Analysis of the 2021 aerial survey of the feral horse population in the Barmah National Park, Victoria

May, 2023

Dr Stuart Cairns[1]

[1]G.E. & S.C. Cairns Pty. Ltd.,
PO Box U21,
University of New England,
Armidale,
NSW, 2351
scairns@une.edu.au

Telephone: 0427454344

Summary

1. The feral horse population in the Barmah National Park (NP) was surveyed in November 2021.
2. The surveys were conducted using a helicopter with the detection and recording of the presence horses being based upon imagery collected using an automated thermal imaging system mounted on the aircraft.
3. The survey area covered was approximately 23,000 ha. The helicopter was flown along 19 east-west orientated transect (254 km of survey effort).
4. Desktop analyses of imagery collected during the survey was undertaken to identify the presence of horses in relation to the survey transects.
5. The data obtained from the survey were analysed using strip transect analysis, conventional distance sampling (CDS) analysis, simple capture-mark-recapture analysis (CMR) and mark-recapture distance sampling (MRDS) analysis. The MRDS analyses failed.
6. Population estimates determined using the CDS methods were relatively close to one another. However, the strip transect analysis proved in this instance to result in the simplest and most suitable detection function model.
7. Based upon the strip transect sampling model, the estimated number of feral horses in the area surveyed in Barmah NP in 2021 was 328 ± 106 individuals. This estimated population was lower than the population estimated in 2019 of 621 ± 144 individuals, which was, in turn, lower than the population estimated in 2018 of 1,088 ± 336 individuals.

1. Introduction

Barmah National Park (35.98° S, 144.99° E) comprises an area of 28,537 ha some 225 km north of Melbourne, Victoria. It lies within the Yorta Yorta traditional boundaries and is sharply defined by its northern boundary, the Murray River; the state border between New South Wales (NSW) and Victoria. Barmah National Park (NP) is part of the Barmah-Millewa Forest, an area of some 66,000 ha that spans the Murray River and forms the largest River Red Gum (*Eucalyptus camaldulensis*) forest in the world. This area constitutes a significant site under the Ramsar Convention (https://www.environment.gov.au/water/wetlands/ramsar), an international convention that provides the framework for national action and international cooperation for the conservation and wise use of wetlands and their resources.

Currently, Barmah NP has within its boundaries a population of feral horses (*Equus caballus*). The management of these horses is part of a strategic action plan that has been developed for the national park and released in February 2020 (https://www.parks.vic.gov.au/projects/barmah-strategic-action-plan). This plan outlines a four-year program to address threats in the Barmah Forest and protect it for current and future generations. In association with the development of this plan, an aerial survey was conducted over parts of Barmah NP in 2018 to estimate the size of the feral horse population. This survey was conducted during winter using thermal imagery technology in the form of a military-grade, high-resolution, forward-looking, imagery system; namely, an L3 Wescam MX10 multi-sensor multispectral system (https://www.wescam.com/products-services/airborne-surveillance-and-reconnaissance/mx-10) mounted on a helicopter. The advantage of using this camera was that it enhanced the detection of horses through the canopy of the red gum forest.

Since the conduct of these two surveys, two further surveys have been undertaken. The first of these, the third in the sequence, was conducted in the winter of 2019, and the fourth survey was conducted the spring of 2021. Both these surveys used the same methodology as outlined above. The results of the first two surveys are reported in Cairns (2019) and the result of the third survey is reported in Cairns (2020). Reported on here is the conduct and outcome of the most recent of the four consecutive surveys. As was the case with the two previous reports, included in this report is the use of different methods of analysis that are available for estimating wildlife density and abundance associated with line transect (distance) sampling. The analysis methods compared were those based upon conventional distance sampling, simple capture-mark-recapture sampling and mark-recapture distance sampling. The results of the previous surveys have been analysed comparatively also using these methods (Cairns 2019, 2020).

In relation to the conduct of these surveys, it should be noted that in order to develop strategies for the management of any animal population, it is crucial that information on population size, changes in numbers and, if possible, population trends and rates of increase or decline be obtained (Durant *et al.* 2011). The changes in numbers between successive (2017-2021) are reported upon here.

2. Survey and Analysis Methods

2.1 Survey Methods

The 2021 survey for feral horses in the Barmah NP was conducted on the 16 November along designated transects using a Eurocopter AS350 Écureuil (*Squirrel*) single-engine light helicopter fitted with a Wescam MX10 automated thermal imaging system. The detection and recording of the presence of individual horses was based upon imagery collected using this system. The data required for estimating the density and abundance of horses were obtained from a desktop analysis of the recorded imagery. In these analyses, an image was identified and confirmed as representing a horse in the landscape based on the infra-red (IR) targets in the video being assessed against a number of criteria (size, shape, movement). This is standard methodology for identifying objects recorded either in full-motion video (FMV) or as other remote-sensed data, where typically an IR target would be in view for around 2.5-3.0 seconds of video; over approximately 100 frames. The geographic location of each object identified as a horse was recorded and its perpendicular distance from the centerline of the survey transect calculated. This distance measurement is required if line transect sampling methodology is to be used to estimate the density of the horse population.

For the conduct of these surveys, the helicopter crew comprised the pilot, a camera operator and a mapping operator. No observers, such as those usually deployed in wildlife surveys using this type of survey platform, were involved (see Cairns 2014). The desktop analysis was conducted by two expert analysts, operating independently from one another.

In conducting the survey, the helicopter was flown along designated survey lines (transects) at a height of 305 m (1,000 ft) above ground level at a ground speed of 112 km hr^{-1} (60 kts). The survey transects were aligned east-west, were parallel to one another, and were spaced 1,000 m apart (Fig. 1). This transect spacing was selected to reduce the likelihood that individual horses would be counted on duplicate, adjacent transects during a single survey session. It also ensured that an adequate coverage of the relatively complex landscape of the survey area was obtained. Systematic surveys such as this one result in smaller variation in density estimation from one realisation to the next than do random surveys, and also avoid any problems associated with overlapping transects (Buckland *et al.* 2001). The NW-SE alignment of transects was, in part, aimed at optimising the performance of the survey platform by minimising the impact of wind buffeting on the aircraft, which can have an effect upon image stability.

The survey was conducted over an area of 26,064 ha and covered some 91% of the total area of the Barmah NP. Nineteen east-west orientated transects, ranging in length from 2.75 km to 26.51 km, were flown during the survey (Fig. 1). The total on-transect effort for the survey was 253.6 km. With a possible nominal half-width of the survey transect on either side of the aircraft of 90 m, coverage was high, being 19.5% of the survey area, or 17.8% of the whole of the park.

Images were captured using the Wescam MX10 imaging system operating in thermal infra-red mode. The field of view was 30° horizontal and 17.2° vertical. Sensor size was 1280 x 738 pixels. Focal length of the imaging system was 17.9 mm. The MX10 was set at 54° below horizontal view angle, with the boresight (image centre point) set to follow the aircraft heading. No image processing was applied post-capture. The Wescam MX10 applies real-time image enhancement to balance the brightness and contrast of the video feed in order to account for the dynamic range within the current field of view.

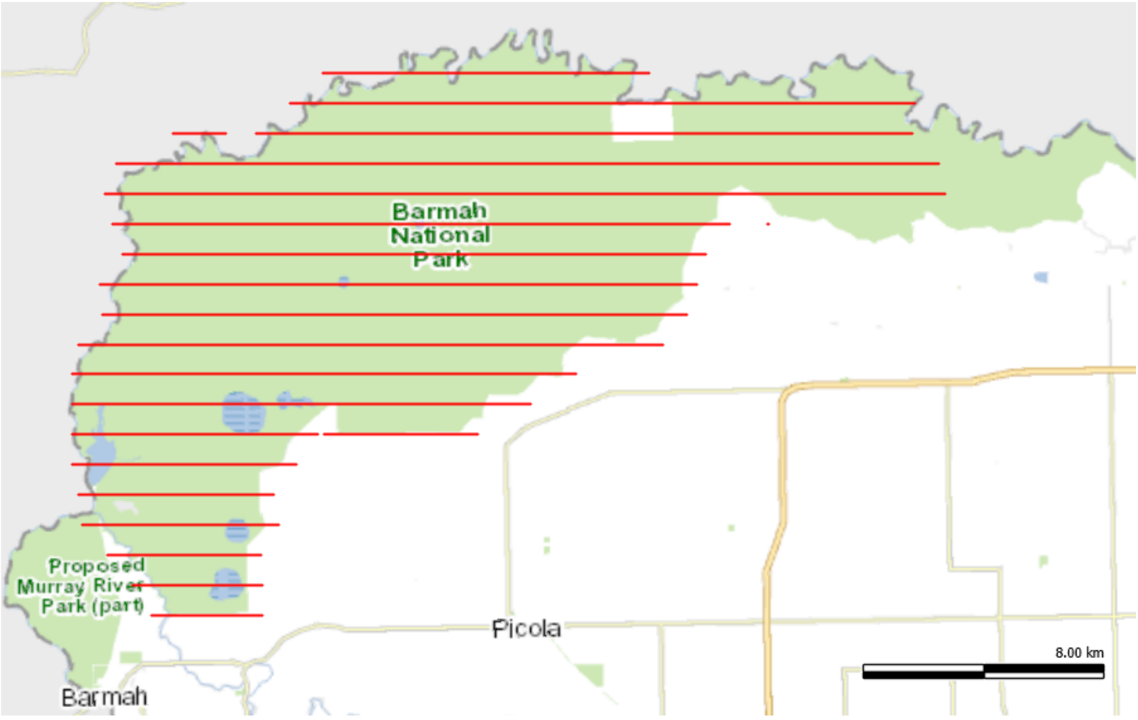

Fig. 1. The flight path showing the19 east-west orientated survey transect flown in the 2021 aerial survey of the Barmah NP.

Data downloaded from the Wescam MX10 video in the form of command video were streamed into AIMS HD software (CarteNav solutions). The video stream was corrected in real time for positional location using a 10 m digital elevation model, a platform and sensor inertial navigation system, and the global navigation satellite system location. Recorded Full Motion Video files conformed to the Motion Industry Standards Board video metadata standards.

Video processing was undertaken using ESRI ArcMap 10.4, and FMV 1.3 plugin. Operating independently from one another, two expert analysts reviewed the video stream, recording information regarding any IR targets each identified. The two target datasets created by the analysts were then merged to form a master-target, spatial dataset showing both those targets which had been identified by both analysts along with those targets which had been identified separately by only one of the two analysts. A target (image) was identified as being representative of a horse if it was clearly a horse (determination being made on size, shape, movement), or if it was an image of an object that was at least a minimum of 1.6 m in length (this being measured in VLC). If the target was not fully visible in a single video frame, it could be measured or identified in relation to the four seconds of video during which that point was visible on the screen during the review process. If the target was still not clearly visible, it could then only be identified as being a horse if it were observed in very close proximity to another target that was clearly identified as a horse, or if the target had an IR halo effect. A target was rejected if the analyst considered that it did not meet any of these identification criteria.

2.2 Density and Abundance Estimation

To estimate the density of feral horses in the survey area, two broad approaches were adopted. The first was based upon transect survey methodology and the second was based upon capture-mark-recapture (CMR) sampling methodology. Transect analyses involved using strip transect analysis in relation to a survey strip the width of which was determined *a priori,* and conventional distance sampling (CDS) analysis in relation to a survey strip that was determined *a posteriori* (Thomas *et al.* 2002). The use of CMR analysis was based upon the duplicate, independent reviews of the imagery data conducted by the two expert analysts. Capture-mark-recapture analysis was first undertaken using the simple Lincoln-Petersen CMR estimator (Seber 1892) and then extended to a hybrid form of analysis that incorporates the perpendicular distance of objects (horses) from the transect centerline as a covariate. This hybrid form of CMR analysis is known as mark-recapture distance sampling (MRDS) (Burt *et al.* 2014). A recent version of the distance sampling software package, DISTANCE 7.3 (Thomas *et. al.* 2010), was used to undertake these analyses.

Due to the alignment of the imaging system with respect to the ground, the imagery captured in each frame has a keystone shape. The narrowest point of the keystone is 180 m (i.e. 90 m either side of the transect centerline); the widest point of it is 220 m (i.e. 110 m either side of the transect centerline). To compensate for this, all analyses were initially restricted to a nominal survey strip with a half-width (*w*) of 90 m (i.e. extending 90 m on either side of the transect centerline). All objects identified as being at distances greater than 90 m from the transect centerline were excluded from the analyses. With the analysis of line transect data, it is often appropriate to remove outlier observations to better manage the required fitting of a detection function model (Buckland *et al*. 2001). This could mean that, for final analyses, the half-width could be further truncated.

For the analysis of the survey results within a transect survey framework, the data were initially analysed as strip transect samples using the method of ratio estimation to account for the different sizes (lengths) of the samplers (transects). The use of this method has a long history in relation to the conduct of aerial surveys of wildlife populations (Caughley 1977; Thompson 2002). This method of analysis is perhaps somewhat redundant now, being based upon the simple assumption that all objects within a designated survey strip are available to be identified and counted. In other words, detection of these objects on the survey strip can be effected with certainty. This is, of course, an assumption that can be easily challenged. However, within the context of the present study, it could perhaps be considered a reasonable assumption given that the target objects (horses) were initially detected using imagery equipment, with the resulting imagery being subsequently reviewed under static conditions.

The alternative to analysing survey results within the framework of strip transect sampling is to analyse them within the framework of line transect (distance) sampling. With this method, the assumption of certain detection of objects on a nominal survey strip can be relaxed. Through the measurement, with the detection of each object, of the perpendicular distance of that object from the centerline of the survey transect, a detection function model can be developed from which the average probability of detection of objects ($P_a \leq 1$) on the nominal survey strip can be determined (Buckland *et al.* 2001). This probability can, in relation to the number of objects sighted and the area of the nominal survey strip (determined *a posteriori*), be used to estimate density, free of the assumption that all objects are available to be detected during the survey process.

The analysis of distance sampling data such as those collected here first involves the estimation of the detection probability of objects within the covered region (the designated survey strip), then the estimation of the density of objects within the covered region given this detection probability and, finally, the estimation of the number of objects in the survey region given the density of animals in the covered region (Borchers & Burnham 2004). In order to estimate the probability (P_a) that an object (horse) within the covered area of width $2w$ (the width of the nominal survey strip) will be observed, the detection function $g(x)$ representing the probability that an object at perpendicular distance x from the survey transect centerline is detected (where $0 \leq x \leq w$ and $g(0) = 1$) needs to be modelled and evaluated at $x = 0$, i.e. at the centreline of the transect (Thomas *et al.* 2002).

The program DISTANCE 7.3 has three different analysis engines that can be used to estimate the detection function and associated statistics, and then the density of the surveyed objects (Thomas *et al.* 2010). Two of these, the conventional distance sampling (CDS) analysis engine and the mark-recapture distance sampling (MRDS) analysis engine were used in these analyses. The results of the analyses conducted using the ranges of detection function model options available within the CDS analysis engine were compared serially in order to determine the most parsimonious (simplest and most suitable) detection function model and, hence, the most likely and accurate estimate of population density. The model with the lowest value for a penalised log-likelihood in the form of Akaike's Information Criterion ($AIC = -2 \times \textit{log-likelihood} + 2K$; where K is the number of estimated parameters in the model) was generally selected as the detection function. In selecting the most parsimonious model, along with comparing AIC values, some consideration was also given to the goodness of fit of the model to the data, and to the shape criterion of the detection function. Although available as an option to improve goodness-of-fit, no manipulation involving the grouping of distances was undertaken.

Following the recommendations of Buckland *et al.* (2001), six detection function models were considered in the analyses using the CDS analysis engine. Each model comprised a key function that, if required, can be adjusted by a cosine or polynomial series expansion containing one or more parameters to ensure a better fit to the survey data being analysed. The different models considered were: a Uniform key function with an optional Cosine or Simple Polynomial series expansion; a Half-normal key function with an optional Cosine or Hermite Polynomial series expansion; and a Hazard-rate key function with an optional Cosine or Simple Polynomial series expansion. The number of adjustment terms incorporated into the model was determined through the sequential addition of up to three terms. Because data truncation can have the effect of simplifying the model fitting process, further truncation beyond the initial truncation to 90 m was given consideration as part of the analyses.

The strip transect and line transect analyses used the combined data drawn from the survey imagery by the two expert analysts. Each individual object (horse) can be identified by the tag of its distance from the transect centerline and therefore be identified as having been either detected by both analysts, or having been detected by Analyst 1, but not by Analyst 2; and *vice versa*. Combining the two data sets would result in a single dataset comprising all objects (horses) that were detected by at least one analyst.

As well as serving to remove any redundancies in this aggregation of results for transect data analyses, this form of object identification is also required if CMR analyses are to be undertaken. Within the CMR analysis framework, there are four possible states in a two-occasion sampling process. Here, the review of the survey imagery by the two analysts can be seen as representing two sampling occasions. The states that exist in the CMR framework take the following forms: an object can be “captured” and then “released” by Analyst 1, to be “recaptured” by Analyst 2; an object can be “captured” and then “released” by Analyst 1, but not successfully “recaptured” by Analyst 2; an object, having failed to be “captured” by Analyst 1, is, however, “captured” by Analyst 2; and, an object can fail to be “captured” by either Analyst 1 or Analyst 2.

The analysis of the CMR data was undertaken using, for the purpose of comparison, models available in the MRDS analysis engine in DISTANCE 7.3. Specifying a detection function in MRDS requires specifying the forms of two functions (Laake & Borchers 2004). The first of these is the unconditional detection function, which is the probability of one or more observer (at least one analyst) detecting an object given its perpendicular distance from the transect centreline. The second is the conditional detection function, which is the probability of one of the two analysts detecting an object given that the other analyst has also detected it given its perpendicular distance from the transect centreline. The first of these two models is the distance sampling (DS) model. The second model is the mark-recapture (MR) model. The forms DS and the MR models are different. The DS model is of the same form as the models used in the CDS analysis engine, except that only the Half-normal and Hazard-rate key functions can be implemented; without the option of including any series expansion terms. The MR model is implemented as a logistic regression model that is fitted to the success/failure data of CMR process in order to provide an estimate of the probability of detection for one observer (analyst), given detection by a second observer (analyst), based upon the perpendicular distance of an object from the transect centerline and, if available, any other explanatory covariates (Burt *et al.* 2014). This model can be viewed as a generalisation of the Lincoln-Petersen (LP) estimator (Seber 1982) that can include explanatory covariates such as the perpendicular distance of an object from the transect centerline.

There are five estimation models in the MRDS analysis engine that can be used to analyse CMR data. However, not all were used here. The first of those that were used was a model identified as the trial, full independence (trial.fi) model with two configurations: one of which fitted only the conventional LP estimator, and the other which fitted the DS and MR logistic regression models in relation to the distance covariate. As well as this, two models identified as trial, point independence (trial.pi) models were also fitted to the data; one using a Half-normal (HN) detection function as the DS model, and the other using a Hazard-rate (HR) detection function as the DS model. In both these models, the MR model is implemented as a logistic regression model, with the single covariate of perpendicular distance. Fitting these models allows the detection probability on the transect centerline to be calculated. This is a step beyond assuming detection to be certain at distance zero from the centerline of the transect, as is the constraining case when using CDS analysis (i.e. where $g(0) = 1$). The point independence (trial.pi) models allow the detection probability, $p(y)$, at distance y from the transect centreline that is obtained from the DS model to differ from the $p(y)$ determined from the mark-recapture data (Laake *et al.* 2011). With the full independence (trial.fi) model, this flexibility is lost, with the detection function shapes and, therefore, the $p(y)$ values being constrained to be identical.

For the purpose of comparison, single-observer, distance sampling models with Uniform, Half-normal and Hazard-rate detection functions were also fitted to the data. These models were fitted to the data pooled for both analysts without the addition of any series expansion terms. The uniform detection function fitted without an additional series expansion term is equivalent to the model fitted to the data for a strip transect analysis.

3. Results and Discussion

3.1 Strip and Line Transect (CDS) Analysis

All 19 of the planned transects were completed in the survey. The recorded imagery was reviewed by two expert analysts, with each positive identification of a horse being tagged with its perpendicular distance in metres from the transect centreline. Fifty-four horses were detected on 12 of these transects by at least one of the two analysts (Table 1). All these horses were detected within 90 m from the transect centreline. These results were analysed separately as both strip transect and line transect survey results. Technically, strip transect analysis would involve the use of the method of ratio estimation when survey transects are of different lengths; as is the case with this survey (Caughley 1977, Thompson 2002). However, the strip transect analysis was able to be undertaken using DISTANCE by fitting to the data, as a sightability model, a Uniform key function without any adjustment (series expansion) terms.

Table 1. The lengths of the 19 survey transects plus the numbers of horses detected by at least one of the expert analysts during the review of the recorded thermal imagery from each transect. Included in these observations are three horses that were detected at a perpendicular distance >80 m from the transect centreline. These were subsequently deleted from the final analyses.

Transect #	Length (km)	No. of detections	Transect #	Length (km)	No. of detections
1	6.79	–	11	15.64	7
2	17.96	–	12	13.64	1
3	18.61	–	13	11.75	12
4	25.31	3	14	6.38	2
5	26.51	3	15	5.47	1
6	19.14	–	16	5.23	3
7	18.45	3	17	4.04	–
8	18.82	–	18	3.25	10
9	18.22	–	19	2.75	1
10	15.67	8			

Examination of the data led to it being truncated at 80 m in order to improve the detection function model fitting process. This reduced the number of recorded detections made by both analyst from 54 to 51 horses. In relation to this, according to Buckland *et al.* (2001), in analysing distance sampling data there are two sampling criteria that, if possible, should be met to achieve a better than suitable outcome. The first of these is that the recommended minimum number of samplers (replicate transect lines) should be at least in the range 10-20 in order to ensure reasonably reliable estimation of the variance of the encounter rate. With there being 19 transects in the survey, it could be assumed that this criterion is being met. The second criterion is that the recommended number of observations, of detections of individual horses in this instance, should be at least in the range 60-80 for reliable modelling of the detection function. With there being only 51 horses in the truncated data set, this criterion is nearly but not really met. Hence, it is anticipated that the estimation of the variance of the encounter rate should be reliable but the modelling of the detection function may not be. Although, given that the detection functions being compared in this analysis are all simple models, fitting them using only 51 detections may not be a problem.

The truncated data were analysed using a Uniform key function without any adjustment (series expansion) terms to represent the strip transect analysis, and again using a Uniform key function with available options of either cosine or simple polynomial adjustments. The data were also analysed using a Half-normal key function with available options of either cosine or Hermite polynomial adjustments, and a Hazard-rate key function with available options of either cosine or simple polynomial adjustments. In the fitting of all the combinations of these detection function models, it was found that none of the resultant key function models incorporated adjustment terms. Hence, final comparisons were between the Uniform, Half-normal and Hazard-rate models.

For final model selection, Akaike's Information Criterion (AIC) was used. To recap in relation to this, the model producing the smallest AIC value was considered the most parsimonious (simplest and most suitable) detection function model for use in estimating horse density and abundance (Burnham & Anderson 2002). With regard to the calculation of the AIC for a particular model, it should be noted that an individual AIC value is, by itself, not interpretable due to its unknown interval scale.

For a given model, the AIC is only comparative, relative to other AIC values in a model set. Although the smallest AIC points to the most parsimonious model, for the purpose of model comparison, it is the differences in AIC values (ΔAIC) that are important. As ΔAIC increases beyond the value of 2.00, there increasing evidence that it is becoming increasingly less plausible that the associated fitted model is the best model, given the data. The converse to this is that when ΔAIC <2.00, it can be said that there is a level of empirical support for a model, given the data. This is a somewhat equivocal result, which, if required, can be accommodated through model averaging (Burnham & Anderson 2002).

Comparing the strip transect model (equivalent to a Uniform key function model without adjustments) with the Half-normal and Hazard-rate models found that this simple model was the most parsimonious of the three. In comparison with the Half-normal model, ΔAIC = 2.00, a value at which the difference between the two models being compared can be considered unequivocal. In comparison with the Hazard-rate model, ΔAIC = 3.99, a value at which the difference between the two models being compared can be considered to be increasingly unequivocal. Based upon these outcomes, analysing the results as data obtained from a strip transect survey is perhaps a suitable approach to adopt, depending upon the outcome of the MCMR analyses (see below). For the population estimates obtained using this method of strip transect analysis see Tables 2 and 3.

There are two things to note about this outcome. The first is that strip transect analysis is predicated on the assumption that the detection of objects is constant and certain across the whole of the survey strip, an assumption that can be relaxed when using CDS analysis to fit alternative models, if required. The second is that truncating the data can have the effect of strengthening this assumption of constant detectability. Truncating the data at 80 m had the intended effect of removing extraneous distant sightings. It also ensured a flattened the sightability curve (Fig. 2).

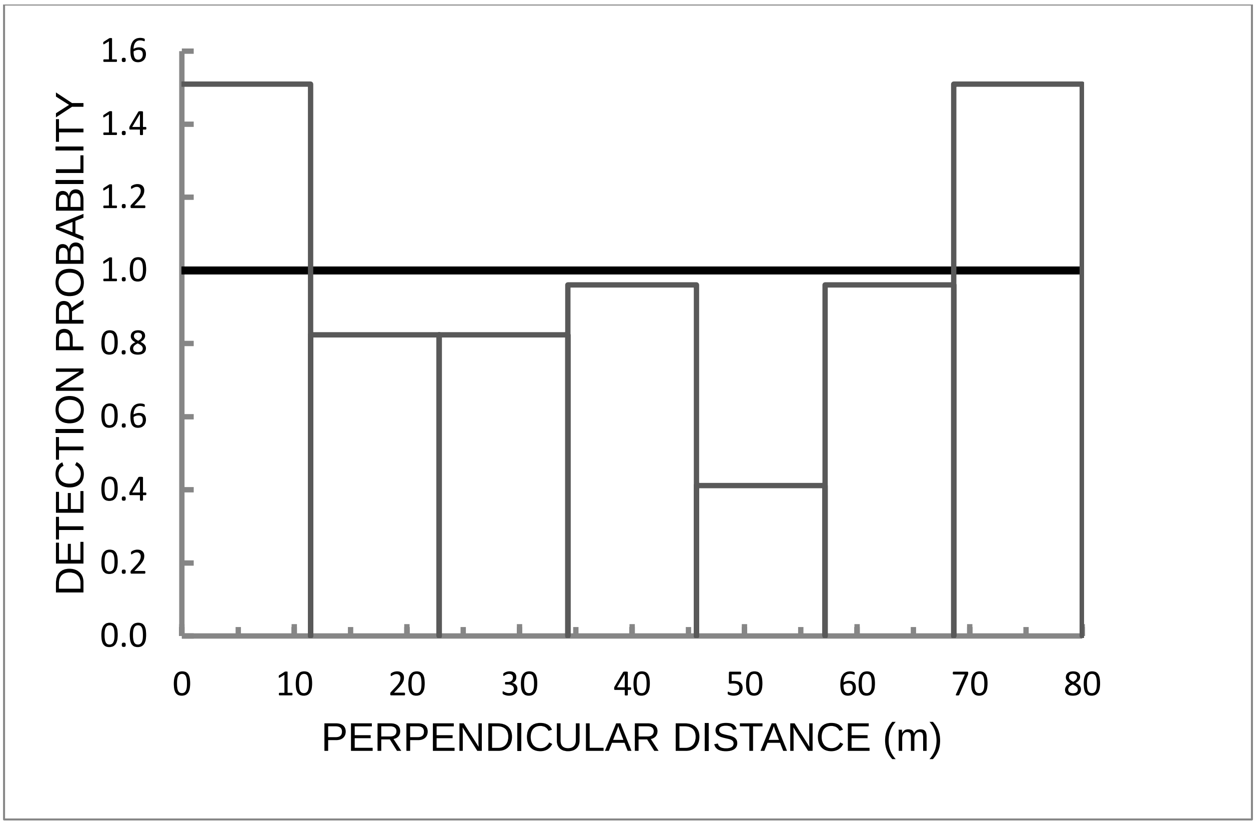

Fig. 2. The Uniform detection function for feral horses detected using thermal imagery during the 2021 survey of the Barmah NP.

Apart from providing density and abundance estimates, line transect analysis provides ancillary information with regard to the dynamics of a survey. This is in the form of two informative statistics, the probability (P_a) that a randomly selected object (horse) in the survey strip will be detected and the associated effective strip width (μ). The probability (P_a) that a randomly selected object (horse) will be detected is a straightforward statistic which is usually expected to be <1.00. If this is the case, then detection of an object would be less than certain across the width of a survey strip. This has been the case for previous surveys for feral horses conducted in Barmah NP (Cairns 2019). However, for the present survey, as was the case for the survey conducted in 2019 (Cairns 2020) with the most suitable detection function being that of a Uniform distribution characteristic of strip transect analysis, P_a =1.00. The associated effective strip width (μ), is interpreted as the perpendicular distance from the transect centreline (i.e. the half-strip width) for which as many objects are detected beyond that distance as remain undetected within that distance (Buckland *et al.* 2001). Hence, a line transect survey can be thought of as effectively covering a survey strip of a total area of $2L\mu$, for some value of $\mu \leq W$ and length L (Borchers

& Burnham 2004). By virtue of the way μ is determined ($\mu = W \times P_a$; where W is the nominal strip width of the survey transect), the higher the value of P_a, the wider will be the effective strip width. In the case of the present survey and analyses, μ turned out to be equivalent to the *a posteriori* truncated survey strip of 80 m.

Comparison of the density and abundance estimates obtained using strip transect and line transect analyses of survey data usually shows that the use of a strip transect approach will result, in most instances, in the underestimation the true density of horses in the survey area. This was the case for the 2018 survey, where detection of the horses was not uniformly certain across the survey strip which meant that the use of strip transect analysis would have resulted in a possible underestimation of the size of the population by some 25% (Cairns 2019). In that instance, this would have been the result of an overestimation of the average probability of detection. However, it was not the case in relation to the present survey, nor the one conducted in 2019 (Cairns 2020), where the detected horses were found to uniformly distributed across the survey strip. Usually, it is assumed that the objects being surveyed are uniformly distributed across the width of the nominal survey strip and that any deviation from this uniform distribution is a function of the detection process and is, accordingly, modelled as part this analysis. However, because of the way in which the data are collected in these surveys, using a thermal imagery camera, it is possible that where the modelled detection function turned out not to be uniform, that, on that occasion, the horses were not uniformly distributed across the survey strip. This could have the result of the recording of the horses during the imagery analysis as individuals rather than as clusters of individuals, which could possibly have been uniformly distributed across the survey strip.

With the use of conventional distance analysis methods, it needs to be noted that the fitting of any of the detection function models available in the CDS analysis engine, those using the Uniform, Half-normal or Hazard-rate key functions, is based upon the assumption that the probability of detection of an object on the transect centreline is certain, i.e. $P_o = 1$. If this is not the case, then even though the method of estimation using conventional distance sampling (CDS) analysis is robust, any resulting density estimates will be negatively biased unless, of course, the probability of detection of an object on the transect centreline is truly equal to one. If double count data that can be treated as capture-mark-recapture (CMR) data is available, then the issue of the true value of P_o can be explored using MRDS analysis engine and any negative bias in estimation can be avoided. This is done in the next section.

3.2 Capture-mark-recapture (MRDS) Analysis

In analysing the detections made by the two analysts in a double-observer configuration, the following three assumptions were made: the analysts are independent observers; all detections recorded are fully independent, with there being no reactive movement on the part of the observed objects (horses); and, the designated second analyst fulfils the role of what is described as the trial observer (tracker). Under this configuration, the function of the tracker is to generate "trials" for the principal observer (analyst), whereby the successes and failures of the principal observer to detect the "trials" set by the tracker (i.e. objects detected by the tracker) generate binary data from which the detection function of the primary observer can be determined (Burt *et al.* 2014). Analysing data in this way has the benefit of being able to estimate the actual probability of detection on the transect centreline ($g(0)$ = P_o). This can also be used as a check of the key assumption of the conventional forms of line transect sampling that P_o is equal to one; an assumption that is almost always not met (Laake & Borchers 2004).

Table 2. Results of the analyses of the 2021 CMR data using the MRDS and the CDS analysis engines. Given for each model are the ΔAIC for model selection, the goodness-of-fit (GOF) of the model to the data, the average probability of detection on the transect centreline (P_o = $g(0)$), the average probability of detection on the nominal survey strip (P_a), the estimated density (± 1 std. error) and population abundance (± 1 std. error) of horses for the survey area and the coefficient of variation of the estimates (cv%).

Model	ΔAIC	GOF	Average P_o	Average P_a	Density (ha^{-1})	Population abundance	cv (%)
MRDS – Petersen	214.85	0.005	0.277	0.523	0.022 ± 0.009	626.4 ± 236.9	37.8
CDS – Uniform (strip)	**0.00**	0.341	–	1.000	0.013 ± 0.004	328.0 ± 105.9	32.3
CDS – Half-normal	2.00	0.239	–	0.999	0.013 ± 0.004	328.0 ± 115.9	35.4
CDS – Hazard-rate	3.99	0.171	–	0.995	0.013 ± 0.004	328.0 ± 107.5	32.7

The MRDS analyses were conducted using the 80 m truncated data from the 2021 survey. Within the CMR framework, of the 51 horses detected by at least one analyst in the truncated (80 m wide) half survey strip on either side of the transect centreline, the principal observer detected 32 of them, missing 19 that were detected by the tracker. The trial (second) observer, the tracker, detected 30 horses, missing 21 that were detected by the principal observer. In the analyses, four MRDS models, including the standard Lincoln-Petersen estimator, were fitted to the data. The other three models were trial models, one being a full independence full independence (trial.fi) model, the other two being point independence (trial.pi) models. These models were compared with single observer models fitted using the Uniform key function (representative of strip transect sampling), the Half-normal and the Hazard-rate key functions. Comparisons among models were made using the AIC statistic. The results of these analyses are given in Table 2.

Of the seven models tested in the analyses, the most parsimonious model, by a clear margin in terms of the ΔAIC, was the strip transect analysis, whereby the Uniform key function was fitted to the 51 detections made by at least one analyst. Of the four MRDS models tested, all except the Lincoln-Petersen estimator failed to converge. The reason for this is unclear, although it might be due to the relatively small sample size and possibly the fact that most of the horses detected by the second (tracker) analyst were wide on the survey strip. The Lincoln-Petersen estimator, a model which does not include the covariate of distance, was, however, far from the simple Uniform distribution model (ΔAIC = 214.85).

Detection of objects (horses) on the transect centreline are assumed to be made with certainty in relation to CDS analyses. This assumption is relaxed and tested with the MRDS analyses. However, because the relevant models failed to converge, an estimate of P_o could not be determined, except for the Lincoln-Petersen estimator, which proved to be a less than suitable model when compared with the three CDS models (Table 2). In relation to this, and perhaps as a cautionary reminder, it should be noted that Laake *et al.* (2008) estimated P_o as being equal to 0.738 (± 0.085) for feral horses in the Australian Alps. The stark difference between this result and the assumption that P_o = 1.00 in relation to the CDS models may be, in part, due to the difference in survey methodology, but, nevertheless, be noted as a cautionary tale.

3.3 Population Estimates and Changes

All three of the CDS models produced essentially the same density and abundance estimates for 2021 (Table 2). The precision (CV%) of these estimates were also similar. The estimated size of the horse population in the 22,967 ha of the park surveyed is given, along with the corresponding estimate obtained from the two previous surveys, in Table 3.

Table 3. Abundance of feral horses (± 1 SD) for estimates obtained from aerial surveys conducted in June 2018, June 2019 and November 2021. Given also are the 95% confidence intervals (95% CI) and coefficients of variation (CV%) for these estimates along with the result of the test of any difference between successive population estimates. For an explanation of this test, see text.

Year	Population abundance	95% CI	CV%	Test statistic	H_o: $N_t = N_{t+1}$
2018	1,087.8 ± 335.6	593.1 – 1995.2	30.8	–	–
2019	621.0 ± 144.4	389.0 – 944.0	23.3	$z = 1.28$	$P = 0.100$
2021	328.0 ± 105.9	169.0 – 635.0	32.3	$z = 1.64$	$P = 0.051$

Examination of the population estimates given in Table 3 would suggest that the number of horses in survey area had declined between 2018 and 2019, and then again between 2019 and 2021. These apparent changes in numbers over the period 2018 to 2021 can be tested by undertaking a simple comparison between the successive abundance estimates using the standard normal z-score as the test statistic (Buckland et al. 2001, pp. 84-86).

The results of these two tests of differences are given in Table 3, where listed, along with the z-score are, under the null hypothesis that was tested, the one-tailed probability for its acceptance in relation to an alternative hypothesis that, in 2019 the size of the horse population was lower than it was in 2018 and in 2021 the population was lower than it was in 2019 (H_A: $N_t \geq N_{t+1}$). Although the use of a one-tailed test is somewhat uncommon, its use here could be countenanced on the basis that the null hypothesis can be framed in relation to an alternative of a decline rather than just a change in abundance. In this case, the outcomes of these analyses would appear to indicate that the number of feral horses in the survey area of

Barmah NP had probably declined between 2018 and 2019 (P = 0.100) and definitely declined between 2019 and 2021.

5. Acknowledgements

As with any project, the job is never completed without the support of others who either wittingly or unwittingly are drawn in to provide assistance. Thanks go to Dale Appleton and the helicopter crew for the conduct of the survey that provided thermal imagery data for analysis. The analyses undertaken in this project were all based on data provided by the two expert analysts who reviewed the video stream recorded during the surveys. Thanks go again to Dale Appleton and Patrick Pigott for ably performing this task. Thanks also go to John Wright for liaising with me on behalf of Park Victoria and for his positive attitude and valuable discussions with regard to the broader issues associated with this project.

6. References

Borchers, D. L. & Burnham, K. P. (2004). General formulation for distance sampling. In *Advanced Distance Sampling* (eds S T. Buckland, D R. Anderson, k. P. Burnham, d. L. Borchers & L. Thomas), pp 6-30. Oxford University Press, Oxford.

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.

Burnham, K. P. & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer.

Burt, M. L., Borchers, D. L., Jenkins, K. & Marques, T. A. (2014). Using mark-recapture distance sampling methods on line transect surveys. *Methods in Ecology and Evolution*, **5**, 1180-1191.

Cairns, S. C. (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, September 2014. 58 pp.

Cairns, S. C. (2019). *Analysis of the 2017 and 2018 aerial surveys of the feral horse population in the Barmah National Park, Victoria.* A report to Parks Victoria. 24 pp.

Cairns, S. C. (2020). *Analysis of the 2019 aerial surveys of the feral horse population in the Barmah National Park, Victoria.* A report to Parks Victoria. 22 pp.

Caughley, G. (1977). *Analysis of Vertebrate Populations*. John Wiley & Sons, London.

Durant, S. M., Craft, M. E., Hilborn, R., Bashir, S., Hando, J. & Thomas, L. (2011). Long-term trends in carnivore abundance using distance sampling in Serengeti National Park, Tanzania. *Journal of Applied Ecology*, **48**, 1490-1500.

Laake, J. & Borchers, D. (2004). Methods for Incomplete Detection at Distance Zero. In: Advanced Distance Sampling, eds. S. Buckland, D. Anderson, K. Burnham, J. Laake, D. Borchers & L. Thomas. OUP, Oxford. Pp. 108-189.

Laake, J. L., Collier, B. A., Morrison, M. L. & Wilkins R. N. (2011). Point-based mark-recapture distance sampling. *Journal of Agricultural, Biological and Environmental Statistics*, **16**, 389-408.

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

Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters.* Macmillan, New York.

Thomas, L., Buckland, S. T., Burnham, K. P., Anderson, D. R., Laake, J. L., Borchers, D. L. & Stringberg, S. (2002). Distance sampling. In: *Encyclopaedia of Environmentrics* (eds. A. H. El-Shaarawi and W. W. Piegorsch). Volume 1, pp. 544-552.

Thomas, L., Buckland, S. T., Burnham, K. P., Anderson, D. R., Laake, J. L., Borchers, D. L. & Strindberg, S. (2002). Distance sampling. In: *Encyclopaedia of Environmentrics* (eds. A. H. El-Shaarawi and W. W. Piegorsch). Volume 1, pp. 544-552.

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. (2010). Distance software: design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology* **47**: 5-14.

Thompson, S. K. (2002). *Sampling*. John Wiley & Sons Inc., New York.