all relevant pixels that are known to be horses
3. Use machine learning software to develop an image recognition model - basically building a model horse in the computers brain if that makes sense. This image recognition model will be for the open plains.
4. Run the model over the entire survey area which will identify all horses and generate a total estimated number of horses in the survey area.
5. Test the model rigorously, tabulate and report all findings.
6. Load raw imagery data onto public portal along with layer showing detected horses
7. Peer review

AIM

The aim in all of this is to develop a peer-reviewed unique and verifiable model that uses aerial image capture and machine learning to identify hoses in an automate, extremely accurate and repeatable process in similar landscape.

PUBLIC PORTAL

As part of their contract services, Airborne Logic will load the raw imagery onto a public portal where anyone can log in and view the data. They will also load a layer showing where horses were machine counted. This means that anyone can zoom in to anywhere in the survey area and identify where a horse was counted and view the actual image.

SELECTED SURVEY AREA

- Based on the annual NPWS spotting counts the distribution of horses in the open terrain is somewhat regular with the exception of The Tantangara/Currango Plain, Nungar Plain and

Boggy plain. Nungar Plain and Boggy plain have generally had a relatively low population density. Tantangara/Currango Plain has routinely had a very high density.

- Tantangara/Currango plain has been subject to a relatively recent cull of a high number of horses. Our team are concerned about the machine counting of dead horses as live ones - that is the machine will not be able to distinguish between an alive horse laying down to a dead horse. While the software could be trained for this purpose, it would add another complexity and cost.
- We think our selected survey area is a good balance of density but ultimately there is plenty of currently available information produced by Parks that illustrates population density distribution. This is both in the parts of the Carins reports and in results of annual spotting counts undertaken by NPWS.
- We have requested the raw data from the flights used to compare our results to the information used in the Cairns report. Providing NPWS release records of cluster locations and size that we can map against our survey area, Cairns own equations can be used to calculate the number of horses he estimates are in our survey area This will enable a direct comparison to our results

SELECTION OF AIRBORNE LOGIC/ALTERNATE FIRMS AND METHODS

- We considered the use of helicopter, drones, fixed wing and satellite for imagery platforms
- We contacted a number of different companies about a range of different potential survey methods
- Challenges:
  - High enough resolution from satellite is not readily available and cost prohibitive to commission.
  - Drone technology is advancing at an incredible rate but for this project, flight times, landing and take-off were an issue. In the future drones will no doubt become more readily available that can fly for longer and at suitable elevations however this is not available yet.
- Fixed wing was the most cost-effective solution we came up with and also provided flight elevations and speeds that would provide the best overall results.
- Airborne Logic were the firm who showed the most interest in the project and had the best suited technology and experience. They were aware of the political nature of the project and are still willing to be involved provided it remained 100% impartial and the work was independently peer-reviewed.

TIMING

- Subject to suitable weather, the imagery flight can be done between now and 28th January and that is what we are aiming for. If suitable weather does not occur in that time frame, the flight crew and plane will be available again from mid-February onwards.
- The team need suitable weather that is primarily safe for flying but also with conditions that will result in the best imagery quality – best results are on clear sunny days or days with high uniform cloud cover. Partly cloudy days that result in strong cloud shadows are not ideal.
- It is critical that our survey is taken in climatic conditions as close as practical to the days that NPWS ran their latest survey. This occurred on 7 of the days between 10th and 19th October 2023. We have requested more detailed data from the Government but are unsure when or if we will get this. Ideally, we would want to know the exact day the NPWS flight

over our survey area occurred. In the absence of that we do have some insights. The Cairns report noted that flights were not run on 3 of these days in October due to poor weather conditions.

- We have reviewed the climatic data from the BOM Cabramurra weather station (only 15km from the survey area) from the dates given by Cairns and the poor weather days can be easily identified. On the remaining days, the maximum daily temperature on all but one day was in the mid-teens.
- Given delaying our survey until October to match NPWS more closely is not viable, we are aiming to match the conditions as closely as practical in as many other aspects as we can. Thankfully the BOM long range forecast of a hot dry summer has not eventuated, and it should be completely viable to pick a day/s with temperature in the mid-teens and this will be our target condition.
- We are aware that a big difference between October and summer affecting horse location is insect activity – particularly those lovely not so little March Flies. When present, March Flies are most active in hot and sunny conditions and many horses will shelter from these biting critters in the shade of the trees.
- To mitigate potential claims that horses are not present in our images due to them sheltering in the trees from March Flies, we are aiming to do imagery with maximum temps in the mid tees along with uniform high cloud cover. This combination will maximise the number of horse on the plains without compromising the quality of the imagery.
- The current survey area will require at least a full day of flying. Running the survey on early mornings on successive days is not only cost prohibitive but early morning light is not ideal for quality of imagery. Statistically data will be more robust if procured in a single day.

CURRENT PROJECT TIMELINE

- Imagery – mid to late January 2024
- Data analysis draft report due 30 March 2024
- Final report including peer review 15th April 2024

EXPANDING THE SURVEY AREA, BOTH NOW AND AT A LATER DATE

- Ideally, we would have loved to survey the entire open terrain area of the northern block
- This would result in a cost increase to our budget of at least 50%
- Imagery capture is more relative to distance travelled however there are economies of scale with the data processing as a large portion of the cost is in machine training. Once the machines are trained to count horses, running additional data is not a lineal cost increase. That is, once the machine learning model has been built it is then a case of running the new data through the model and letting it work.
- If further funds were raised, or indeed if NPWS adopted the methods, additional areas could be surveyed and the data processed, not only in KNP but in any other similar terrain including the Victorian Alps.

AUTONOMY AND INDEPENDENCE

We have had extensive correspondence with many individuals and groups interested in the project and have taken very seriously all the input we have received. While we completely understand that many would like more direct involvement in project planning we believe that continuing in our roles as independent intermediaries is the only workable approach.

Claire and Rocky are jointly making all key decisions based on the advice of our professional team after taking into consideration community input. We remain open to queries and input and will continue updates over the GoFundMe platform as the project progresses.

Airborne Logic have made it crystal clear that they will be undertaking a completely unbiased scientific analysis and will not be pressured in any way in this respect.

****

Note: "Ideally, we would survey the same entire area using the NPWS Wild Horse Surveys (2,745km2) however being privately funded and tasked with the cost of training software, this is beyond reach. The NPWS Cairns 2023 survey asserts that there are 17,432 horses in the whole of the KNP with 13,163 of these existing in a 1299km^2 survey area referred to as the Northern Block. Given the raw information of both the size and location of the clusters is not available (although please note it has been formally requested but not yet provided) we can only extrapolate that they estimate 11,238 horses are in the 618km^2 open plains area of this Block. That is 64.5% of the entire horse population living in quite a small area equating to 18.17 horses for every square kilometre."

entropy

MDPI

Article

How the Outliers Influence the Quality of Clustering?

Agnieszka Nowak-Brzezińska * and Igor Gaibei

Institute of Computer Science, Faculty of Science and Technology, University of Silesia, Bankowa 12, 40-007 Katowice, Poland;
* Correspondence:

Abstract: In this article, we evaluate the efficiency and performance of two clustering algorithms: AHC (Agglomerative Hierarchical Clustering) and $K - Means$. We are aware that there are various linkage options and distance measures that influence the clustering results. We assess the quality of clustering using the Davies–Bouldin and Dunn cluster validity indexes. The main contribution of this research is to verify whether the quality of clusters without outliers is higher than those with outliers in the data. To do this, we compare and analyze outlier detection algorithms depending on the applied clustering algorithm. In our research, we use and compare the LOF (Local Outlier Factor) and COF (Connectivity-based Outlier Factor) algorithms for detecting outliers before and after removing 1%, 5%, and 10% of outliers. Next, we analyze how the quality of clustering has improved. In the experiments, three real data sets were used with a different number of instances.

Keywords: clustering; outlier detection; clustering quality indexes; AHC; k-Means

check for updates

Citation: Nowak-Brzezińska, A.; Gaibei, I. How the Outliers Influence the Quality of Clustering? *Entropy* **2022**, *24*, 917. https://doi.org/10.3390/e24070917

Academic Editor: Gergely Palla

Received: 22 May 2022
Accepted: 28 June 2022
Published: 30 June 2022

1. Introduction

Data clustering is one of the most effective tools for dealing with large amounts of data [1]. When there is a lot of data, we cannot manage it or extract valuable knowledge. By creating clusters of similar data in large data sets, we naturally divide them into homogeneous groups, which allows us to quickly search for groups of objects best suited to what we are currently looking for. When we have an extensive database of fingerprint images and try to classify a currently examined fingerprint, we have to browse through large amounts of data (a complete review of the entire repository) to find the most suitable image. The idea of clustering assumes that we first analyze the similarity of objects and then combine the most similar objects into groups. These groups contain representatives reflecting the group's content. Then, in the search process, it is enough to browse the representatives of the groups to find a group most similar to the information sought. Furthermore, only a selected group is analyzed. Whenever the real data is analyzed, creating a good quality group is not always possible. Consequently, it can threaten the effectiveness of searching for information in the group structure. Data outliers are factors that hinder the creation of coherent and separable clusters. That is why the problem of outliers in the data is so significant. In our research, we check how the outliers in data affect the difficulties in creating cohesive and well-separated groups. We use the methods known in the literature to assess the quality of clusters. Therefore, we can compare the quality of clusters containing outliers with the quality of clusters after a prior removal or omission outliers. We expect that the measured quality of the cluster should improve after the outliers have been removed. So, after removing or omitting outliers, the cluster should be more consistent internally and well-separated externally. In other words, the similarity of objects within groups should be even more significant (than before removing outliers). In contrast, the similarity of the groups relative to each other should be small. We should remember that when we collect large amounts of data, one of the most valuable techniques is dividing data into consistent groups and analyzing the created groups. We use two popular clustering algorithms: AHC (a hierarchical type) and $K - Means$ (a partitional type). Both of them are easy to understand and implement. A result is a set of data clusters. These clusters are searched to find accurate information, for example, a given data that matches a query in the best way. As we search only within the representatives of these clusters, we may omit the most relevant data even if they exist in this data set. That is why it is essential to verify the quality of the clusters [2,3]. There are many possible quality indexes for measuring the quality of created clusters. We decided to use the two most popular: *Dunn* and *Davies–Bouldin* indexes.

Similar studies are, of course, carried out by scientists around the world. We decided to test two clustering algorithms with different parameters because, as we know, they always have a significant impact on clustering quality. Our goal is to check which parameters impact a better quality of the created clusters. We also want to check which method allows outliers to be detected. We want to know whether a type of input data affects the effectiveness of detecting outliers and improves the quality of clusters after removing outliers. Does the character of the input data influence the effectiveness of clustering or outlier detection processes? Do the clustering parameters impact the quality of the clusters? Finally, we will confirm that the more outliers we detect, the more the quality of the clusters improves.

The structure of the paper is as follows. Section 2 contains an analysis of the existing knowledge related to the problems of clustering algorithms, methods of detecting outliers, and their impact on the quality of the created clusters. This section contains both the references to the research analyzed by the authors devoted to those particular issues by other scientists and the references to the authors' works on this topic. Section 3 describes the most important aspects (the definition of clustering, distance measures, and the quality indexes) of the analyzed clustering methods. Outlier detection algorithms (*LOF* (Local Outlier Factor) and *COF* (Connectivity-based Outlier Factor)) are introduced in Section 4. The essential research value of the paper is presented in Section 5 containing the research methodology, the description of the used data set, and the results of experiments. It also describes the programming environment used to implement the selected algorithms and the planned experiments. The paper winds up with a summary containing the interpretation of the retrieved results.

2. State of the Art

In the literature, one can find a lot of papers on either the comparison of the $k-Means$ and the AHC algorithm, the use of different distance measures or methods of combining clusters, methods of detecting outliers, or, finally, methods of analyzing the quality of clustering. In [4], the authors discuss and compare clustering algorithms and methods of cluster quality assessment (F-measure, Entropy) for different values of the number of clusters. However, they do not investigate the influence of outliers on the clustering results. In [5], the authors compared the clustering times for AHC and $k-Means$. However, their research does not cover the existence of outliers in the data or the study of the quality of clusters. The authors of [6] present the comparison of the $k-Means$ and AHC algorithms in terms of the number of clusters, the number of objects in clusters, the number of iterations, and clustering times for small and large data. However, the impact of outliers on clustering results or cluster quality research is not included there. A very interesting study was carried out in the paper [7], which compared various clustering algorithms with respect to the size of the data, noise resistance in the data, data types, or the number of input parameters. However, the research in the searched range, i.e., the impact of outliers on the quality of clustering, was not included. The paper [8] presents, in turn, a comparison of dozens of different approaches based on clustering and outlier detection but without any research details. Although, it is impossible to find papers that combine these issues into one study. In [9], we compared the clustering algorithms, outlier detection algorithms, and the methods for assessing the quality of created clusters, but wenever before merged all the issues in one study. We wanted to investigate whether the clustering algorithm we chose (AHC or $k-Means$ algorithm) influences the efficiency of data clusters containing outliers. Moreover, we wanted to find out whether the clustering parameters impact the obtained results.

3. Clustering Data Containing Outliers

It is known that clustering algorithms are designed to find objects similar to each other and put them into groups [1]. The more similar objects are, the easier it is to create a group from them. However, it is crucial that in the data we cluster, there is a part of the objects very similar to each other but simultaneously not similar to objects from other groups. If such a condition is met, we receive clusters consistent internally and well-separable externally. Such structures have high quality, most commonly assessed by measures of an internal cohesion assessment (the smallest possible distances within clusters) and an external separation (the highest possible distance between clusters). If outliers appear in the data, they significantly deteriorate the quality of clusters. It is worth emphasizing at this point that the outlier can be both a given error or information noise and real outlier data. Of course, we would like to eliminate these possible data errors just at the stage of data preprocessing because they do not contribute any information to the system and even disturb the created consistency of groups. In turn, the rare data, in reality, can bring a piece of significant, new knowledge to the system, and hence they should not be deleted or combined with all data because we may not see them. Such data should be distinguishable and further analyzed. That is why we propose to discover the outliers before the clustering process. For further analysis, such outliers should be introduced to domain experts, and the clustering process should proceed without outliers. Only then the searching within clusters is efficient.

3.1. Clustering Definition and Distance Measures

The main idea of the clustering process is to assign the objects to the created clusters considering their distance or similarity. The greater the distance, the less similar to each other the objects are, and thus they should not belong to the same group. Good-quality clustering requires the created groups to be as internally homogenous and externally distinct as possible. Using a proper distance or similarity measure by a given data type (quantitative, qualitative, or binary) is essential. There are many available measures of the distance or the similarity of data. We can distinguish between measures typically dedicated to numerical data (e.g., Euclidean) and typically linked to categorical data (e.g., Simple Matching Coefficient). In this paper, we use, analyze and compare the *Euclidean* and *Chebyshev* distance measures because we analyze numerical data in the experiments. Having two objects x and y in a p-th multidimensional space $(i = 1, 2, \ldots, p)$, the distance between these objects can be determined as *Euclidean* (Equation (1)) or *Chebyshev* (Equation (2)).

$$d(x,y) = \sqrt{\sum_{i=1}^{p} (x_i - y_i)^2} \tag{1}$$

$$d(x,y) = \max_{i}(|x_i - y_i|) \tag{2}$$

The choice of these measures has an impact on the results obtained. In the conducted experiments, the distance measure we have chosen in regard to clustering or detecting outliers will further influence whether we will achieve a better quality of the groups.

3.2. Clustering Algorithms: Hierarchical vs. Partitional

One of the most general classifications of clustering algorithms defines hierarchical and non-hierarchical clustering algorithms. Hierarchical clustering creates a tree of clusters by identifying and merging similar objects. The primary purpose of hierarchical clustering is to cluster such similar objects.

We used the agglomerative hierarchical clustering (AHC) algorithm. Given a D data set of N instances, this algorithm (AHC), recursively merges two clusters at each step until all instances are merged into one cluster. The conventional procedure of the AHC, also called the stored dissimilarities approach, takes a pairwise dissimilarity D matrix of an N size as input, initializes a binary tree with N leaves (singletons) with null height values, and iteratively adds new nodes (merged clusters) by fusing a pair of clusters (C_i, C_j) determined as follows:

$$(C_i, C_j) = \arg\min_x D(C_k, C_l) \tag{3}$$

AHC can be computationally costly. For the usual AHC procedure described above, the time complexity is $O(N^3)$. In AHC we use the *Lance–Williams* (*LW*) formula to calculate the dissimilarity between the initial cluster and a cluster formed by joining two other clusters [10].

We compute the distance between pairs of clusters using the following popular methods: Single Linkage (SL), Complete Linkage (CL), and Average Linkage (AL) [11]. In SL, the distance between two clusters is computed as the shortest possible distance between two points in the clusters, in CL as the distance between two data points furthest apart belonging to different clusters, respectively. At the same time, the AL uses an average distance between each point in the first and the second cluster. The AHC algorithm works as follows:

1. In the first step, each object constitutes a separate cluster. So there are $k = N$ clusters, and we must calculate the distance between each pair of points.
2. Find and join the two most similar clusters reducing the number of clusters by one.
3. Repeat the second step until obtaining the declared final number of clusters (k) or combining all objects into one big cluster.

The clustering process should be interrupted when the halted an optimal number of clusters is reached so that their quality is highest.

In each iteration of the $K - Means$ algorithm, we try to divide N objects into k groups so well that each object belongs to the group to which it is most similar. That division, however, means that if an object in the data set is essentially dissimilar to any cluster, the algorithm will try to include it in one and as a result breaking its internal consistency. Each cluster contains a representative. The representative plays a significant role because the object we want to include in the group is compared with it. Let us suppose it turns out that the distance of a given object from this particular representative is the smallest compared to the distances to the representatives of other groups. In that case, we include this object in this particular group. Then, this object also participates in forming a group representative.

The main idea of the algorithm is as follows:

1. Select the number of clusters (k) and assign k hypothetical centers at random;
2. For each observation of the original set, a nearest cluster center is determined;
3. The centroids are calculated-these are vectors, the components representing the average values of the particular features, calculated overall records of the cluster;
4. The center of the cluster is shifted to its centroid. Then, the centroid becomes the center of the new cluster;
5. Steps 2–4 are repeated iteratively;
6. The algorithm ends when no cluster changes occur at some iteration.

When analyzing both algorithms, we notice that the AHC algorithm seems more resistant to outliers. Unfortunately, this algorithm, in turn, requires more memory occupation. However, everyone can deal with outliers by seeing that they are those objects in the data set that do not match the created clusters, making them challenging to form.

3.3. Clustering Quality Indexes

Cluster validity is a way of assessing the quality of clustering results [2]. We obtain a different partition of the data into groups for the two different clustering algorithms or the same clustering algorithm but with different parameters. We use the cluster quality indexes without knowing which partition is the best. In this work, the *Dunn* and *Davies–Bouldin* indexes were used to validate the clustering quality. The *Dunn* index defines compact groups of clusters, the objects of which are well-grouped together, and the clusters themselves are located as far away from each other as possible. Higher values of the *Dunn* index indicate good quality of clustering-the higher the index, the better [3]. The *Dunn* index for k clusters is defined as Equation (4).

$$D(u) = \min_{1 \le i \le k} \{ \min_{1 \le j \le k, j \ne i} \{ \frac{(\delta(X_i, X_j)}{\max_{1 \le c \le k} \{\Delta(X_c)\}} \} \} \tag{4}$$

where $\delta(X_i, X_j)$ is an inter-cluster distance between cluster centroids C_i and C_j, and $\Delta(X_c)$ is an intra-cluster distance of cluster X_c. The indexes itself is sensitive to noise and outliers in the data. The index's modifications reduce this error with different methods of measuring inter-cluster distance. The quality of clustering performed using the quantitative and qualitative characteristics of the data set is shown by the *Davies–Bouldin* index. Since clusters must be compact and well-separated, the lowest possible index value indicates high-quality clustering. The *Davies–Bouldin* index for k clusters is defined as Equation (5).

$$DB(u) = \frac{1}{k} \sum_{i=1}^{k} \max_{i \ne j} \{ \frac{\Delta(X_i) + \Delta(X_j)}{(\delta(X_i, X_j))} \} \tag{5}$$

for $\Delta(X_i)$ being an average distance between the points within a X_i cluster ($\Delta(X_j)$, respectively, for cluster X_j). As can be seen from the definition, the *Davies–Bouldin* index determines the average similarity between each cluster and the cluster closest to it. The clustering process becomes more complicated when the data contain outliers. It is then much more difficult to form internally coherent and externally separable clusters. However, when we want to detect outliers in data, clustering algorithms are probably best. If the clustered data contain outliers, the quality of the created clusters decreases significantly. Outliers do not fit into the groups.

4. Outlier Detection

This section presents a definition of outliers in data in the context of literature in contrast with our definition.

4.1. Outlier'S Definition

Outlier detection is finding data points that behave very differently from what is expected [12,13]. An outlier is an object in a data set that deviates significantly from the remaining data, has values far from the estimated or average values, or is not similar to any other object in its characteristics. The definition says, "An outlier is an observation that is far removed from the rest of the observations" [14]. An observation in a data set is called an outlier if at least one of the following conditions is met:

- It deviates from standard or known data behavior;
- It has values that are far from estimated or average values;
- It is not related or similar to any other element in the group in terms of its characteristics.

Outliers can contain valuable data about abnormal parameters of systems. Recognizing such non-standard parameters provides valuable information with specific applications. Some examples are as follows: earth science, medical diagnosis, intrusion detection systems, prevention of credit-card fraud [12]. One of the main problems is that no single scoring method would assess the similarity of two data points and how much they differ from each other in the data set.

If the data set contains outliers, we may have data preprocessing or statistical analysis errors. Any outlier in the data set can skew the test results and lead to an erroneous interpretation of the data. Thus, the removal of outliers is an essential task in the analysis and processing of data. Analyzing algorithms and models for removing outliers is of particular interest to scientists. Because not every outlier object must be an error in the data, we assume in our research that we do not remove the discovered outliers but only skip them in the further analysis-e.g., clustering. We introduce the outliers to the domain experts to look at these outliers closer.

4.2. Outlier Detection Algorithms

The issues of outliers detection in data are broadly analyzed in the literature. The methods proposed so far can be allocated to one of four groups: statistical-based, cluster-based, distance-based, and density-based methods.

The statistical-based outlier detection method assumes that the data have a specific regular distribution and we use the probability distribution to find out the data which deviate from the statistical distribution curve. Such data is an outlier when an occurrence probability is lower than a threshold value. It is necessary to know the characteristics of the data in advance to select a suitable distribution model. However, in practical applications, the data is unknown and complex. The data is mainly multi-dimensional. The time complexity of this method is very high, so it is not suitable for high-dimensional data.

The cluster-based detection method detects objects that do not belong to any cluster or a small cluster. This method focuses on the overall distribution of the data and performs outlier detection after clustering a data set. A certain number of clusters is formed and the clusters whose data points are significantly smaller than other clusters constitute an output in the form of outliers.

The concept of distance-based outlier definition is based on the following assumption: by calculating the distance from an object to its neighbors and sorting, the object with the largest value in the order is marked as an outlier.

Density-based outlier detection is proposed to overcome the shortcomings of distance-based detection of global outliers. The most known density-based method in the literature is the Local Outlier Factor (LOF) algorithm.

Local Outlier Factor (LOF) algorithm uses a density-based approach [15]. We detect anomalies by measuring a local density deviation at a given data point concerning the data points near it. We calculate a local density for all objects in the data set. We can identify data points with the same density as their neighbors and ones with a lower density by comparing the calculated density. Those with lower density are considered outliers. Density-based approaches distinguish the following two parameters that define the concept of density:

- The $MinPts$ parameter (minimum points) indicating a minimum number of points;
- A parameter Eps defining a considered volume.

These parameters allow determining a density threshold for the algorithm, which decides whether or not a particular point is an outlier.

The idea of the LOF algorithm follows the 5 steps:

1. In the first step an *Euclidean* distance between each pair of objects is calculated.
2. In the next step, we calculate a distance $dist\_k(o)$ between a given data point o and its k-th nearest neighbor, using a so called *reachability-distance*.
3. In step 3, for each data point o the k-distance neighborhood of o is calculated.
4. In step 4 it is necessary to calculate reachability distances to all k-nearest neighbors of a point in order to determine a local reachability density of that point, which is computed by back-calculating the sum of all reachable distances of all k-nearest neighbors.
5. Finally, in step 5, it is necessary to calculate an LOF for every data point.

LOF values are sorted and the highest $LOF_k(o)$ value is selected as possible outlier.

The Connectivity-based Outlier Factor (COF) algorithm is a variation of the LOF algorithm. The distinction lies in a different approach to assessing the density of cluster objects. The algorithm assigns a degree of outlier value to each data point. Unlike the LOF, the COF algorithm calculates k-nearest neighbors ($k - NN$) using a chain distance. This approach is based on the location of the data points. The objects in the cluster have a linear distribution. Chain distances represent the minimum total distance (between the first and last data points). Objects with high COF values are considered outliers.

In order to determine the COF value for each data point, we execute the following process:

1. At first we find k nearest neighbors of the data point o. For each data point o we find the $N_k(o)$ set of its k nearest neighbors.
2. Then we need to find a closest set-based path (SBN), which is an ordered sequence of k nearest data points starting with the point under consideration.
3. Next it is necessary to find the cost of an SBN trail. We represent the trail as a set of weights of the respective edges.
4. We consider the weight of an edge to be a distance between the two data points.
5. After that we need to find an average chaining distance of the data point and finally a COF value of the data point.
6. In the last step the COF values are sorted and the highest $COF(o)$ value is selected.

4.3. The Concept of Outlier Detection Based on the Lof and Cof Algorithms-Our Approach

In our research, we want to check how the occurrence of outliers affects the quality of clusters. We assume that the quality of clusters with outliers is worse than without these outliers. Thus, it should be evident that we first examine the quality of the clusters with outliers, then ignore the identified outliers and redetermine the quality of the formed clusters. We will use the most known algorithms in the literature to discover the outliers: the LOF and the COF algorithms. Usually, the LOF and COF algorithms result in the same outliers detected.

The scheme of our approach is shown in Figure 1.

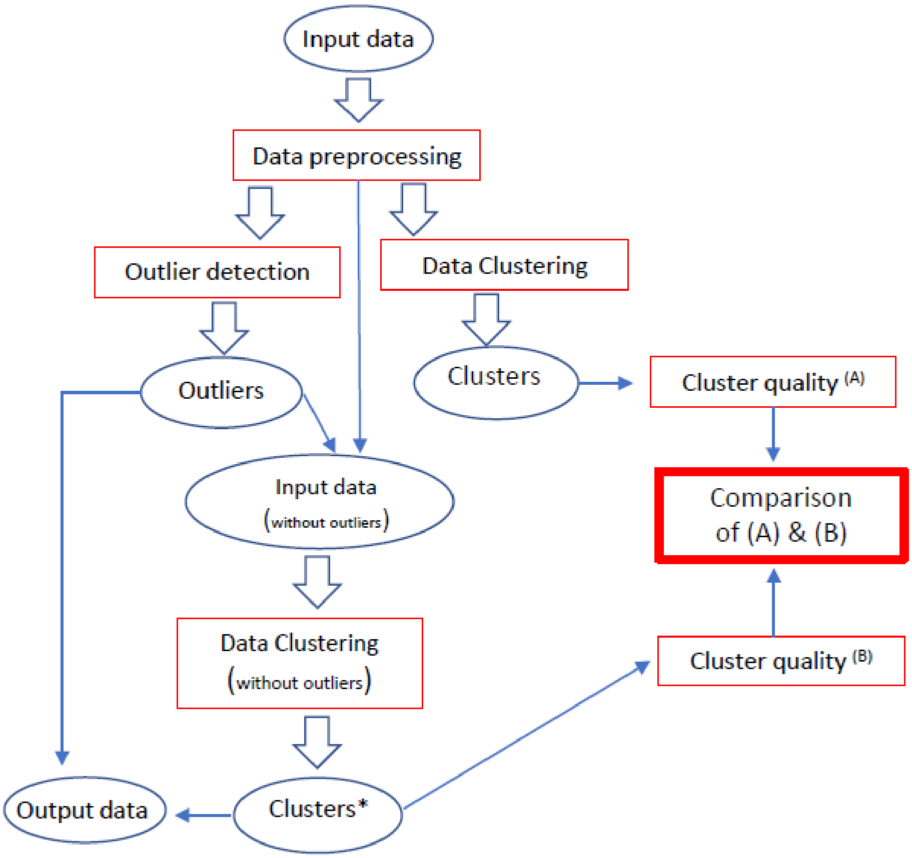


Figure 1. The concept of outlier detection based on *LOF* and *COF* algorithms—our approach. Clusters * mean clusters after removing discovered outliers.

Input data undergoes necessary data preprocessing operations and then clustering (we choose the AHC or $k - Means$) algorithm. Then we evaluate the quality of clusters obtained in this way. We detect outliers (select the LOF or COF algorithm) and return them as one of the two elements constituting the data's output. Bypassing the previously detected outliers, we cluster the input dataset again (and return the created clusters as the second of two elements of the output data) and assess the quality of the resulting clusters. We can compare the quality of data clusters containing outliers (A) with the quality of the clusters of the same input data without outliers (B). We expect that by excluding the outliers from the input dataset the quality of created clusters will increase.

5. Experiments

The experiments aimed at checking the impact of the clustering algorithms, clustering methods, and the selected distance measures on the effectiveness of outlier detection, measured by the response of cluster quality assessment indexes to remove outliers from the set. We wanted to see if the clustering algorithms and the outlier detection algorithms contributed similarly to improving the quality of clusters after detecting and removing outliers. We performed experiments on three different real datasets. We modified the number of detected outliers three times, using 1%, 5%, and 10% of the entire dataset as the number of outliers. We wanted to recognize the differences in the results. Our goal was to answer the question ″is it true that the more outliers we discover the better the quality of clusters without selected outliers will be″. In other words, we assume that if we discover outliers first and then cluster the data excluding the outliers, the quality of such clusters will be better than if we cluster the data including outliers. We analyzed two indexes for cluster quality assessment: the *Dunn* and the *Davies–Bouldin*. We measured the quality of clusters for the original data set in which potential deviations may occur. Then we look for outliers and omit them in the clustering process. In this way, we can compare the quality of clusters before and after removing outliers. Improving the quality of a cluster occurs if, after removing outliers, the quality measured by the *Dunn* index will increase as compared to the quality of clusters in which the outliers were not omitted in the clustering process. It is the completely opposite when the *Davies–Bouldin* index is concerned. Quality improvements occur when the quality of clusters measured by the *Davies–Bouldin* index decreases after removing outliers. Based on the experiments' results, we can count in how many cases, after removing the outliers, the quality of the clusters has improved (i.e., the Dunn index increased, and the Davies–Bouldin index decreased).

5.1. Data Description

The source of the databases is the UCI Machine Learning Repository [16], a collection of databases and data generators used by the machine learning community to analyze machine learning algorithms empirically. A brief description of each data set is shown in Table 1.

Table 1. Description of the data sets.

Dataset	Type	Number of Instances	Number of Features
A	numeric	740	21
B	mixed	6321	13
C	numeric	78,095 (First 12,000 used in experiments)	38

The created databases differ in the number of instances and attributes and the types of attributes. *A* (Absenteeism at work Dataset) is the database created with absenteeism records from July 2007 to July 2010 at a courier company in Brazil. The set contains 740 instances, each consisting of 21 numeric (Integer, Real) attributes [17]. *B* (Shill Bidding Dataset) contains information about bidders, auctions, bids, prices, and auction duration. This dataset contains 6321 instances, each consisting of 13 mixed numeric attributes [18]. *C* (MoCap Hand Postures Dataset) a dataset containing 78,095 instances, with each instance consisting of 38 numeric (Integer, Real) attributes [19]. To record 12 users performing five hand gestures with markers attached to a left-handed glove, a Vicon motion capture camera system was used. It is worth mentioning that the dataset B originally was of mixed type. Only one feature was qualitative, but this feature has only one value, and we decided to exclude it in this analysis. Therefore, finally, all datasets were numeric. Qualitative data research will be the basis of our research in the future.

5.2. Methodology

The purpose of the experiments was to compare various clustering methods, clustering algorithms, and distance measures, which makes it possible to determine how changes in these parameters affect the final clustering results and how much the quality of outlier detection is improved. The steps involved in the experiments are described below. For each of the three datasets, the following experiments were carried out:

1. Loading the dataset and preparing it correctly before applying clustering algorithms: preprocessing data using standarization, normalization, etc.;
2. Data clustering using two different algorithms: $k - Means$ with various number of clusters and AHC with different clustering methods (single, complete, average) and two different ways of measuring distance (*Euclidean* and *Chebyshev*). The tests were carried out with a different number of clusters in the range of k. Iteratively, starting with $i = 1$ and increasing an i-th parameter by one at each step, the number of clusters k is calculated as $k \approx \sqrt{N} \pm i\%N$ until the condition that $k \geq 2$ and $k < N$ is satisfied;
3. Assessing the clustering quality using the *Dunn* and the *Davies–Bouldin* indexes;
4. Finding 1%, 5%, and 10% of all outliers in the dataset using the LOF and COF. Removing the selected outliers and reclustering and recalculating the quality of clusters.

In total, we performed 686 experiments which are presented in this paper. The number of 686 experiments comes from the following calculation:

- There are two clustering algorithms: $k - Means$ and AHC;
- In case of AHC algorithm we may set the following values of the distance measures and clustering methods. For distance measure we have two options: *Euclidean* and *Chebyshev* distance measures. From the clustering methods we may choose one of three methods: single linkage (SL), complete linkage (CL), and average linkage (AL). Thus, using the AHC clustering algorithm we have 6 different combinations of given input parameters (see Figure 2).
- We adjust the number of created clusters to the size of the dataset. It means that for three used datasets we have various number of clusters. We do not want to check every possible value of k parameter because this would not be an efficient solution. The classical $k - Means$ clustering algorithm requires multiple repetitions pf the clustering process for an iteratively changed (most often by 1) the number of clusters, starting from the value $k = 2$. For a large data set, this process would be very ineffective. In the literature we can also come across an idea to divide the dataset into $\sqrt{N}$ of clusters. In our case, for the A dataset containing 740 of objects it would be 27 of clusters. Our idea is to adjust the number of different test values of the k parameter proportionally to the size of the analyzed data sets. Instead of that we propose to change the value of k iteratively according to the following formula. Starting with $i = 1$ and increasing it by one at each step, the number of clusters k is calculated as

$$k \approx \sqrt{N} \pm i\%N \tag{6}$$

until the condition that $k \geq 2$ and $k < N$ is met. The calculated values of k parameter are included in Table 2.

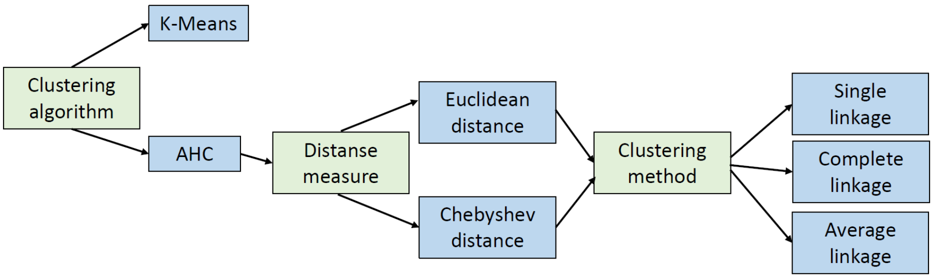


Figure 2. Clustering parameters used in the research.

Table 2. Description of the databases.

Dataset	N	k Values
A	740	5, 12, 19, 27, 34, 42, 49, 56
B	6321	16, 79, 142, 205
C	12,000	109, 229

For example, in case of the A dataset the calculation of k will be following:

- For $i = 1$ $k = \sqrt{740} \pm 1\% \cdot 740 = 34$ and 19;
- For $i = 2$ $k = \sqrt{740} \pm 2\% \cdot 740 = 42$ and 12;
- For $i = 3$ $k = \sqrt{740} \pm 3\% \cdot 740 = 49$ and 5;
- For $i = 4$ $k = \sqrt{740} \pm 4\% \cdot 740 = 56$ (here we can not continue the process of calculationg k values because we met the stop criteria which is in this case $k \geq 2$ and $k < N$).

This solution will allow us to check different k parameter values adapted to the size of the input dataset.

- The number of experiments is 686 as there are 8 versions of k parameter for the A dataset, 4 versions for the B dataset and 2 versions of k for the C dataset. We have 14 versions, and we repeat them for each of 6 different concepts of the AHC algorithm and 1 version of the $k - Means$ algorithm. Adding all these combinations together, we reach 98 experiments.
- Choosing two outlier detection algorithms LOF and COF accordingly and for each of the three different variants of the number of outliers 1%, 5% and 10% we obtain the final number of experiments equal to 686.
- Every experiment contains the value of clustering quality indexes *Dunn* and *Davies–Bouldin* which are essential for comparing before and after excluding potential outliers from a given dataset.

5.3. Experimental Environment

To analyze clustering algorithms before and after removing the outliers, the Spyder programming environment (Python 3.8) was used, as well as the following libraries: Pandas for data processing and analysis [20], NumPy to perform basic operations on n-arrays and matrices: addition, subtraction, division, multiplication, transposition, calculating determinant, etc. [21], PyCaret to prepare the data for modeling, create an unsupervised anomaly detector, and prepare the model for predictions on unseen data [22] and Scikit-learn, one of the most widely used Python packages for data science and machine learning, which allows many operations and provides a great variety of algorithms for data processing, reduction in dimensions, model selection, regression, classification and cluster analysis [23].

The algorithms described in Sections 3 and 4 have been implemented using Python and tested on the datasets described herein. We use Python 3.8 and the Anaconda package in this work, which includes many of the libraries required to run machine learning models, data mining, and output data in various formats. Existing Scikit-learn library models were used to implement the AHC and $K - Means$ clustering algorithms, the *Dunn* and *Davies–Bouldin* indexes and the Pycaret library to implement the outlier detection algorithm. The program operates in the following way:

1. Import Python analytical libraries Scikit-learn, NumPy, Pandas, PyCaret, and libraries to perform operations related to time.
2. Implementation of algorithms:
   (a) AHC ($algorithm\_of\_clustering$) with parameters: *k* denoting the selected number of clusters, *linkage* denoting the type of linkage used in clustering, *affinity* denoting distance measures;
   (b) $K - Means$ (kmeans) with a *k* parameter denoting the selected number of clusters;
   (c) *Dunn* and *Davies–Bouldin* algorithms ($dunn\_validator$, $davies\_validator$);
   (d) *LOF* and *COF* algorithms with parameter percent denoting the percentage of removed outliers.

   Data preparation functions:
   (a) $df.replace$—a function to replace the missing values with other values dynamically;
   (b) $df.fillna$—a function to replace Null values in Pandas data frame;
   (c) $\_normalize\_databases$—a function to normalize and standardize values in the data frame.
3. Uploading and reading all three datasets.
4. Execution of AHC, $K - Means$, *LOF*, *COF*, *Dunn*, and *Davies–Bouldin* algorithms on datasets.
5. Transfering results to the Excel file.

5.4. Results

First, the impact of the percentage of detected outliers for both the *LOF* and *COF* algorithms was examined with regard to a frequency of improvement in the quality of clusters after removing the detected outliers. The results are presented in Table 3. We can see that using the *Davies–Bouldin* index was much likelier to improve the quality of clusters than the *Dunn* index, regardless of how many outliers were detected. It is essential to explain that all results presented in this Section are the average values of the analyzed parameters for each of the 686 experiments performed in this research.

All experiments present the number of cases in which there has been improvement, deterioration, or no changes in the values of the quality of clusters. The percentage values we see in the tables do not mean to represent an average value but the exact number of cases reflecting the event. It is expressed in percentage compared to all experiments from a given group. For example, in Table 4, we can see that when 1% of outliers are discovered and removed (there are 196 such cases), in 129 of these 196 cases, which is 65.82%, the quality of clusters measured by the *Davies–Buldin* index has improved. In 102 cases in this group, the quality of clusters measured by the *Dunn* index has improved. The Tables are extended by a piece of additional information (the number of cases confirming a given event).

Table 3. The impact of the number of outliers of the frequency of increase in cluster quality.

% Of Outliers/ # Cases	Increase in Quality Dunn	Increase in Quality Davies–Bouldin	Decrease in Quality Dunn	Decrease in Quality Davies–Bouldin	No Change Dunn
1%	52.04% 102	65.82% 129	32.14% 63	34.18% 67	15.82% 31
5%	63.78% 125	78.57% 154	31.12% 61	21.43% 42	5.10% 10
10%	67.86% 133	82.14% 161	32.14% 63	17.86% 35	0.00% 0
Chi ^{2} Pearson		Dunn index: $p = 0.00000$		Davies–Bouldin index: $p = 0.00136$	

Then we decided to check whether any of the clustering algorithms used contributed more to improving the quality of clusters than the other after removing outliers. The results are presented in Table 4.

Table 4. The impact of the number of outliers of the frequency of increase in/decrease in the cluster quality.

Clustering Algorithm	Increase in Quality Dunn	Increase in Quality Davies–Bouldin	Decrease in Quality Dunn	Decrease in Quality Davies–Bouldin	No Change Dunn
AHC	60.91% 307	73.81% 372	31.15% 157	26.19% 132	7.94% 40
$K - Means$	63.10% 53	85.71% 71	35.71% 30	14.29% 12	1.19% 1
Chi ^{2} Pearson		Dunn index: $p = 0.07329$		Davies–Bouldin index: $p = 0.01882$	

It turns out that taking into account all the experiments performed, the quality of the clusters was higher after removing the outliers, more often for the $k - Means$ algorithm than for *AHC*. Furthermore, this is regardless of whether the *Dunn* or *Davies–Bouldin* index was used. We see that not all the differences studied are statistically significant. At the level of statistical significance, $p < 0.05$, we will say that in the case of the *Davies–Bouldin* index, the use of the $K - Means$ algorithm for clustering data has much more often led to a record improvement in the quality of clusters after removing deviations. In other words, the $K - Means$ algorithm is not resistant to the presence of outliers. Therefore, no statistically significant differences in the quality of clusters were noticed when we eliminated outliers using the *Dunn* index.

An important task was to examine the impact of using the outlier detection method on the frequency of improvement of the quality of clusters after removing outliers. Table 5 contains the results. There is an interesting tendency there.

Using the *COF* outlier detection algorithm the increase in quality of created clusters is achieved much more often than using the *LOF* algorithm. It means that *COF* algorithm depends more significantly on the occurence of outliers. We notice that using the *COF* algorithm statistically significantly ($p < 0,05$, Chi^2 Pearson Test) more often leads to improving the quality of clusters after eliminating the outliers. Therefore, the *COF* algorithm tends to discover more significant outliers. After removing them, the quality of the clusters improves.

Table 5. LOF and COF algorithms for cluster quality indexes.

Outlier Detection Algorithm	No Change	Dunn Increase	Dunn Decrease	Davies–Bouldin Increase	Davies–Bouldin Decrease
LOF	13.95% 41	53.40% 157	32.65% 96	68.03% 200	31.97% 94
COF	0% 0	69.05% 203	30.95% 91	82.99% 244	17.01% 50
Chi ^{2} Pearson		Dunn index: $p = 0.00000$		Davies–Bouldin index: $p = 0.00014$	

We also wanted to check if and how the distance measures contribute to improving the quality of the clusters. It turns out that when using the *Euclidean* distance measure, the improvement of cluster quality is more often achieved for the *Davies–Bouldin* index, while for the *Chebyshev* measure, the quality of the clusters is more often improved by using the *Dunn* index. As Table 6 indicates, there are no statistically significant differences ($p > 0.05$) in the effectiveness of improving the quality of clustering after removing outliers depending on what distance measure (*Euclidean* or *Chebyshev*) we use.

Table 6. Distance mesures for cluster quality indexes.

Distance Measure	No Change	Dunn Increase	Dunn Decrease	Davies–Bouldin Increase	Davies–Bouldin Decrease
Euclidean	7.94% 20	58.33% 147	33.73% 85	76.19% 192	23.81% 60
Chebyshev	7.94% 20	63.49% 160	28.57% 72	71.43% 180	28.57% 72
Chi ^{2} Pearson		Dunn index: $p = 0.44332$		Davies–Bouldin index: $p = 0.22409$	

Knowing that the analyzed datasets are real datasets that differ with respect to the size and type of the analyzed data, we also decided to investigate the differences in the frequency of increase or decrease in clustering quality depending on the input data source. The results are presented in Table 7.

Table 7. The frequency of improving the quality of clusters according to the type of data.

Dataset	No Change	Dunn Increase	Dunn Decrease	Davies–Bouldin Increase	Davies–Bouldin Decrease
A	10.12% 34	60.71% 204	29.17% 98	78.57% 264	21.43% 72
B	2.38% 4	60.12% 101	37.50% 63	74.40% 125	25.60% 43
C	3.57% 3	65.48% 55	30.95% 26	65.48% 55	34.52% 29
Chi^2 Pearson	Dunn index: $p = 0.00727$			Davies–Bouldin index: $p = 0.04103$	

Types of data sets we analyze significantly impact how effective the process of outlier detection is and consequently impact the quality of the created clusters. There are statistically significant differences for each of the analyzed datasets in the frequency of improvement in the quality of clusters after removing previously found outliers.

Table 8 also presents interesting results. We can see that depending on which set was analyzed, the quality of clusters did not constantly improve as the number of detected deviations increased. It is also impossible to unequivocally determine whether any of the measured indexes of the quality of clusters always allows obtaining an improvement in the quality of clusters. This confirms that the size and type of analyzed data have a significant impact on the effectiveness of deviation detection and the quality of clustering.

Table 8. The frequency of improving the quality of clusters according to the type of data and the number of discovered outliers.

	Increase in Clustering Quality Indexes					
	Dunn			Davies–Bouldin		
Dataset	**1%**	**5%**	**10%**	**1%**	**5%**	**10%**
A	43.75% 49	64.29% 72	74.11% 83	68.75% 77	83.04% 93	83.93% 94
B	60.71% 34	62.50% 35	57.14% 32	64.29% 36	76.79% 43	82.14% 46
C	67.86% 19	64.29% 18	64.29% 18	57.14% 16	64.29% 18	75.00% 21
Chi ^{2} Pearson	Dunn index: $p = 0.00000$			Davies–Bouldin index: $p = 0.00136$		

We see a trend in which the more deviations we detect and turn off from clustering, the more often the quality of the clusters improves. We should point out that in the end, the analyzed dataset with a specific type of data determines the effectiveness of outlier detection and improves the quality of clusters.

The last analyzed clustering parameter, which can affect the improvement of the quality of clusters after removing the outliers, is the cluster combinination method. Table 9 indicates that there are statistically significant differences between the clustering methods (single, complete, average) in the frequency of improvement of cluster quality after removing outliers. We can see that outliers removal improves the quality of clusters by less than 30 percent of cases (using the *Dunn* index to assess the quality) while using the single method. In the case of the complete or average method, this effect is obtained much more often (about 80%).

Table 9. The frequency of improving the quality of clusters according to the clustering method.

Dataset	**Dunn**			**Davies–Bouldin**	
	No Change	**Increase**	**Decrease**	**Increase**	**Decrease**
Single	13.69% 23	29.76% 50	56.55% 95	58.33% 98	41.67% 70
Complete	2.98% 5	79.76% 134	17.26% 29	88.10% 148	11.90% 20
Average	7.14% 12	73.21% 123	19.64% 33	75.00% 126	25.00% 42
Chi^2 Pearson		$p = 0.00000$		$p = 0.00000$	

5.5. Discussion

The research concludes that the *COF* algorithm more often improves the quality of clusters than *LOF* by removing the outliers. In the context of clustering algorithms, the $K-Means$ algorithm reacts much more actively to the outlier's presence and skipping. Probably the reason is that this algorithm is much less resistant to the appearance of outliers in the set than the hierarchical algorithm. The research also confirmed the original assumption that the more outliers we remove from the set, the better the quality of the clusters would be. However, an essential conclusion seems to be that the input data type significantly affects the results achieved: the quality of the clusters created for data containing potential outliers.

6. Summary

In this research, we assessed the influence of the clustering parameters, the clustering algorithms, and the outliers detection methods on the quality of created clusters. Several hundred experiments were performed, where individual clustering parameters (distance measures, number of clusters, clustering algorithms) and outliers detection parameters (number of outliers and outlier detection algorithm) changed for three different data sets. We checked which factors responded positively to outliers. It turned out that a vast majority of experiments confirmed the thesis that if a data set contains outliers, it will negatively affect the quality of created clusters. Therefore, this should prompt us to search for outliers before clustering large and real data sets. By removing the outliers first, we will be able to form good-quality clusters from the data and, therefore, achieve a greater efficiency in exploring such datasets. An additional benefit of outlier detection will be a reduced clustering time (as there is no longer any difficulty in a cluster formation). Consequently, a better quality of the created clusters will translate into improved quality of explored knowledge. Of course, the detected outliers are, by definition, passed on to field experts who have a chance to explore knowledge in a previously underexplored area.

Author Contributions: Conceptualization, A.N.-B.; formal analysis, A.N.-B.; investigation, I.G.; methodology, A.N.-B.; project administration, I.G.; software, I.G.; validation, A.N.-B.; visualization, A.N.-B.; writing–original draft, A.N.-B.; writing–review and editing, A.N.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LOF	Local Outlier Factor
COF	Connectivity-based Outlier Factor
AHC	Agglomerative Hierarchical Clustering
k-means	$k - Means$ clustering algorithm

References

1. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2005.
2. Legany, C.; Juhasz, S.; Babos, A. Cluster validity measurement techniques. In Proceedings of the 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, AIKED'06, Madrid, Spain, 15–17 February 2006; World Scientific and Engineering Academy and Society (WSEAS): Stevens Point, WI, USA, 2006; pp. 388–393.
3. Dunn, J.C. Well Separated Clusters and Optimal Fuzzy Partitions. *J. Cybern.* **1974**, *4*, 95–104. [CrossRef]
4. Steinbach, M.S.; Karypis, G.; Kumar, V. *A Comparison of Document Clustering Techniques*; Technical Report; University of Minnesota: Minneapolis, MN, USA, 2000.
5. Karthikeyan, B.; Dipu, J.G.; Manikandan, G.; Thomas, T. A Comparative Study on K-Means Clustering and Agglomerative Hierarchical Clustering. *Int. J. Emerg. Trends Eng. Res.* **2020**, *8*, 1600–1604. [CrossRef]
6. Saleena, T.S.; Sathish, A.J. Comparison of K-Means Algorithm and Hierarchical Algorithm using Weka Tool. *Int. J. Adv. Res. Comput. Commun. Eng.* **2018**, *7*, 74–79.
7. Vadgasiya, M.G.; Jagani, J.M. An enhanced algorithm for improved cluster generation to remove outlier's ratio for large datasets in data mining. *Int. J. Adv. Eng. Res. Dev.* **2014**, *1*, 203–208.
8. Jabbar, A.M. Local and Global Outlier Detection Algorithms in Unsupervised Approach: A Review. *Iraqi J. Electr. Electron. Eng.* **2021**, *17*, 76–87. [CrossRef]
9. Nowak-Brzezińska, A.; Horyń, C. Outliers in rules-the comparision of LOF, COF and KMEANS algorithms. *Procedia Comput. Sci.* **2020**, *176*, 1420–1429. [CrossRef]

10. Lance, G.N.; Williams, W.T. A general theory of classificatory sorting strategies. II Clustering systems. *Comput. J.* **1967**, *10*, 271–277. [CrossRef]
11. The CLUSTER Procedure: Clustering Methods. In *SAS/STAT 9.2 Users Guide*; SAS Institute: Cary, NC, USA, 2009; pp. 1250–1260.
12. Kishan, G.M.; Chilukuri, K.M.; HuaMing, H. *Anomaly Detection Principles and Algorithms*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 23–38.
13. Ranga Suri, N.N.R.; Narasimha Murty, M.; Athithan, G. *Outlier Detection: Techniques and Applications*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 3–9.
14. Maddala, G.S. Outliers. In *Introduction to Econometrics*, 2nd ed.; MacMillan: New York, NY, USA, 1992.
15. Breunig, M.; Kriegel, H.P.; Ng, R.T.; Sander, J. LOF: Identifying Density-Based Local Outliers. In Proceedings of the ACM SIGMOD 2000 International Conference on Management of Data, Dallas, TX, USA, 16–18 May 2000.
16. UCI Machine Learning Repository. Available online: https://archive.ics.uci.edu/ml/ (accessed on 10 October 2021).
17. Martiniano, A.; Ferreira, R.P.; Sassi, R.J. Universidade Nove de Julho-Postgraduate Program in Informatics and Knowledge Management. Available online: https://archive.ics.uci.edu/ml/datasets/Absenteeism+at+work (accessed on 5 April 2018).
18. Alzahrani, A.; Sadaoui, S. Shill Bidding Dataset. Available online: https://archive.ics.uci.edu/ml/datasets/Shill+Bidding+Dataset (accessed on 10 March 2020).
19. Gardner, A.; Selmic, R.R.; Kanno, J.; Duncan, C.A. MoCap Hand Postures Data Set, Louisiana Tech University, Quinnipiac University. Available online: https://archive.ics.uci.edu/ml/datasets/MoCap+Hand+Postures (accessed on 2 November 2016).
20. Wes McKinney and the Pandas Development Team. Pandas: Powerful Python Data Analysis Toolkit Release 1.3.3. Available online: https://devdocs.io/pandas~0.25/ (accessed on 20 November 2021).
21. NumPy Reference, Release 1.21.0, Written by the NumPy Community. Available online: https://numpy.org/doc/stable/numpy-ref.pdf (accessed on 20 November 2021).
22. Anomaly Detection Tutorial (ANO101). Available online: https://github.com/pycaret/pycaret/blob/master/tutorials/Anomaly%20Detection%20Tutorial%20Level%20Beginner%20-%20ANO101.ipynb (accessed on 20 November 2021).
23. An Introduction to Machine Learning with Scikit-Learn. Available online: https://scikit-learn.org/0.21/tutorial/basic/tutorial.html (accessed on 10 October 2021).

Distance sampling

Len Thomas, Stephen T. Buckland, Kenneth P. Burnham, David R. Anderson, Jeffrey L. Laake, David L. Borchers & Samantha Strindberg

Volume 1, pp 544–552

in

Encyclopedia of Environmetrics
(ISBN 0471 899976)

Edited by

Abdel H. El-Shaarawi and Walter W. Piegorsch

Distance sampling

Distance sampling is a widely-used group of closely related methods for estimating the density and/or **abundance** of biological populations. The main methods are *line transects* and *point transects* (also called *variable circular plots*). These have been used successfully in a very diverse array of taxa, including trees, shrubs and herbs, insects, amphibians, reptiles, birds, fish, marine and land mammals. In both cases, the basic idea is the same. The observer(s) perform a standardized survey along a series of lines or points, searching for objects of interest (usually animals or clusters of animals). For each object detected, they record the distance from the line or point to the object. Not all the objects that the observers pass will be detected, but a fundamental assumption of the basic methods is that all objects that are actually on the line or point are detected. Intuitively, one would expect that objects become harder to detect with increasing distance from the line or point, resulting in fewer detections with increasing distance. The key to distance sampling analyses is to fit a *detection function* to the observed distances, and use this fitted function to estimate the proportion of objects missed by the survey. From here we can readily obtain point and interval estimates for the density and abundance of objects in the survey area. The basic methods (sometimes called *conventional* or *standard distance sampling*) are described in detail in [5], which is an updated version of [4]. Free software, Distance [19], provides for the design and analysis of distance sampling surveys, and is being updated to include much of the work mentioned in the section on Current Research below.

Distance sampling is an extension of quadrat-based sampling methods. Two forms of quadrat sampling are *strip transects*, in which the observer travels along a line, counting all objects within a predetermined distance of the line, and *point counts*, in which numbers of objects (usually birds or plants) in a circle about a point are counted. Population density is then estimated by dividing the total count by the total area surveyed. A fundamental assumption of these methods is that all objects within the strip or circle are counted. This assumption is difficult to meet for many populations, and cannot be tested using the survey data. Furthermore, for scarce species, the methods are wasteful, because detections of objects beyond the strip or circle boundary are ignored. If the width of the strip or the radius of the circle is made sufficiently small that detection of any object within the surveyed area is almost certain, then perhaps 50% or more of detections are outside the surveyed area and so are ignored. Distance sampling extends quadrat-based methods by relaxing the assumption that all objects within the circle or strip are counted. By measuring distances to the objects that are observed, the probability of observing an object within the circle or strip can be estimated.

Another approach to estimating wildlife abundance involves **capture–recapture methods**. These are often more labour-intensive and more sensitive to failures of assumptions than distance sampling. However, they are applicable to some species that are not amenable to distance sampling methods, and can yield estimates of survival and recruitment rates, which distance sampling cannot do. Capture–recapture methods can be useful for populations that aggregate at some location each year, whereas distance sampling methods are more effective on dispersed populations. They should therefore be seen as different tools for different purposes (see also trapping webs under Related Methods below).

In **fisheries** applications, *catch per unit effort*, *catch-at-age* and *catch-at-length* are all commonly used to estimate abundance [10], as they require that the commercial catch is sampled, which is more cost-effective than sampling the living fish. Acoustic surveys of fish schools can provide data amenable to distance sampling methods.

Alternative methods for estimating animal abundance are reviewed and compared in [15]–[18] and [21].

Line-transect Sampling

In **line-transect sampling**, a series of straight lines (tracklines) is traversed by an observer. This may be achieved in various ways, depending on the study species. In terrestrial studies, these include walking, horseback, all-terrain vehicle, aeroplane and helicopter. Transect surveys in aquatic environments can be conducted by divers with snorkels or SCUBA gear, from submarines, surface vessels, aircraft, or from sleds with mounted video units pulled underwater by a surface vessel. In the case of large observation platforms, there is typically a team of observers.

Estimation

Perpendicular distances x are measured from the line to each detected object of interest. In practice, detection distances r and detection angles θ are often recorded, from which perpendicular distances are calculated as $x = r\sin\theta$. Suppose k lines of lengths $l_1, \ldots, l_k$ (with $\sum l_j = L$) are positioned according to some randomized scheme, and n animals are detected at perpendicular distances $x_1, \ldots, x_n$. Suppose in addition that animals further than some distance w from the line (the truncation distance) are not recorded. Then the surveyed area is $a = 2wL$, within which n animals are detected. Let P_a be the probability that a randomly chosen animal within the surveyed area is detected, and suppose an estimate $\widehat{P}_a$ is available. Then animal density D is estimated by

$$\widehat{D} = \frac{n}{2wL\widehat{P}_a} \tag{1}$$

To provide a framework for estimating P_a, we define the detection function $g(x)$ to be the probability that an object at distance x from the line is detected, $0 \le x \le w$, and assume that $g(0) = 1$. That is, we are certain to detect an animal on the trackline. If we plot the recorded perpendicular distances in a histogram, then conceptually the problem is to specify a suitable model for $g(x)$ and to fit it to the perpendicular distance data. As shown in Figure 1, if we define $\mu = \int_0^w g(x)\,\mathrm{d}x$, then $P_a = \mu/w$. The parameter μ is called the effective strip (half-) width; it is the distance from the line for which as many objects are detected beyond μ as are missed within μ (Figure 1). Thus

$$\widehat{D} = \frac{n}{a\widehat{P}_a} = \frac{n}{2wL\hat{\mu}/w} = \frac{n}{2\hat{\mu}L} \tag{2}$$

We now need an estimate $\hat{\mu}$ of μ. We can turn this into a more familiar estimation problem by noting that the probability density function (pdf) of perpendicular distances to detected objects, denoted $f(x)$, is simply the detection function $g(x)$, rescaled so that it integrates to unity (*see* **Frequency curves**). That is, $f(x) = g(x)/\mu$. In particular, because we assume $g(0) = 1$, it follows that $f(0) = 1/\mu$ (Figure 2). Hence

$$\widehat{D} = \frac{n}{2\hat{\mu}L} = \frac{n\hat{f}(0)}{2L} \tag{3}$$

The problem is reduced to modeling the pdf of perpendicular distances, and evaluating the fitted function at $x = 0$. The large literature for fitting density functions is now available to us. The Distance program uses the methods of [3], in which a parametric 'key' function is selected and, if it fails to provide an adequate fit, polynomial or cosine series adjustments are added until the fit is judged to be satisfactory by one or more criteria.

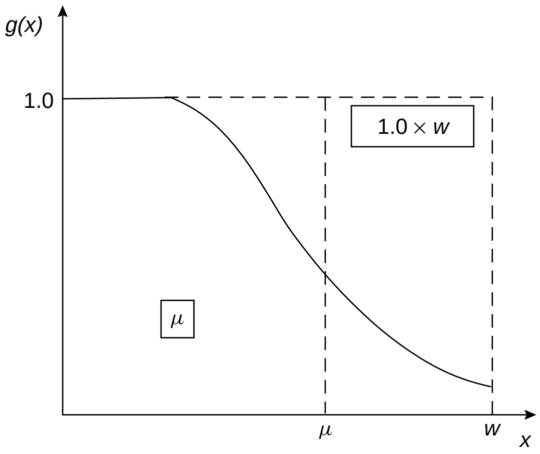


Figure 1 The area μ under the detection function $g(x)$, when expressed as a proportion of the area w of the rectangle, is the probability that an object within the surveyed area is detected; μ is also the effective strip width, and takes a value between 0 and w. Reproduced from Buckland, S.T., Anderson, D.R., Burnham, K.P. & Laake, J.L. (1998). Distance sampling, in *Encyclopedia of Biostatistics*, P. Armitage & T. Colton, eds, Wiley, Chichester, Figure 2, p. 1192 by permission of John Wiley & Sons, Ltd

Often, the perpendicular distances are recorded by distance category, so that each exact distance need not be measured, or data are grouped into distance categories before analysis. Standard likelihood methods for multinomial data are used to fit such 'grouped' data.

Variance and Interval Estimation

The variance of $\widehat{D}$ is well approximated using the formula [5]:

$$\widehat{\mathrm{var}}(\widehat{D}) = \widehat{D}^2 \left[\frac{\widehat{\mathrm{var}}(n)}{n^2} + \frac{\widehat{\mathrm{var}}[\hat{f}(0)]}{[\hat{f}(0)]^2} \right] \tag{4}$$

The variance of n generally is estimated from the sample variance in encounter rates, n_j/l_j, weighted by the line lengths l_j. When $f(0)$ is estimated by

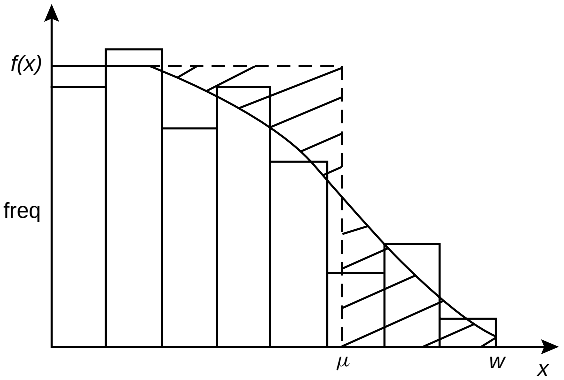


Figure 2 The pdf of perpendicular distances, $f(x)$, plotted on a histogram of perpendicular distance frequencies (scaled so that the total area of histogram bars is unity). The area below the curve is unity by definition. Because the two shaded areas are equal in size, the area of the rectangle, $\mu f(0)$, is also unity. Hence $\mu = 1/f(0)$. Reproduced from Buckland, S.T., Anderson, D.R., Burnham, K.P. & Laake, J.L. (1998). Distance sampling, in *Encyclopedia of Biostatistics*, P. Armitage & T. Colton, eds, Wiley, Chichester, Figure 3, p. 1192 by permission of John Wiley & Sons, Ltd

maximum likelihood, its variance is estimated from the **information matrix**.

If we assume that $\widehat{D}$ is lognormally distributed, approximate 95% confidence limits are given by $(\widehat{D}/C, \widehat{D}C)$ where

$$C = \exp\{1.96[\widehat{\mathrm{var}}(\ln \widehat{D})]^{0.5}\} \tag{5}$$

with

$$\widehat{\mathrm{var}}(\ln \widehat{D}) = \ln\left[1 + \frac{\widehat{\mathrm{var}}(\widehat{D})}{\widehat{D}^2}\right] \tag{6}$$

Often, **bootstrap resampling** for variance and interval estimation is preferred. Resamples are usually generated by sampling with replacement from the lines, so that independence between the lines is assumed, but independence between detections on the same line is not. If the model selection procedure for the detection function is applied independently to each resample, the bootstrap variance includes a component due to model selection uncertainty.

Cluster Size Estimation

Animals often occur in groups, which we term clusters. These may be flocks of birds, pods of whales, etc. If one animal in a cluster is detected, then it is assumed that the whole cluster is detected, and the distance to the center of the cluster is recorded. Equation (3) then gives an estimate of the density of clusters. To obtain the estimated density of individuals, we must multiply by an estimate of mean cluster size in the population, E(s):

$$\widehat{D} = \frac{n\hat{f}(0)\widehat{\mathrm{E}}(s)}{2L} \tag{7}$$

Probability of detection is often a function of cluster size, so that the sample of detected cluster sizes exhibits size bias (larger clusters are easier to detect and so are over-represented in the sample). In the absence of size bias, we can take $\widehat{\mathrm{E}}(s) = \bar{s}$, the mean size of detected clusters. Several methods exist for estimating E(s) in the presence of size bias [5] (*see* **Size-biased sampling**). One that works well in practice is to regress $\log s$ on $\hat{g}(x)$, the estimated probability of detection at distance x ignoring the effect of cluster size, and then predict $\log s$ when detection is certain, $\hat{g}(x) = 1$, as there can be no size bias in that circumstance. The prediction is back-transformed using a bias adjustment.

Assumptions

The physical setting for line-transect sampling is idealized as follows:

1. N objects are distributed through an area of size A according to some stochastic process with average rate parameter $D = N/A$.
2. Lines, placed according to some randomized design, are surveyed and a sample of n objects is detected.

It is not necessary that the objects be randomly (i.e. Poisson) distributed. Rather, it is critical that the line or point be placed randomly with respect to the local distribution of objects. This ensures that objects in the surveyed strip are uniformly distributed with respect to distance from the line. Thus, if the strip has half-width w, object-to-line distances available for detection are uniformly distributed between zero and w.

Three assumptions are essential for reliable estimation of density using standard line-transect methods:

1. Objects directly on the line are always detected, $g(0) = 1$.
2. Objects are detected at their initial location, prior to any movement in response to the observer.
3. Distances are measured accurately (for ungrouped distance data), or objects are correctly allocated to distance interval (for grouped data).

A fourth assumption is made in many derivations of estimators and variances: whether an object is detected is independent of whether any other object is detected. Point estimates of D are robust to the assumption of independence, and robust variance estimates are obtained by taking the line to be the sampling unit, either by bootstrapping on lines, or by calculating a weighted sample variance of encounter rates by line.

It is also important that the detection function has a 'shoulder'; that is, the probability of detection remains at or close to one initially as distance from the line increases from zero. This is not an assumption, but a property that allows more reliable estimation of object density.

Given the above, the point and interval estimates of D are extremely robust to variation in $g(x)$ due to other factors such as observer, habitat, etc. Large variations in density over the study area are also not a problem, although if areas of differing density can be defined in advance then stratification of survey effort could be used to increase precision.

Point-transect Sampling

In point-transect sampling, an observer visits a number of points, the locations of which are determined by some randomized design. The method is usually (but not exclusively) used for songbird populations, in which typically many species are recorded and most detections are aural. By recording from points, the observer can concentrate on detecting the objects of interest, without having to navigate along a line, and without having to negotiate a randomly positioned line through possibly difficult terrain. The principal disadvantages are that detections made while travelling from one point to the next are not utilized, a problem especially for scarce species, and the method is unsuited to species that are generally detected by flushing them, or to species that typically change their location appreciably over the time period of the count (see below).

Estimation

Detection distances r are measured from the point to each detected object. Suppose the design comprises k points, and distances less than or equal to w are recorded. Then the surveyed area is $a = k\pi w^2$, within which n objects are detected. As for line-transect sampling, denote the probability that an object within the surveyed are a is detected by P_a with estimate $\widehat{P}_a$. Then we estimate object density D by

$$\widehat{D} = \frac{n}{k\pi w^2 \widehat{P}_a} \tag{8}$$

We now define the detection function $g(r)$ to be the probability that an object at distance r from the point is detected, and we again assume that $g(0) = 1$. For line transects, the area of an incremental strip at distance x from the lines is $L\,\mathrm{d}x$, independently of x, which leads to the result that the pdf of distances differs from the detection function only in scale. By contrast, an incremental annulus at distance r from a point has area $2\pi r\,\mathrm{d}r$, proportional to r, so that the pdf of detection distances is $f(r) = 2\pi r g(r)/\nu$, where $\nu = 2\pi \int_0^w rg(r)\,\mathrm{d}r$. The respective shapes of the two functions $g(r)$ and $f(r)$ are illustrated in Figure 3. If we define an effective radius ρ, analogous to the effective strip width of line-transect sampling, then $\nu = \pi\rho^2$ is the effective area surveyed per point (Figure 4). Hence

$$\widehat{D} = \frac{n}{a\widehat{P}_a} = \frac{n}{k\pi w^2 \pi\hat{\rho}^2/\pi w^2} = \frac{n}{k\hat{\nu}} \tag{9}$$

The area of the triangle in Figure 4 is $\rho^2 f'(0)/2$ where $f'(0)$ is the slope of $f(r)$ at $r = 0$. Since this is equal to the area under $f(r)$, which is unity, it follows that $\nu = \pi\rho^2 = 2\pi/f'(0)$, and

$$\widehat{D} = \frac{n\hat{f}'(0)}{2\pi k} \tag{10}$$

We therefore need to model the pdf of detection distances, and evaluate the slope of the fitted function at $r = 0$. The program Distance does this using the same set of models for the detection function as for line-transect sampling.

Variance and Interval Estimation

The methods for variance and interval estimation for line-transect sampling also apply to point transects

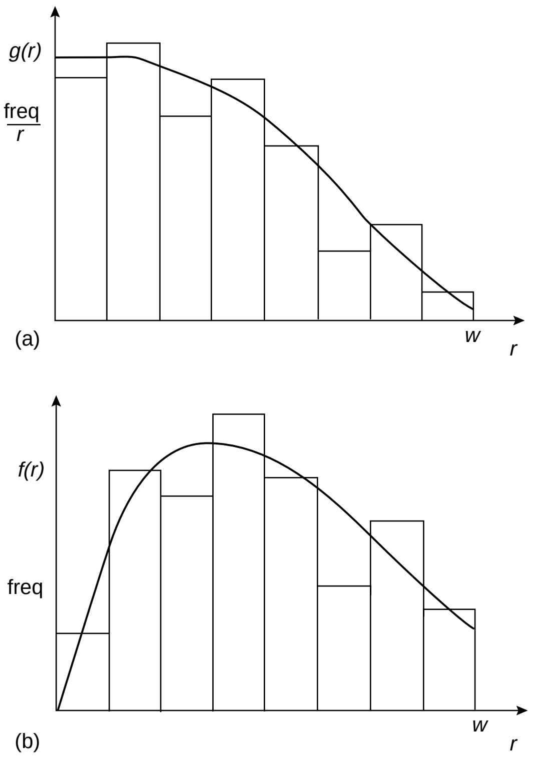


Figure 3 Histograms of detection distances from a point-transect survey. In (a) each histogram frequency has been scaled by dividing by the midpoint of the corresponding group interval. Also shown are the corresponding fits of the detection function [$g(r)$ in (a)] and the pdf of detection distances [$f(r)$ in (b)]. Reproduced from Buckland, S.T., Anderson, D.R., Burnham, K.P. & Laake, J.L. (1998). Distance sampling, in *Encyclopedia of Biostatistics*, P. Armitage & T. Colton, eds, Wiley, Chichester, Figure 4, p. 1194 by permission of John Wiley & Sons, Ltd

with minor modifications. The variance of n is usually estimated from the sample variance in encounter rates between points. However, point-transect surveys are often designed by defining a series of lines, as if a line-transect survey is to be carried out, then locating a series of points along each line. If the distance between neighboring points on the same line is smaller than the distance between neighboring points on different lines, then the data for all points on the same line should be pooled and the variance of n estimated from the sample variance in encounter rates between lines, weighted by the number of points on each line. Similarly, in this situation, bootstrap variance estimates should be calculated by resampling lines with replacement, rather than individual points.

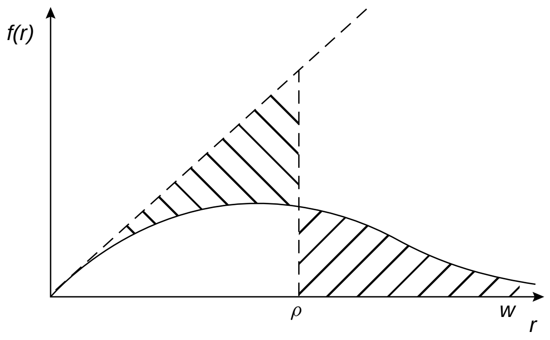


Figure 4 The pdf of detection distances, $f(r)$. The area under the curve is unity by definition. Because the two shaded areas are equal in size, the area of the triangle, $\rho^2 f'(0)/2$, is also unity. Hence $\nu = \pi\rho^2 = 2\pi/f'(0)$. Reproduced from Buckland, S.T., Anderson, D.R., Burnham, K.P. & Laake, J.L. (1998). Distance sampling, in *Encyclopedia of Biostatistics*, P. Armitage & T. Colton, eds, Wiley, Chichester, Figure 5, p. 1195 by permission of John Wiley & Sons, Ltd

Assumptions

Assumptions are virtually unchanged from those given for line-transect sampling. As there, the standard analyses are very robust to failure of the assumption of independent detections. Point-transect sampling is, however, more subject to bias than line-transect sampling when objects move through the area around a point. In principle, we try to obtain a snapshot, locating each object at the position it occupied at one instant in time. However, the count is not instantaneous, because the observer needs time to detect all objects close to that point. If, during that time, movement brings new objects into the neighborhood of the point, then object density will be overestimated. To minimize bias, we recommend that the amount of time spent at the point before and after the snapshot instant be fixed in advance, and be as small as possible, given the requirement to detect all objects close to the point.

Related Methods

Trapping webs [5, 22] provide an alternative to traditional capture–recapture sampling for estimating animal density. They represent the only application of distance sampling in which trapping is an integral part, and where data are taken passively. Traps are placed along lines radiating from randomly chosen points; the traditionally used rectangular trapping grid cannot be used as a trapping web. Here detection by an observer is replaced by animals being caught in a trap at a known distance from the center of a trapping web. The trap could be a camera or other similar device. Trapping continues for several occasions and data from either the initial capture of each animal or all captures and recaptures are analyzed. To estimate density over a wider area, several randomly located webs are required.

Cue counting [9] was developed as an alternative to line-transect sampling for estimating whale abundance from sighting surveys. Observers on a ship or aircraft record all sighting cues within a sector ahead of the platform and their distance from the platform. The cue used depends on species, but might be the blow of a whale at the surface. The sighting distances are converted into the estimated number of cues per unit time per unit area using a point-transect modeling framework. The cue rate (usually corresponding to blow rate) is estimated from separate studies, in which individual animals or pods are monitored over a period of time.

Indirect methods are often used when the animals are rare, cryptic or tend to move away before being detected. Instead of counting the animals, the objects counted are something produced by the animals, for example animal dung (e.g. deer dung [11]) or nests (e.g. great apes [12]). To convert object density to animal density one must then estimate two further parameters: object production rate and object disappearance rate, from separate studies.

Related techniques sometimes used by botanists to estimate densities (and sometimes also termed distance sampling) are **nearest neighbor methods** and point-to-nearest object methods [6]. These approaches do not involve modeling the detection function, and so are outside the definition of distance sampling used here.

Current Research

The basic theory of distance sampling is now well established, as are the standard estimation and field methods [5]. Most research is now focused on methods for increasing precision and relaxing the assumptions of the standard methods, and on advanced design issues. There is still much to be done in these areas, so the subject is still a lively one for statistics and ecology.

Generally, probability of detection is a function of many factors other than distance of the object from the line or point. We have considered briefly one other factor, cluster size, because if we do not allow for size bias in detection when objects occur in clusters then our object density estimator may be biased. Other sources of heterogeneity contribute little to bias, provided $g(0) = 1$. Nevertheless, higher precision might be anticipated if additional covariates are recorded and their effects on $g(x)$ modeled. One approach, first used by [14], is to allow covariates to affect the scale of the detection function but not its shape. Marques and Buckland (unpublished) have extended the detection function estimation methods outlined in the section on line-transect sampling above to allow the scale parameter of the key function to be a function of covariates. This approach is implemented in the software Distance.

In some surveys, detection on the trackline is not certain ($g(0) < 1$), perhaps because some animals are underground or under water, or simply hidden by vegetation, when the observer passes. In this case, capture–recapture methods may be combined with distance sampling, through the use of two observation platforms [2]. The platforms might be treated as mutually independent so that, provided that animals detected by both platforms (duplicate detections) can be identified, two-sample capture–recapture methods that incorporate covariates can be used. Bias in such methods is typically large enough to be of concern unless heterogeneity in detectability is well-modeled. However, it is seldom possible to record covariates that reflect this heterogeneity adequately. For example if a whale produces a blow that is particularly visible from one platform, due to light conditions or some other factor in the environment that is difficult to measure, then it will tend to be more visible from the other platform too, and abundance will be underestimated. These problems may be reduced by separating the areas of search for the two platforms, and using one to set up trials for the other. The resulting **binary data** may then be modeled using **logistic regression** [1]. In some studies, the platform that sets up the trials could be provided, for example, by a radio-tagging study, where locations of animals are known, or by an underwater acoustic array (so long as species could be identified accurately). In double-platform methods, Horvitz–Thompson-like estimators are used to estimate density, given the estimated probability of detection for each observation (*see* **Sampling, environmental**).

Spatial modeling of distance sampling data is potentially useful for several reasons: animal density may be related to habitat and environmental variables, potentially increasing precision and improving understanding of factors affecting abundance; abundance may be estimated for any subregion of interest, by integrating under the fitted spatial density surface; and a model-based approach allows data collected from nonrandom surveys (platforms of opportunity) to be used. One approach [7] is to conceptualize the distribution of animals as an inhomogeneous **Poisson process**, in which the detection function represents a thinning process. If, in the case of line-transect sampling, the data are taken to be distances along the transect line between successive detections, this allows us to fit a spatial surface to these data. We can refine this further by conceptualizing the observations as waiting areas, i.e. the effective area surveyed between one detection and the next, where the effective width of the surveyed strip varies according to environmental conditions and observer effort [7, 8].

Geographic information systems (GISs) are now widely available. This makes it possible to implement automated design algorithms that generate survey designs with known properties rapidly and simply. The software Distance has a built-in GIS and implements methods developed by Strindberg (unpublished). It can generate surveys based on a range of point- and line-transect designs, as well as performing simulations to compare the efficiency of different designs and to investigate design properties such as probability of coverage. For complex surveys in which coverage probability is not uniform, but has been calculated analytically or by simulation, Horvitz–Thompson-like estimators can be used to estimate abundance. This avoids the biased estimates that result from standard estimation methods, which assume that coverage probability is even. For example, ship-board surveys typically use continuous zig-zag survey lines, so that costly ship time is not wasted in traveling from one line to the next. For convex survey regions or strata, a design with approximately even coverage probability can be obtained by defining a principal axis for the design and adjusting the angle of the survey line with respect to this axis as the ship progresses through the area. By contrast, fixed-angle or fixed-waypoint zig-zag designs do not give even coverage probability unless the survey region is rectangular (Figures 5 and 6). If the survey region or stratum is not convex, a combination of splitting the region into a number of almost convex sub-regions and placing a convex hull around the sub-regions can be used.

Adaptive sampling [20] (*see* **Adaptive designs**) offers a means of increasing sample size, and hence increasing precision, by concentrating survey effort where most observations occur. Standard adaptive sampling methods can readily be extended to distance sampling surveys [20]. For example, for point transect sampling we can define a grid of points, randomly superimposed on the study region, and randomly or systematically sample from the grid to form the primary sample. When a detection is made at a primary sample point, points from the

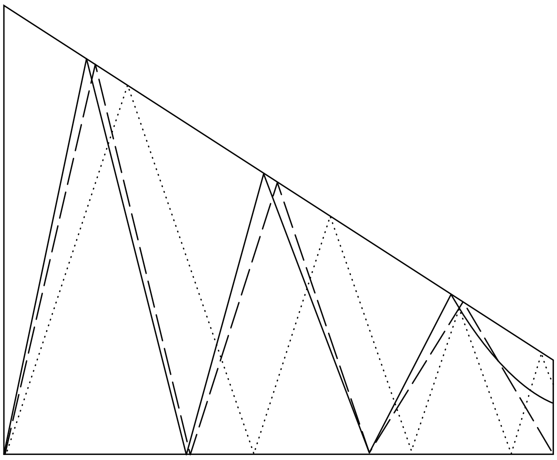


Figure 5 A trapezoidal survey region illustrating three zig-zag designs: equal-angle (dotted line); fixed-waypoint (dashed line); and even-coverage (solid line). The principal axis of the design is parallel to the base of the trapezium in this example, and for the fixed-waypoint design, waypoints are equally spaced with respect to distance along the principal axis, alternating between the top boundary and the base. Reproduced from Buckland, S.T., Thomas, L., Marques, F.F.C., Strindberg, S., Hedley, S.L., Pollard, J.H., Borchers, D.L. & Burt, M.L. (2001). Distance sampling: recent advances and future directions, in *Quantitative Methods for Current Environmental Issues*, V. Barnett, A. El-Shaarawi, C. Anderson & P. Chatwin, eds, Springer-Verlag, New York, Figure 8, by permission of Springer-Verlag

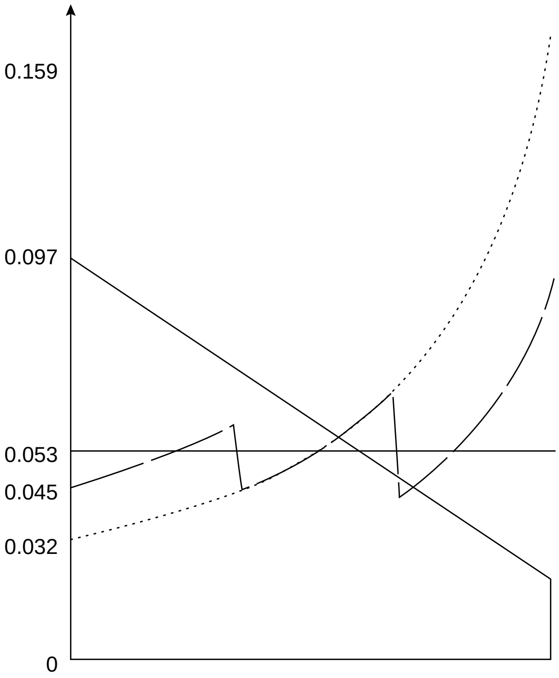


Figure 6 Coverage probability against distance along the principal axis for the three designs of Figure 5. Also shown is the height of the trapezium as a function of distance along the principal axis, which indicates that the fixed-angle design has too low coverage where the study region is wide, and too high where it is narrow. For the fixed-waypoint design, coverage probability changes at each waypoint, and between waypoints varies smoothly in the same manner as the fixed-angle design. Reproduced from Buckland, S.T., Thomas, L., Marques, F.F.C., Strindberg, S., Hedley, S.L., Pollard, J.H., Borchers, D.L. & Burt, M.L. (2001). Distance sampling: recent advances and future directions, in *Quantitative Methods for Current Environmental Issues*, V. Barnett, A. El-Shaarawi, C. Anderson & P. Chatwin, eds, Springer-Verlag, New York, Figure 9, by permission of Springer-Verlag

grid that surround the primary sample point are sampled. If detections are made at these extra points, then further sampling is triggered. A major practical problem of adaptive sampling is that the required survey effort is not known in advance. This is particularly problematic for shipboard surveys, in which the ship is available for a predetermined number of days. A method has been developed [13] that avoids this problem. When additional effort is triggered, the ship changes to a zig-zag (and hence continuous) course, centered on the nominal trackline. The angle of the zig-zag is a function of how far the ship is ahead or behind schedule. Unlike standard adaptive sampling, the method is not design-unbiased, but simulations indicate that the bias is small. An experimental trial on a survey of harbor porpoise in the Gulf of Maine yielded substantially more detections and better precision than did conventional line transect sampling [13].

References

[1] Borchers, D.L., Buckland, S.T., Goedhart, P.W., Clarke, E.D. & Hedley, S.L. (1998). Horvitz–Thompson estimators for double-platform line transect surveys, *Biometrics* **54**, 1221–1237.

[2] Borchers, D.L., Zucchini, W. & Fewster, R.M. (1998). Mark-recapture models for line transect surveys, *Biometrics* **54**, 1207–1220.

[3] Buckland, S.T. (1992). Fitting density functions using polynomials, *Applied Statistics* **41**, 63–76.

[4] Buckland, S.T., Anderson, D.R., Burnham, K.P. & Laake, J.L. (1993). *Distance Sampling: Estimating Abundance of Biological Populations*, Chapman & Hall, London, reprinted (1999) by Research Unit for Wildlife Population Assessment, St Andrews.

[5] Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2001). *Introduction to Distance Sampling*, Oxford University Press, Oxford.

[6] Diggle, P.J. (1983). *Statistical Analysis of Spatial Point Patterns*, Academic Press, London.

[7] Hedley, S.L. (2000). Modelling Heterogeneity in Cetacean Surveys, PhD dissertation, University of St Andrews.

[8] Hedley, S.L., Buckland, S.T. & Borchers, D.L. (1999). Spatial modelling from line transect data, *Journal of Cetacean Research and Management* **1**, 255–264.

[9] Hiby, A.R. (1985). An approach to estimating population densities of great whales from sighting surveys, *IMA Journal of Mathematics Applied in Medicine and Biology* **2**, 201–220.

[10] Hilborn, R. & Walters, C.J. (1992). *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*, Chapman & Hall, New York.

[11] Marques, F.F.C., Buckland, S.T., Goffin, D., Dixon, C.E., Borchers, D.L., Mayle, B.A. & Peace, A.J. (2001). Estimating deer abundance from line transect surveys of dung: sika deer in southern Scotland, *Journal of Applied Ecology*.

[12] Plumptre, A.J. (2000). Monitoring mammal populations with line transect techniques in African forests, *Journal of Applied Ecology* **37**, 356–368.

[13] Pollard, J.H. & Buckland, S.T. (1997). A Strategy for Adaptive Sampling in Shipboard Line Transect Surveys, Report of the International Whaling Commission **47**, 356–368.

[14] Ramsey, F.L., Wildman, V. & Engbirng, J. (1987). Covariate adjustments to effective area in variable-area wildlife surveys, *Biometrics* **43**, 1–11.

[15] Schwarz, C.J. & Seber, G.A.F. (1999). A review of estimating animal abundance III, *Statistical Science* **14**, 427–456.

[16] Seber, G.A.F. (1982). *The Estimation of Animal Abundance and Related Parameters*, 2nd Edition, Macmillan, New York.

[17] Seber, G.A.F. (1986). A review of estimating animal abundance, *Biometrics* **42**, 267–292.

[18] Seber, G.A.F. (1992). A review of estimating animal abundance II, *International Statistical Review* **60**, 129–166.

[19] Thomas, L., Laake, J.L., Derry, J.F., Buckland, S.T., Borchers, D.L., Anderson, D.R., Burnham, K.P., Strindberg, S., Hedley, S.L., Burt, M.L., Marques, F.F.C., Pollard, J.H. & Fewster, R.M. (1998). Distance 3.5. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK. `http://www.ruwpa.st-and.ac.uk/distance/`.

[20] Thompson, S.K. & Seber, G.A.F. (1996). *Adaptive Sampling*, Wiley, New York.

[21] Thompson, W.L., White, G.C. & Gowan, C. (1998). *Monitoring Vertebrate Populations*, Academic Press, San Diego.

[22] Wilson, K.R. & Anderson, D.R. (1985). Evaluation of a density estimator based on a trapping web and distance sampling theory, *Ecology* **66**, 1185–1194.

(*See also* **Ecological statistics**)

LEN THOMAS, STEPHEN T. BUCKLAND,
KENNETH P. BURNHAM, DAVID R. ANDERSON,
JEFFREY L. LAAKE, DAVID L. BORCHERS &
SAMANTHA STRINDBERG

Impacts and management of wild horses in the Australian Alps Submission

By Joanne Canning, BSc Equine Sports Science, HND Equine Sports Coaching,

BHSII **Summary of points**

Reference a)

- Lack of independent appropriate research and review of some research papers currently used against the wild horses
- Biologically impossible population estimates
- Threatened Fauna the true story
- Threatened Flora the true story
- Short review of some papers quoted regularly in documents not in support of wild horses
- Positive impacts of wild horses
- The real causes of damage
- Management techniques, adherence to Standard Operating Procedures and humaneness assessment

Reference b)

- Obligations

Reference c)

- Funding

Reference d)

- Repair, restore and avoid true damage

Introduction

The wild horse management issue has been going on for many years now.
To resolve the issue there has to of course be compromise on BOTH SIDES.

Wild horse advocates have tried to offer many compromises - in fact anything that AVOIDS LETHAL MANAGEMENT. This includes - removals for rehoming, relocating to less sensitive areas, contraceptive population control and Reserve Design. But as these are not the cheapest or easiest options they are pushed to one side in favour of only one option of lethal control.

Of EXTREME IMPORTANCE is to first of all TRULY establish the effects of the wild horses in this environment. Also of EXTREME IMPORTANCE is not to lay the blame on the wild horses for damage done by other species! And also of EXTREME IMPORTANCE is to be transparent, honest and manage the issue with integrity and empathy because of the sensitivity of the issue and also the fact that horses are of course sentient beings. This is now scientifically very well established.

So here are considerations that need to be taken into account :-

Lack of independent research studies done on wild horses in the

Australian Alps

“While studies on herbivory are widespread, there is less information specifically on the effects of horses. This is because controlled experimental studies are rare, and most rely on a correlational approach and are often complicated by the presence of other herbivores (Beever & Brussard 2000). Exclosure plots that exclude all grazing herbivores are likely to exaggerate the impacts of horses (Linklater et al. 2002). Some studies fail to find an effect, or may even find a positive impact (e.g. Fahnestock & Detling 1999).”
“Some methodological problems make interpretation difficult. Exclosure plots are often positioned to record impacts in very specific habitat types, which are not representative of damage across the range, and exclosure plots typically exclude other large grazers like deer (e.g. Linklater et al. 2002).”
“The ITRG concludes that there are significant knowledge gaps in our understanding of horses in KNP.”
“Surprisingly little is known about the ecology of horses in KNP. “

Reference
https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Animals-and-plants/Pests-and-weeds/Kosciuszko-wild-horses/final-report-independent-technical-reference-group-supplementary-wild-horse-management-plan-160221.pdf?la=en&hash=26B121837E28A2C68514B34D99A21C18F7557E16

“Research on feral horses is surprisingly lacking in Australia and would benefit from national leadership and direction. It should: accurately map the distribution and abundance of feral horses; quantify feral horse impact in relation to density and control; evaluate the humaneness and suitability of control techniques; document community and stakeholder perceptions on feral horse impact and management; and assess whether feral horse impacts threaten native species or communities.” “There is a scarcity of published peer reviewed research on feral horses in Australia. The relationship between feral horse density and damage remains to be quantified in any area, which impedes effective management. Much of the evidence on environmental impacts and population ecology is anecdotal” (36) “The assumption that an introduced animal is causing damage may not be correct. Feral horses in some places may in fact be beneficial. They may be reducing bushfire fuel loads by removing grass. They may be exerting no negative impact at all and simply providing pleasure for those that love seeing wild horses living free in the bush. Control in this case would be a waste of time and money.” (36)
“In general, although a significant amount of studies have been done on the effects of herbivores on the environment, there are still relatively few studies specifically about the effect of wild horses. (Beever, EA and Brussard, PF 2000, ‘Examining ecological consequences of feral horse grazing using exclosures’, Western North American Naturalist 60: 236–254.)” (36)

Establishing a truly accurate, biologically and scientifically possible population estimate :-

Population increase rates in the AANP from previous research

Population dynamics of wild horses in the Australian Alps National Parks, was studied in detail at several sites each spring and autumn, between 1999-2002. The sites were Big Boggy, Cowombat and Currango. The sites were chosen because wild horses could readily be found in these areas. They were independent (too far apart for the horses to move in-between) and there was no management done in these areas for the period of the surveys. The annual population increases for these 3 areas were found to be - Big Boggy – **7%**, Cowombat – **3%** and Currango – **9%**. (1)

Surveys have been done almost annually in the Big Boggy area of the Kosciuszko National Park using the mark-recapture method from 2006-2019. In 2012 the estimate of horses in this area was 195. In 2019 the estimate was 220. The annual population increase for this period therefore is approx 5% per year. (2)

Page 11 of the Kosciuszko National Park Draft Wild Horse Management Plan 2016 states –
"In Kosciuszko National Park, populations are likely to increase by between 6% and 17% per year". (3)

Page 12 of the Kosciuszko National Park Horse Management Plan 2008 states –
"The horse population can increase by up to 20% per year when conditions are good, but the population growth rate in Kosciuszko is expected to be closer to 8% (Dobbie and Berman 1992; NPWS 2003)." (4)

This follows many other scientific papers on this matter and previous and other reports by the NPWS.

Current population situation Kosciuszko National Park

In the new plan of management for the Kosciuszko National Park Wild Horse Population, maintenance of the wild horse heritage values has been based on leaving 3,000 horses in the areas of the Park that are designated as wild horse retention areas. (5)
The major problem with this is, that this has been based on a starting figure taken from the estimation done in the 2020 survey of 14,380 (6). This is therefore assuming reduction of 11,380 horses over the period of the plan. If attempt is made to remove this amount of horses there will actually be none left at all in the whole Park. This would of course not be in keeping with the legislation of 2018.

I would like to clarify this statement. I have concentrated on the North Kosciuszko region as this is where the majority of horses are - approx 85% according to the official 2020 survey by Cairns (6) and these are the horses that have been subjected to massive removals in 2020-2022.

Previous surveys, using the Distance Software (applied by Cairns) have produced estimates that are **scientifically and biologically not possible for the species.** For instance the reported annual increase of 37% (7), (which in actual fact calculates at 41% after accounting for removals and this was a major error when calculations were done) in the North Kosciuszko area between 2014 and 2019, which is approximately DOUBLE the accepted scientific maximum (refer to above normal rates of increase). This was pointed out in the peer review by St Andrews

University – **“The block specific finite rates of growth are likely at the centre of the 2019 report. The high rate of growth reported for the North Kosciuszko block are of particular interest as it appears to exceed published maximum growth rates for the species.**” (8)

No feasible explanation was ever given, because there isn’t one. This anomaly was also pointed out by the Scientific Advisory Panel in their final report to the Minister. – **“The rate of increase in north KNP between 2014 and 2019 is above the biologically possible rate of reproduction (Garrott et al. 1991**).” (9)

The only reason given as a possibility was movement of horses, but in the document - 2019 Australian Alps Feral Horse Aerial Survey: Summary Report (10), it states – “Feral horses (Equus caballus) also known as ‘wild horses’ or ‘brumbies’ occur in three large and currently **separate distribution areas** of the Australian Alps National Parks (AANP) and adjoining State Forest areas.”

In spite of this ludicrous estimate that is not biologically possible, the surveys have continued to be done in the same way. And, in fact, the purported annual increase in the Southern Kosciuszko region between 2020-2022 calculates at approx **95% - even more ludicrous than before and of course totally impossible!** Whereas in the Northern Region, the increase calculates at approx 6%, which is a realistic increase and a far cry from the once again ridiculous purported 41% annual increase from the previous surveys!
There is a similar situation with the kangaroo surveys, which have shown supposed increases of over 400% - also absolutely not biologically possible for the species. This was brought up in the recent Kangaroo Enquiry AND YET, when doing both the new trials of survey methods in the localised areas of planned removals in 2020 and the new official published surveys again done by Cairns in 2020 and 2022, the choice was made to use this software yet again, in the full knowledge that it is producing scientifically and biologically impossible purported annual increases and therefore numbers. I have personally written to NPWS, the Minister, all NSW Members of Parliament and St Andrews University in Scotland pointing out this very serious problem. So they have all been made aware of this.

The scientific advisory panel recommended doing more localised counts and trialling different methods of counting. They recommended doing this prior to removals starting and then regularly monitoring the population during removals. The localised counts done prior to removals were done only in the 3 areas planned for removals and the opportunity was therefore available to conduct a real head count, rather than using this software again, which had already proven to show biologically impossible estimates. However, once again, the “Distance Software “ was applied to elevate the numbers!

Comparison of numbers estimated by the previous Cairns survey and the trial surveys done in 2020, only go to show how ludicrous the results being produced are. I have made comparison here of some of the supposed population estimates in the 3 areas that were designated for removals under the post fire wild horse control.

So, in the 3 blocks where removals have taken place- (figures obtained in GIPA release)(11)(12)

Zone 1 - Nungar
Cairns estimate was **2191 horses**
The SAP estimated a significantly lower amount of 700 in the area.
However results from the new helicopter survey estimated only **163 horses**
And the helicopter survey with thermal imaging estimated **135 horses**

Zone 2 - Cooleman Plain
Cairns estimate 2464
SAP estimated 3600
Helicopter normal estimated 2400
Helicopter with thermal imaging- 1630

Zone 3 - Kiandra
Cairns estimate **2707**
SAP estimate 320
Helicopter normal - **824**
Helicopter with thermal imaging- **1051**

Comparison of estimates from the localised counts done in the 3 areas of removals

	No of clusters seen	Average cluster size No of horses seen Estimated Increase by computer software
Cairns		7362
Helicopter	122	3.8 **463** 3387 630%
Helicopter with Thermal Imaging	72	2.88 **207** 2816 1250%

On the “normal helicopter” survey, they saw 122 clusters of horses. Average cluster size- 2.9-4.3. So actually saw approx. 463 horses. In all 3 areas. The rest were computer generated. Elevated to 3,387
On the helicopter with thermal imaging, only 72 clusters were seen. Approx. 207 horses. Elevated to 2,715
UAV survey on Kiandra, spotted 32 clusters - approx. 106 horses. Cairns total for these areas **7,362! More than double the other surveys that still used the Distance Software**. Absolutely ludicrous and more proof that using the software to estimate these numbers is not working by any stretch of the imagination. A measure

of the scientific soundness of these surveys has to of course be based on whether the results are even biologically possible for the species. The results are not biologically possible and therefore this should have been taken to **realise that these surveys are not working as they are not scientifically sound.**

All surveys conducted with helicopter would involve double counting because, for the most part, horses will most certainly run from one transect to another. This was also brought up in a previous peer review by St Andrews. One assumption of the software is of course No Movement.

See below for horses actually seen in the 2014, 2019, 2020 and 2022 surveys with percentage increases by use of the software and comparison to the headcounts done annually by Parks. (6, 7, 13)

Comparison of numbers of horses actually seen to survey estimates and headcounts - Northern Kosciuszko Region

	No of clusters seen	**Average cluster size**	**No. of Estimate Increase by NPWS horses computer actual seen software headcount**
2014 Cairns	**104**	**3.55**	**369 3255 970% 1637**
2019 Cairns	**269**	**4.82**	**1298 15687 1110% 3120**
2020 Cairns	**164**	**4.41**	**723 12511 1630% 2468**
2022	**288**	**4.36**	**1255 12714 913% ?**

Please take note of the figure from the **NPWS actual headcount (spotting count) done in 2020. (see below**). This showed a figure of 2468, before drastic removals! Not 12,511!

As it is stated in the Cairns survey document done in 2020, that 85% of the horses in the Park are in the Northern Region, it is absolutely clear that there are already less than the 3,000 horses proposed to remain in the Park.

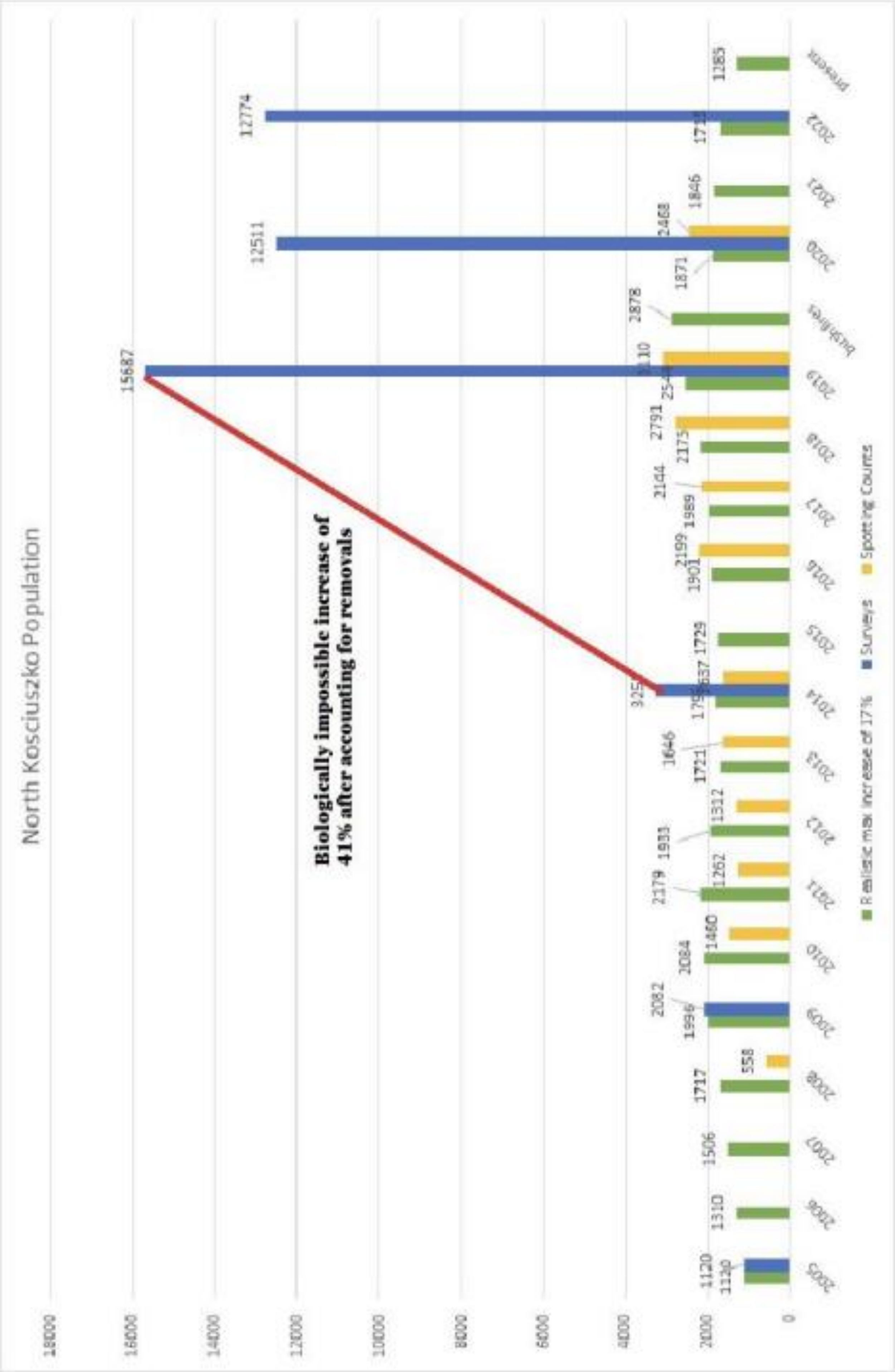
A helicopter spotting count was undertaken in 2021, in the Northern Region of the Park by the Snowy Mountain Brumby Sustainability and Management Group at their own expense. They counted 838 horses in this area of the Park. Although clearly some may not have been spotted, it would be impossible for them to have missed 11,673 horses!

I have calculated in the below table a realistic wild horse population estimate for the Northern Region, starting with the population figure taken from the well-respected survey done in 2005 by Montague Drake, that is quoted in many official documents. I have then calculated annual population increases of 17% per annum which is at the upper end of the scale of recognized scientific increases for wild horses and much higher than the previous estimations of increase shown above for the Kosciuszko National Park. I have then compared this to some of the more recent results from the annual spotting counts done by NPWS. It’s absolutely clear that the helicopter spotting counts already undertaken annually, are giving far more realistic estimates of numbers than the Distance Software surveys.

Calculations made starting with the results of the 2005 survey done by Montague Drake and then using a realistic scientific wild horse population increase per year of 17%

Year	start pop	17% inc	Removals	End Pop
2005	1120	190	0	1310
2006	1310	223	27	1506
2007	1506	256	45	1717
2008	1717	292	13	1996
2009	1996	339	251	2084
2010	2084	354	260	2179
2011	2179	370	616	1933
2012	1933	329	541	1721
2013	1721	293	218	1795
2014	1795	305	372	1729
2015	1729	294	122	1901
2016	1901	323	235	1989
2017	1989	338	152	2175
2018	2175	370	0	2544
2019	**2544**	433	99	2878
bushfires	**2878**		1007	1871
2020	**1871**	318	343	1846
2021	**1846**	314	444	1715
2022	1715	292	722	1285
2023	1285			

For a more visual effect of how ludicrous the survey numbers are, I have done a graph with the computed increases at 17%, plus the results from the annual spotting counts and compared to the “official” survey results done by Cairns.(see below).

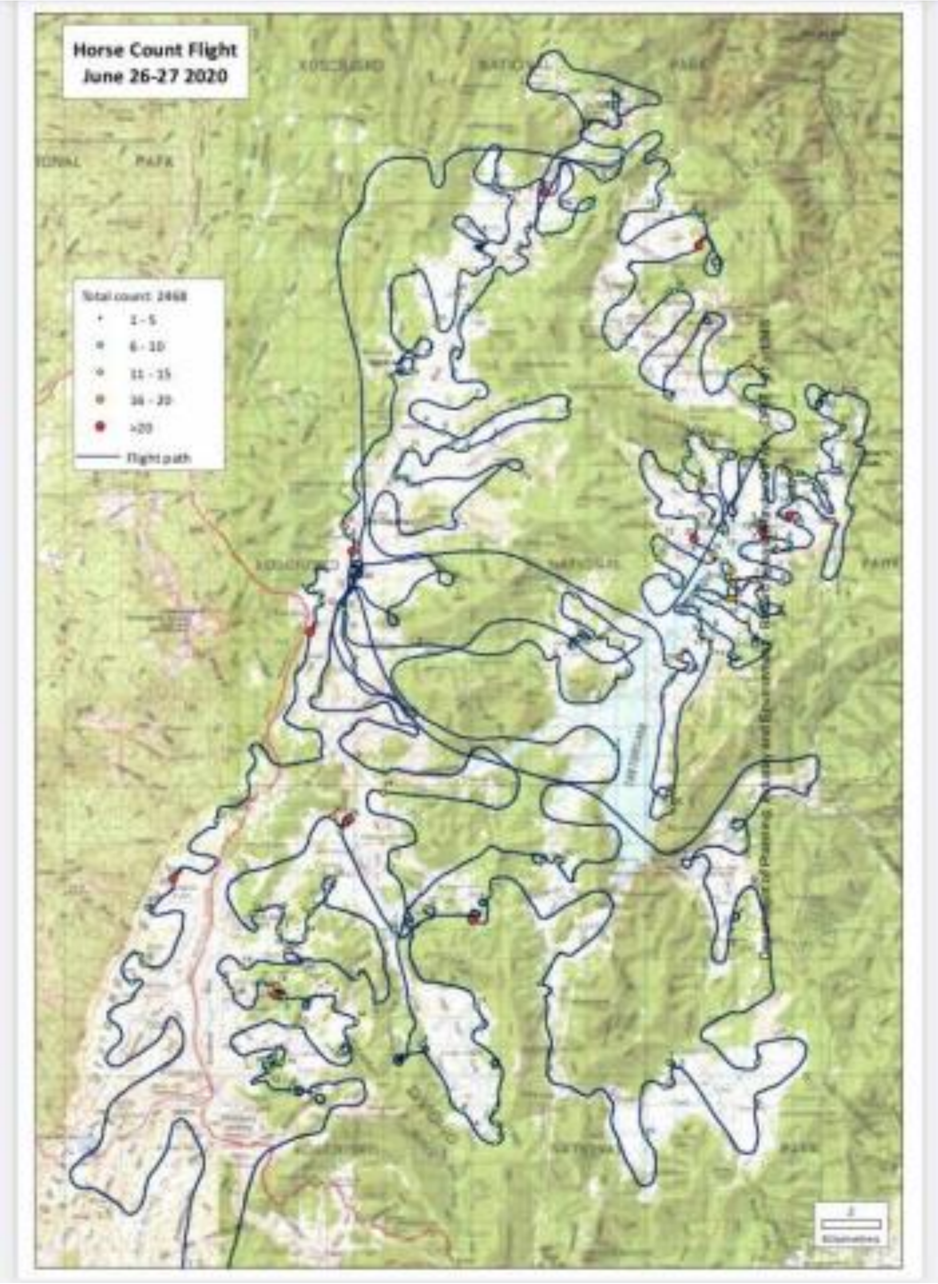


It's clear from the above, that before ANY further management is undertaken, it is absolutely imperative to know more accurately, how many horses are currently remaining in all of the Parks. It is certain that there is already less than 3,000 horses remaining in the Kosciuszko National Park.

A new ACTUAL headcount therefore needs to be undertaken as soon as possible, on both the Kosciuszko National Park and the Alpine National Park, Victoria and the Barmah National Park (where similar ludicrous estimates have been made using this software.This could be done using the same helicopter spotting count that is done every year by NPWS, BUT THIS MUST INVOLVE WILD HORSE ADVOCATES, as recommended by the Scientific Advisory Panel. To get a fair and impartial count, there definitely has to be people from both sides involved. As we know that 85% of the horses are in the Northern Region of the KNPark and these spotting counts have been successfully undertaken every year, we know it is possible to do this successfully and then it is easy to estimate the other 15% in the rest of the Park.

All surveys done using the Distance Software and other similar equations or computer modelling software should be halted, because, as shown by all of the above, they have given ludicrous estimates that are not scientifically or biologically possible and with estimates that are varying so enormously that it is absolutely ludicrous to continue with them. Not to mention the waste of funding involved!

Both the ITRG and the CAP and SAP always recommended involvement of the community in all and any population surveys. This has never been done and needs to be remedied immediately. Deliberate continued use of methodology that is producing estimates that are not biologically or scientifically sound is irresponsible of course. Especially when the lives of possibly thousands of wild horses are involved.



Results of annual headcount in Northern Region of Kosciuszko National Park 2020

Victoria National Parks

There is a similar situation with the assessment of population numbers in the Victoria National Parks. As follows:-

Barmah National Park

In the Barmah National Park, for many years, numbers were reported as being fairly constant. For instance - from 2006 to 2017 the following figures were given:- 2006 "100 – 150 horses occupy the Barmah State Park" (cited by Dawson et al. 2006)(36)

2010 - "Mr McCormack said that in the aerial survey four mobs of horses were observed, the largest being a mob of 15. He said the number of horses in the park was greater than the 38 counted so far. ''It's really hard to quantify it without some proper verification, but we think it's probably in the order of around 150,'' he said." (46)
2012 - 144 horses counted. (47)
2017 - Barmah National Park east of Echuca on the Murray River has a currently

estimated population of over 200 animals.(48)
June 2017 - counted 134 horses
June 2018 - counted 147 horses - covered 80% of the Park (other 20% very low population). Using Distance Software this number altered to 620-730 horses. This is "horses assumed unseen".(49)
June 2019 - again 80% of Park surveyed as above. Estimated 540 horses using Distance Software. Actual number of horses seen was not reported! September 2019 - again 80% of the Park surveyed but this time with transects only 300m apart! Horses counted were 282 horses.(50)
There was a further survey done in 2021!!! Results STILL NOT PUBLISHED TO DATE!!! It's now 2023 of course!

Observations
The Distance Software is the software used for the surveys of the AANP. Results from these surveys have produced figures that are biologically and scientifically impossible. As mentioned previously, estimates show computed annual population increases of 41% and even 95%!!! The scientific accepted maximum being 20%.

In the September 2019 survey, the transects flown were only 300m apart! So ALL horses should have been able to be counted in this survey. But as the transects were then so close, there is an extremely high chance that horses would be double counted because of movement from one transect to another! So it's extremely likely that the 282 would actually be an OVERESTIMATE!

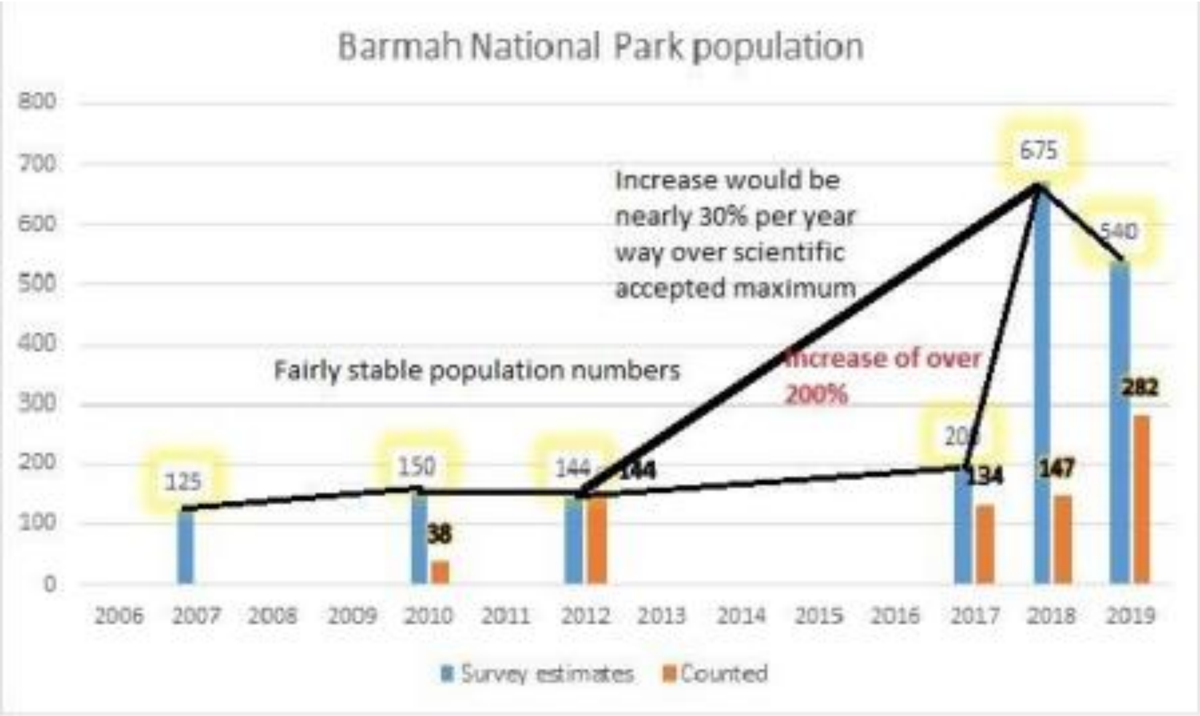
Using a figure of 200 for 2017 - increasing by 17% per year (expected population increases would be between 6-17%) but allowing for the deaths of horses in the 2018/2019 disaster, the total population would be a MAXIMUM OF approx 220 horses in 2022, BEFORE the shooting of many by Parks Victoria contractors (in spite of there being a purpose built sanctuary awaiting them and offers to actually trap and remove the horses to go to live in this approved sanctuary!) and the flooding of the Park that caused the death of many wild horses. During this flooding the wild horses ended up congregating on a small "island" area that remained dry. This was of course the ideal opportunity to see that were under 200 wild horses present in the Park.
It is extremely unlikely that the population increases would be more than 17%! Especially in the Barmah Park where conditions can be hard with flooding and droughts. Parks own papers give the figures of 6-17%. This would be scientifically accepted increases.

It is not a difficult task to use drones to locate and video approx 200 horses. As Parks are of course well aware of most likely locations of them. This would provide visible proof to all parties involved.

Visible video proof of the FLIR counts done in 2017-2019 has never been provided and determining the differences in animal species with Infra Red in tree covered areas would be EXTREMELY DIFFICULT - especially from a higher altitude of a helicopter rather than drones. Viewing the footage taken would at least allow us to see if species differentiation is perfect. Or deer could easily be mistaken for a horse.

There are local companies able to do this using drone and video technology, which would give much better chance of species differentiation.

To get from approx 200 horses in 2017 to 540 in 2019 after accounting for losses in early 2019, would be an annual population increase of OVER 60% per annum!!!!!! Beyond ludicrous!



ALPINE NATIONAL PARK - VICTORIA

In 2003 the estimated population in the ANP was 1067.
In 2009 the estimated population in the ANP was 3442 (51)
This involves an increase in population per year of approx 24.5% when accounting for removals. (once again above the scientific accepted maximum.)

The 2014 survey estimated population however was 2350. Therefore a significant **reduction in population**.

In 2019 the estimated population was approximately 5,000. This involves an increase of approx 22%. The most recent survey, done in December 2021(but only published in 2023!!) - The estimate is 2456, which is of course a dramatic reduction! The reason given for this NOW is the 2020 wildfires. Although it had been reported previously that not many horses were lost in the wildfires! There were 78 clusters at 3.36 estimated cluster size = 262 Brumbies ACTUALLY SEEN (52)

So according to these surveys, the population increased for 6 years at 24.5% per annum (higher than scientific accepted maximum and the mean is 12%). Then decreased significantly over the next 5 years (no wildfires). Then suddenly increased over the next 5 years again at approx 22%. This is absolutely not feasible! Although there may have been significant losses from the wildfires in 2020, considering the official reports of there being not so many, it is more likely that the estimates from

2019 were way exceeding the actual numbers in the first place!

Bogong Region of the ANP

In 2021 - Having flown in a helicopter at 150m high, flying transects only 1km apart, they carefully HEADCOUNTED AND PHOTOGRAPHED - 121 Brumbies in the Bogong High Plains. Seeing and exactly head counting 121 horses with photo ID is of course unbelievably easy and would ensure no double counting -some were seen twice as it was over 2 days but these were then discounted. So, a perfect headcount with photographic proof! 121 Brumbies! (53)

But, in spite of doing this very accurate assessment, they decided to apply yet another “estimator “ !! And then the estimated population number became 252 Brumbies!!!! More than double the ACTUAL RECORDED NUMBER of 121!!!

Flying transects only 1,000m apart (500m either side) from a helicopter would enable ALL WILD HORSES TO BE SEEN AND COUNTED ! And the helicopter hovered over the groups of horses to enable them to take photographic evidence and then they more than doubled the number!
This has ZERO TO DO WITH SCIENCE!

They even state that this would be an annual increase of 29% after accounting for removals. Once again - the scientific accepted maximum is 20%. So once again publishing papers with ludicrous estimates that are way over the scientific accepted maximum!

These discrepancies giving ludicrous estimates of populations going from biologically impossible increases to significant decreases (without a wild fire incident) only go to prove that there is a MASSIVE PROBLEM with using this software or any similar software! Headcounts are absolutely possible- as has been shown in the northern region of Kosciuszko National Park and the Bogong region of the ANP. So headcounts ONLY should be used for further assessment of population numbers!!

Threatened Fauna and Flora - the true story

I absolutely dispute the inference that the wild horses are threatening to the native Fauna and Flora in the Park. The threats are from other major contributors! As below: –

Threats to Corroboree Frog

Corroborate Frog threats -

- chytrid fungus
- climate change
- droughts
- wildfires
- predation by carnivores

“The spread and persistence of chytrid fungus in the population is facilitated by a species living alongside the Corroboree Frog, the Common Eastern Froglet (Crinia signfera). This species appears to sustain high infection levels, but doesn’t develop the disease. As a result, it acts as a reservoir host, sustaining the disease in the ecosystem and allowing transmission to other species.”

“An additional threat to the Southern Corroboree Frog is climate change. Reduced precipitation and warmer temperatures are likely to eventually affect breeding pools and vegetation around them. Droughts already result in egg and tadpole deaths, and as the frequency of droughts increases with climate change, the capacity for the Southern Corroboree Frog to recovery greatly reduces.” (38)

“There are few peer-reviewed studies of the impacts of feral horses on ecosystems in this region.” (39)

“We surveyed the scene, calling out: “Hey, frog!”. At ponds not severely burnt, reasonable numbers of northern corroboree frogs responded. At badly burnt sites where frogs had been found for 20 years, we were met with silence. The adults there had likely died.” “After the fires, heavy rain in denuded burnt catchments produced water runoff laden with sediment. Some frog breeding habitat was eroded and filled with silt and ash. Once-mossy ponds were now gravel and ash.” (40) “They contained a fascinating series of photos. Some revealed how a number of ponds largely escaped the fires, only to be destroyed afterwards by flooding.” (40)

The Smoky mouse threats

“Major threats to the species include predation by introduced carnivores, habitat changes due to altered fire regimes and dieback caused by the Cinnamon Fungus (Phytophthora cinnamomi), and loss, modification and fragmentation of habitat due to road construction and intensive timber harvesting.” (41)

The Broad toothed rat threats

Main threats -

- Predation by wild cats and foxes
- competition and grazing by rabbits
- Competition, disease transmission and habitat degradation by wild Pigs
- Catastrophic fire events
- Global warming causes loss of snow cover resulting in increased exposure to foxes and cats.
- Climate change resulting in loss of sub-alpine and alpine habitat, and; spread of the plant root fungus Phytophthora cinnamom
- Habitat loss, fragmentation and degradation from roads, ski runs, buildings and recreational activities. (42)

The Alpine Skink

Main Threats –

“Wildfire has the potential to eliminate the species”
“Historically, large tracts of habitat have been lost as alpine resort villages have been constructed and expanded. Construction of dams has destroyed habitat that was almost certainly occupied by the species. Concurrent development of infrastructure such as roads, tracks and ski runs have also destroyed and fragmented habitat. Development of ski runs may have a greater than expected effect on habitat for Alpine She-oak Skinks, as it is more favourable to build ski runs in large, continuous grassy areas that provide a uniform surface. These large grassy areas are the optimal habitat of the Alpine She-oak Skink. “
“Predation by Rats (Rattus rattus), Foxes (Vulpes vulpes), Cats (Felis catus) and Wild Dogs (Canis lupus familiaris) is a current threat.”
Also mentioned are weed invasion, climate change and trampling of habitat by ALL animals. (43)

Threatened Flora

Results from surveys did not reveal a significant effect of wild horse grazing on plant community composition, species richness, diversity, evenness, or dominance. And the effects of horses did not vary by site, indicating that different precipitation levels do not drive differences in grazing effects (27) (29)

In fact plant species richness was higher in horse grazed compared to ungrazed areas. Butterfly and bumblebee habitat use, as well as feeding and resting activities were also higher in grazed areas. (28)

A study showed that horse grazing resulted in overall increases in heather and herbaceous cover and decreases in gorse cover and height, with scarce differences among vegetation types. Floristic diversity increased more over time in grazed than in ungrazed paddocks. . Some herbaceous species characteristic of heathlands were favoured by horse grazing. Horse grazing reduced gorse dominance, controlling excessive accumulation of combustible material and **REDUCING FIRE RISK**, and promoted the presence of species of conservation interest, so is a promising management tool for restoration of heathlands and their biodiversity.” (26)
“The peatlands have been badly affected by cattle grazing and fire since European settlement of the region commenced in 1823. Some recovery is evident within

protected areas but serious fires in 2003 reversed this process in many areas." (30)

"These bog communities are also critical habitats for species such as the endangered Northern Corroboree Frog and rare Broad-toothed Rat." "**CLIMATE CHANGE** is going to have a significant impact on the Sphagnum bogs and fens, through drought, increased temperatures and increased incidence of wildfire." (32)

"As far as the possible role of endozoochory for conservation of plant diversity in grassland is concerned, the results emphasise the importance of large herbivores as potentially strong seed dispersal vectors." (30)

Review of some papers purporting horse damage

Reference to:
An Assessment of Feral Horse Impacts on Treeless Drainage Lines in the Australian Alps - December 2015 (wordpress.com)

"The proportion of horse-present sites that showed evidence of recent fire was lower than that for horse-free sites (46% versus 76%)"

Therefore showing that sites where horses are present are less likely to be prone to fire.

"Two vegetation-related variables were assessed; projected foliage cover and the proportion of foliage cover that is native. No significant differences were detected among horse-present and horse-free sites for either of these variables"

NO DIFFERENCE in vegetation cover and proportion of native foliage.

"To remove the potential influence of wombats, macropods and exotic grazing and browsing mammals on soil, stream stability and vegetation cover, the data were re-analysed excluding sites where evidence of the presence of any of these species was detected."

The influence of wild pigs and deer were therefore not removed.

"Stream channel width (m)Mean – signs of horses not present 1.0 Signs of horse presence 0.9"

As above – the influence of deer and wild pigs not taken into account. Horse presence does not mean horse damage.

Reference to:
Assessment of Impacts of Feral Horses (Equus caballus) in the Australian Alps (wordpress.com)

"The Experimental Monitoring Programs"
"Two studies were established at each of two sites, as described below. The aims of these experiments are to compare the effects of removal of grazing with continued grazing by feral horses on floristic composition and structure of favoured grazing areas (grasslands), and on bank condition and disturbance of two small streams at the study sites" "Methods"

"Replicated exclosure experiments were established at two sites (Cowombat Flat and Native Cat Flat). These sites were selected by the AALC and Friends of the Cobberas as areas that support permanent populations of feral horses but are not currently grazed by cattle (although both have been grazed by cattle in the past)." "fenced to exclude horses but not other grazers (rabbits, wombats, macropods), and unfenced so that grazing is unrestricted."

To show differences in areas excluding horses, the exclosure plots would HAVE TO EXCLUDE ONLY HORSES! Very small animals like rabbits and wombats were not excluded but other large grazing animals like DEER AND WILD PIGS were excluded! - mentioned in the ITRG report

Reference to:
2013 Observations of Pest Horse Impacts in the Australian Alps (sqspcdn.com)

"This "Observations" Report is produced for general information and is a record of personal observations made by the authors for the Mt Pilot area of Kosciuszko National Park in 2013. "

***Observations only. And again using exclosure plots and therefore excluding all larger animals including deer and pigs therefore damage that cannot be attributed only to horses*.**

Reference to:
feral-horse-impact-thesis.pdf (wordpress.com)

"Exotic species colonised tracks, but not at the expense of the native species."

Very old paper from 1990. Much of the paper is concerning dung piles relative to population but then also concentrates on study of tracks. Native species of flora even on tracks not affected !

"However, when crossing streams they avoided the Sphagnum in favour of the more solid ground under grassy and herbaceous vegetation, if this was available."

Confirming what we all know that horses will avoid these soft areas if at all possible!

"Soil is often in a highly compacted state on tracks used by horses or other animals."

No differentiation between horses and other animals, therefore cannot be associated only with horses! Refers to studies done with shod domestic horses to show trampling effects on tracks! Tracks frequented not only by many different animals but would be used by humans also!

"However, in the current study, many of the pre-European Sphagnum Bogs would have already disappeared because of cattle grazing which was phased out in the early 1970's, leaving the open herbaceous and grassy areas."

Speaks for itself!

Reference to:
feral-horse-exclusion-plot-monitoring-and-analysis_final_1.pdf

(wordpress.com) "However, these impacts cannot be separated from those of

deer"

Again the use of exclusion plots that excluded all of the larger animals including deer and pigs! The paper is supposed to be concerning horse exclusion but does not only exclude horses in these plots!

Reference to:
feral-horse-exclusion-plot-monitoring-and-analysis_final_1.pdf (wordpress.com)

"Our study is, to the extent of our knowledge, the first that has used BFAST to generate fPAR derivatives to explore the link between feral herbivore impacts on the phenology of vegetation."

Again a study done on the impact of all herbivores, where horses were present but not necessarily the cause of the damage. Deer and in fact pigs again would also be present in these areas!

Reference to:
Assessing the Impacts of Feral Horses on the Bogong High Plains, Victoria (amazonaws.com)

Again only observation study assuming that anywhere that horses were present, that horses were doing any damage seen! Any one of all the wildlife population in this area could be the cause of any damage seen!

Positive impacts of Wild Horses

There has been increasing scientific study worldwide showing the ecological benefits of wild horses and they are being used for regeneration of degraded ecosystems including forested and alpine areas. Wild horses reduce bushfire fuel loads. Research has shown that when herbivore populations deplete, catastrophic wildfires take over because excessive amounts of ground fuels act as kindling. There was scant evidence of fire until the megafauna disappeared. (14)(15)(16) Native animals and their habitat have been decimated by fire more than anything else. And waterways adversely affected. Not to mention the damage to human's health and property.

Wild horses reseed native grasses by intact seed dispersal in their droppings (17), are soil builders/fertilisers of soil via their droppings, which build the humus content of soil, allowing soil to gain more texture and retain more water. Hence ground water tables are replenished, feeding more seeps and springs more continuously. They create more abundant and dependable water sources in dry areas and break up ice and snow, helping other wildlife to survive.(14)

Large wild herbivores are crucial to ecosystems and their removal has a cascading effect on other species. Evidence from Australia suggests that rainforest was converted to sclerophyll vegetation in the aftermath of the loss of megafauna. (29) Herbivores greatly accelerate the nutrient cycle in ecosystems by returning nutrients to soil at high rates. (7)

The damage is being done by other species and blamed on the wild horses

Some research and live video supporting the positive impacts of Wild Horses and disputing the negative effects

My Big Backyard Productions - YouTube

Experimental rewilding enhances grassland functional composition ...

https://www.horsetalk.co.nz/2020/09/08/justice-equine-scapegoats-australia-brumbydebate/?fbclid=IwAR1cWANnOymexy-8uDP_LwamBM_jw3heJBX-7r8p-_5BJMKwtufhrsAQl4Y

Horses keystone species regeneration earth

Spotlight on the overlooked role of horses as carbon sequesters

Brumbies can fill a useful role in Australian ecosystems, says ecologist

https://www.eurekalert.org/pub_releases/2020-03/uots-cal031820.php?fbclid=IwAR0Dm8iIt6omw_KHIMsJs_d2mz4ej3zN14HPI4bD8kzwWQW5ljoRTfbElkU#

https://earthledger.one/feral-horses-gallop-to-the-rescue-of-butterflies-in-

distress/ In defence of australias brumbies

Interest payments in wild horses | Rewilding Europe

https://rewildingeurope.com/search/wild+horses/page/1/

Rewilding horses in Europe. Background and guidelines - a living ...

Wild Horse Fire Brigade

https://mobile.twitter.com/BBCSpringwatch/status/1285549837308645377?s=07

A Geographic Assessment of the Global Scope for Rewilding ... - PLOS

https://m.facebook.com/groups/1165599150138501?view=permalink&id=2442680152430388

https://m.facebook.com/groups/1165599150138501?view=permalink&id=2459073827457687

https://snowybrumby.wordpress.com/2014/08/26/brumbies-facts-you-should-know/

https://www.bellingencourier.com.au/story/5425985/letter-brumbies-in-national-parks a-new-perspective/

The real causes of damage - OTHER ANIMAL AND HUMAN DAMAGE

In spite of the millions of other "introduced" animals like deer, pigs, wild dogs, goats and rabbits who are well known to do massive damage, the wild horses are singled out for elimination.

Impacts to the environment by humans include building of dams, tourism, recreational vehicles, ski runs and facilities and climate change. And yet the wild horses are the scapegoats. Most ecologists and scientists confirm global warming to be the main cause of species decline and extinction.

Wild pigs damage

Wild pigs cause serious habitat degradation by rooting in the soil in search of food, dispersing seeds of weeds, regular wallowing and digging of dust-beds can impact on terrestrial and aquatic systems through erosion, siltation and increased turbidity.

“While searching for populations of threatened plant species in the Kiandra area between 1999 and 2001, we noticed considerable damage to sub-alpine treeless vegetation by pigs. Most damage was recorded in dry grassland communities and was evidenced by denuded circles up to 20 m in diameter. Some of these bare circles appeared to have been scoured more than once, judging by the varying amount of regeneration within them.”

“A group of 15 pigs and piglets was observed in Nungar Plain during the survey. Damage to vegetation by pigs is obvious and extensive. Herb-rich grassland communities are the worst-affected. Rooting is localised but very thorough.” Reference https://www.researchgate.net/profile/Neville_Walsh/publication/237372947_The_flora_of_Nungar_Plain_a_treeless_sub-alpine_frost_hollow_in_Kosciuszko_National_Park/links/53e166810cf2d79877a952b1/The-flora-of-Nungar-Plain-a-treeless-sub-alpine-frost-hollow-in-Kosciuszko-National-Park.pdf?origin=publication_detail

“Riparian and wetland habitats are attractive to feral pigs where they can cause serious habitat degradation by rooting in the soil in search of food, and can also prey on the eggs and chicks of nesting water birds in wetlands.”

Reference

https://www.parks.vic.gov.au/-/media/project/pv/main/parks/documents/managementplans/barmah-national-park-and-barmah-forest-ramsar-site-strategic-action-plan-2020-2023.pdf?la=en&hash=E130D0B2C8E9C9E22B349EC9148F9AA99ED09BAA

“Feral pigs disperse seeds of weed species, and in the process of rooting up the ground they trample vegetation and extensively disturb the soil. In addition, regular wallowing and digging of dust-beds can impact on terrestrial and aquatic systems through erosion, siltation and increased turbidity.”

Reference:-

https://pestsmart.org.au/wp-content/uploads/sites/3/2020/06/Managing-vertebrate-pests-feral-pigs.pdf https://www.environment.nsw.gov.au/Topics/Animals-and-plants/Threatened-species/ NSW-Threatened-Species-Scientific-Committee/Determinations/Final-determinations /2004-2007/Predation-habitat-degradation-disease-transmission-by-feral-pigs-key-thr eatening-process-listing

Deer damage

Deer are in plague proportions and much of the damage purported to be wild horse damage is in fact from deer. Research methodology has meant that no distinction has been made.

"Deer degrade ecosystem quality through grazing, browsing and trampling of vegetation, ringbarking trees, as well as dispersing weed seeds and enriching nutrient levels. They also cause soil disturbance in creeks, wetlands and swamps, where they wallow in mud."

Reference:-
https://www.parks.vic.gov.au/-/media/project/pv/main/parks/documents/managementplans/barmah-national-park-and-barmah-forest-ramsar-site-strategic-action-plan-2020-2023.pdf?la=en&hash=E130D0B2C8E9C9E22B349EC9148F9AA99ED09BAA

"Over a million deer are wreaking havoc in Victoria's state forests and national parks, and instead of being managed as a serious pest, deer are oddly protected under the Wildlife Act 1975 in order to support hunting interests."

Other papers/articles relevant to damage by Deer

https://www.abc.net.au/news/2017-07-21/harrietville-project-looks-at-how-to-deal-wit h-deer/8732414

https://www.abc.net.au/news/2017-03-31/deer-hunters-cull-sambar-deer-in-alpine-national-park/8396774

https://www.theage.com.au/national/victoria/feral-deer-damaging-alpine-national-par k-to-be-culled-in-parks-victoria-trial-20150722-gii2th.html

https://www.theland.com.au/story/3872462/deer-plague-hits-hip-pocket/

https://www.environment.nsw.gov.au/topics/animals-and-plants/threatened-species/nsw-threatened-species-scientific-committee/determinations/final-determinations/2004-2007/herbivory-and-environmental-degradation-caused-by-feral-deer-key-threatening-process-listing

Foxes and Cats damage

"Foxes and cats have already contributed to the extinction of a number of small native marsupials and are threat to many remaining threatened species"

Reference

https://onlinelibrary.wiley.com/doi/abs/10.1111/mam.12080

Rabbits damage

https://www.environment.nsw.gov.au/topics/animals-and-plants/threatened-species/nsw-threatened-species-scientific-committee/determinations/final-determinations/2000-2003/competition-and-grazing-by-the-feral-european-rabbit-key-threatening-process-listing

Goats damage
https://www.environment.nsw.gov.au/topics/animals-and-plants/threatened-species/nsw-threatened-species-scientific-committee/determinations/final-determinations/2004-2007/competition-and-habitat-degradation-by-feral-goats-capra-hircus-key-threatening-process-listing

The Hard Hooved Story

The hard hooves story, that the Australian Alps are not adapted to the pressures of hard hooves is incorrect. The Procoptodon was a relative of the kangaroo that was 3 times the size of the current kangaroos and on each foot they had a single large claw similar in appearance to a horse's hoof. There were many other megafauna and their extinction was caused by human hunting. A new study by an international team of ecologists revealed that introduced herbivores have restored many important ecological traits. Losses of megafauna had a profound effect on ecosystems (24)(25)

Damage by Humans

Impacts to the environment by humans include building of dams, tourism, recreational vehicles, ski runs and facilities and climate change. And yet the wild horses are the scapegoats. Most ecologists and scientists confirm global warming to be the main cause of species decline and extinction.

This landscape is susceptible to a dry, warm climate and it will struggle to survive the effects of global warming. Especially the wetlands and snow patch vegetation, which are dependent on the current temperatures. These Alpine and subalpine wetlands are likely to be impacted through longer drought periods, increased temperatures and an increase in incidence and severity of wildfire. Climate change is already affecting the Australian Alps, as evidenced by a 30% reduction in snow cover. The flora and fauna is partly determined by snow presence, depth, and persistence, so climate change is expected to have substantial impacts on alpine biodiversity. The low-temperature conditions that have created the alpine and subalpine ecosystems make them highly sensitive to climate change. Even a small increase in mean ambient temperature is likely to result in the loss of wetlands. (18) Reduced snow cover is expected to have a detrimental effect on alpine fauna. The mountain pygmy-possum is likely to suffer a contraction in suitable habitat and local populations of broad-toothed rat are likely to be impacted by seasonal reduction in available habitat and increased predation by foxes. Climate change may also affect the breeding success of alpine frog species because the pools have the potential to dry before the tadpoles reach metamorphosis. Climate change, and the associated reduction in

snow, is also likely to affect water production. (18)
Tourism and recreation in protected areas results in a range of indirect impacts on the environment, including facilitating the spread of weeds. Research found that out of 156 exotic taxa recorded in 18 vegetation surveys between 1986 and 2004, 152 were associated with tourism infrastructure. Many exotics become invasive environmental weeds so there is a need to limit both introduction of exotic propagules and disturbance to natural vegetation during the construction, maintenance and use of tourism infrastructure in protected areas. (19) Infrastructure such as walking tracks have negative effects on vegetation including in mountain regions. In the alpine area there is a range of paved, gravel and raised steel mesh walking tracks in addition to an extensive network of informal/non-hardened tracks. A study showed that gravel and paved tracks had distinct verges of bare ground and exotic species. Regular bike and motor vehicle activity disturbance is also a serious issue. A study showed that weeds readily colonise gravel track verges and road disturbance sites. They found problems with the spread of new tracks and erosion areas caused through overuse. Sensitive plants are being trampled, such as the wetland and short alpine herbfield. (20)(21) Tourism to the alpine area is having a range of negative environmental impacts. Direct impacts include; compaction of soil, erosion, trampling of vegetation, urine and faecal contamination of waterways, particularly glacial lakes, disturbance to wildlife, noise pollution. (21)

“Human waste contributes to increased nutrification and contamination of pristine waterways, and has negative impacts on the tourism experience. Temporary toilets at Rawson Pass have helped to deal with some of the problems of human waste, but not adequately. The withdrawal of camping from within the catchment areas of the glacial lakes has also helped, but increasing usage of other areas by campers may result in new areas being impacted.”(21)

“It is estimated that the human footprint has affected 83% of the global terrestrial land surface and has degraded about 60% of the ecosystems services in the past 50 years alone. Land use and land cover (LUCC) change has been the most visible indicator of the human footprint and the most important driver of loss of biodiversity and other forms of land degradation. “(22)

It would appear that any amount of development is allowed in the Park if it is to bring in large amounts of money – like the tourism, that is being massively expanded in spite of the well-researched and recorded damage that it causes to the park (as mentioned above)! There is also of course the massive destruction of the Park by the construction of Snowy 2. But again, payment of massive amounts of money to purchase environmental credits means this gets ignored. Very contradictory. There is a trap that has been there for several years at Carol’s Creek. This trap has been set up right next to a waterway, attracting horses into the area, in spite of accusations of them damaging the waterways! Why on earth would they do that? Especially considering that salt blocks are used to attract the horses into the traps!

Management of the wild horses

Before deciding whether any management is even necessary, it is imperative to know the population size. Then also have independent (not government funded in any way) unbiased research done on the effects of ONLY horses in the environment.

After that, if management is found to be necessary, the most humane methods available to manage the horse population are of course those that are NONE LETHAL.

In my opinion that is either relocation to areas that are not "sensitive ", Reserve Design used to keep sensitive areas inaccessible, or passive trapping and rehoming. With also possibly fertility control in preference to any lethal methods. Fertility control was not considered to be suitable where there are large numbers. But at this point you have no idea of the actual numbers. And fertility control is used in the USA on large herds of horses successfully.

Management by shooting

Animals are recognised as sentient beings. Ground shooting of the horses is of course NOT a humane method to manage a population of wild horses. From the recent footage from an Australian Abattoir, it is blatantly obvious that even when contained in stocks and with "professionals", that the horses are not able to be euthanised without often several attempts!!! The suffering of these horses is unimaginable! And therefore ground shooting of loose horses humanely would be an even more impossible task! They could end up having half their face blasted off! And aerial shooting even worse! Strict standards of animal welfare cannot be upheld with either ground shooting or aerial shooting. Horses will suffer extremely.

The Standard Operating Procedures that have been "put together " for ground shooting of horses would NOT BE POSSIBLE TO ADHERE TO. As I am sure you are aware from photographs taken by local photographers, these SOPs have NOT been adhered to. Just as I knew they wouldn't. It's impossible.

It is stated that -
"Any government-managed program will require that all appropriate Codes of Practice and Standard Operating Procedures are followed for all management activities."

Also, the ITRG assessment of humaneness of management techniques was based on adherence to the SOPs. As they were not adhered to, this assessment is not viable. As it was, the "assessment of mode of death" for chest shots was already MODERATE in the LEVEL OF SUFFERING assessment. The fact then that SOPs were not adhered to would increase this up to SEVERE or EXTREME LEVEL OF SUFFERING. Particularly as the horses were not even shot in the required area for a correctly placed chest shot.

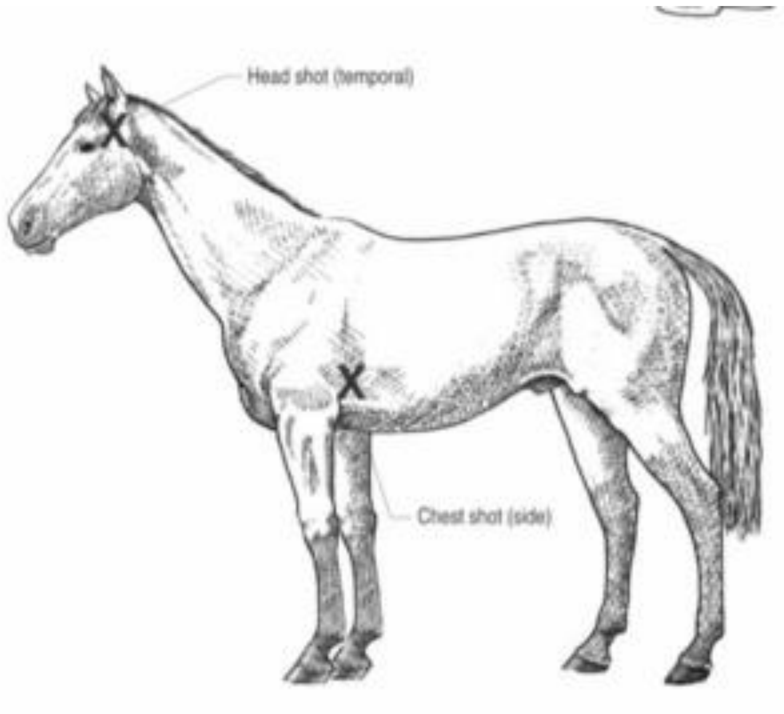
INDEPENDENT TECHNICAL REFERENCE GROUP ASSESSMENT OF HUMANENESS OF CONTROL METHODS

"Each assessment was based on a number of specific assumptions including that the method is carried out by skilled, competent and experienced operators in accordance with best practice through compliance with a SOP. Where no SOP existed, the panel used the

best available information to guide the assessment. It is important to note these assumptions when considering the relative humaneness for any given method, as any deviation from them will alter the outcome of the method”
“Some methods, such as those that include the risk of free-running animals becoming injured without being able to be followed up, have the potential to result in significant adverse impacts if best practice is not followed. It is likely that those methods that do not meet the requirements of best practice will result in poorer animal welfare outcomes than indicated here.” (54)

ASSESSING HUMANENESS DOCUMENT

“Chest shots do not render the animals instantaneously insensible and are likely to result in a higher incidence of wounding. Shooting at other parts of the body is unacceptable.” (55)
“Chest Shot - Side view • The horse is shot from the side so that the bullet enters the chest at a point behind the foreleg, slightly above and immediately behind the elbow joint.” (55)



Recommended shot placements for feral horses

FROM SHOOTERS THEMSELVES

Reference –
https://steamcommunity.com/app/518790/discussions/0/1692669912394162665/

“I think every single one of us has taken shots (including double lung shots) and seen: Animals that have dropped immediately. Animals that have jumped, walked away and died slowly/quickly. Animals that have jumped and run varying distances and died slowly/quickly. Animals that are wounded and lie down, get back up, take a few steps, lie back down and slowly die. That covers the range of what we'd see in normal life.”

ANIMAL WELFARE FOR WILD HERBIVORE MANAGEMENT

Reference –
Hampton2017.pdf (murdoch.edu.au)

“Shooting, in particular, has been subjected to little transparent, published research (Caudell et al. 2009). Shooting is commonly used worldwide to professionally and recreationally manage wild herbivores. However, there is ongoing concern about the animal welfare outcomes of many wildlife shooting programs (Bradshaw and Bateson 2000; Brook et al. 2015). Despite this societal concern, there have been few research methods published to allow quantifiable assessment of animal welfare outcomes. Poorly quantified Type 1 animal welfare measures from shooting programs include the frequency of non-fatal wounding (Aebischer et al. 2014), the frequency of animals requiring repeat shooting (Lewis et al. 1997), the accuracy achieved by different methods (Daoust and Caraguel 2012), and the role of manipulable variables in influencing welfare outcomes (Hampton et al. 2014). Many procedural documents (standard operating procedures (SOPs) etc.) have recently been developed for wild herbivore control (Sharp and Saunders 2004) but few cite quantified welfare data. Regulatory approaches often rely on procedural documents, rather than aspiring to welfare standards, combined with the use of qualitative assessment models (e.g. the ‘Five Domains’ model). However, under this approach, there is rarely any requirement for regular monitoring or quantification of welfare outcomes in operational herbivore management (Hampton et al. 2016). For example, under the current approach to wild herbivore welfare regulation in Australia, a hypothetical herbivore shooting program may generate a frequency of non-fatal wounding exceeding 50% or another hypothetical program may generate a frequency of accidental mortalities during capture exceeding 50%. As long as these programs complied with the conditions specified in their relevant procedural documents, and in the absence of AEC oversight, there would be no impediment to their continued operation. I suggest that animal welfare regulation of wild herbivore management in Australia could undergo considerable refinement. Alternative methods are commonly used to manage overabundant feral horses (e.g. mustering and translocation, trapping and euthanasia, and ground-based shooting; Nimmo and Miller 2007), however few studies have quantified animal welfare impacts for
these methods. In particular, we are unaware of any empirical evidence that has been published for ground shooting despite the existence of a national model standard operating procedure in Australia (Sharp 2011b). “

THE RSPCA STATE

Reference –
https://kb.rspca.org.au/knowledge-base/rspca-policy-g1-humane-killing/

1.2 RSPCA Australia defines humane killing as when an animal is either killed instantly or rendered insensible until death ensues, without pain, suffering or distress.

They also state - “If the correct firearm and ammunition are used, a well-placed head shot (with the brain as the point of aim) will result in immediate unconsciousness. When there is adequate damage to the brain and the animal does not regain consciousness there will be no suffering. In contrast, with chest shots (which cause damage to the heart and lungs), the time to unconsciousness can range from seconds up to a few minutes. When an animal is shot in the chest, the time to loss of consciousness and the time to death will depend on which tissues are damaged and, in particular, on the rate of blood loss and hence how long it takes for the brain to have insufficient oxygen. Loss of consciousness and death is likely to be quicker when animals have been shot in the heart. A phenomenon called ‘hydrostatic shock’, where a pressure wave from the bullet causes damage to internal organs, can contribute to ‘bringing down an animal’ quicker and causing a more rapid loss of consciousness in some instances when animals are shot in the chest. However, compared with head-shot animals, those that are chest shot have a higher risk of remaining conscious and suffering for a short period prior to death – though the extent of suffering will vary depending on which tissues are damaged and the rate of blood loss. During severe bleeding they are likely to feel a sense of breathlessness and potentially some anxiety and confusion before they lose consciousness. Unfortunately, it is not uncommon for shooters to aim for the chest as it presents a larger target area compared to the head, thereby increasing the likelihood of shooting the animal, especially for less skilled shooters. To avoid suffering, shooters should be required to demonstrate competency in killing an animal instantly using a head shot.”
Reference –

What is the difference between head shooting and chest shooting? – RSPCA Knowledgebase

The management plans state that SOPs will be followed :-

STANDARD OPERATING PROCEDURES
“Ground shooting is best suited to accessible and relatively flat areas where there are low numbers of horses.” (23)
(The plan in Victoria Alpine National Park is to remove 500 in the first year in the Eastern Alps and remove 100 from Bogong. Even larger numbers than this are being removed from the Kosciuszko National Park! These are NOT low numbers! Relatively flat and accessible????)

“Ground shooting is time consuming and labour intensive, and is therefore not considered an effective method for large scale control.” (23)
(So not an effective method in this case then, as it is large scale)

“Ground shooting as a means of population control is not suitable in inaccessible, wooded or rough terrain where sighting of target animals and accurate shooting is difficult, or when wounded animals cannot easily be followed up and killed.” (23)
(So again not suitable)

“Only head (brain) or chest (heart/lung) shots must be used. Shots to the head are preferred over chest shots as they are more likely to cause instantaneous loss of consciousness. Chest shots do not render the animals instantaneously insensible and are likely to result in a higher incidence of wounding. Shooting at other parts of the body is unacceptable.” (23)

(For a group of say 10 horses, there would therefore need to be 10 shooters all perfectly positioned to deliver a perfect head shot at exactly the same time! Absolutely ludicrous! Or be able to position 2 shooters to possibly deliver 2 perfect shots at exactly the same time before the rest all gallop off. And that’s if they can even get close enough to deliver even one single perfect shot!)

“Group flight response is a limiting factor for humane and instantaneous killing of horses.” (23)

(EXACTLY MY POINT)

“If possible, all horses in a group should be killed before any further groups are targeted.”(23)

(No chance)

“Wounded horses must be located and killed as quickly and humanely as possible with a second shot, preferably directed to the head. If left, wounded animals can escape and suffer from pain and the disabling effects of the injury.” (23)
(Unless very badly injured and not able to run away there would be no chance of finding the wounded animals with maybe horrendous serious life threatening injuries!)

“Horses must NOT be shot from a moving vehicle or other moving platform, as this can significantly detract from the shooter’s accuracy.” (23)

(But aerial shooting is approved in Victoria)

“The objective is to fire at the closest range practicable in order to reduce the risk of non-lethal wounding. Accuracy with a single shot is important to achieve an immediate and therefore humane death. A horse should only be shot at when: * it is stationary and can be clearly seen and recognised
* it is within the effective range of the firearm and ammunition being used
* a humane kill is probable. If in doubt, do NOT shoot.
* Ensure there are no other horses behind the target animal that could be wounded by the shot passing through the target.
* Although horses are large animals, the vital areas targeted for clean killing are small.” (23)

(If all the above directives are followed, NO SHOTS would be delivered. So they would have to shoot and injure and hope they can follow up and reshoot - which is not following the SOPs)

“Shooting of individuals should stop when the flight response of the herd limits further accurate shooting.” (23)

(So that’s either immediately or after one shot only)

“If the stallion is shot first the mares might panic and escape with their foals.” (23)

(Shouldn’t be shooting anyway when mares have foals at foot)

AERIAL SHOOTING

In a study done in Australia after aerial shooting wild horses it was found
The Instant Death Rate was 63% - Absolutely not acceptable! (37)

Up to 6 bullets per horse peppered all over their bodies!(37)

In total, 35% of horses displayed bullet-wound tracts affecting the cranium, 50% the cervical spine, and 57% the thorax, whereas 3% of horses displayed bullet-wound tracts affecting the forelimbs and 8% the abdomen.(37)
Horses not rendered immediately insensible (37%)!!!! Absolutely NOT acceptable. (37)

Aerial shooting - Lets be realistic. Trained people in the slaughter houses with trained domestic horses held in stocks, are still not able to euthanise horses in a humane way successfully! So how anyone can possibly believe that it is possible to shoot a galloping horse from a moving helicopter and manage to get a shot direct in the very narrow spots to achieve instant humane euthanasia is beyond me. Hence the scenes we had from the Guy Fawkes incident and other similar incidents. And the results from the above report.

And in any case lethal control is not necessary and is cruel. These are sentient beings and they deserve life. Non-lethal control may be more expensive and more difficult, but surely no one actually wants the horses to be slaughtered when there are perfectly feasible alternatives? Let’s be honest, human overpopulation is the cause of most of the world’s problems, but lethal control would never be considered.

Obligations by NPWS - Kosciuszko National Park

Let’s remind ourselves of the intent of the legislation-

“The new wild horse heritage plan will prioritise supporting populations in less sensitive areas and resources will be allocated to relocating brumbies to those areas. Lethal culling of brumbies will not occur.”
“This bill will end the uncertainty as to whether or not brumbies will be shot: They will not be shot.”
“the focus will now be on identifying areas where a population can be protected without significant environmental harm. The priority will be to move brumbies to those areas. However, trapping and rehoming will be considered if the number of brumbies increases too much, and that poses an environmental threat or safety risk.”
“The new framework of managing brumbies in the Kosciuszko National Park also will involve a number of new approaches, including brumbies found in "highly-sensitive" alpine areas of the national park being relocated by authorities”

“The previous draft plan set an aggressive target for reducing the population. That would have resulted in a horrific mass slaughter of the iconic brumby—600 horses—in the Kosciusko National Park. Culling is cruel and barbaric.” “The 2016 Kosciuszko National Park Draft Wild Horse Management Plan set an aggressive target of reducing the population to 600 horses. In that draft plan, lethal methods were considered, including ground shooting. That is not an acceptable situation and this legislation will ensure that doesn’t happen.”

“The heritage management plan will specifically prohibit lethal culling of the brumby, aerial or otherwise, and will identify those areas in the park where brumbies can roam without causing significant environmental harm, . . . If brumbies are found in highly-sensitive alpine areas of Kosciuszko National Park, resources will be allocated towards relocation first, followed by re-homing, should population numbers grow too high.” (33)(34)(35)

SO THE INTENT OF THE LEGISLATION WAS CLEAR! NO LETHAL MANAGEMENT! This was what was voted through in Parliament.

“The National Parks and Wildlife Service (NPWS) is required by law to reduce the number of horses in Kosciuszko National Park from over 14,000 to 3,000 horses.”

NPWS are required by law to retain 3,000 horses in the Park! They need to therefore show irrefutable proof that there are more than 3,000 horses currently in the Park! By doing an actual headcount, just as they have every year! Not by using computer software that has given BIOLOGICALLY IMPOSSIBLE ESTIMATES. The headcount should be done with video footage and a Wild horse Advocate present - as has been suggested many times!

“To meet this legal requirement, NPWS must undertake a range of control measures, including trapping and rehoming and ground shooting.”

There is NO “MUST” about using a range of control measures! They can JUST use trapping and rehoming, if even necessary AFTER THEY HAVE PROVIDED THE ABOVE PROOF that there are still over 3,000!

“The RSPCA has investigated the 11 dead horses reported in the media this week and confirmed there is no evidence of any breach of animal welfare legislation. These horses were culled (ground shooting) by NPWS in meeting its legal obligations under the plan.”

“All control measures, including ground shooting, are undertaken in accordance with the highest animal welfare standards with strict requirements developed in consultation with a range of experts.”

Any government-managed program will require that all appropriate Codes of Practice and Standard Operating Procedures are followed for all management activities. This is not just about the Animal Welfare Legislation. The SOPs were NOT FOLLOWED and they are there for a reason! The assessment of humaneness was done accounting for ADHERENCE to the SOPs.

The horses were not shot in the required areas.

FUNDING

Funding for the “official surveys” should be totally removed. The annual headcounts should then continue in the North Kosciuszko region. Barmah surveys - could be done as headcounts but NO APPLICATION of the Distance software to inflate numbers!

The Bogong region of the ANP - an accurate headcount was done. No doubling by application of software was necessary!

In Kosciuszko NP, there are already MUCH LOWER NUMBERS than required to be retained by the new management plan! So save on funding by doing no further management until they are back to the required retention number of 3,000. If the numbers increase in the future to above this amount and after INDEPENDENT and unbiased STUDIES ARE DONE on the effects of wild horses, including their good effects, if it should be deemed necessary to remove some, funding should be allocated to rehomers to help rehome and reduce numbers.

No more money should be wasted on doing surveys that are producing results that are totally ludicrous and biologically impossible!

There are now extremely low numbers of Brumbies remaining in Barmah National park after the substantial losses due to floods and the large number shot, so no further reduction is necessary, saving funding there as well!

In the ANP (including Bogong) after the MASSIVE WILDFIRE LOSSES of over 50% in the Eastern Alps and the very large number already shot! - no further reduction should be necessary again saving massive funding! As with Kosciuszko NP, remaining wild horses should stay in the Park. Independent and unbiased research should then be undertaken to assess if any actual damage is being done EXCLUSIVELY BY wild horses. I strongly dispute that any real damage is being done by them - a few hoof prints made by the wild horses when going to drink is, I am sure, the extent of their imprint on the environment and compared to the human imprint and that of other animals, it’s totally ludicrous to even assess this as damage! If it’s deemed necessary to remove any more, then only non lethal techniques should be used to remove them.

STRATEGIES TO HELP RESTORE THE ENVIRONMENT

Tackle global warming/climate change effects, which is one of the main problems as shown above. Horses are carbon sequesters!

Reduce chance of wildfires by every means possible including retaining all the currently remaining wild horses - who reduce the risk of fire as mentioned previously.

Prevent human activity damage – of which there is way more than any other animal.

Concentrate on management of the species that are really doing the damage as specified above.

REFERENCES

1. https://theaustralianalpsnationalparks.org/wp-content/uploads/2013/12/feral-horses.pdf
2. https://drive.google.com/file/d/12PlCz3WGyepA2pinKw58rzR_-A2IeS44/view?usp=sharing
3. https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/

Animals-and-plants/Pests-and-weeds/Kosciuszko-wild-horses/kosciuszko-national-park-draft-wild-horse-management-plan-160271.pdf

4. https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Animals-and-plants/Pests-and-weeds/Kosciuszko-wild-horses/kosciuszko-national-park-horse-management-plan-080254.pdf
5. https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Parks-reserves-and-protected-areas/Parks-management-other/kosciuszko-national-park-wild-horse-heritage-management-plan-210549.pdf
6. https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Animals-and-plants/Pests-and-weeds/Kosciuszko-wild-horses/kosciuszko-national-park-wild-horse-populations-survey.pdf
7. https://theaustralianalps.files.wordpress.com/2019/12/feral-horses-in-the-australian-alps-the-analysis-of-aerial-surveys-conducted-in-2014-and-2019-cairnss-2019.pdf
8. https://drive.google.com/file/d/12PlCz3WGyepA2pinKw58rzR_-A2IeS44/view?usp=sharing
9. https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Animals-and-plants/Pests-and-weeds/Kosciuszko-wild-horses/kosciuszko-national-park-wild-horse-scientific-advisory-panel-report.pdf
10. https://theaustralianalpsnationalparks.org/wp-content/uploads/2019/12/2019-australian-alps-feral-horse-aerial-survey-summary-report-and-faqs.pdf 11. https://drive.google.com/file/d/1FvXG4DMd9Wop88sY15lS-XnPctUu2Dib/view?usp=sharing
12. https://drive.google.com/file/d/1tirD8W5rNKXG86lSSkqQ7VA7FX0JMxTg/view
13. https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Animals-and-plants/Pests-and-weeds/Kosciuszko-wild-horses/kosciuszko-national-park-wild-horse-population-survey-2022.pdf?fs=e&s=cl
14. https://b32d739d-d185-4dd2-8388-9b640855e5bd.filesusr.com/ugd/6a30c6_39c677396bba4fbea3e2b49c94d3b97c.pdf
15. https://www.science.org/doi/10.1126/sciadv.1400103
16. https://www.researchgate.net/publication/275657970_Collapse_of_the_world'_s_largest_herbivores
17. https://www.sciencedirect.com/science/article/abs/pii/S1439179104000866
18. https://www.ecologyandsociety.org/vol19/iss2/art13/
19. https://www.researchgate.net/profile/Wendy-Hill-5/publication/29465288_Indirect_Impacts_of_Nature_Based_Tourism_and_Recreation_The_Association_Between_Infrastructure_and_the_Diversity_of_Exotic_Plants_in_Kosciuszko_National_Park_Australia/links/0a85e53b4f477bb9c2000000/Indirect-Impacts-ofNature-Based-Tourism-and-Recreation-The-Association-Between-Infrastructure-and-the-Diversity-of-Exotic-Plants-in-Kosciuszko-National-Park-Australia.pdf?origin=publication_detail
20. https://pubmed.ncbi.nlm.nih.gov/16099588/
21. https://www.researchgate.net/publication/228604806_Managing_the_Kosciuszko_Alpine_Area_Conservation_milestones_and_future_challenges
22. https://sustainabledevelopment.un.org/content/documents/1124landuse.pdf
23. https://pestsmart.org.au/toolkit-resource/ground-shooting-of-feral-horses/
24. https://www.eurekalert.org/news-releases/911712
25. https://en.m.wikipedia.org/wiki/Procoptodon
26. https://www.publish.csiro.au/rj/RJ17079
27. https://bioone.org/journals/western-north-american-naturalist/volume-77/issue

-4/064.077.0412/Effects-of-Feral-Horse-Herds-on-Rangeland-Plant-Communities-across/10.3398/064.077.0412.short
28.https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/1365-2664.13338
29.https://www.fs.fed.us/rm/pubs_journals/2016/rmrs_2016_baur_l001.pdf
30.https://www.sciencedirect.com/science/article/abs/pii/S1439179104000866
31.https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Parks-reserves-and-protected-areas/Types-of-protected-areas/peat-forming-bogs-and-fens-snowy-mountains-technical-report-120257.pdf
32.https://www.environment.act.gov.au/__data/assets/pdf_file/0009/576819/Technical_Report_20_Sphagnum_Bog_Mapping_and_Recovery_Plan.pdf
33.https://www.timebase.com.au/news/2018/AT04712-article.html
34. https://www.nsw.gov.au/news/protecting-kosciuszkos-wild-horses-from-culling
35.https://www.parliament.nsw.gov.au/Hansard/Pages/HansardResult.aspx#/docid/'HANSARD-1820781676-76718'
36.https://pestsmart.org.au/wp-content/uploads/sites/3/2020/06/FeralHorse.pdf
37.https://www.researchgate.net/publication/314197502_Assessment_of_animal_welfare_for_helicopter_shooting_of_feral_horses
38.https://theconversation.com/australian-endangered-species-southern-corroboree-frog-16189
39.https://bioone.org/journals/wildlife-research/volume-46/issue-2/WR18093/Feral-horse-impacts-on-corroboree-frog-habitat-in-the-Australian/10.1071/WR18093.short
40.https://theconversation.com/our-field-cameras-melted-in-the-bushfires-whenwe-opened-them-the-results-were-startling-139922
41. https://www.environment.gov.au/resource/smoky-mouse-pseudomys-fumeus
42.https://www.environment.act.gov.au/__data/assets/pdf_file/0006/995532/Technical-Report-35-BT-Rat-survey.pdf
43.https://www.environment.gov.au/cgi-bin/sprat/public/publicspecies.pl?taxon_id=64721
44.https://livrepository.liverpool.ac.uk/3006632/1/ES-2013-6393.pdf
45.https://www.jstor.org/stable/26269528
46.https://www.google.com/amp/s/amp.smh.com.au/environment/sustainability/floods-put-the-barmah-brumbies-down-for-the-count-20101005-1664u.html
47.Wehner, B., Pigott, J.P. & Carboon, J. (2012). 2012 Barmah National Park Wild Horse Survey. Report for the Barmah Horse Advisory Committee, Parks Victoria. October 2012
48.https://www.parks.vic.gov.au/-/media/project/pv/main/parks/documents/visitorguides-and-publications/alpine-national-park/alpine-np-feral-horse-sap-2018-21.pdf?la=en&hash=CA8AA077676F2089DF663551527F010B607A6F3D
49.https://www.parliament.vic.gov.au/images/stories/committees/paec/Inquiry_into_Auditor-Generals_report_No._202_Meeting_Obligations_to_Protect_Ramsar_Wetlands_2016/Submissions/310a._Australian_Brumby_Alliance_appx_1.pdf
50.https://www.parks.vic.gov.au/-/media/project/pv/main/parks/documents/management-plans/barmah-national-park-and-barmah-forest-ramsar-site-strategic-action-plan-2020-2023.pdf?la=en&hash=E130D0B2C8E9C9E22B349EC9148F9AA99ED09BAA
51.https://theaustralianalps.files.wordpress.com/2013/12/2009feralhorsealpssurvey.pdf

52. https://www.parks.vic.gov.au/-/media/project/pv/main/parks/documents/get-into-nature/conservation-and-science/feral-horse-survey-reports/eastern-alps-feral-horse-survey-2021-final-report.pdf?la=en&rev=45173129f8d4432fb50645ddeda94243&hash=5AAB3F40184BB4464C2E9921A27953DFDDDA0340
53. https://www.parks.vic.gov.au/-/media/project/pv/main/parks/documents/get-into-nature/conservation-and-science/feral-horse-survey-reports/2021-feral-horse-survey-report-final.pdf?la=en&rev=9bea421a8b1e4aa0a5f6755f6e3d3aee&hash=F8C7A6B7C67D59751CF9B2E3A51C4E1F16551CA5
54. https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Animals-and-plants/Pests-and-weeds/Kosciuszko-wild-horses/final-report-independent-technical-reference-group-supplementary-wild-horse-management-plan-160221.pdf?la=en&hash=26B121837E28A2C68514B34D99A21C18F7557E16
55. https://www.agriculture.gov.au/sites/default/files/documents/humaneness-pest-animals.pdf