CSIRO PUBLISHING

Wildlife Research, 2008, **35**, 320–330 www.publish.csiro.au/journals/wr

A comparison of mark–recapture distance-sampling methods applied to aerial surveys of eastern grey kangaroos

Rachel M. Fewster[A,C] *and Anthony R. Pople*[B]

[A]Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand.
[B]Biosecurity Queensland, Department of Primary Industries and Fisheries, GPO Box 46, Brisbane, Qld 4001, Australia.
[C]Corresponding author. Email: r.fewster@auckland.ac.nz

Abstract. Aerial surveys of kangaroos (*Macropus* spp.) in Queensland are used to make economically important judgements on the levels of viable commercial harvest. Previous analysis methods for aerial kangaroo surveys have used both mark–recapture methodologies and conventional distance-sampling analyses. Conventional distance sampling has the disadvantage that detection is assumed to be perfect on the transect line, while mark–recapture methods are notoriously sensitive to problems with unmodelled heterogeneity in capture probabilities. We introduce three methodologies for combining together mark–recapture and distance-sampling data, aimed at exploiting the strengths of both methodologies and overcoming the weaknesses. Of these methods, two are based on the assumption of full independence between observers in the mark–recapture component, and this appears to introduce more bias in density estimation than it resolves through allowing uncertain trackline detection. Both of these methods give lower density estimates than conventional distance sampling, indicating a clear failure of the independence assumption. The third method, termed point independence, appears to perform very well, giving credible density estimates and good properties in terms of goodness-of-fit and percentage coefficient of variation. Estimated densities of eastern grey kangaroos range from 21 to 36 individuals km^{-2}, with estimated coefficients of variation between 11% and 14% and estimated trackline detection probabilities primarily between 0.7 and 0.9.

Introduction

The commercial harvest of kangaroos (*Macropus* spp.) in Australia is managed through an annual quota system, where quotas are set as a percentage of the estimated population size (Pople and Grigg 1998). Most harvesting occurs over vast, relatively open regions of arid and semiarid Australia, so population estimates are largely obtained from aerial surveys (e.g. Caughley and Grigg 1981). Traditionally, aerial surveys used strip transects with fixed-wing aircraft (Caughley *et al*. 1976), but more recently there has been increasing use of line-transect surveys with helicopters (Clancy *et al*. 1997; Southwell and Sheppard 2000). Roughly 50% of the Australian kangaroo harvest is taken in Queensland, where the size of the kangaroo population is determined almost solely from helicopter surveys (Lundie-Jenkins *et al*. 1999). Helicopter surveys are more expensive than fixed-wing surveys, but have the advantage of substantially improved visibility (Pople *et al*. 1998*a*, 1998*b*).

Bias in estimates of population size can lead to over- or underharvesting, so there has been considerable research effort in estimating kangaroo population size accurately (Pople 2004). Two methodologies for estimating population size are line-transect distance sampling (Buckland *et al*. 2001), and mark–recapture methods using two simultaneous observers. Both methods rely on assumptions that might be difficult to meet for aerial surveys of kangaroos.

Conventional line-transect distance sampling makes the assumption that all animals directly on the transect line are detected with certainty. This assumption is unlikely to be true for aerial surveys, because of the difficulties of detection from high altitude. In general, this method will overestimate detection probability, and therefore underestimate density of kangaroos. This was noted by Clancy *et al*. (1997), who compared results from helicopter line-transect surveys with those from walked line-transect surveys. Estimates of red kangaroo (*Macropus rufus*) and eastern grey kangaroo (*Macropus giganteus*) density were found to be similar between helicopter surveys and walked line transects, but helicopter estimates were 2–3 times lower than ground-based estimates for common wallaroos (*Macropus robustus*). Furthermore, walked line-transect surveys are themselves prone to underestimate density, because of evasive responsive movement by the kangaroos away from the observers (Southwell 1994).

Mark–recapture methods have been used in aerial surveys of both aquatic and terrestrial fauna in Australia (Marsh and Sinclair 1989; Bayliss and Yeomans 1989; Pople *et al*. 1998*c*), including kangaroos (Choquenot 1995). They are often referred to as 'double-count' methods. The methods involve two teams of observers counting independently along the same transect. A protocol is established for identifying which animals were seen by both observers (known as duplicates), and which were seen by only the first or only the second observer. The mark–recapture method relies on an assumption that all factors affecting detectability can be identified, and that their influence

 10.1071/WR07078 1035-3712/08/040320

on detectability can be correctly modelled. If there is unmodelled heterogeneity in detection probabilities, density estimates will be biased (Caughley and Grice 1982).

Owing to the known shortcomings of both conventional distance sampling and mark–recapture analysis methods for aerial and other surveys, new methods that combine the strengths of both approaches have recently been developed. The new methods are collectively known as mark–recapture distance-sampling methods. Roughly speaking, they use the mark–recapture data gained from the independent observers to estimate each observer's detection probability on the transect line. This removes the need for the conventional distance-sampling requirement that detection is certain on the line. The distance-sampling data can then be used in conjunction with the estimated on-line detection probabilities to estimate animal density.

The basic principle above is common to all mark–recapture distance-sampling methods, but there are many different ways in which the mark–recapture and distance-sampling data may be combined. This has led to a confusing array of analysis methods, all of which have individual strengths and weaknesses. Effort is now required to establish firm analysis guidelines for mark–recapture distance-sampling data.

In this article, we describe the implementation of mark–recapture distance-sampling survey methods for aerial surveys of kangaroos in Queensland. Using the survey data for eastern grey kangaroos in four survey blocks, we aim to showcase the flexibility of mark–recapture distance-sampling analysis techniques for estimating kangaroo density. We also aim to compare results from three different approaches to analysis of mark–recapture distance-sampling data, and to assess the suitability of the different approaches for future surveys.

The first of our analysis methods can be described as a mark–recapture analysis in which distance is included as a covariate, and is described in full by Borchers *et al*. (1998*a*, 1998*b*). This method is based on a mark–recapture likelihood, and distance of the animals from the transect line enters the model only as a covariate. We will refer to this as our Method *M*.

The second method is the full-likelihood analysis of Borchers *et al*. (1998*a*), in which the likelihood includes components based on mark–recapture, distance sampling, and other covariates in addition. We refer to this as Method *F*. Method *F* has the advantage that the distance-sampling data are included explicitly in the likelihood, but it has the disadvantage that it is necessary to specify likelihood components for quantities about which there might be very little information (for example, the distribution of animal group size in the population).

The third method (Method *P*) is the point-independence method of Borchers *et al*. (2006), which includes the mark–recapture and distance-sampling components of the likelihood, but not the other components. Additionally, the point-independence method relaxes an important assumption of the other two methods, namely that the two observers make detections independently at all distances, once the observed set of covariates is accounted for. Under point independence, it is only necessary for the observers to be independent at a single point or distance, which is generally taken to be zero distance from the transect line. This addresses the key problem of unmodelled heterogeneity in mark–recapture analyses. If there are covariates that simultaneously make both observers more likely to detect an animal group (for example, a movement cue), but which are not included in the model, the result is that detections from the individual observers are no longer independent, given only the set of covariates that the model does include.

The point-independence Method *P* is a new method and may represent an important advance over previous techniques. As such, it is important to evaluate the methodology relative to other techniques in a range of applications. The point-independence method has been applied to aerial surveys of pack-ice seals and penguins in Antarctica (Southwell *et al*. 2007, 2008) and to Australian wild horses (Laake *et al*. 2008). Laake *et al*. (2008) provide a more thorough description of the various analysis approaches than we give here, and compare the point-independence Method *P* with conventional distance-sampling and mark–recapture approaches.

This article is among the first to explore the performance of Method *P* relative to the more sophisticated of the previous techniques for combining distance-sampling and mark–recapture data. Borchers *et al*. (1998*a*) compared Methods *M* and *F* together, and demonstrated that they did not always give equivalent answers. Borchers *et al*. (2006) briefly displayed problems with the mark–recapture Method *M* compared with Method *P*, although they did not examine Method *F*. Laake and Borchers (2004) also compared Methods *M* and *P*, together with another mark–recapture distance-sampling method, using a known population of golf tees. By comparing Methods *M*, *F* and *P* together for a real aerial survey application, we aim to validate previous expectations about their relative performance, and assist in providing clear guidelines for the appropriate analysis of aerial survey data.

Materials and methods

Study area

The survey was conducted in two survey blocks of ~10 000 km^2 centred on the southern Queensland towns of Roma, in the brigalow belt bioregion, and Charleville, in the mulga lands bioregion. The Roma block is a mixture of tussock grasslands mostly sown to cereal crops, and *Eucalyptus*, *Casuarina* and conifer open grassy woodlands. Eastern grey kangaroos are abundant in the Roma district, whereas red kangaroos and common wallaroos occur at relatively low densities. The Charleville block is dominated by low *Acacia* and *Eucalyptus* woodlands, with a lower stratum of tall shrubs and tussock grasses. Large areas of the original vegetation around Charleville and much of the mulga lands have been 'pulled' with tractors and chains, but not cleared, leaving areas of fallen timber and regenerating vegetation. All three species of macropod are common in the Charleville district. Sheep grazing is the principal form of land use in both blocks, with cereal crops being grown around Roma.

Helicopter mark–recapture distance-sampling methods

Independent-observer surveys were undertaken during the annual aerial survey of 10 blocks in 2000. Clancy *et al*. (1997) give a full description of the independent-observer procedures, also known as double counting. Briefly, a Robinson R44 helicopter with the doors removed was flown at a ground speed of 93 km h^{-1} (50 kts), at 61m (200ft) above ground level. The helicopter housed a double-observer team sitting in the front and back left positions, simultaneously recording sightings to the left side of the transect line. Additionally, there was a third observer in the back right position, recording sightings to the right side of the aircraft. The right-side sightings do not contribute to the double-observer scheme, so they are not included in this analysis.

The two back-seat observers (A and B) rotated positions between flights, while the front-seat observer (C) retained the same position throughout the survey. The two observer teams, C–A and C–B, operated for roughly equal amounts of survey effort. Due to differences in individual observer patterns, we analyse data from the different observer teams separately. This provides data from four survey blocks, which we call Roma1 (Roma block with team C–A), Roma2 (Roma block with team C–B), Char1 (Charleville block with team C–A) and Char2 (Charleville block with team C–B).

Observers searched ahead of the helicopter and to the side to detect clusters of kangaroos. Detected clusters were placed into 25-m distance classes up to 125 m perpendicular to the transect line, measured from directly below the observer. The distance classes were delineated on aluminium poles extending perpendicularly from either side of the helicopter. The front-seat observer used a separate pole extending out from the helicopter. Distances were measured to the position at which the animal was first seen, to mitigate problems of responsive movement. Species and cluster size were also recorded with each sighting.

Sightings were recorded into microcassette recorders. The sightings of the two left-side observers constituting the double-observer team were recorded independently in continuous time into a dual-channel tape recorder. Following the survey, tapes were replayed and sightings were identified as being made either by the front observer only, the rear observer only, or both observers (duplicates). All three observers had over 100 h of experience in helicopter surveys of kangaroos using line-transect sampling.

Helicopter surveys in both Roma and Charleville followed parallel east–west transect lines, ~80 km long and 10 km apart. The exact distances were determined by a global positioning receiver. Four lines were completed in three of the survey blocks, and three lines were completed in the Char2 block. The observers counted in 5-minute units with a 30-s break between them, so effort was divided into consecutive transect segments each separated by ~800 m. Surveys were conducted during the 3 h after sunrise and the 2 h before sunset in late May 2000.

Statistical analysis

A complete likelihood for mark–recapture distance-sampling data is provided by Laake and Borchers (2004) and Borchers *et al*. (1998*a*, 2006). Laake and Borchers (2004) and Borchers *et al*. (2006) outline several of the possible options for analysing these data, and show that the different options arise from using different components of this full likelihood. Within each option there are generally submethods arising from different analysis choices. To avoid making our discussion too complicated, we select one method from each of the three broad analysis choices. We use the notation of Laake and Borchers (2004) and Borchers *et al*. (2006), and leave most of the mathematical detail to these sources.

Let y_i be the perpendicular distance of a kangaroo cluster from the transect line. Let ω_i be the observed capture history for detected cluster *i*. This means that ω_i is either (1, 0), (0, 1), or (1, 1), respectively denoting detection of cluster *i* by Observer 1 only, Observer 2 only, or both observers (a duplicate). Let z_i be the cluster size of group *i*. For the analysis of eastern grey kangaroos, z_i is the only recorded covariate assumed to have an influence on detection probability in addition to y_i. The unknown number of clusters in the search strips, which we wish to estimate, is written N_c. The total number of sightings made by both observers combined is $n_\bullet$.

The full likelihood for mark–recapture distance sampling can be written as follows:

$$L(N_c,\phi,\theta) = L_{n_\bullet}(N_c,\phi,\theta) \times L_z(\phi,\theta) \times L_{y|z}(\theta) \times L_\omega(\theta). \qquad (1)$$

The individual components are explained below. Here, N_c is treated as an unknown parameter, and it is the only parameter that is of real interest. The other parameters, (ϕ, θ), are nuisance parameters that are needed to explain how our survey data relate to the parameter of interest, N_c. The parameters θ control the detection process. This means that θ determines the detection functions of Observer 1 and Observer 2, which give the probability that the observer will detect a cluster with specified size and at specified distance from the line. The parameters ϕ control the unknown statistical distribution of cluster size in the population. For example, we will model cluster size using a negative binomial distribution with parameters ϕ to be estimated. We need a term in ϕ because the distribution of cluster size among the detections is not likely to be the same as that in the overall population, because larger clusters are usually more detectable and are therefore over-represented in the detections.

The likelihood formulation is explained as follows. Each of the four components builds upon the previous ones using a conditional hierarchy. For a cluster to be included in the data, it must be detected by at least one observer. Overall detection is modelled by the binomial component $L_{n_\bullet}(N_c,\phi,\theta)$, which relates the total number of detections $n_\bullet$ directly to the number present, N_c, taking into account the overall detection probabilities controlled by the nuisance parameters ϕ and θ.

Given overall detection, the observed cluster size *z* has a distribution that depends on the distribution of cluster sizes in the population at large (controlled by ϕ), and their relative detectabilities (controlled by θ). The component $L_z(\phi,\theta)$ models the distribution of detected cluster sizes. This component is a nuisance component, with a distinct disadvantage that it forces us to articulate a statistical model such as negative binomial for cluster size. If *z* contained extra covariates in addition to cluster size, this component might become intractable, because we would then have to specify a model for the joint distribution of all covariates in the population at large. For example, a model might be needed for the joint distribution of cluster size and weather conditions. Such a model is likely to be fanciful at best.

Given both overall detection and cluster size, the detection distance *y* has a distribution captured by $L_{y|z}(\theta)$, which takes into account the fall-off of detections with distance from the line, for this cluster size. The component $L_{y|z}(\theta)$ is the distance-sampling component. It depends on θ only, and requires the assumption that distances are uniform in the population at large, before detection.

Finally, given all three of overall detection, cluster size z, and detection distance y, the mark–recapture component $L_\omega(\theta)$ describes the detailed capture history ω: whether the detection is made by Observer 1 only, Observer 2 only, or both. Each observer's sightings serve as 'marks' for the other observer, who 'recaptures'. The component $L_\omega(\theta)$ depends only upon detection parameters θ, and makes no assumptions about the distribution of either y or z in the population at large.

All four of the components involve the detection parameters θ, which enter via modelled detection functions specified in Models 1–4 in the detection model section below. In particular, the mark–recapture component $L_\omega(\theta)$ and the distance-sampling component $L_{y|z}(\theta)$ show how the combined mark–recapture distance-sampling analysis uses two different aspects of the data to give us strengthened evidence about the detection process θ.

From the discussion above, it is clear that some likelihood components are more desirable than others. The component $L_z(\phi,\theta)$ will often be problematic to formulate, and the binomial component, $L_{n_\bullet}(N_c, \phi,\theta)$, relies on a binomial detection process. However, these less desirable components each involve the detection parameters θ, so we risk losing information about θ if they are omitted and θ is estimated only from the other components. This conspires to ensure that the best choice of analysis method is not clear-cut.

The various different analysis options operate by using only selected components of the likelihood, and using empirical methods (primarily Horvitz–Thompson-based methods) to fill in the remaining pieces of missing information. For example, if the detection process θ is estimated using the components $L_{y|z}(\theta)$ and $L_\omega(\theta)$ only, we do not gain an estimated distribution for cluster size because $L_z(\phi,\theta)$ is missing, but it can be estimated empirically. Specifically, if it is estimated that a cluster of size 1 is detected with overall probability 0.5, then we estimate that for every detected cluster of size 1, there were two such clusters in the population. This process is repeated for clusters of every different size, to form an empirical distribution of cluster size. Estimates of N_c follow similarly. The Horvitz–Thompson-like approach on which these methods operate is discussed in detail by Borchers *et al*. (1998b).

The analysis options that we examine in this paper are as follows.

Method M*: mark–recapture component*

This method uses only the mark–recapture component $L_\omega(\theta)$ to estimate θ, and all other quantities are estimated by the Horvitz–Thompson-like approach. We implement this method exactly as in Borchers *et al*. (1998*a*), although with some different detection models described below.

Method *M* can be described as a mark–recapture analysis with distance as a covariate. It is one of the better known methods in the literature. However, it is known to be extremely sensitive to unmodelled heterogeneity in capture probability, and it does not exploit the full power of mark–recapture distance-sampling methods because it omits the distance-sampling component $L_{y|z}(\theta)$. An advantage of this is that it does not rely on an assumption of uniform distances in the population at large, which is needed for the $L_{y|z}(\theta)$ component and can be violated if there is persistent responsive movement of the animals towards or away from the observer. Fewster *et al*. (2008) discuss in detail the relative merits of the uniform assumption against the assumptions underlying Method *M*.

In Method *M*, the two observers are assumed to detect animals independently at all distances, given the known covariates. Method *M* has recently been implemented in the software Distance 5 (Thomas *et al*. 2006) under the name 'Full Independence' in the MRDS engine. Our implementation differs from that of Distance 5 only in the formulation of some of the detection functions (described below), and in variance estimation, for which we use the bootstrap.

Method F*: full likelihood*

Method *F* uses all components of the full likelihood in Eqn (1). Our implementation of Method *F* follows that in Borchers *et al.* (1998*a*), although with some different detection models. Method *F* is not implemented in Distance 5. It is tractable in the kangaroo case because we have only one covariate, cluster size. As discussed above, it is not likely to be suitable for analyses with complex covariate sets. However, we include it in the present analysis to check on the influence of the omitted likelihood terms in Method *M*.

Our implementation of Method *F* assumes that the two observers detect animals independently at all distances, given the known covariates. Unlike Method *M*, this assumption is not necessary for Method *F*, and a point-independence application of Method *F* would be possible.

Method P*: point independence*

Method *P* is a new method described by Laake and Borchers (2004) and Borchers *et al*. (2006), and now implemented in Distance 5. Method *P* uses the two likelihood components $L_{y|z}(\theta) \times L_\omega(\theta)$. It therefore exploits the combined strength of mark–recapture and distance sampling, but omits the more controversial components $L_z(\phi,\theta)$ and $L_{n_\bullet}(N_c,\phi,\theta)$. As discussed above, this omission has both advantages and disadvantages.

The principal advance of Method *P*, however, is the introduction of a new idea called point independence. This provides a solution to the problem of unmodelled heterogeneity in capture probabilities that plagues all mark–recapture methods. Unmodelled covariates that simultaneously affect the detection probabilities of both observers have the effect of inducing dependence between the observers. If Observer 1 sees the animal, it is possible that Observer 2 also sees it for the same, unknown, reason. The assumption of full independence of observers, which underpins all previous mark–recapture distance-sampling analyses, is therefore violated.

Method *P* circumvents this problem by allowing a different form for the detection function in the likelihood component $L_{y|z}(\theta)$ from that in component $L_\omega(\theta)$. The mark–recapture component, $L_\omega(\theta)$, is interpreted as estimating the *conditional* probability that Observer 1 detects the animal, *given* that Observer 2 has detected it, and *vice versa*. This contrasts with previous interpretations of this component, which assume (often incorrectly) that $L_\omega(\theta)$ involves the unconditional probability that each observer detects the animal, and therefore uses the same detection functions as the distance-sampling component.

Borchers *et al*. (2006) show that the distance-sampling component can actually be modelled with a separate detection function, which removes the need for the assumption that $L_\omega(\theta)$ is estimating the unconditional detection functions. As long as the observers can be assumed to be acting independently at a single value of y (usually the point $y = 0$, on the transect line), it is possible to use different forms for the conditional detection functions in $L_\omega(\theta)$ and the unconditional distance-based detection function in $L_{y|z}(\theta)$. Conceptually, the two methods only have to agree at a single point to enable us to 'hook' the distance-based detection function onto an estimable intercept, and from there on it can be estimated separately.

Our implementation of Method *P* is exactly the same as that in Distance 5, except that we use the bootstrap to estimate variance. Because Method *P* already involves two different detection functions, one for component $L_\omega(\theta)$ and one for component $L_{y|z}(\theta)$, we simplify our approach to this method by exploring only one functional form for each.

Detection models

An appealing aspect of mark–recapture distance sampling is the ability to accommodate many different forms of dependence of detectability on distance, cluster size, and other covariates. To showcase this flexibility, we fitted Methods *M* and *F* using four different detection models. Their functional forms are given below. Examples of the resulting dependence of detection on distance and cluster size are shown in Fig. 1.

Let $p_o(y,z)$ be the probability that observer o detects a cluster of size z located at distance y from the transect line, for observers $o = 1, 2$. All four likelihood components in (1) involve the functions $p_o(y,z)$ and expressions derived from them (Borchers *et al*. 2006). The different detection models specify different functional forms for $p_o(y,z)$.

Model 1: 'Logit'

This model was used in Borchers *et al*. (1998*a*, 1998*b*, 2006) and is the only option currently implemented in Distance 5. It involves parameters α_o, β_o, and γ_o, to be estimated for each observer.

$$p_o(y,z) = \left(1 + \frac{1}{\exp(\alpha_o + \beta_o y + \gamma_o z)}\right)^{-1}. \quad (2)$$

Model 2: 'Shoulder'

Fig. 1 shows that the logit model does not give the detection functions a 'shoulder' close to $y = 0$ when the cluster size is small. To allow fully shouldered detection functions, which

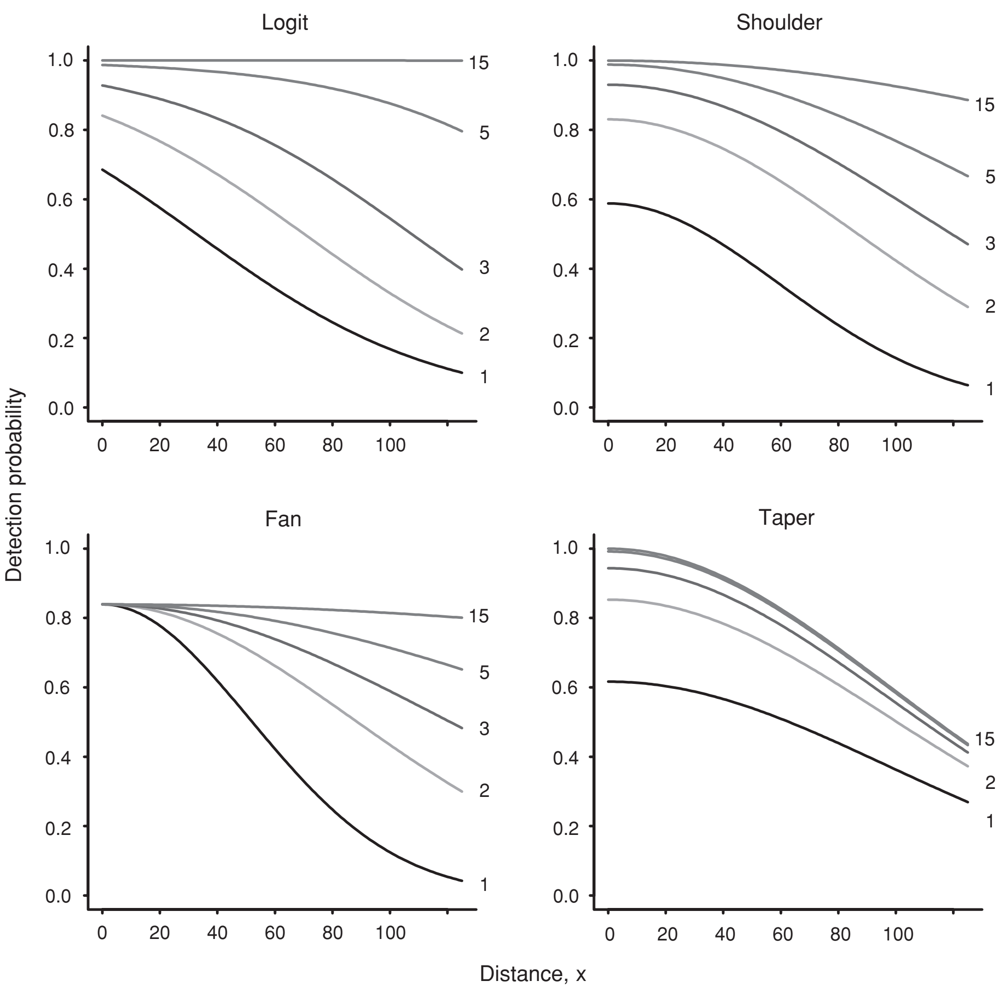

Fig. 1. Fitted detection functions from the four detection models (Logit, Shoulder, Fan, and Taper). Results are shown for the front-seat observer in the Roma1 survey block using the full-likelihood with full-independence method (*F*). Each figure shows the fall-off of detection probability with distance for a range of different cluster sizes. The cluster size is displayed to the right of each curve.

otherwise resemble the logit model, we suggest the following functional form. The shoulder comes at the expense of an extra parameter, δ_o:

$$p_o(y,z) = \alpha_o \exp\left(-\frac{y^2}{\beta_o z^{\gamma_o}}\right)\left\{1-\exp(-\delta_o z)\right\}. \quad (3)$$

Model 3: 'Fan'

This form describes the case where cluster size does not affect detectability close to the transect line, but has an increasing effect as distance increases. Fig. 1 shows the resulting fan-like detection pattern.

$$p_o(y,z) = \alpha_o \exp\left(-\frac{y^2}{\beta_o z^{\gamma_o}}\right). \quad (4)$$

Model 4: 'Taper'

The 'taper' model is the opposite of the 'fan' model, in which cluster size has the greatest impact on detectability at low distances, and the smallest impact at high distances. The detection function is gained by omitting the term z^{γ_o} in (3):

$$p_o(y,z) = \alpha_o \exp\left(-\frac{y^2}{\beta_o}\right)\left\{1-\exp(-\delta_o z)\right\}. \quad (5)$$

For Method P, we restrict attention to the logit detection model (2) for the $L_\omega(\theta)$ likelihood component, and we note that the detection function has a different interpretation under Method P. It is no longer the unconditional probability that observer o detects a cluster of size z located at distance y, but is rather the conditional probability given that the other observer detects this group. For Method P, we therefore write:

$$p_{1|2}(y,z) = \left(1+\frac{1}{\exp(\alpha_1+\beta_1 y+\gamma_1 z)}\right)^{-1};$$

$$p_{2|1}(y,z) = \left(1+\frac{1}{\exp(\alpha_2+\beta_2 y+\gamma_2 z)}\right)^{-1}, \quad (6)$$

where $p_{1|2}(y,z)$ is the conditional probability that Observer 1 detects a cluster of size z located at distance y, given that it is detected by Observer 2, and similarly for $p_{2|1}(y,z)$.

For Method P, we also have to specify an overall detection function for the distance-sampling component $L_{y|z}(\theta)$. We use the half-normal model available in Distance 5:

$$g(y,z) = \exp\left[-\left\{\frac{y}{\exp(\theta_1+\theta_2 z)}\right\}^2\right],$$

where $g(y,z)$ specifies the shape of the overall detection function with distance, while its intercept is obtained from the conditional capture–recapture detection functions (6) at their point of independence ($y = 0$).

Variance estimation

We use bootstrap variance estimation for all methods, with transect segments as the resampling unit. Segments were ~8 km long and separated by a relatively small gap of ~800 m, which raises possible concern about non-independence between segments. We used a Durbin–Watson test to verify that there was no evidence of dependence between adjacent segments. Segments were resampled at random with replacement until the resampled survey effort first exceeded the effort in the original dataset.

Model selection and goodness-of-fit

We used AIC to assess the relative merits of different models. Within a single method, such as Method M or Method F, AIC can be used to select between the different detection models (2) to (5). Comparing the different methods M, F and P is more complicated because they are based on different likelihood components. This means that the raw AIC values should be expected to differ by large amounts, without providing any implication as to the relative merits of the different models. We resolve this by evaluating the AIC for each method relative to the full likelihood expression given in Eqn (1), following Borchers *et al.* (2006: 375). For example, Method M provides only an estimate of θ from the likelihood, but this estimate is used to construct estimates of N_c and the distribution of z using the Horvitz–Thompson-like method described above. These Horvitz–Thompson-based estimates can be fed into the likelihood expression (1) just as they are for Method F, so that all three methods can be assessed on the same scale. In principle, this procedure puts both Methods P and M at a disadvantage relative to Method F, because Method F returns the parameter estimates that genuinely maximise the full likelihood (1), while neither P nor M are expected to do so. In practice, however, the Horvitz–Thompson estimates have greater flexibility than the models chosen to formulate (1), so Methods P and M can perform better than Method F.

We assess goodness-of-fit for each model by plotting the overall detection distances, for both observers combined, against the estimated overall detection functions. We use a Chi-square goodness-of-fit test to assess whether the observed numbers of detections in the five distance classes are consistent with the numbers expected from the fitted detection functions.

Results

Identification of duplicate sightings

Sightings were categorised as duplicates when they were near simultaneous, of the same species, and the same cluster size. However, because kangaroos were often moving and observers recorded sightings at slightly different times, some joint sightings were placed in different distance classes by each observer (13% of all macropod sightings at Charleville, 9% at Roma). Kangaroos form small, loose aggregations, so some joint sightings were also recorded with different cluster sizes by the two observers (7% of all sightings at Charleville, 10% at Roma). Misidentification also occurred, particularly with mixed species groups, and some duplicate sightings were recorded as different species (5% of all sightings at Charleville, 2% at Roma).

If near-simultaneous sightings by the two observers were in adjoining distance classes, they were categorised as being the same cluster. However, if the cluster size differed by more than one, the sighting was not considered a duplicate. If different species were recorded near simultaneously by the two observers in the same distance class and within one of the same cluster size, the sighting was considered a duplicate. In all cases, the front observer's records for duplicate sightings were used in the analysis. This treatment of errors is conservative, intended to lead to an overestimate of sighting probability on the line and therefore an underestimate of density.

Detection models for Methods M and F

Fig. 1 shows fitted detection functions from all four detection models (2) to (5) for the front-seat observer in the Roma1 block using Method *F*. The figure shows the flexibility of the methods, and also demonstrates that the choice of detection model can have an appreciable influence on the final result. The results in Fig. 1 are all obtained from the same dataset.

Distance data for the kangaroo surveys were collected at relatively low resolution, by grouping the data into five distance classes. The power of the data to discriminate between different shapes of detection functions is therefore relatively low. All four of the proposed models (2) to (5) were selected by AIC as the favoured model for at least one of the four survey blocks and either Method *M* or Method *F*. The logit model (2) was selected most frequently (3 out of 8 analyses), and was the only model that never performed poorly. All the other models had at least one result with $\Delta AIC > 10$, while the worst result for the logit model was $\Delta AIC = 4.3$. We conclude that, for these data, the logit model is consistently the best of the four forms proposed. This affirms the suitability of this model as the only option currently available in Distance 5.

Comparison of Methods M and F

Fig. 2 shows a comparison of fitted detection functions for the front-seat observer in Roma1 and Roma2 using Methods *M* and *F*. The two sets of results from Roma1 are quite similar. This suggests that the mark–recapture component $L_\omega(\theta)$ used in Method *M* reflects much the same information about the detection parameters θ as the other components used in the full likelihood, for the Roma1 data.

However, the results for the Roma2 survey block are very different. Detection under Method *M* is estimated to increase with distance, while Method *F* indicates a clear decrease with distance. These data provide an excellent illustration of the potential pitfalls of restricting to the mark–recapture Method *M*. The reason for the discrepancy can be seen in Fig. 3, in which the first barplot for Roma2 shows the underlying data on which Method *M* relies. A small sample of five 'marks' from Observer 2 in the furthest distance class, all of which were, by chance,

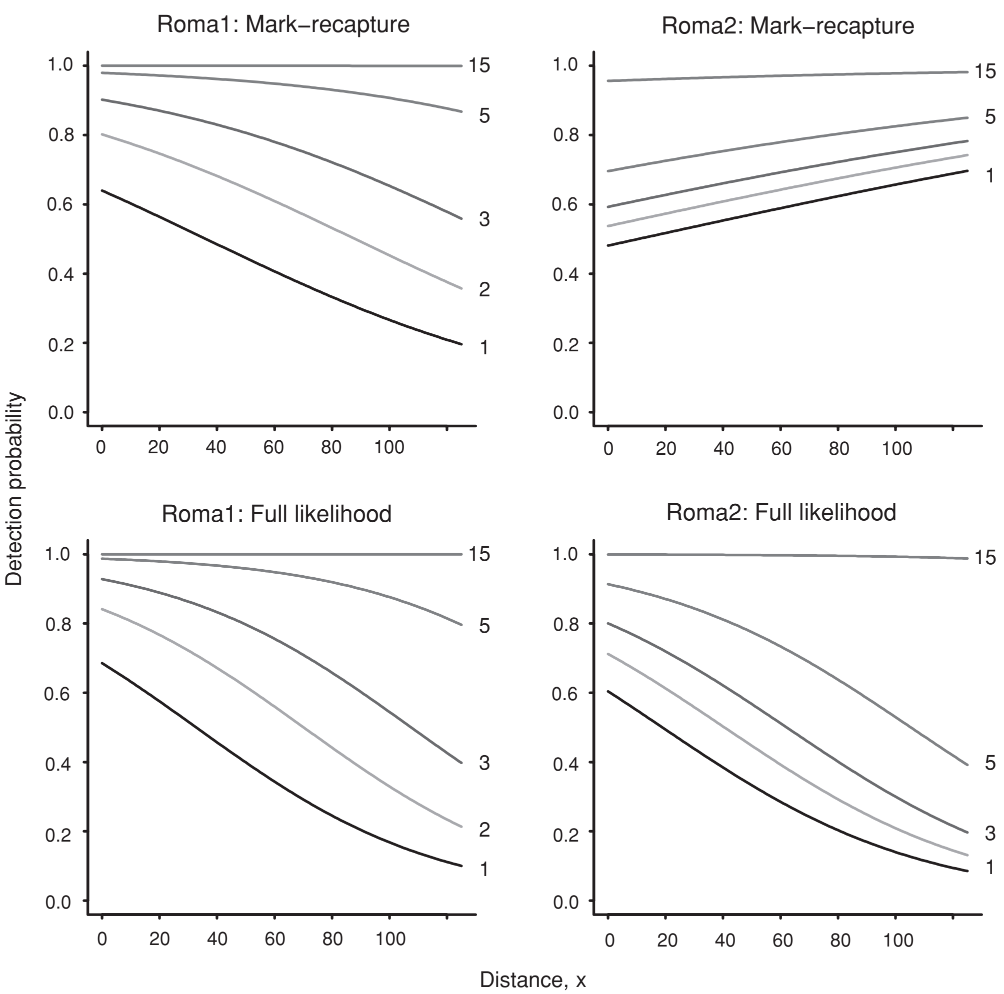

Fig. 2. Comparison of fits using the mark–recapture method (*M*) and the full-likelihood method (*F*) for the front-seat observer in the Roma1 and Roma2 survey blocks. The logit detection model was used in each case. Each figure shows the fall-off of detection probability with distance for a range of different cluster sizes. The cluster size is displayed to the right of each curve.

'recaptured' by Observer 1, creates an impression that Observer 1 has very high detection probability in the furthest distance class. Method *M* is unable to attribute this chance result to dependence between the observers, and the result is a somewhat implausible shape for the detection functions, leading to a severe underestimate of density. Method *F* performs much better for the Roma2 block, showing the advantage of incorporating extra information about the detection parameters θ through the additional likelihood components.

Point-independence method (P)

Fig. 3 shows the more complicated output obtained from the point-independence method (*P*) for both observers in the Roma1 and Roma2 survey blocks. Each point-independence analysis produces three detection functions, compared with two for Methods *M* and *F*. The first two detection functions are the conditional detection functions, $p_{1|2}$ and $p_{2|1}$, which are shown in Fig. 3 as functions of distance averaged over all cluster sizes. The third detection function is the overall probability of detection of both observers as a function of distance from the line, which is shown in the third panel of Fig. 3. The power of the point-independence method in combining different sources of data is seen from the Roma2 results, in which the conditional detection function $p_{1|2}$ increases with distance, while the overall detection function is still able to decrease with distance and provides a good fit to the observed pattern of distance detections. The results show clearly that the distance-sampling component of the likelihood, written $L_{y|z}(\theta)$ and underlying the third panel of Fig. 3, adds important extra information to the analysis that would be lost by restricting to the mark–recapture component $L_{\omega}(\theta)$, as in Method *M* and the first two panels of Fig. 3.

Comparison of Methods M, F, and P

Fig. 4 shows estimated numbers of individuals in the covered area, from each of the three methods in each of the four survey blocks, using the logit detection model in each case. The estimated number of individuals is gained from the estimated number of clusters, $\hat{N}_c$, and the estimated distribution of cluster size in the population at large. Bootstrapped 95% confidence intervals, using the percentile method with 1000 bootstrap replicates, are shown around each estimate. The same pattern repeats in each of the four survey blocks: Method *P* gives the highest estimates, followed by Method *F*, and Method *M* gives the lowest. In three of the four blocks, the lower 95% confidence limit from Method *P* is comparable with the upper 95% limit from Method *M*.

The results in Fig. 4 are strongly suggestive of unmodelled heterogeneity in Method *M*, which is mitigated by the extra likelihood components in Method *F* and, to a much greater extent, by Method *P*, even though Method *P* includes only one more likelihood component than Method *M*. The likely reason for the results is that a high proportion of sightings at far distances were

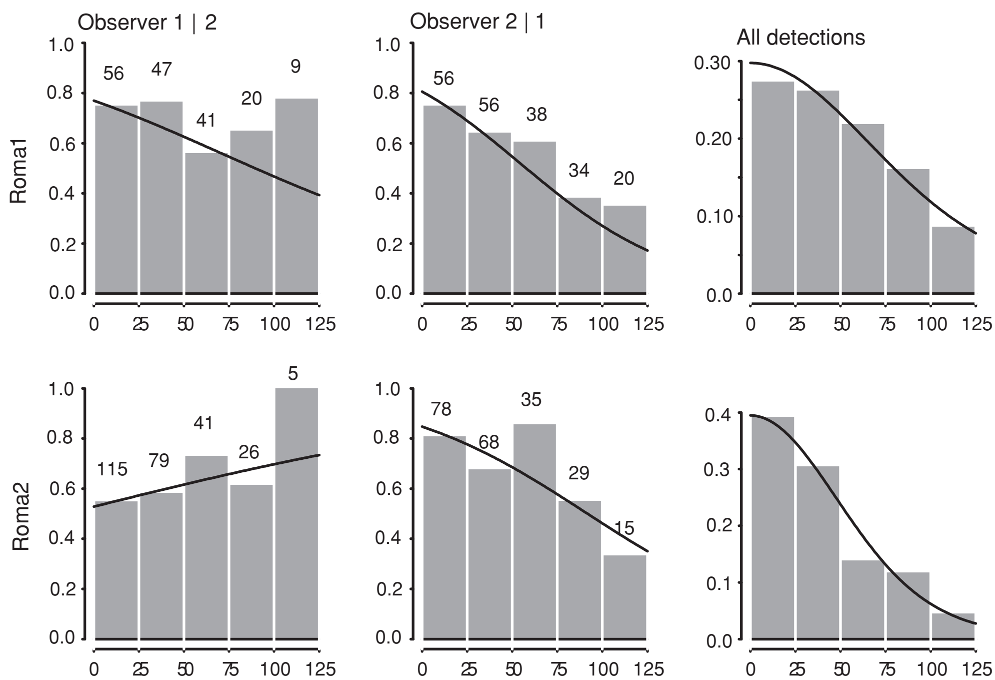

Fig. 3. Fitted detection functions for the Roma1 and Roma2 survey blocks from the point-independence method (*P*). The first panel shows the fitted conditional detection functions of Observer 1 given Observer 2, $p_{1|2}(y)$, averaged over all values of cluster size *z*. The observed proportions of conditional detections are shown as barplots beneath the curves. The number above the bar is the number of sightings made by Observer 2, and the height of the bar is the proportion of these sightings that were also detected by Observer 1. There should be a good fit of the curve to the bars, unless the sample size (number above the bar) is low. The second panel shows the same concept for Observer 2 given Observer 1, $p_{2|1}(y)$. The third panel shows the overall detection function for both observers, $p_{\bullet}(y)$, averaged over all values of cluster size *z*. The histogram of all observed detection distances is shown beneath the curve, and the fitted curves and histograms are scaled to have the same area underneath them.

very detectable to both observers for some reason that was not evident in the model. The proportion of duplicate sightings at high distances is therefore higher than it should be if observers were genuinely independent, so detectability is overestimated and density is underestimated. The effects of the spuriously increasing detection function for Observer 1 in the Roma2 analysis under Method *M* are particularly evident in Fig. 4.

Table 1 shows estimated densities and coefficients of variation (CV) from the three methods in the four survey blocks. In each case, the CV for Method *P* is comparable with, or better than, those for Methods *M* and *F*. Model selection using AIC on the full likelihood (1) conclusively favoured Method *P* in each case. For three of the four cases, Method *F* was ranked above Method *M* by AIC, but Method *M* was ranked above Method *F* in the Roma1 block.

Overall fitted detection functions, and the underlying detection distance data, are shown in Fig. 5. The *P*-value from the Chi-square goodness-of-fit test is shown in each case. Visual inspection of the fits demonstrates that Method *M* is very poor, while Method *F* is reasonable and Method *P* is very good. Method *P* explicitly fits a detection function to these distance data, so its good performance is expected. Goodness-of-fit was satisfactory in every case for Method *P*, marginal for Method *F*, and very poor for Method *M*.

We conclude that both AIC and goodness-of-fit tests show a clear preference for Method *P* over both of Methods *F* and *M* for these aerial survey data.

Discussion

Our analysis has provided a strong endorsement of the new point-independence method of Laake and Borchers (2004) and Borchers *et al*. (2006) for mark–recapture distance-sampling analyses. The point-independence method is conclusively favoured by AIC and goodness-of-fit tests over the other two methods examined. Additionally, its theoretical foundations are stronger and more credible.

Both of our Methods *M* and *F*, which rely on full independence between observers at all distances, provided lower estimates of kangaroo density than the point-independence Method *P*. The full likelihood Method *F* appears to be a statistical halfway house between Methods *M* and *P*, in which the extra components included in the likelihood function mitigate some of the problems of unmodelled heterogeneity in capture probability, but not to the extent of the point-independence Method *P*. Borchers *et al*. (1998a) also found that Method *F* gave higher estimates than Method *M* for double-observer surveys of minke whales.

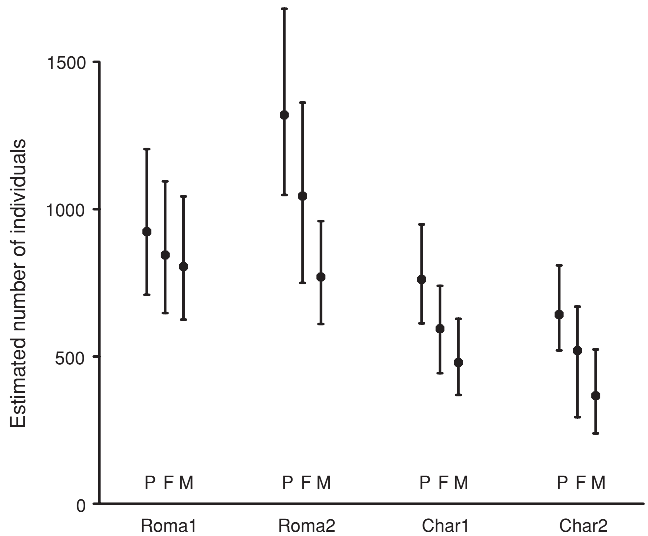

Fig. 4. Comparison of estimated number of individuals in the covered area from the three mark–recapture distance-sampling methods: point-independence (*P*), full-likelihood with full independence (*F*), and mark–recapture likelihood with full independence (*M*). Point estimates and 95% bootstrapped confidence intervals are shown for each of the four survey blocks.

Examination of the estimated detection functions and underlying mark–recapture data for Method *M* highlights the potential for serious problems with this method. We recommend that Method *M*, which is essentially a mark–recapture analysis with distance as a covariate, should not be used in practice. There might be justification for using Method *M* if responsive movement is known to be a problem with the survey, because *M* is the only method that does not rely on the assumed uniformity of distances in likelihood component $L_{y|z}(\theta)$ (Fewster *et al*. 2008). However, the bias caused by unmodelled heterogeneity in Method *M* will often be worse than that caused by responsive movement using Method *P*. In general, responsive movement is probably easier to address by field methods than unmodelled heterogeneity, which, by definition, is due to unknown causes. Fewster *et al*. (2008) attempted to investigate the extent of responsive movement for the kangaroo data, but were unable to draw conclusions because any effects of responsive movement were swamped by effects of unmodelled heterogeneity in Method *M*, which gives the only valid testing framework.

Table 1. Results from the four survey blocks for estimated density, $\hat{D}$, in individuals km^{-2}, together with estimated percentage CV and the difference between the AIC of the fitted method and the AIC of the best method for the survey block

Methods are point-independence (*P*), full-likelihood with full independence (*F*), and mark–recapture likelihood with full independence (*M*), using the logit detection model in each case

Method	Roma1			Roma2			Char1			Char2		
	$\hat{D}$	CV	ΔAIC	$\hat{D}$	CV	ΔAIC	$\hat{D}$	CV	ΔAIC	$\hat{D}$	CV	ΔAIC
P	27	14	0	36	12	0	21	11	0	28	12	0
F	24	14	93	29	14	350	17	13	200	23	20	170
M	23	13	74	21	12	400	13	14	220	16	20	200

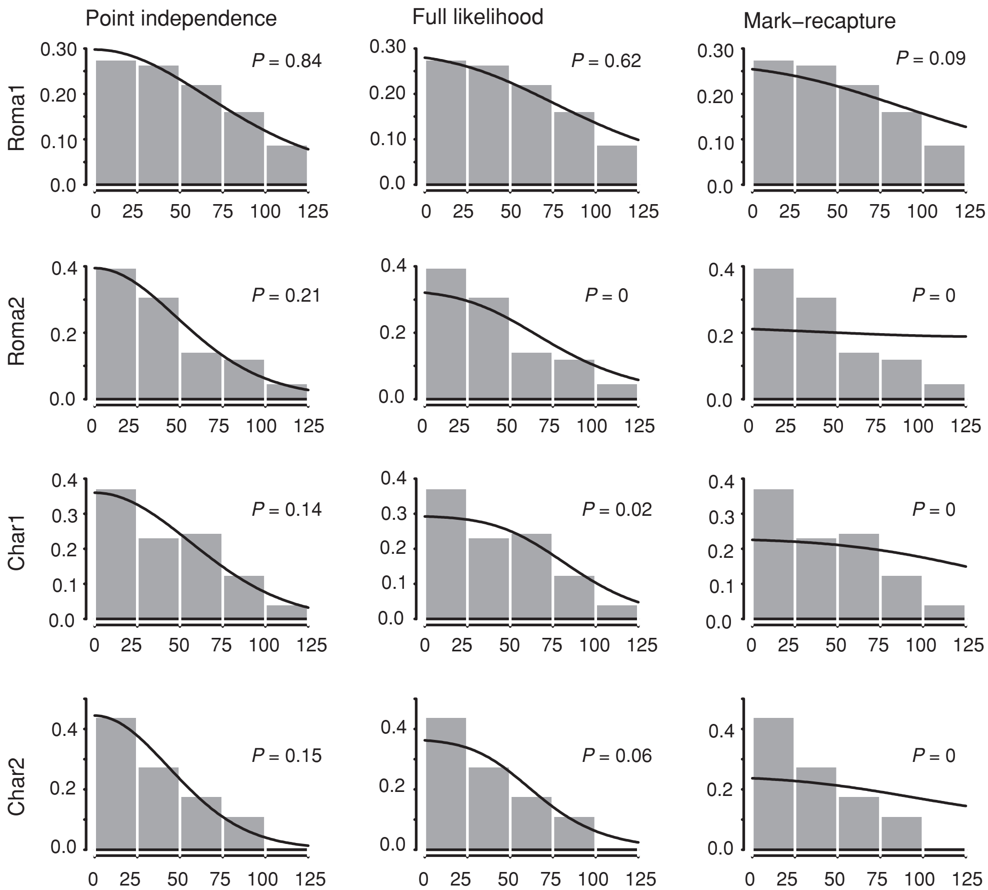

Fig. 5. Comparison of the overall detection functions from the three mark–recapture distance-sampling methods: point-independence (*P*), full-likelihood with full independence (*F*), and mark–recapture likelihood with full independence (*M*). The histogram of observed detection distances is shown in each plot, and the fitted curves and histograms are scaled to have the same area underneath them. The four rows correspond to survey blocks Roma1, Roma2, Char1 and Char2 respectively.

We reproduced the analyses of this paper in the new Mark–Recapture Distance Sampling analysis engine of Distance 5 (Thomas *et al*. 2006), where applicable. Our results are in agreement with those from Distance 5, except that our bootstrapped variance estimates were consistently higher than the analytic variance estimates from Distance 5. Distance 5 reported estimated CVs in the range 6–9% for Method *P*, while our bootstrapped CVs ranged from 11% to 14%. The CVs from Distance 5 seem on the low side for surveys of this type, and further investigation of analytic variance estimation methods for these analyses may be warranted.

Although we do not report the results here, we also analysed the kangaroo data using conventional distance-sampling methods, by pooling data from both observers together. In every case, the resulting point estimate of abundance lay between that of Methods *F* and *P*, and was usually closer to that of Method *P*. This is a strong indication that Methods *M* and *F* have failed, because they are aimed at *correcting* the negative bias in conventional distance sampling caused by imperfect trackline detection, yet themselves provide lower estimates than conventional distance sampling. We do not recommend conventional distance-sampling analyses for these data, because the assumption of perfect trackline detection is certainly violated. However, the bias caused by this assumption appears to be less than that caused by unmodelled heterogeneity in mark–recapture distance-sampling Methods *M* and *F*. A similar effect was noted by Laake *et al*. (2008), Laake and Borchers (2004:149), and Laake (1999). In this respect, it appears that the point-independence Method *P* is a considerable advance over the previous methods for combining mark–recapture and distance-sampling methodologies.

Estimated probabilities of detection on the transect line (g_0 in conventional notation) ranged from 0.53 (Observer 1 in the Roma2 block) to 0.94 (Observer 1 in the Char1 block), with most estimates in the range 0.7–0.9. It is interesting that the density estimates from one observer team (Roma2 and Char2) were both higher than those from the other observer team (Roma1 and Char1), as seen in Table 1. This might be coincidence, or it might reflect some systematic problem with one of the observer teams, which might be caused, for example, by consistent under- or overestimation of measured distances, or a consistent policy of ignoring responsive movement and measuring distances when the group is abeam of the aircraft, rather than where it was first detected. However, distances were always those measured by Observer 1, who was the same in both teams, so this explanation is unlikely.

Density estimates of eastern grey kangaroos were comparable for all four survey blocks, ranging from 21 to 36 individuals km^{-2} using Method *P*. The reported coefficients of variation (11–14% from Method *P*) were very good, despite being considerably higher than those reported from Distance 5. Our analysis indicates that the mark–recapture distance-sampling methods can be successfully implemented by aerial surveys in the field, and even strong effects of unmodelled heterogeneity in capture probability appear to be overcome by the point-independence Method *P*. Conventional distance-sampling methods should always be expected to underestimate density when detection on the transect line is not perfect, so methods that supposedly correct for imperfect trackline detection yet report lower density estimates than conventional distance sampling should be treated with high suspicion. In our analysis, this was the case for Methods *M* and *F*, but not for Method *P*. Conversely, the estimates from Method *P* were not a straightforward upscaling of the conventional distance-sampling estimates by the shortfall of g_0 on the transect line, indicating that the model has achieved more than a simple scaling factor. The combination of the mark–recapture and distance-sampling data has strengthened both methods.

The point-independence methodology does require that observers detect animals independently at some distance from the transect line, usually $y = 0$. It is possible that this assumption, even though it is much weaker than that of full independence, is still too strong for some survey situations. If this is the case, the point-independence methods would be expected to underestimate density overall; however, they would necessarily perform better than the equivalent methods based on full independence.

Acknowledgements

We thank two referees and the editor for their constructive comments on this manuscript. We are grateful to the Queensland Parks and Wildlife Service for financial and logistical survey support, and to Geoff Lundie-Jenkins and Gordon Maag for acting as observers and organising survey logistics. We thank Mark Read for piloting the surveys, and Josie Carwardine for transcribing tapes. This project was also supported by an Australian Research Council Linkage grant to Hugh Possingham.

References

Bayliss, P., and Yeomans, K. (1989). Correcting bias in aerial survey population estimates of feral livestock in northern Australia using the double-count method. *Journal of Applied Ecology* **26**, 925–933. doi:10.2307/2403702

Borchers, D. L., Zucchini, W., and Fewster, R. M. (1998*a*). Mark–recapture models for line transect surveys. *Biometrics* **54**, 1207–1220. doi:10.2307/2533651

Borchers, D. L., Buckland, S. T., Goedhart, P. W., Clarke, E. D., and Hedley, S. L. (1998*b*). Horvitz–Thompson estimators for double-platform line transect surveys. *Biometrics* **54**, 1221–1237. doi:10.2307/2533652

Borchers, D. L., Laake, J. L., Southwell, C., and Paxton, C. G. M. (2006). Accommodating unmodeled heterogeneity in double-observer distance sampling surveys. *Biometrics* **62**, 372–378. doi:10.1111/j.1541-0420.2005.00493.x

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2001). 'Introduction to Distance Sampling.' (Oxford University Press: Oxford.)

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 Grigg, G. C. (1981). Surveys of the distribution and density of kangaroos in the pastoral zone of South Australia and their bearing on the feasibility of aerial survey in large and remote areas. *Australian Wildlife Research* **8**, 1–11. doi:10.1071/WR9810001

Caughley, G., Sinclair, R., and Scott-Kemmis, D. (1976). Experiments in aerial survey. *Journal of Wildlife Management* **40**, 290–300. doi:10.2307/3800428

Choquenot, D. (1995). Species- and habitat-related visibility bias in helicopter counts of kangaroos. *Wildlife Society Bulletin* **23**, 175–179.

Clancy, T. F., Pople, A. R., and Gibson, L. A. (1997). Comparison of helicopter line transects with walked line transects for estimating densities of kangaroos. *Wildlife Research* **24**, 397–409. doi:10.1071/WR96103

Fewster, R. M., Southwell, C., Borchers, D. L., Buckland, S. T., and Pople, A. R. (2008). The influence of animal mobility on the assumption of uniform distances in aerial line-transect surveys. *Wildlife Research* **35**, 275–288. doi:10.1071/WR07077

Laake, J. L. (1999). Distance sampling with independent observers: reducing bias from heterogeneity by weakening the conditional independence assumption. In 'Marine Mammal Survey and Assessment Methods'. (Eds G. Amstrup, S. Garner, J. Laake, B. Manly, L. McDonald and D. Robertson.) pp. 137–148. (Balkema: Rotterdam.)

Laake, J. L., and Borchers, D. L. (2004). Methods for incomplete detection at distance zero. In 'Advanced Distance Sampling'. (Eds S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers and L. Thomas.) pp. 281–306. (Oxford University Press: Oxford.)

Laake, J. L., Dawson, M. J., and Hone, J. (2008). Visibility bias: mark–recapture, line-transect, or both? *Wildlife Research* **35**, 299–309. doi:10.1071/WR07034

Lundie-Jenkins, G., Hoolihan, D., and Maag, G. W. (1999). An overview of the Queensland macropod monitoring programme. *Australian Zoologist* **31**, 301–305.

Marsh, H., and Sinclair, D. F. (1989). Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. *Journal of Wildlife Management* **53**, 1017–1024. doi:10.2307/3809604

Pople, A. (2004). Population monitoring for kangaroo management. *Australian Mammalogy* **26**, 37–44.

Pople, A. R., and Grigg, G. C. (1998). Commercial harvesting of kangaroos in Australia. Environment Australia, Canberra. http://www.ea.gov.au/biodiversity/trade-use/wild-harvest/kangaroo/harvesting/

Pople, A. R., Cairns, S. C., Clancy, T. F., Grigg, G. C., Beard, L. A., and Southwell, C. J. (1998*a*). An assessment of the accuracy of kangaroo surveys using fixed-wing aircraft. *Wildlife Research* **25**, 315–326. doi:10.1071/WR97077

Pople, A. R., Cairns, S. C., Clancy, T. F., Grigg, G. C., Beard, L. A., and Southwell, C. J. (1998*b*). Comparison of surveys of kangaroos in Queensland using helicopters and fixed-wing aircraft. *The Rangeland Journal* **20**, 92–103. doi:10.1071/RJ9980092

Pople, A. R., Thompson, J. A., Clancy, T. F., and Boyd-Law, S. (1998*c*). Aerial survey methodology and the cost of control for feral goats in western Queensland. *Wildlife Research* **25**, 393–407. doi:10.1071/WR97123

Southwell, C. (1994). Evaluation of walked line transect counts for estimating macropod density. *Journal of Wildlife Management* **58**, 348–356. doi:10.2307/3809401

Southwell, C., and Sheppard, N. (2000). Assessing harvested populations of the euro (*Macropus robustus erubescens*) in the Barrier Ranges of western NSW. *Australian Mammalogy* **21**, 165–171.

Southwell, C., Borchers, D., Paxton, C. G. M., Burt, L., and de la Mare, W. (2007). Estimation of detection probability in aerial surveys of Antarctic pack-ice seals. *Journal of Agricultural Biological & Environmental Statistics* **12**, 41–54. doi:10.1198/108571107X162920

Southwell, C., Paxton, C. G. M., and Borchers, D. L. (2008). Detectability of penguins in aerial surveys over the pack-ice off Antarctica. *Wildlife Research* **35**, 349–357. doi:10.1071/WR07093

Thomas, L., Laake, J. L., Strindberg, S., Marques, F. F. C., Buckland, S. T., *et al.* (2006). Distance 5.0. Research Unit for Wildlife Population Assessment, University of St Andrews, UK. http://www.ruwpa.st-and.ac.uk/distance/

Manuscript received 29 June 2007, accepted 14 December 2007