

# The influence of animal mobility on the assumption of uniform distances in aerial line-transect surveys

Rachel M. Fewster<sup>A,E</sup>, Colin Southwell<sup>B</sup>, David L. Borchers<sup>C</sup>,  
Stephen T. Buckland<sup>C</sup> and Anthony R. Pople<sup>D</sup>

<sup>A</sup>Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand.

<sup>B</sup>Australian Antarctic Division, Channel Highway, Kingston, Tas. 7050, Australia.

<sup>C</sup>CREEM, University of St Andrews, The Observatory, Buchanan Gardens,  
St Andrews KY16 9 LZ, Scotland.

<sup>D</sup>Biosecurity Queensland, Department of Primary Industries and Fisheries, GPO Box 46,  
Brisbane, Qld 4001, Australia.

<sup>E</sup>Corresponding author. Email: r.fewster@auckland.ac.nz

**Abstract.** Line-transect distance sampling is a widely used method for estimating animal density from aerial surveys. Analysis of line-transect distance data usually relies on a requirement that the statistical distribution of distances of animal groups from the transect line is uniform. We show that this requirement is satisfied by the survey design if all other assumptions of distance sampling hold, but it can be violated by consistent survey problems such as responsive movement of the animals towards or away from the observer. We hypothesise that problems with the uniform requirement are unlikely to be encountered for immobile taxa, but might become substantial for species of high mobility. We test evidence for non-uniformity using double-observer distance data from two aerial surveys of five species with a spectrum of mobility capabilities and tendencies. No clear evidence against uniformity was found for crabeater seals or emperor penguins on the pack-ice in East Antarctica, while minor non-uniformity consistent with responsive movement up to 30 m was found for Adelie penguins. Strong evidence of either non-uniformity or a failure of the capture–recapture validating method was found for eastern grey kangaroos and red kangaroos in Queensland.

## Introduction

Line-transect sampling (Buckland *et al.* 2001, 2004) is a popular method for estimating the density of objects (usually animals or plants) in a region. A set of line transects is located in the region according to some randomised scheme. Observers traverse the lines, searching for objects in a strip of width  $w$  to either side of each line. For each detected object, the observer measures the perpendicular distance from the line to the object. Object density is estimated by analysing the fall-off of detections as distance increases. In the simplest form of the methodology, known as conventional line-transect sampling, detection is assumed to be certain on the line itself. At larger distances, the method conceptually fills in the missing sightings to reach the level on the line, thereby estimating the density of objects in the whole strip.

Unlike plot sampling, line-transect sampling does not require all objects in sample plots to be detected, so large regions can be surveyed in a cost-effective manner. In many situations, conventional line-transect sampling is the most efficient method available for estimating animal density in a large region, in terms of the amount of effort needed to reach a desired precision. However, it relies on certain assumptions that cannot always be met in the field. These are: (1) objects directly on the transect line are detected with certainty; (2) objects are detected at their initial locations, before any movement in response to the observer; and (3) distance measurements from

the line to the object are made without error (Buckland *et al.* 2001: 29–37).

Extensions to the conventional scheme allow some assumptions to be relaxed at the expense of extra effort in implementation. For aerial surveys, the assumption of certain detection on the line is often impossible to meet, partly because the aircraft bodywork can obstruct the line of vision and partly because detection is taking place from a height. The assumption is often relaxed by introducing a capture–recapture component using two separate observer teams, which in turn demands new assumptions regarding observer independence (e.g. Borchers *et al.* 2006).

In addition to the assumptions stated above, most methods for analysing line-transect data rely on a further requirement that the statistical distribution of object distances from the line is uniform. This requirement refers to the distances of all objects, both detected and undetected, so it cannot be verified directly from the observed distances. If the uniform requirement fails, the analysis step in which the missing sightings are filled in uniformly up to their levels on the line is flawed. The uniform requirement is often not stated as an explicit assumption of line-transect sampling, because in a correctly designed and implemented survey it is satisfied automatically as a consequence of the other assumptions. The issues surrounding this requirement are subtle, and have led to some confusion in the literature.

In this article we aim to clarify issues relating to the uniform assumption, and to investigate the validity of the assumption for aerial surveys of five animal species with different behaviours. At first, it is not obvious that the uniform assumption should be fulfilled automatically by the other assumptions. It might seem that objects with a patchy or trended spatial distribution do not satisfy the requirement. This confusion is two-fold. First, it is the distances from the lines to the objects that need to be uniform, not the distribution of objects themselves. Second, we need to establish the exact statistical procedure over which we require uniform distances, and our purpose for the requirement, which we do now.

For any procedure of statistical estimation, we need some concept of repeated sampling according to which our estimate is (ideally) unbiased with low variance. In distance sampling, a suitable concept of repeated sampling is called the 'design framework' (e.g. Fewster and Buckland 2004). In the design framework, we conceptualise that objects stay still, while we repeatedly cast down transect lines at random and conduct surveys along them. If we cast down transect lines according to a uniform distribution, then the distances from the fixed objects to the randomly positioned lines will be uniform. This uniformity applies no matter how unevenly the objects are located. The uniformity refers to the distribution of all possible line-to-object distances, over the infinitely many surveys that we can conceptualise where objects stay still and lines are located at random. Our purpose in making this requirement is that it ensures that the distance-sampling density estimate is approximately correct, on average, in the design framework. We conceptualise survey repeats in order to define the concept of the 'average' estimate.

If a design does not ensure that the uniformity requirement is met, there is no assurance that the density estimates will be correct, on average, under this design. The design is under the control of the investigators, who have the responsibility to ensure that the design fulfils the uniformity requirement. Suitable designs for line placement include systematic designs with a random start-point, stratified designs, and completely random designs.

Fig. 1 illustrates the rationale behind the uniform assumption. In Fig 1(a), a fixed strip is shown in which the distances of objects from the line appear to be decidedly non-uniform, as demonstrated by the histogram in Fig. 1(b). We then randomise the position of the line and measure the distances from the same fixed objects to the randomised lines for 10000 lines generated uniformly between the limits shown by the arrow in Fig. 1(c). The resulting distribution of distances over the randomised lines is satisfactorily uniform (Fig. 1(d)).

The procedure illustrated in Fig. 1 (c) and (d) is clearly not implemented in reality: practitioners do not have the opportunity to repeat a randomised transect thousands of times while the objects beneath them stay stationary. In its most basic form, the uniform requirement applies to conceptual survey repeats, and its purpose is to assure the practitioner that the distance-sampling procedure is statistically valid in the sense of giving approximately the correct answer 'on average' in this defined sense. However, getting the correct answer 'on average' is of little consolation if the actual answer obtained from the practitioner's single survey has the potential to be poor. This means

that it is also important that the density estimate should have low variance in the design framework. An unbiased procedure with low variance guarantees that *every* answer obtained from the procedure will be close to the correct answer.

There are many issues that impact on the design-variance of the density estimate (Fewster *et al.* 2008). One of these issues is the extent to which uniformity holds *within* the single survey. If the set of line-to-object distances in the single realised survey is appreciably non-uniform, this might lead to a poor estimate of the detection function and a poor estimate of overall density. A chance surplus of large distances will lead to more detections at high distances, overestimation of detectability, and consequent underestimation of object density. Conversely, a chance paucity of large distances will lead to overestimation of object density. Practitioners must therefore attempt to facilitate uniformity *within* a single survey, as well as using a random design that guarantees uniformity across the conceptual survey repeats. Most importantly, surveys should consist of a large number of lines so that any unevenness in object distances within a line is balanced out when lines are pooled together (Buckland *et al.* 2001). Many short lines are preferable to a few long lines. A survey with many lines is a way of imitating in reality the conceptual replication shown in Fig. 1. Aggregations of objects or features of terrain that can cause imbalances of object distances within a single line are balanced by the distribution in other lines.

The preceding discussion about the uniform requirement is summarised as follows.

- The uniformity of line-to-object distances is required in a conceptual framework to ensure the approximate unbiasedness of the density estimator. This requirement is fulfilled by selecting the transect lines according to an appropriately randomised design, even though only one realisation of the design (the actual survey) will usually be carried out.
- The uniformity requirement is not necessarily violated when there is a non-uniform distribution of distances within a single survey, for example when the non-uniformity is due to spatial clustering of objects or other sampling variability (chance imbalances). However, to get a *good* estimate from every survey, uniformity is desirable within every survey.
- The primary way of facilitating uniformity within the survey, if all other assumptions of distance sampling hold, is to use as many transect lines as possible within the survey. This will help to lower the variance in density estimation due to chance non-uniformity within the survey. It also has other effects in improving our ability to estimate the variance (e.g. Buckland *et al.* 2001: 241).

The discussion above is based on the understanding that Assumptions (1) to (3) of distance sampling hold. The uniform requirement is often omitted from the list of assumptions, because it is automatically satisfied by a randomised design if all other assumptions hold. Within a survey, non-uniformity caused by sampling variability is undesirable, but is not, in itself, a violation of the uniform requirement. However, there are situations where the uniform requirement can be explicitly violated, notably if the design is not correctly randomised, or if the other assumptions of distance sampling do not hold. We now examine these violations of the uniform requirement.

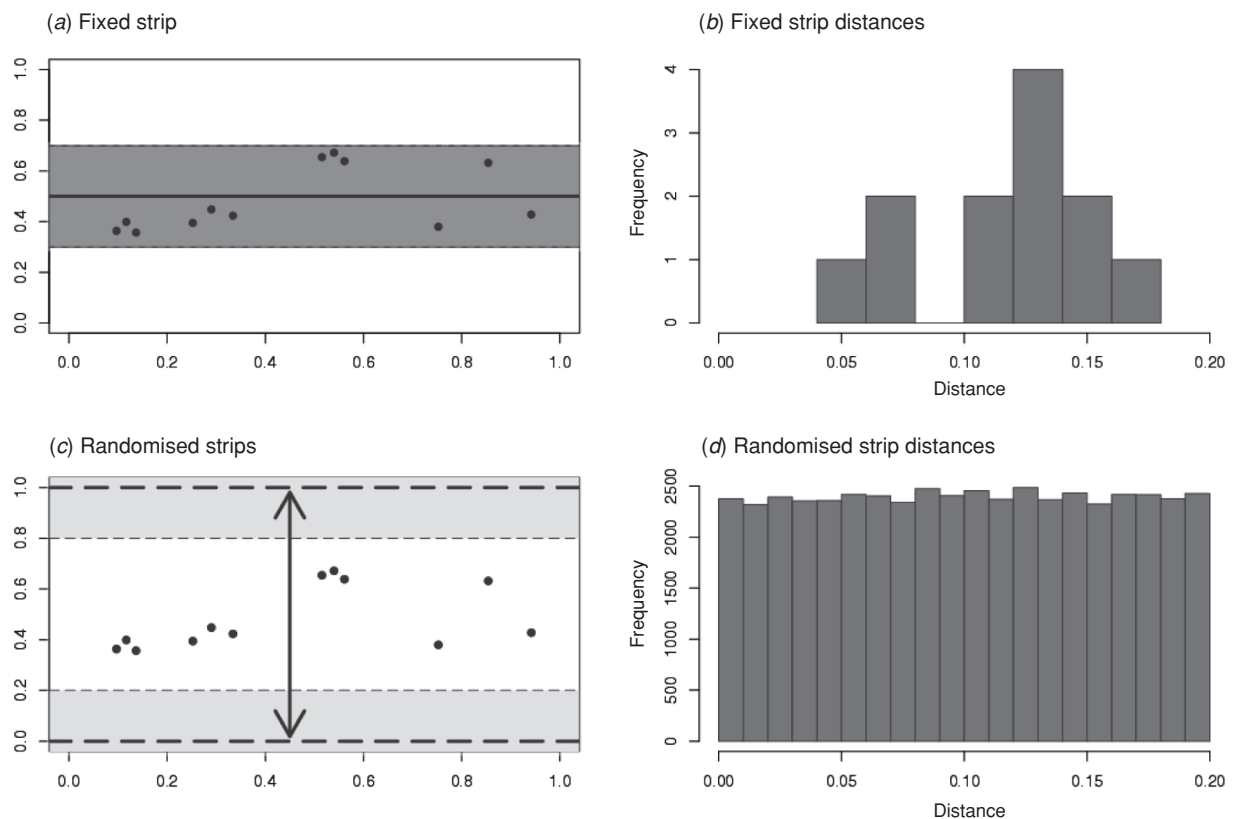
Non-random placement of transects is a violation of the uniform requirement, and may cause severe bias if transects are

deliberately located along features such as tracks or waterways. The distribution of animals or plants is commonly affected by such features. For example, animals may occur at higher density near to tracks for ease of mobility, or at lower density due to disturbance, or density might be affected by different vegetation types along these features. It should generally be possible to avoid non-random placement of transects in aerial surveys of homogeneous regions, although safety issues can impose constraints leading to non-randomness in mountainous terrain. Melville and Welsh (2001) reported bias in density estimation due to violation of the uniform requirement, and the cause of this violation was non-randomised transect lines. Fewster *et al.* (2005) repeated their analyses, and showed that the bias was negligible when the design was correctly randomised. Fewster *et al.* (2005) did note some remaining bias in variance estimation even with the randomised design, and progress on this issue has been made in Fewster *et al.* (2008). Care must be taken when evaluating distance sampling or other methodologies by simulation, and some of the pitfalls are discussed in Fewster and Buckland (2004).

The other primary mechanism for violating the uniform requirement is responsive movement of the animals before detection. This is an explicit violation of distance-sampling Assumption (2) above. Systematic responsive movement before detection can cause severe bias, and it is often difficult to deter-

mine whether it has occurred. No matter how many conceptual survey repeats are carried out (as in Fig. 1), or transect lines are used in the survey, the uniform requirement will never be satisfied if animals systematically change their positions with respect to the transect line, after the line has been selected, and before detection. Responsive movement commonly occurs in a direction away from the observer on the line, although sometimes animals are attracted to the observer and move towards the line.

Inaccurate measurement of distances (Assumption (3)) does not directly violate the uniform requirement, because uniformity is required of the true, rather than the measured, object distances. Examples of inaccurate measurements are the heaping of observed distances close to round numbers, a consistent tendency to overestimate or underestimate distances, and a failure to record distances to the centre of a group of animals if observations are of groups rather than of individuals. In helicopter surveys, inaccurate distance measurements may result if distance classes are delineated by fixed poles and the helicopter is unable to maintain a constant altitude, especially due to undulating terrain. Consistent aircraft tilt must also be avoided. While not in itself a violation of the uniform requirement, inaccurate measurement of distances interferes directly with estimation of the detection function, and might also lead to apparent non-uniformity of the underlying distances where none exists.



**Fig. 1.** Demonstration of the uniformity of object distances under randomisation of the search strips. (a) A single strip covers 12 objects, both detected and undetected. (b) Distances of the objects from the transect line in the fixed strip do not appear to be uniformly distributed. (c) The strip locations are randomised within the bounds shown by the arrows, and object distances are recorded when they lie within the search strip (distance  $w = 0.2$  from the line). (d) Distances from the 12 objects to the randomised lines combined across 10000 line positions are uniformly distributed ( $P = 0.33$  from the Kolmogorov–Smirnov test).

Inaccurate measurement may therefore interfere with our ability to test for uniformity within a survey.

In view of the discussion above, we hypothesise that undetected responsive movement poses the greatest challenge to the uniform requirement, for surveys that are properly designed and executed with good sample size. We therefore aim to assess evidence for non-uniformity within aerial surveys covering a spectrum of species from fairly immobile to highly mobile. Under our hypothesis, we should expect non-uniformity to be more evident within surveys of highly mobile species than within surveys of less mobile species, if it is evident at all.

We examine the uniform assumption for aerial surveys of five species: crabeater seal (*Lobodon carcinophaga*), emperor penguin (*Aptenodytes forsteri*), Adelie penguin (*Pygoscelis adeliae*), eastern grey kangaroo (*Macropus giganteus*), and red kangaroo (*M. rufus*). Crabeater seals, Adelie penguins and emperor penguins were surveyed simultaneously by helicopter over the pack ice off East Antarctica in 1999 (Southwell *et al.* 2007, 2008). Eastern grey kangaroos and red kangaroos were surveyed simultaneously by helicopter in Queensland, Australia in 2000 (Fewster and Pople 2008). Both surveys employed double-platform distance-sampling methodology, which allows us to conduct a test of uniformity within the surveys (Southwell *et al.* 2004).

A previous analysis of observed response behaviour of seals and penguins from the East Antarctica survey (Southwell 2005) showed that seals exhibited very little responsive movement to the helicopter, even in the closest distance category, with mean movements less than 1 m, 90% of movements less than 5 m, and 56% of groups stationary in this category. This reflects the limited movement capability of seals on ice. Penguin groups moved greater distances and more frequently, with mean movements of ~4 m, 90% of movements less than 12 m, and 30% of groups stationary in the closest distance category. Where responsive movement was observed to occur, measured distances were corrected by visual estimation of the movement (Southwell 2005). However, we anticipate that some level of responsive movement went undetected, and is therefore uncorrected for in the current analysis. We classify seals as the least mobile of the taxa considered, penguins as being of intermediate mobility, and kangaroos as being highly mobile.

It is important to emphasise that there is no completely rigorous way of testing the uniform assumption on real data. Most methods for analysing line-transect data rely on the uniform assumption, but in order to test the assumption we must use a method that does not. A suitable validating method is a capture–recapture analysis of the double-platform survey data, following Southwell *et al.* (2004) and Borchers *et al.* (1998a, 1998b). However, it is important to recognise that the assumptions behind the validating method may be more questionable than the uniform assumption we are attempting to validate. In particular, the capture–recapture analysis assumes a level of independence between the observers that has been shown to be inadequately satisfied by the crabeater seal survey (Borchers *et al.* 2006). We take steps to mitigate this problem by conducting our analyses on the closest distance categories where violation of the independence assumption is less evident. Despite this, the investigations of this article must be taken as exploratory rather than definitive. Apparent failure of the

uniform assumption must be interpreted *either* as failure of the uniform assumption, or as failure of the independence assumption underlying the validating method. We emphasise that this validation problem plagues any attempt to investigate the uniform assumption from real data. Without additional data on the locations of both detected and undetected objects, no method can give the right answer for the undetected object distances without making assumptions that are themselves subject to doubt.

Although the independence problem in the validating method is, by definition, caused by factors unknown to the observer, it is possible that it is more prevalent in surveys of more mobile species, because a movement cue is a likely mechanism for simultaneous detection by both observer teams, even if the movement is not consciously noted by the observers. For the more mobile species where we hypothesise greater potential violation of the uniform distribution, therefore, there might also be greater violation of the independence assumption underlying the validating method.

## Methods

### *East Antarctica aerial survey for seals and penguins*

Detailed descriptions of the survey methods for the East Antarctica pack-ice survey are given by Southwell (2005) and Southwell *et al.* (2008). The survey region comprised 1.5 million km<sup>2</sup> of pack-ice between longitudes 64°E and 150°E. Two helicopters (Sikorsky S76) were flown, each maintaining as closely as possible an altitude of 130 m and a speed of 90 kn. Speed and altitude were maintained by visual inspection of aircraft instruments. A digital record of altitude taken at intervals of 10 s showed that pilots remained within 20 m of the required altitude for 95% of the survey effort.

Each helicopter housed a team of four observers: a front and back observer for each of the left and right sides of the transect line. Front and back observers searched the same area simultaneously and were not able to see or hear each other. Each flight provided data for two double-platform distance-sampling analyses: one from the left-side front and back observers, and one from the right-side front and back observers. In total, 12 observers were involved in the survey. Data were collected on 10 survey days in December 1999. The survey involved a total of 37 transects.

Observers searched ahead of the helicopter and up to 800 m from the line for groups of seals and penguins. Vision was partly obscured for groups less than 100 m from the line, so any groups recorded at less than 100 m are discarded. At the instant that the group passed abeam of the helicopter, the angle of declination from the horizon was measured using a sighting gun. Perpendicular distances were calculated using the angle of declination and altitude data corresponding to the time of the sighting. Distances were therefore recorded as continuous data, not binned into intervals. Where responsive movement was detected before the group passed abeam, the perpendicular distance moved was estimated visually and subtracted from the perpendicular distance measured when the group was abeam of the helicopter (Southwell 2005).

Flights lasted up to 2.5 h and were broken into 20-min observation sessions separated by rest periods of 2–5 min. The time since the start of the flight ('flight-time') and time since the start

of the session ('session-time') were included as covariates possibly indicative of observer alertness. The time on-effort since the start of the survey ('experience') was also included as a measure of observer experience. Additional covariates recorded with each sighting were: species; group size; visibility (excellent, good, or poor); glare (present or absent); shadow (present or absent); observer; and observer position in the helicopter (front or back). Sightings and covariates were recorded digitally via an automated data-entry system (Southwell *et al.* 2002), and duplicate sightings for the front and back observers of each team were identified by matching the times and details of sightings. Extensive observer training was undertaken before the survey commenced.

### Queensland aerial survey for kangaroos

The survey region consisted of two survey blocks of ~10000 km<sup>2</sup> centred on the southern Queensland towns of Roma, in the brigalow belt bioregion, and Charleville, in the mulga lands bioregion. One helicopter (Robinson R44) with the doors removed was flown at an altitude of 61 m and a speed of 50 knots. The helicopter housed one team of double observers searching to the left side of the helicopter from the front and back. All surveys followed parallel east-west transect lines, ~80 km long and 10 km apart. In total, eight lines were surveyed in Roma, and seven in Charleville.

Observers searched ahead of the helicopter and up to 125 m outward. Three observers were involved in the survey, all of whom had over 100 h experience in helicopter surveys of kangaroos using line-transect sampling. Distances were measured in five binned distance categories each of 25 m width, from 0 m to 125 m. The distance classes were delineated on aluminium poles extending perpendicularly from either side of the helicopter. Distances were measured to the position at which the animal was first seen, to mitigate problems of responsive movement. Front and back sightings were recorded independently in continuous time into a dual-channel tape recorder, and duplicate sightings were identified afterwards from the recordings. Species and group size were included with every sighting. Neither observer was able to hear or see when the other observer made a sighting. Observers counted in 5-min units separated by breaks of 30 s.

### Capture-recapture analysis for the uniform assumption

Borchers *et al.* (2006) outline several different analysis options for data collected from double-platform line-transect surveys. The options are based on different combinations of the likelihoods from the line-transect data and from the double-observer capture-recapture data. Of these options, all except one rely on the uniform assumption. The only valid option for testing the uniform assumption is to use the single exception, which involves only the capture-recapture component of the likelihood. In the notation of Borchers *et al.* (2006), this is  $L_{\omega}(\theta)$ , given in their Eqn (1). We note that this method has known shortcomings, seen for example in Borchers *et al.* (1998a, 2006) and in Fewster and Pople (2008), so we proceed with caution and take steps to mitigate these shortcomings where possible.

Let  $\omega_i$  be the observed capture history for detected group  $i$ . This means that  $\omega_i$  is either (1, 0), (0, 1), or (1, 1), respectively denoting detection of group  $i$  by Observer 1 only, Observer 2

only, or both observers (a duplicate). For detected group  $i$ , we know the perpendicular distance,  $x_i$ , and the set of all other covariates, which we call  $z_i$ . We model the detection probability separately for the two observers. The probability that Observer  $j$  detects a group, given that the group is located at distance  $x$  and has covariate vector  $z$ , is modelled as.

$$p_j(x, z) = \frac{\exp\left(\alpha_j + \beta_j x + \sum_{r=1}^R \gamma_{jr} z_r\right)}{1 + \exp\left(\alpha_j + \beta_j x + \sum_{r=1}^R \gamma_{jr} z_r\right)} = \text{logit}^{-1}\left(\alpha_j + \beta_j x + \sum_{r=1}^R \gamma_{jr} z_r\right) \quad (1)$$

where  $j = 1$  or  $2$ , where  $z = (z_1, \dots, z_R)$ , and where the parameter vector to be estimated is  $\theta_j = (\alpha_j, \beta_j, \gamma_{j1}, \dots, \gamma_{jR})$ . For example, if the covariate vector contains only the group size,  $s$ , then the detection probabilities for the two different observers are modelled as:

$$p_1(x, z) = \text{logit}^{-1}(\alpha_1 + \beta_1 x + \gamma_1 s),$$

$$p_2(x, z) = \text{logit}^{-1}(\alpha_2 + \beta_2 x + \gamma_2 s),$$

and the parameters to be estimated are  $\theta = (\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2)$ . These parameters are estimated via the following likelihood:

$$L_{\omega}(\theta) = \prod_{i=1}^n \frac{P(\text{detected by the observer combination in } \omega_i | x_i, z_i)}{P(\text{detected} | x_i, z_i)} \quad (2)$$

where  $n$  is the total number of groups detected. The problem with this method lies in computing the numerator and denominator in Eqn (2), because we are forced to assume that observers detect groups independently at all covariate values  $(x, z)$ . This means that every possible covariate that impacts on detectability has to be included in the model, which is generally asking the impossible. If an important covariate is left out, but has the effect that each observer is more likely to detect the animal group (for example, a movement or behaviour cue), the result is that detections are no longer independent given only the covariates that are recorded. This shortcoming is known to be an important problem with this method, in common with all capture-recapture analyses. The impact of unmodelled covariates, or equivalently non-independence of observers, is thought to be minor at close distances but severe at large distances. The usual result is an over-estimation of detection probability at large distances, because both observers tend to see the same animal groups for unknown reasons, creating the illusion that each observer is seeing a higher proportion of the available groups than they really are. In terrestrial aerial surveys, a primary cause of observer dependence at large distances is the fact that both observers have the same line of sight to animal groups, so obstructions by scrub or other features act simultaneously for both observers but are not readily included in the detection model.

Under the independence assumption, termed 'full independence' by Borchers *et al.* (2006), the expressions in (2) are given by:

$$\begin{aligned} P(\text{detected with } (0, 1) | x, z) &= (1 - p_1)p_2, \\ P(\text{detected with } (1, 0) | x, z) &= p_1(1 - p_2), \\ P(\text{detected with } (1, 1) | x, z) &= p_1p_2, \\ P(\text{detected} | x, z) &= p_1 + p_2 - p_1p_2, \end{aligned} \quad (3)$$

where  $p_1 = p_1(x, z)$  and  $p_2 = p_2(x, z)$ . The parameters  $\theta = (\alpha_1, \beta_1, \dots)$  enter the likelihood (2) through the expressions for  $p_1$

and  $p_2$  in (1), and are estimated by maximising the likelihood. The estimated parameters can then be fed into the expressions in (3) to give estimates of detectability that do not depend on the uniform assumption, and do not assume that detectability necessarily decreases with distance from the line.

To provide even more flexibility in the shape of the detection function, we also extended the logistic formulation in (1), to include smooth terms similar to those in generalised additive models, for every continuous covariate. The smooth terms were fitted with smoothing splines with up to three degrees of freedom, and allow detectability to depend upon the covariate in a very flexible, non-prescribed way.

We used an exhaustive selection procedure for model selection to select between different covariates for the detection functions  $p_1(x, \mathbf{z})$  and  $p_2(x, \mathbf{z})$ . The likelihood (2) was maximised using every combination of available covariates. Categorical variables (e.g. visibility and glare) were either included or omitted, while continuous variables (e.g. observer experience) were either included as a smooth term with three degrees of freedom, included as a linear term as in Eqn (1), or omitted. The selected model was chosen with AIC. Following Southwell *et al.* (2007), a secondary model selection was carried out in which each selected term was allowed to include or omit an interaction with perpendicular distance. AIC was used to select the final model from this secondary procedure.

#### *Test for uniformity*

We test for uniformity using the method of Southwell *et al.* (2004: 608), who evaluated the uniform assumption for ship-board surveys of crabeater seals in Antarctica. Using the estimated detection probabilities  $P(\text{detected} | x, \mathbf{z})$ , we can estimate the overall probability  $p(x) = P(\text{detected} | x)$  that a group is detected at distance  $x$ , averaged over all other covariates  $\mathbf{z}$ . The averaging uses covariates in the proportions in which they are estimated to occur in reality (e.g. see Borchers *et al.* 1998b, eqn 17), so that highly detectable covariate sets are not over-represented. The search strip is then divided into  $K$  distance classes. The overall probability that a group is detected by at least one observer, given that it is in distance class  $k$ , is then estimated as  $\int_l^u p(x) dx$ , where  $l$  and  $u$  are the lower and upper boundaries of class  $k$ .

Under the null hypothesis of uniform distances within a survey, the expected number of animal groups (detected and undetected) in each distance class 1, ...,  $K$  is equal. Among the detected groups only, the expected numbers typically decline with distance, because detectability typically declines with distance. Under the null hypothesis of uniformity, the expected number of detected groups in each distance class is affected only by the detection probability  $p(x)$ . For example, if it is estimated that animal groups in distance class  $k$  have half the detection probability of groups in distance class 1, then the expected number of detections in class  $k$  should be half the expected number of detections in class 1. Under the alternative hypothesis of non-uniformity, there might be different numbers of animal groups available for detection in class  $k$  and class 1, so we could not assume the same relationship between expected number of detections and detection probability for any class.

The test for uniformity is illustrated in Fig. 2, and with real data in Figs 3, 4, 6, and 7. Out of the  $n$  sightings made in total, the number that should be expected to fall into distance class  $k$  under the uniform hypothesis is  $n \times \int_l^u p(x) dx / \int_0^w p(x) dx$ , where  $w$  is the search strip half-width. These estimated expected frequencies are shown as bold lines on Figs 2–7. The histograms in these figures give the observed numbers of sightings in each distance class. To test uniformity, we test whether the observed frequencies in the  $K$  distance classes are consistent with the expected frequencies (in other words, whether the histograms are consistent with the predictions from the bold lines). A standard  $\chi^2$  goodness-of-fit test is used, with  $(K - 1)$  degrees of freedom.

Fig. 2 shows four different scenarios for the uniform test. Case A is obtained when the null hypothesis of uniformity holds, and the observers have full independence so the analysis based on the capture–recapture likelihood (2) is valid. Case D shows the outcome under the null hypothesis of uniformity, when the observers do not have full independence so the capture–recapture analysis is invalid. In Case D, we expect detectability to be overestimated in high distance classes due to failure of independence (too many duplicates) at these distances. The expected frequencies will therefore be too high in the far-distance classes. The test shares out the  $n$  available animal groups among the  $K$  classes, so the expected frequencies in the near-distance classes are consequently too low. Case D is the typical scenario obtained when the capture–recapture independence assumption is violated: there appear to be too many observations in the near-distance classes, and too few in the far-distance classes. This may lead to failure of the uniformity test due to a failure in the capture–recapture independence assumption, even though there is no violation of the uniform assumption.

The same pattern may also arise when the uniform assumption is genuinely violated by responsive movement towards the observer (Fig. 2, Case C). It is impossible to know whether a test result similar to Case C or D in Fig. 2 is due to responsive movement towards the observer, or failure of the capture–recapture independence assumption with no implication about violation of uniformity, or both.

Violation of the uniform assumption due to responsive movement away from the observer would tend to cause the reverse pattern (Fig. 2, Case B), in which the observed frequencies appear too low in the near-distance classes, and too high in the far-distance classes. It is less likely that this pattern would be caused by failure of the capture–recapture independence assumption.

Fig. 2 demonstrates that the uniformity test proposed here cannot distinguish between genuine non-uniformity and failure of the capture–recapture independence assumption. Genuine non-uniformity may be caused by responsive movement, or alternatively by a chance imbalance of distances within the survey, also known as sampling variability. Chance imbalances are possible for any survey, but are most likely to occur when the number of transects is small, the population is aggregated, or the terrain is inhomogeneous. If the mode of failure of the test is consistent among many independent surveys of the same species, however, the systematic effects of responsive movement or failure of the capture–recapture assumption are more likely to be responsible.

For the East Antarctica survey, distances were measured exactly rather than in intervals. For these analyses we may also conduct a Kolmogorov–Smirnov (KS) test to test the null hypothesis of uniformity. The KS test compares the empirical distribution function of detected distances with the distribution function that would be expected under the uniform hypothesis, which can be calculated from the bold lines of Fig. 2. It is complementary to the  $\chi^2$  test, with different strengths and weaknesses, and will not always give the same result. The  $\chi^2$  test can be sensitive to the choice of the  $K$  distance classes underlying the test, especially if there are inherent problems in the data such as rounding errors. The KS test does not have this problem, but tends to lack power when the distributions differ primarily near the boundaries of the search strip at 0 and  $w$ . As a rule, the  $\chi^2$  test is more powerful, but it is always useful to conduct both tests.

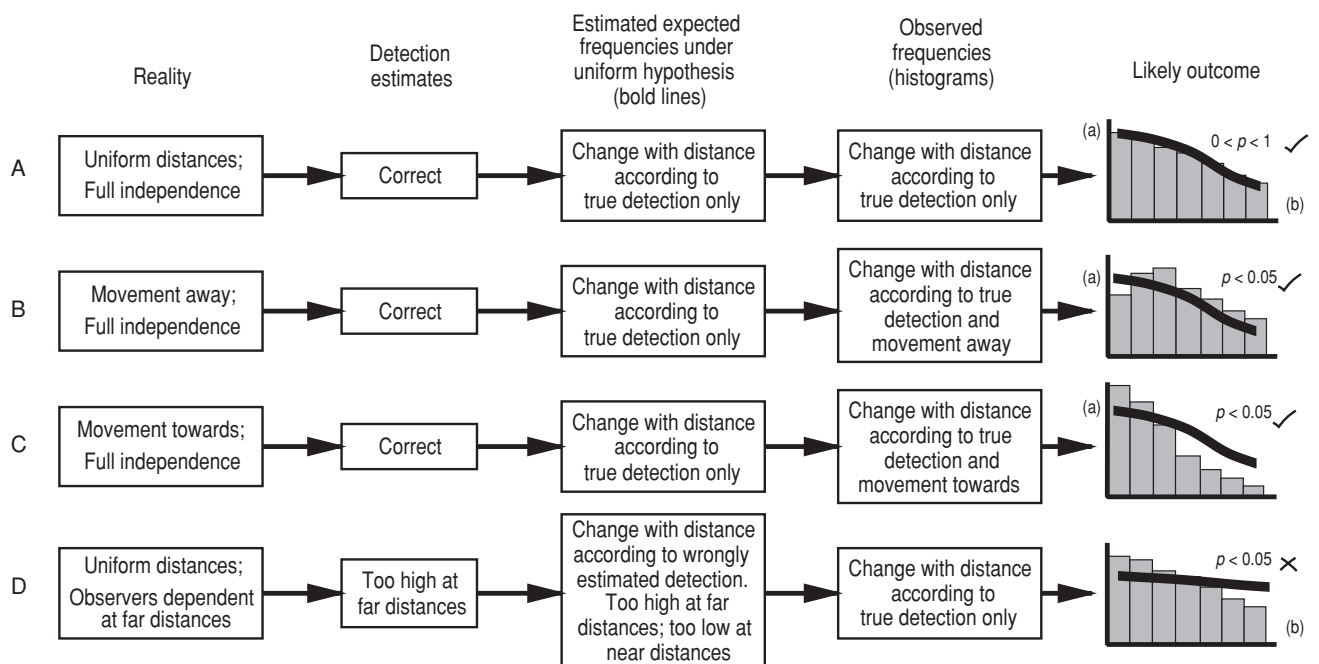
*Analysis of the seal data*

The crabeater seal dataset was very large, with 1740 seal groups detected. A previous analysis of these data highlighted marked observer differences in the detection functions (fig. 2 of Southwell *et al.* 2007). This suggests that we should be cautious, because unmodelled or incorrectly modelled heterogeneity will invalidate the capture–recapture analysis for our uniform test. Because we are not attempting to estimate seal density, maintaining the sample size is not a priority, so we partition the data and analyse different observer teams separately. Each analysis

involves a single pair of observers in a single front–back configuration. Model selection is carried out separately for each analysis, so detectability effects important to individual observer teams can be isolated.

Southwell *et al.* (2007) also showed that the independence assumption on which the validating capture–recapture analysis is based might not be satisfactory for the crabeater seal data at far distances. Fig. 4 of their paper shows that the independence assumption creates a severe discrepancy between the observed distance data and the estimated conditional detection functions (their dashed lines). While we cannot be sure that this effect is due to failure of independence rather than non-uniformity, the magnitude of the effect makes this very likely given the size of the dataset and the anticipated immobility of seals on ice. The discrepancy is not severe at near distances, however, and it is at near distances that non-uniformity due to responsive movement is most likely to arise. By restricting our analyses to near distance classes only, we have reasonable confidence in the capture–recapture independence assumption, and we are able to test for uniformity in the realm where it is most likely to be violated. We decided upon 400-m truncation as a compromise between minimising the independence problems and ensuring sufficient sample size for analyses.

We present results from the four observer teams with the largest sample sizes after truncation, ranging from  $n = 93$  to  $n = 136$  detected groups. Because different observer teams oper-



**Fig. 2.** Performance of the test for uniformity under four possible scenarios. In cases A, B and C there is full independence between observers (no unmodelled covariates) and the test is expected to give the correct outcomes of  $0 < P < 1$  when distances are genuinely uniform, and  $P < 0.05$  when uniformity is violated by responsive movement. In case D the assumption of independent observers is violated in the validating capture–recapture method (for example, due to unmodelled covariates) and the test gives the wrong outcome of  $P < 0.05$  even though the uniform assumption is valid. The three bold lines marked (a) denote the true pattern of detection, and are all identical. The two histograms marked (b) result from uniformity combined with true detection and are identical. The outcomes C and D are not distinguishable: a test result with the bold line too low at near distances and too high at far distances could be due either to violation of uniformity due to responsive movement towards the observer, or to violation of the independence assumption in the validating capture–recapture analysis. Note that the detection curve should be estimated correctly in Cases B and C despite the violation of uniformity, because the validating method does not rely on the uniform assumption.

ated over different spatial areas, the data partitioning has the additional advantage of providing four independent surveys of crabeater seals, so that we can check on the consistency of our results. However, it also creates relatively small samples, and this will affect the power of the test for uniformity. For this reason, we also apply the test to the full dataset of all observer combinations, although still truncating by distance.

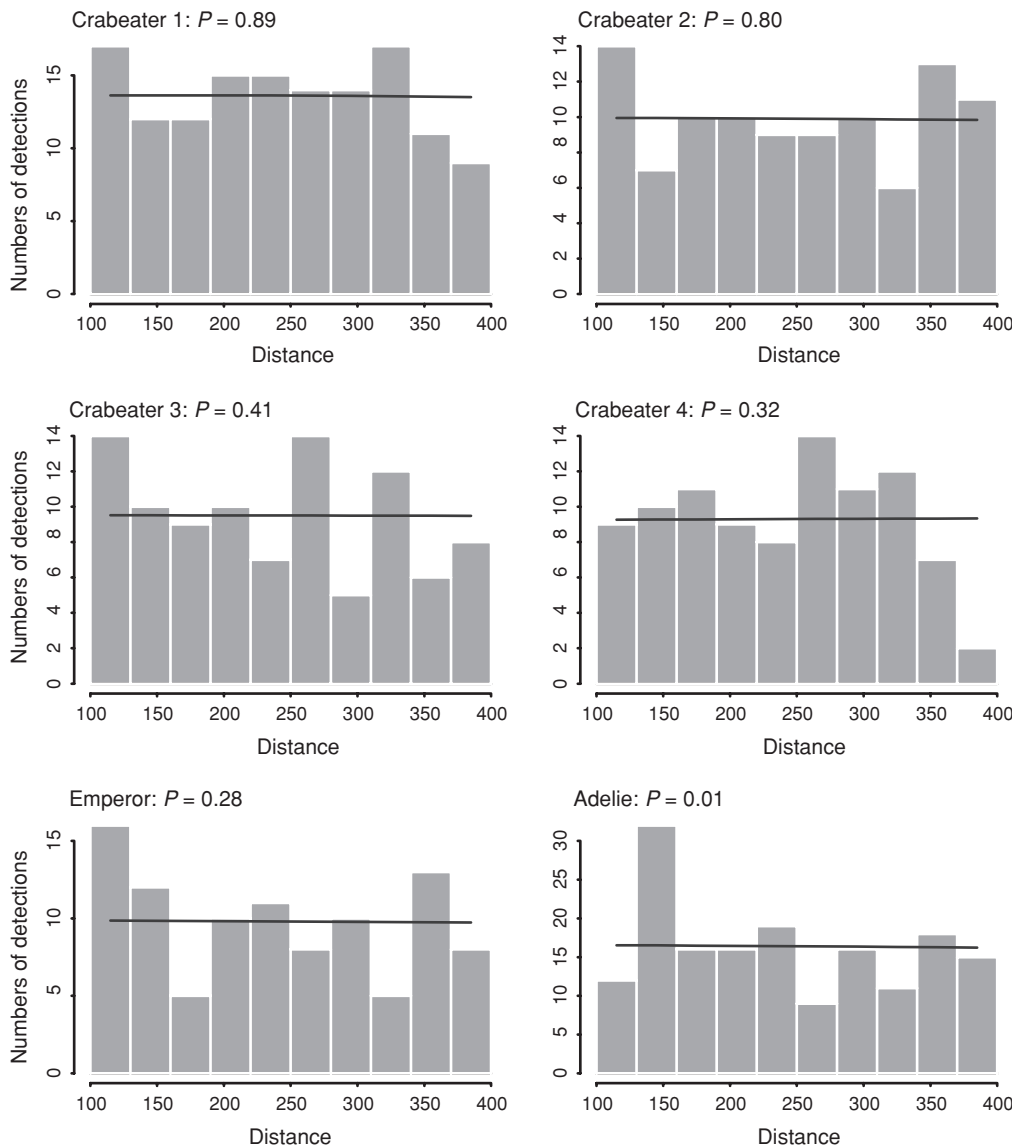
*Emperor penguins and Adelie penguins*

Available sample sizes were considerably smaller for penguins, so it was not possible to partition the data according to different observer teams. Instead, model selection for the detection functions in (1) allowed categorical variables to be included for observer and observer position (front or back), together with the interaction between observer and position. We used the same distance truncation (400 m) as for the seal data. The remaining sample sizes were 98 and 164 for emperor and Adelie penguins respectively. The two species were analysed separately.

*Eastern grey kangaroos and red kangaroos*

The Queensland survey differed from the East Antarctica survey because distances were recorded in binned intervals, rather than continuously, and only group size and observer were recorded as covariates. We analysed the data from different observer teams separately, creating two independent analyses for the Roma survey block, and two for the Charleville survey block. Red kangaroos were too sparse in Roma to provide an analysis. Each of the analyses is based on four transects of ~80 km, except for the second Charleville analysis, which is based on three transects.

Although we expect that the independence problem at greater distances will be a severe influence for the kangaroo data, as for the seals and penguins, truncation is harder for the kangaroos because we have only five distance classes. We therefore present results from the uniform test for all surveys using all five distance classes to 125 m, and separately using three distance classes to 75 m. Model fitting was carried out separately for the



**Fig. 3.** Results from the  $\chi^2$  test for uniformity using data for crabeater seals (four observer combinations), emperor penguins, and Adelie penguins. Ten distance classes from 100 m out to 400 m are used for the  $\chi^2$  test (d.f. = 9), and *P*-values from the test are shown. The histograms show observed numbers of detections in each distance class, and the thick lines show the estimated expected number of detections under the null hypothesis of uniformity. Sample sizes are respectively 136, 99, 95, 93, 98 and 164.



two cases. Sample sizes ranged from 62 to 256 for the 125-m analyses, and from 48 to 193 for the 75-m truncated analyses.

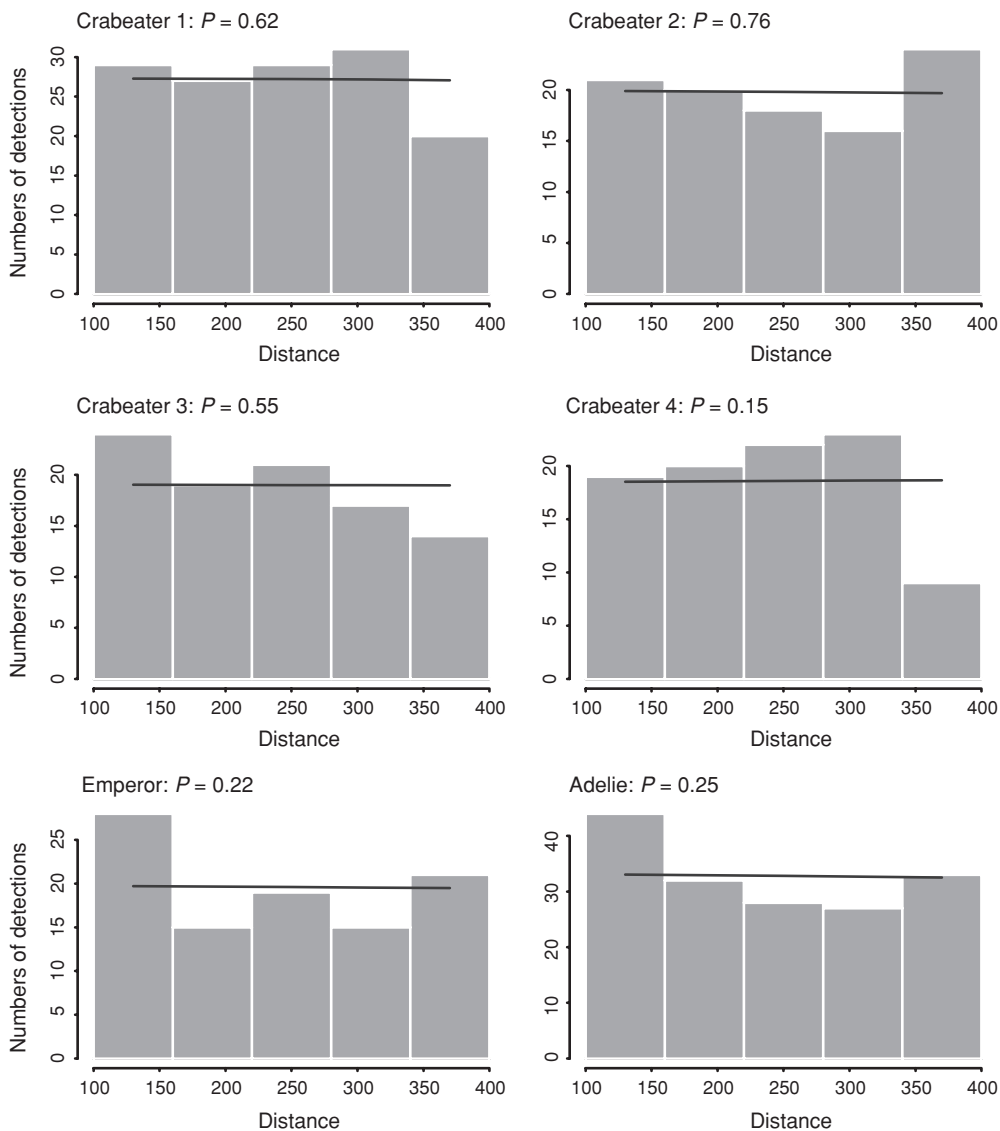
**Results**

Fig. 3 shows results from the  $\chi^2$  goodness-of-fit test using 10 distance classes from 100 m to 400 m for the crabeater seal, emperor penguin, and Adelie penguin analyses. No evidence is found against the uniform assumption for any of the seal analyses ( $P = 0.89, 0.80, 0.41, 0.32$ ). Visual inspection suggests that there might be some systematic departure at higher distances in the Crabeater 4 analysis, but it does not reach statistical significance.

Selected capture–recapture models ranged widely in the covariates included for different observer teams, ranging from only perpendicular distance and observer experience in the Crabeater 2 analysis, to a complex model including six covariates, three smooth terms and two interaction terms in the Crabeater 1 analysis. Interestingly, even the complex models still produced expected number of sightings that were nearly constant across the 10 distance classes (bold lines in Fig. 3),

suggesting that detectability was nearly constant within 400 m of the line. This was checked against sightings histograms for each individual observer in a team, confirming that the flatness did not arise due to observer dependence, as in Case D of Fig. 2. We conclude that the truncation scheme has been effective, leaving little heterogeneity in detection throughout the strip, and we therefore have confidence in the validating capture–recapture analysis.

The results above are consistent for four analyses constituting four independent surveys of crabeater seals. To check whether the results are affected by the small sample sizes, we also applied the uniformity test to the whole crabeater seal dataset, combining all observers in the same analysis. We did this by adding a categorical variable for observer combination to each of the four selected models above. With distances truncated at 400 m, there was evidence that either uniformity or the validating independence assumption were violated ( $P = 0.0067$  for each model; sample size  $n = 1087$ ). However, with distances truncated at 300 m, there was no evidence against uniformity ( $P = 0.36$  for

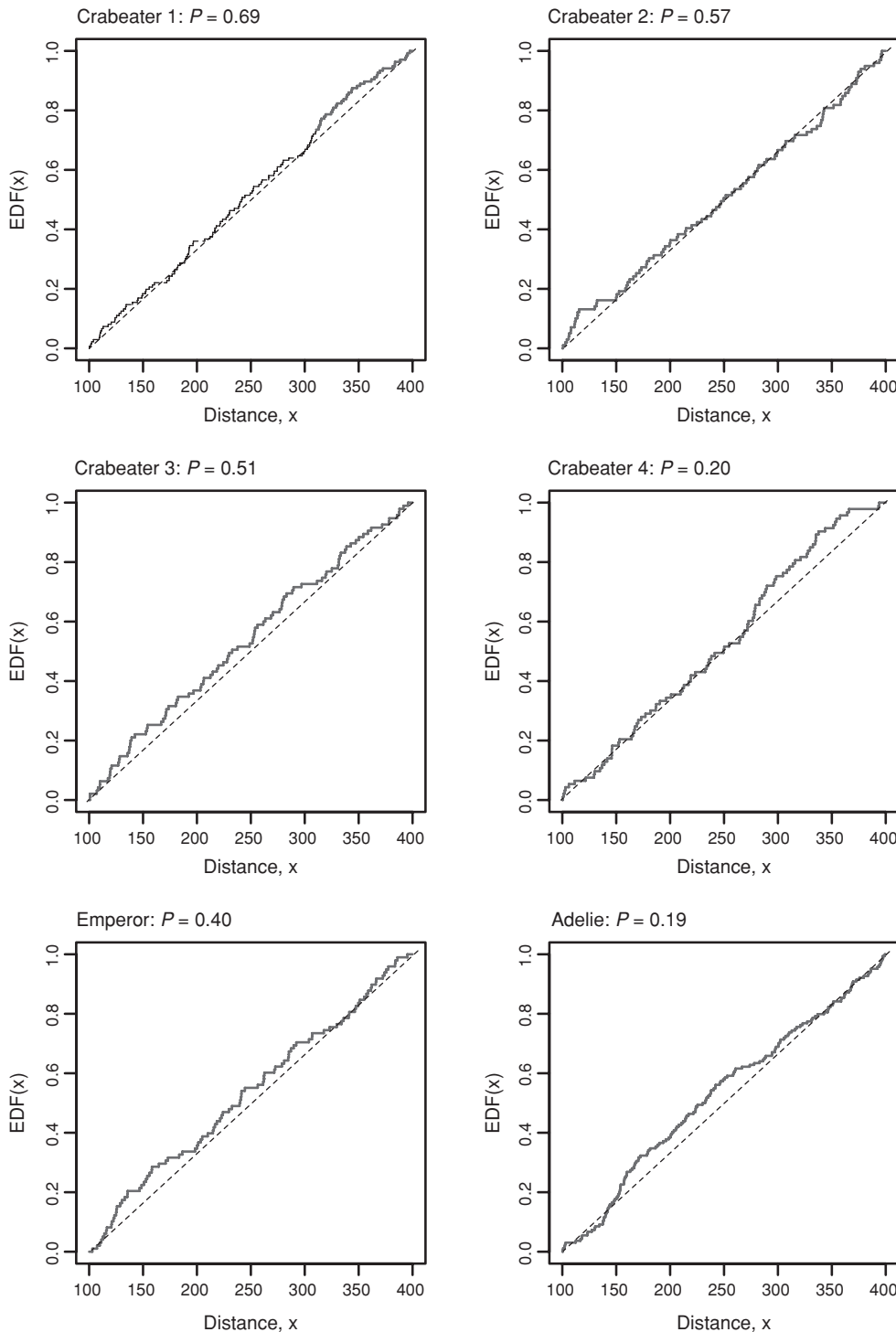


**Fig. 4.** Results from the  $\chi^2$  test for uniformity using data for crabeater seals (four observer combinations), emperor penguins, and Adelie penguins. Five distance classes from 100 m out to 400 m are used for the  $\chi^2$  test (d.f. = 4). Sample sizes are as for Fig. 3.

each model;  $n = 779$ ). Given that there is no evidence against uniformity in the closest distance classes up to 300 m, where responsive movement is likely to be most problematic, we suspect that the significant result for 400 m is due to a breakdown of observer independence between 300 m and 400 m. This is corroborated by inspection of the histograms in Figs 3 and 4, where departures from the bold lines become somewhat evident in the final distance classes between 300 m and 400 m.

The emperor penguin analysis did not present evidence against the uniform assumption ( $P = 0.28$ ). A simple model involving distance, observer position (but not observer), and a smooth term for observer experience was selected.

For Adelie penguins, significant evidence against the uniform assumption was found ( $P = 0.01$ ). From Fig. 3 it is evident that there are too few sightings in the nearest distance class, and too many in the next class. This is strongly suggestive



**Fig. 5.** Results from the Kolmogorov–Smirnov test for uniformity using data for crabeater seals (four observer combinations), emperor penguins, and Adelie penguins. The dashed lines give the estimated expected distribution function for detected distances under the null hypothesis of uniformity. The jagged lines give the empirical distribution function of actual detected distances. The test measures the maximum vertical distance between the dashed lines and the jagged lines. Sample sizes are as for Fig. 3.

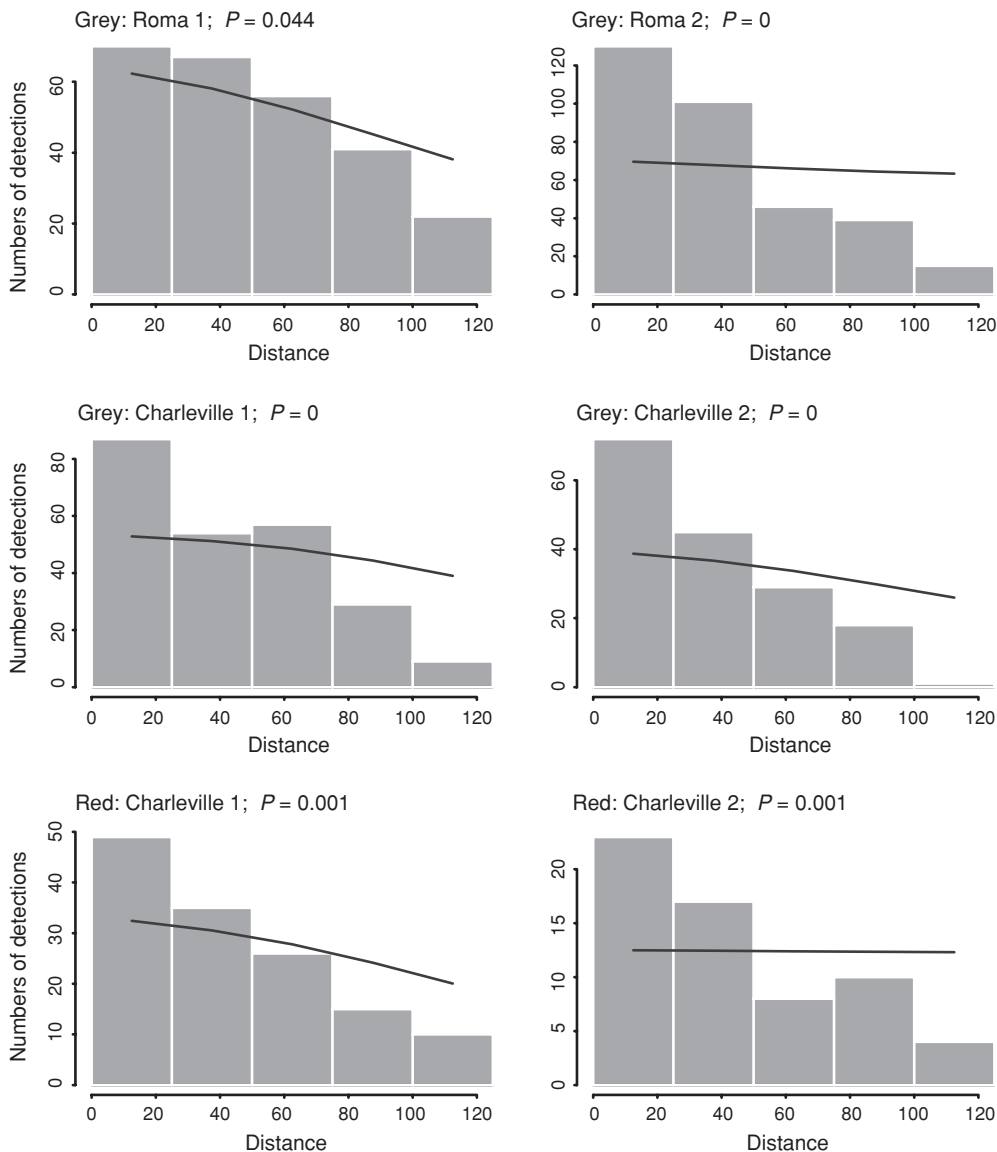
of responsive movement away from the helicopter for the penguins closest to the helicopter. The selected model for Adelie penguins was reasonably complex, involving distance, experience, flight time, observer, observer position, and an interaction between distance and observer position. This may reflect the fact that Adelie penguins are smaller and less distinctive than emperor penguins, and more cryptic in the light–dark conditions encountered on the pack-ice.

Fig. 4 shows results from the same analyses grouped into five distance classes instead of ten. The results are largely in agreement with Fig. 3, with the Crabeater 4 analysis becoming a little more marginal. The most interesting result is that there is now no evidence against uniformity for Adelie penguins ( $P = 0.25$ ). This may be explained by the wider distance classes masking the effect of responsive movement, where penguins move outwards away from the transect line. Non-uniformity was detected over the 30-m distance classes in Fig. 3, but not over the 60-m classes in Fig. 4. This is consistent with the findings of

Southwell (2005), who observed responsive movement of up to 20 m in penguins. Undetected movement leading to a deficit inside the nearest 30-m band but not inside the wider 60-m band is a reasonable finding for these penguins.

Fig. 5 shows results from the KS test for the seal and penguin analyses. While mostly in agreement with the previous results, the KS test fails to identify the discrepancy with the Adelie data ( $P = 0.19$ ). This is consistent with the low power of the KS test at the upper and lower boundaries of the search strip. The KS test measures the maximum vertical distance between the jagged lines and the dashed lines, and this distance is restricted at the boundaries. For most analyses, the jagged line follows the dashed line closely.

Figs 6 and 7 show results from the kangaroo analyses with respectively 125-m and 75-m truncation. Here we see severe lack of fit in every case except for the 75-m analyses in Roma 1 (eastern grey kangaroo;  $P = 0.95$ ) and Charleville 1 (red kangaroo;  $P = 0.27$ ). It is clear that either the uniform assumption is

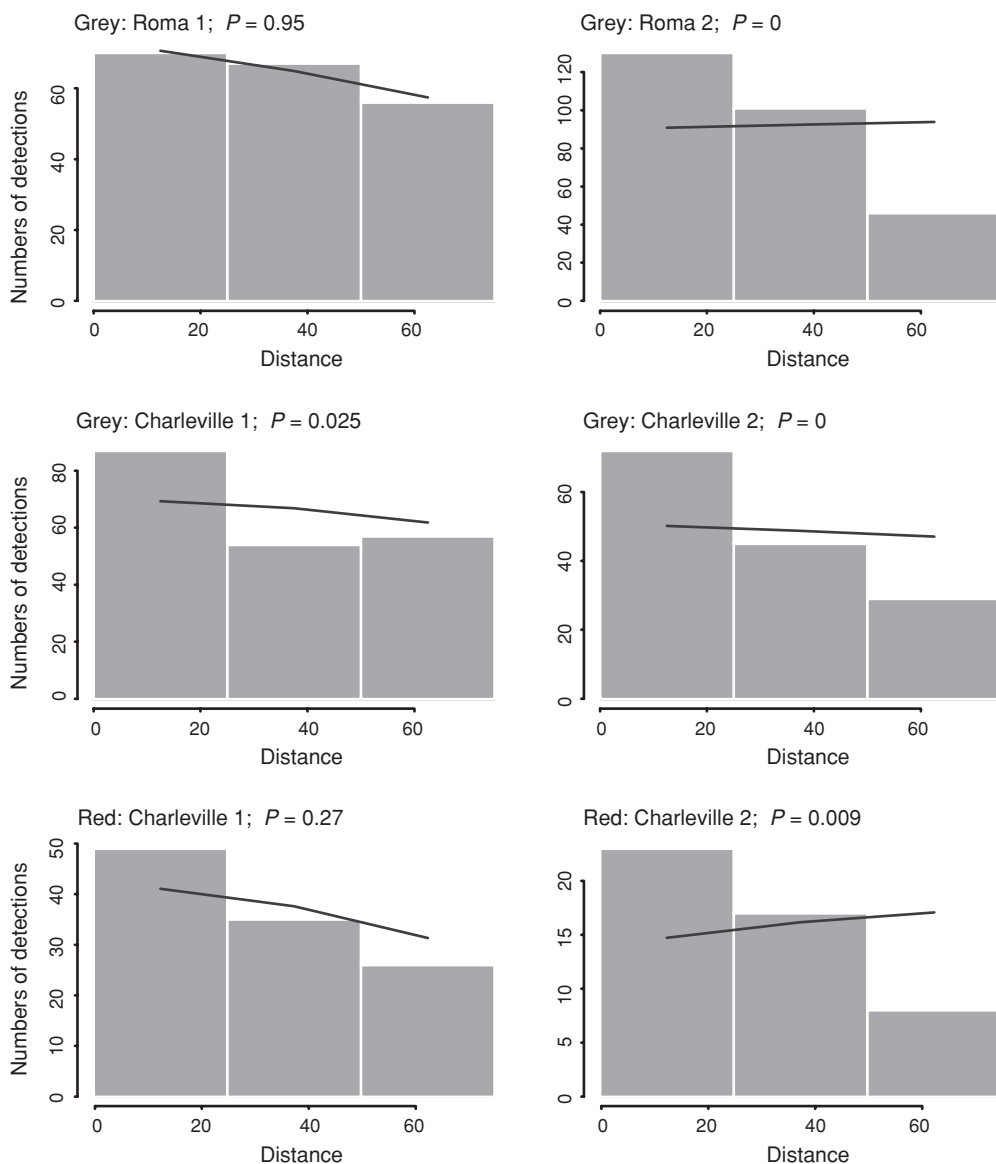


**Fig. 6.** Results from the  $\chi^2$  test for uniformity using data from eastern grey and red kangaroos in four survey blocks in Queensland. Five distance classes from 0 m out to 125 m are used for the  $\chi^2$  test (d.f. = 4). Sample sizes are respectively 256, 331, 236, 165, 135 and 62.

violated by these data, or the independence assumption underlying the validating capture–recapture analysis is violated, or both. In every case, the line denoting expected frequencies lies below the observed frequencies in the near classes, and above the observed frequencies in the far classes. This pattern may arise from severe responsive movement towards the transect line, but it is also characteristic of a failure of the capture–recapture independence assumption, as in Fig. 2, Case D. Unknown cues in far distance classes that cause both observers to detect the animal group lead to too many duplicate sightings at far distances. Each observer detects a high proportion of the other’s sightings, which overestimates detectability at far distances. In our uniform test, this causes the expected number of detections at far distances to be too high, and consequently the expected number at near distances to be too low.

If the severe lack of fit in Figs 6 and 7 is due to genuine non-uniformity for these data, the possible explanations are severe undetected responsive movement towards the helicopter, or

chance imbalance (sampling variability), perhaps attributable to the small number of transect lines in each survey block. The explanation of chance imbalance is unlikely, because the same pattern in the histograms is seen across four independent survey blocks and two different species. Although the number of lines was small (three or four), the line lengths were long (80 km), over-reaching the extent of local spatial aggregations or features of the terrain. The consistent pattern across the different survey blocks suggests that the explanation is a systematic effect: either genuine non-uniformity due to responsive movement, or failure of the capture–recapture independence assumption. Consistent and extensive movement of kangaroos towards the flight path is strongly counter to our field observations, which suggest that kangaroos close to the line may move in any direction, while movement tends to be outwards from the line for more distant kangaroos. It seems more plausible that the severe lack of fit in Figs 6 and 7 is due to a failure of the capture–recapture independence assumption and cannot be taken as evidence for



**Fig. 7.** Results from the  $\chi^2$  test for uniformity for kangaroos, using three distance classes from 0 m out to 75 m for the  $\chi^2$  test (d.f. = 2). Sample sizes are respectively 193, 277, 198, 146, 110 and 48.

non-uniformity. We can only conclude that evidence for non-uniformity in the kangaroo case is confounded with issues of non-independence in the validating method, and that the test results in Figs 6 and 7 cannot be used to draw conclusions about non-uniformity or responsive movement.

It is interesting that detectability was occasionally estimated to increase with distance in kangaroo surveys, notably for red kangaroos in the Charleville 2 analysis (Fig. 7). This can be caused by observer dependence at far distances, coupled with the small sample size ( $n = 48$ ). A detailed explanation is shown in fig. 3 of Fewster and Pople (2008).

## Discussion

We have shown that the assumption of uniform distances in line-transect sampling follows automatically from the survey design if all other assumptions are met. The chief problem with uniformity is likely to be movement in response to the observer, before detection. The extent of this effect is highly dependent on the behavioural tendencies and intrinsic mobility of the taxa concerned. Our analyses revealed little or no evidence against uniformity for crabeater seals, which are largely immobile. For the species of intermediate mobility, no evidence against uniformity was found for emperor penguins, but some evidence was found for Adelle penguins, probably reflecting undetected responsive movement of up to 30 m away from the helicopter. This finding is consistent with our field observations of the two species, which have noted that Adelle penguins are more responsive to the aircraft than emperor penguins and tend to move more quickly, thereby covering greater distances before detection.

For the highly mobile kangaroos, it was impossible to distinguish effects of non-uniformity from problems with the independence assumption in the validating method. The analysis appears to indicate a strong net movement of kangaroos towards the helicopter, which runs counter to our experience that kangaroos more commonly move outwards away from the flight path. For this reason, we assume that the dominant influence on the kangaroo results is the failure of the capture–recapture independence assumption, rather than attraction of kangaroos to the helicopter. It is likely that responsive movement is problematic for aerial surveys of kangaroos, but our analysis is unable to measure its extent or direction.

Failure of the uniform requirement has serious consequences for density estimates of animal populations. The only way to remove reliance on the uniform assumption is to use an analysis based completely on capture–recapture, as we used for our validating method. However, capture–recapture methods are notoriously vulnerable to unmodelled heterogeneity, or equivalently to failure of the assumption that observers are detecting animals independently given the covariates included in the model. The point-independence methodology of Borchers *et al.* (2006) relaxes the full independence assumption to an assumption of independence only at a single point, which is usually located on the transect line. However, it reverts to the uniform assumption to enable this move.

For most aerial surveys, we believe that the uniform assumption is a weaker assumption than that of full observer independence. In the absence of responsive movement, the uniform requirement can be fulfilled by using a randomised design and employing large numbers of transects. If responsive movement

is suspected, the problem is identifiable, and there is the potential to develop field methods addressing the issue. By contrast, the causes of unmodelled heterogeneity, leading to a violation of full observer independence, are, by definition, unknown. Addressing responsive movement may pose serious challenges in the field, but the problem is at least known and specific.

We therefore favour analysis methods based on the uniform assumption for estimating animal density, notably the point-independence method of Borchers *et al.* (2006), which combines line-transect data with double-observer capture–recapture data in a robust way. However, when using these methods, practitioners should be aware of the importance of detecting and correcting for responsive movement. In aerial surveys, it is important to search ahead of the aircraft and fix locations of animals before responsive movement begins. The perpendicular distance to the group must be that at first detection rather than the distance observed when the group is abeam of the aircraft. Protocols for sighting and for distance measurement should be established during observer training, and the influence of undetected responsive movement on the final result should be stressed. Measurement of distances before the group passes abeam presents practical problems for aerial surveys, and development of instruments to facilitate this would be an important advance. For example, accurate measurements of animals ahead of the aircraft could be obtained by developing a sighting gun that measures both declination from the horizon and planar angle from the transect line simultaneously.

## Acknowledgements

We thank the observers, pilots, and support staff of both surveys. The East Antarctica survey was conducted under the Antarctic Pack Ice Seal program supported by the Australian Antarctic Division. The Queensland survey was supported by the Queensland Parks and Wildlife Service. We thank two referees and the editors for their detailed and constructive comments.

## References

- Borchers, D. L., Zucchini, W., and Fewster, R. M. (1998a). 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. (1998b). 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 unmodelled 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.)
- Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (Eds) (2004). ‘Advanced Distance Sampling.’ (Oxford University Press: Oxford.)
- Fewster, R. M., and Buckland, S. T. (2004). Assessment of distance sampling estimators. 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.)
- Fewster, R. M., and Pople, A. R. (2008). A comparison of mark–recapture distance sampling methods applied to aerial surveys of eastern grey kangaroos. *Wildlife Research* **35**, 320–330. doi:10.1071/WR07078

- Fewster, R. M., Laake, J. L., and Buckland, S. T. (2005). Line transect sampling in small and large regions. *Biometrics* **61**, 856–859. doi:10.1111/j.1541-0420.2005.00413\_1.x
- Fewster, R. M., Buckland, S. T., Burnham, K. P., Borchers, D. L., Jupp, P. E., Laake, J. L., and Thomas, L. (2008). Estimating the encounter rate variance in distance sampling. *Biometrics* doi:10.1111/j.1541-0420.2008.0108.x.
- Melville, G. J., and Welsh, A. H. (2001). Line transect sampling in small regions. *Biometrics* **57**, 1130–1137. doi:10.1111/j.0006-341X.2001.01130.x
- Southwell, C. (2005). Response behaviour of seals and penguins to helicopter surveys over the pack ice off East Antarctica. *Antarctic Science* **17**, 328–334. doi:10.1017/S0954102005002798
- Southwell, C., de la Mare, W. K., Underwood, M., Quatararo, F., and Cope, K. (2002). An automated system to log and process distance sight–resight aerial survey data. *Wildlife Society Bulletin* **30**, 394–404.
- Southwell, C., de la Mare, B., Borchers, D., and Burt, L. (2004). Shipboard line transect surveys of crabeater seal abundance in the pack-ice off East Antarctica: evaluation of assumptions. *Marine Mammal Science* **20**, 602–620. doi:10.1111/j.1748-7692.2004.tb01182.x
- 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 East Antarctica. *Wildlife Research* **35**, 349–357. doi:10.1071/WR07093

Manuscript received 29 June 2007, accepted 28 March 2008