

special issue:

Advances in tools for bird population
monitoring in New Zealand

Detection probability for estimating bird density on New Zealand sheep & beef farms

Florian Weller^{1*}, Grant Blackwell^{1,2} and Henrik Moller¹¹Centre for Study of Agriculture, Food & Environment, University of Otago, PO Box 56, Dunedin 9054, New Zealand²Present Address: Parliamentary Commissioner for the Environment, PO Box 10-241, Wellington 6143, New Zealand*Author for correspondence (Email: florian.g.weller@gmail.com)

Published on-line: 30 July 2012

Abstract: Factors influencing detection probability in line transect distance sampling were investigated to estimate the abundance of four common farmland birds on 12 sheep & beef farms in the South Island of New Zealand. Our primary aim was to evaluate the necessity of employing distance methods to correct for heterogeneity in detection probability. Detections of skylark *Alauda arvensis*, blackbird *Turdus merula*, song thrush *Turdus philomelos*, and Australian magpie *Gymnorhina tibicen* were recorded using ten 500-m unbounded line transects on each of 9–10 visits, and modelled using multiple covariate distance sampling methods. Covariates of detectability played a strong role in model fitting, but showed few consistent directional trends within species. Increased woody vegetation cover on farms greatly decreased detectability, while few seasonal or geographical effects were found. No detectability differences were found between farms using certified organic, integrated management or conventional farming systems, indicating that bird population dynamics might be compared between systems using simpler index counts. However, unless detectability parameters can be standardised to a high degree within a survey, we recommend the use of analysis methods that incorporate heterogeneity in detection probability.

Keywords: detectability; distance sampling; farm management systems; farmland birds; multiple covariate modelling

Introduction

Most bird monitoring in New Zealand has been done in protected areas, on native species of conventional conservation interest (e.g. O'Donnell 1996; Hooson & Jamieson 2003; Powlesland et al. 2003). More recently, environmental managers and researchers have emphasised a need to increase our knowledge of the composition and health of species assemblages in the 'forgotten 60%' – the large part of New Zealand that is in agricultural or other production landscapes (Norton 1998; Perley et al. 2001; Moller et al. 2005; Macleod et al. 2008). Introduced species now dominate production landscapes and form an integral part of their ecological networks. Nevertheless many native species persist in the predominantly low lying and fertile agricultural landscapes – the very habitats where biodiversity might flourish most with appropriate land use and habitat protection (Blackwell et al. 2008; Meadows et al. 2008; Moller et al. 2008; Blackwell et al. 2011). Among potential indicators of agricultural ecosystem health, bird abundance and diversity stand out as likely candidates because birds are near the top of food chains, relatively easy to monitor, well recognised and familiar to consumers and farmers alike. The latter makes them potentially useful as flagship species to incentivise environmental care, especially if maintaining

bird abundance is incorporated into market Quality Assurance programmes for securing market access and premium prices for farm produce (Moller et al. 2005; Coleman et al. 2009). Reliable and efficient survey methods are needed to determine the state and dynamics of bird populations. The overall aim of the present study was to explore the viability of distance sampling as a methodology for assessing populations of common birds on farmland.

Several survey methods suitable for the monitoring of bird populations are available, and their appropriateness for different circumstances has been widely discussed (e.g. Buckland et al. 2000; Borchers et al. 2002; Thompson 2002; Johnson 2008). For the present survey, focus was placed on the type of methodology that aims to establish an estimate of detectability for use in the adjustment of raw counts into estimates of absolute abundance (Buckland et al. 2000; Morrison et al. 2001). Raw counts and similar related 'relative indices' are inexpensive and their relative simplicity can encourage participation of local stakeholders in monitoring, an important opportunity to build support for environmental care (Agrawal 2005). However, relative indices can only reliably indicate changes in bird abundance between habitats or places, or trends in abundance at the same habitats and places, if detectability remains about constant. This study

This special issue reviews the advances in tools for bird population monitoring in New Zealand. This issue is available at www.newzealandecology.org/nzje/.

New Zealand Journal of Ecology (2012) 36(3): 0-0 © New Zealand Ecological Society.

is part of the Agriculture Research Group On Sustainability (ARGOS) project that seeks to compare biodiversity between certified organic, integrated management or conventional farming (Moller et al. 2005). Relative indices, like raw bird counts or 'five-minute bird counts' (Dawson & Bull 1975), could only reliably compare bird abundance between these farming systems if detectability remained similar. This paper focuses on understanding variation in detectability to guide agricultural industry agents and policymakers on the most efficient way of monitoring birds to engender environmental care, but we expect the general issue of relative detectability to be important for all conservation agendas where bird abundance is compared across habitats and for long-term trend analysis.

Line transect distance sampling (Buckland et al. 2001, 2004) was chosen to measure changes in detectability because: (1) it is relatively easy to implement in the field; (2) well-proven computer software is available for analyses; (3) the central assumption of declining detectability with increasing distance from a transect holds well for relatively open farm landscapes; (4) the method has been intensively documented and tested (Norvell et al. 2003; Newson et al. 2005, 2008; Buckland 2006). By implementing suitable stratification, potential sources of imprecision can be reduced at the design stage (Buckland et al. 2001; Johnson 2008). Also a highly useful feature of modern distance analysis is the ability to introduce likely parameters into the analysis as covariates of detectability (Marques & Buckland 2003; Marques et al. 2007). Knowledge of the covariates and parameters that affect detectability can help researchers at both planning and analysis stages to derive more reliable density estimates.

Our survey was limited to four common focal species in order to enhance precision of sightings and achieve higher efficiency and replication in the field. The species chosen for monitoring were skylark *Alauda arvensis*, blackbird *Turdus merula*, song thrush *Turdus philomelos*, and Australian magpie *Gymnorhina tibicen*. Of these, the former three are European introductions; magpies are self-introduced from Australia (Heather & Robertson 2000). These species are among the most common birds on sheep & beef farms in New Zealand, preferring open paddocks interspersed with vegetation blocks. They were chosen because of their ubiquity, the relative ease of identification in the field both visually and by call/song, and because previous surveys on the same farms have shown that density could be estimated with higher precision than was possible for other common species (MacLeod et al. 2012).

We determined the extent and causes of variation in detectability for the surveyed species, and whether distance sampling provided an effective means of adjustment. Temporal variability (time of day or year), habitat differences (vegetation type, weather conditions) and variability introduced by the mechanisms of field monitoring (observer identity and detection by sight or sound) have been found to influence detectability elsewhere (e.g. Sauer et al. 1994; Newson et al. 2005, 2008; Alldredge et al. 2007; MacLeod et al. 2012).

Methods

Study areas

This study focused on 12 of the 36 sheep & beef farms that are participating in the ARGOS project (www.argos.org.nz). Our study farms occur in four clusters of three farms each, located near Outram (Otago), Oamaru (Otago), Owaka (Southland) and on Banks Peninsula (Canterbury) (see MacLeod et al. (2012) for a location map). Each cluster contained one representative of each of three different farm management types ('panels'): conventional (CON), integrated management (INT), and certified organic (ORG).

Data collection

There were 10 rounds of repeat visits (circuits) between November 2005 and August 2007 during which attempts were made to sample all farms (Table 1). Circuits were categorised into three seasons: 'breeding' (September–January), 'post-breeding' (February–May) and 'winter' (June–August). On each farm visit, 10 randomly placed 500-m line transects were selected, with a minimum spacing of 200 m between transect lines. Monitoring using a laser range finder and a GPS unit was carried out following standard procedures for line transect distance sampling (Buckland et al. 2001). Two observers (the 'main observer' and one of a changing roster of seven other observers) walked these transects in a North–South direction. Each observer usually performed five consecutive transects in a fieldwork day starting from c. 3.5 h after dawn.

Distance sampling methodology

Distance sampling theory is presented in detail in Buckland et al. (2001, 2004). In distance sampling with line transects, observers record the distance from a randomly placed line to all birds detected within a truncation distance w . Not all the birds within distance w will be detected, but a fundamental assumption is that all birds at zero distance are detected or, failing that, the proportion of detections at zero distance is known. Overall detection probability is expected to decrease with increasing distance from the line or point. The distribution of the observed distances is used to estimate a 'detection function' $g(y)$ that describes the probability of detecting a bird at distance y perpendicular to the centreline of the transect. Given that various basic assumptions hold, this function allows the estimation of the average probability P of detecting a bird given that is within width w of the line. Bird density can then be estimated as $D = n / (P * a)$, where n is the number of birds detected and a is the size of the region covered (the total length of the transect multiplied by $2w$). In program Distance, the detection function is modelled by combining a robust key function with an optional flexible series expansion (Buckland et al. 2001). We also included covariates that might have influenced detection probability, such as habitat parameters, weather conditions and observer identity (Marques & Buckland 2003; Marques et al. 2007).

Table 1. Spacing of monitoring circuits over the survey period.

Year	2005		2006								2007												
Month	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	
Circuit	1		2	3		4			5		6		7	8			9						10
Season	BREED		BREED		POST		POST		WINTER		BREED		BREED		BREED		POST						WINTER

Detectability estimation using distance sampling

Distance model building

All calculations for the estimation of detection functions were carried out in Distance 6.0 Release Beta 1 (Thomas et al. 2010); additional computations were performed in GenStat Version 9.1 (Lawes Agricultural Trust 2006). Datasets were analysed per species and farm – 4 (species) \times 12 (farms) sets of models, minus 3 sets where insufficient detections were available for a reliable density estimate – giving a total of 45 sets. Datasets were truncated such that the minimum detection probability per set was approximately 0.15 (Buckland et al. 2001). Data were generally grouped into 9–13 intervals to correct for rounding/heaping errors and to smooth irregular datasets. Of 45 datasets, only five that had exceptionally high sample sizes were analysed ungrouped.

Covariates of detectability entered the model in the form of a scaling parameter (σ) that controls the width of the detection function, thus influencing its scale (Marques & Buckland 2003). The parameters selected were: season, observer, wind speed, heard/seen, habitat, and minutes since sunrise (Table 2). The first three of these parameters were applied equally to all detections of a transect, while the last three were applied separately for each detection. Observer described whether the transect was carried out by the main observer, any one of the other observers, or two observers at the same time (a small percentage, used for training new observers). Wind speed was recorded three times per day using a Kestrel 400 portable weather meter, and applied to the transects closest in time to each occasion. Heard/seen was recorded as ‘heard and/ or seen’ if the bird was detected visually while establishing its position, regardless of whether first detection was by ear or eye, and as ‘heard only’ if the bird was heard but could not be detected visually (e.g. hidden in vegetation). Habitat was recorded in three broad categories at time of detection (Table 2). Factor covariates (all except wind speed and minutes since sunrise) were constrained to a maximum of three levels to avoid

unbalancing models based on small datasets. (See Appendix 1 for an example demonstrating the effect of the parameter wind speed on a detection function.)

Model selection and averaging

We generated all possible models for each species and farm with combinations of (1) either of the two recommended key functions (half-normal or hazard-rate; Buckland et al. 2001), (2) up to any two of the six covariates, and (3) with or without one of the three available series expansions. Any covariates resulting in a conspicuously improved model fit were trialled in up to fourth-level combinations. Model selection was performed using Akaike’s Information Criterion (AIC) and Akaike weights (Akaike 1973). AIC ranks models for fit and identifies the most parsimonious models (best model fit with the least number of added parameters). A second-order derivation (AICc) that compensates for small sample sizes (c. <40) has often been used to rank Distance models (Burnham & Anderson 1998). We used AICc scores as ranking criteria, and searched for models with a delta AICc below 4 as having reasonable support (Burnham & Anderson 1998). Models falling under the following additional criteria were also eliminated: (1) convergence failure of the fitting algorithm (excluding termination of algorithm after 500 iterations); (2) $g(0)$ [detection probability = 1] exceeded first histogram bar by more than 5%; (3) series expansion levels were fitted to obvious artefacts or incidental features; (4) model variance was conspicuously larger than in similarly scored models; (5) model was subject to ‘pretending variables’ (Burnham & Anderson 1998). In our models the AIC differences between top models were generally slight, so a clear best model was rarely apparent. Hence, estimates of density and detection probability were derived from averaging over each set by default (Burnham & Anderson 1998). The averaging of estimates and variances using Akaike weights was performed manually in OpenOffice Calc 2.2.1 (www.openoffice.org).

Table 2. Covariates used in constructing detection function models and general linear models.

Covariate	Type	Levels	Comments
<i>Detection function modelling</i>			
Season	Factor	Breed, post, winter	‘Breed’ = Sep–Jan; ‘post’ = Feb–May; ‘winter’ = Jun–Aug
Observer	Factor	Main, other, main & other	‘Other’ covers 7 observers
Heard/seen	Factor	Heard, seen	Whether the bird was detected by ear only or was seen as well
Habitat	Factor	Veg, open, other	‘Vegetation’ covers all types of woody vegetation (native and introduced, incl. shelterbelts; ‘open’ covers paddocks/pasture and crops; ‘other’ covers bare ground, utility areas and wetlands
Minutes since sunrise	Continuous	-	Recorded at start of transect
Wind speed	Continuous	-	Recorded three times over the course of 10 transects
<i>General linear modelling</i>			
Panel	Factor	Cont, int, org	Farm management type: ‘con’ = conventional; ‘int’ = integrated management; ‘org’ = certified organic
Season	Factor	Breed, post, winter	‘Breed’ = Sep–Jan; ‘post’ = Feb–May; ‘winter’ = Jun–Aug
Cohort	Factor	1, 2	Timespan from beginning of ‘breed’ season to end of ‘winter’ season in each of the two years
Vegetation cover	Continuous	-	% woody vegetation incl. shelterbelts in total farm area (1 - ~ describes % of open area)
Introduced cover	Continuous	-	% introduced woody vegetation incl. shelterbelts in total farm area
Percent introduced	Continuous	-	% introduced woody vegetation incl. shelterbelts in total woody vegetation 1 - ~ describes % of native woody vegetation in total woody vegetation

Covariate influence estimation

To compare the relative importance of individual covariates for model fit in a set, the 'score' of each covariate was computed for each species-by-farm set of models as the sum of the Akaike weights of all models that contained this covariate (Burnham & Anderson 1998). Since a single Akaike weight value describes the probability that a given model is the best model in a set, this sum can be treated as a measure of the covariate's contribution to the total fit of all relatively well fitting detection functions in the set. 'Sigma ratios' (Marques & Buckland 2003) were derived from the scale factor sigma that controls the width of the detection function (see Appendix 2 for the full list of covariate scores and explanation of estimation methods for sigma ratios).

Influence of farm-level drivers on detectability

Estimates of 'effective strip width' (ESW) were generated for each model. If all objects were detected out to this distance to both sides of the transect and none beyond, the expected number of objects would be the same as for the actual total survey out to the chosen truncation distance (Buckland et al. 2001). It follows that the area $\text{transect length} \times 2 \times \text{ESW}$ can be considered equivalent to the completely censused area. This makes effective strip widths a flexible spatial derivative of the transect's average detection probability that incorporates the truncation distance used.

Since detections from all 10 circuits were pooled in the interest of larger sample sizes, species-by-farm ESWs spanned the entire survey duration. For the three farm-species cases where no model could be fitted (two thrush and one blackbird sets), the ESWs were averaged from the values of the other two farms in the cluster. To test for seasonal differences in detectability, farm-specific seasonal ESWs (sESWs) for each species were then computed from the model estimates by evaluating the full detection function for the individual three levels of the covariate 'season', and integrating the function (Thomas et al. 2010). This was done only for those models where the 'season' covariate weight in the set was ≥ 0.1 (to filter out species-by-farm combinations where season had already been shown to have little influence). Seasonal ESWs that included less than 10 sightings in that season, or for which the model curve did not pass general quality criteria (above), were removed. In cases where no seasonal ESW was computed, the all-year ESW was used. Out of a maximum 36 cases per species (12 farms \times three seasons), 18 seasonal ESWs were constructed for skylarks, 11 for blackbirds, six for thrushes, and 12 for magpies.

Several farm-level explanatory parameters were fitted to the sESWs in a general linear model to test their influence on estimated detection probability (Table 2). Hierarchical general linear modelling (Bryk & Raudenbush 1992) as implemented in GenStat 9.1 was employed, as this allowed the inclusion of farm identity as a random factor to account for the fact that each farm had been sampled 9 or 10 times.

Interest here was in effects on farm-scale estimates of density, so parameters that applied to the entire farm were chosen (Table 2). Season was the same parameter used as a distance modelling covariate, while panel described the three recognised farm management regimes. The three vegetation measures used were area percentages derived from map data using ArcGIS, based on land-use maps created by the ARGOS project. The measures described the percentage of all woody vegetation on total farm area (*vegetation cover*), percentage

of introduced woody vegetation on total farm area (*introduced cover*), and percentage of introduced woody vegetation in total woody vegetation (*percent introduced*). All three measures include shelterbelts. Percentage of native vegetation in total farm area was found not to explain variation in ESW during preliminary model building and therefore was omitted from consideration.

Strong collinearity was found between vegetation cover and percent introduced, so the two parameters were not included in the same models together. All parameters were centred (mean subtracted from each instance) to avoid collinearity of lower-order terms with interaction terms in the hierarchical general linear model (following Quinn & Keough 2002). The resulting models were selected using the same criterion as with the Distance models (a threshold value of AICc difference ≤ 4 from the best model).

Results

Role of covariates in modelling bird detectability

The overall highest scored covariate was 'minutes since sunrise' (Fig. 2). It was selected most frequently in blackbird models, followed by skylark and magpie models, and still moderately frequently (23%) in thrush models. There was no real tendency towards either detectability increase or decrease with time since sunrise in three of the species, but there was a clear positive correlation in skylarks (Table 3).

'Wind speed' had moderate to high explanatory power for detectability of all species. It increased detectability in blackbirds and decreased it in magpies. The effect of wind speed was inconclusive (no distinct direction) in thrushes and skylarks. The 'heard/seen' ratio had low to moderate overall influence (Fig. 2). There was no clear effect of the two levels on detectability in skylarks and thrushes, but there was a distinct positive effect of 'heard only' over 'seen' in blackbirds and magpies (Table 3). 'Observer identity' featured strongly in thrushes, but less in the other species (Fig. 2). The relative order of effect strength of the three covariate levels (main observer, other observer, and main+other observer) differed between species but, except in blackbirds (where no clear direction was visible), the presence of the main observer was always associated with the highest detectability (Table 3). 'Habitat' was a lot more important in blackbirds than in any of the other species (Fig. 2). In skylarks and blackbirds, levels 'open' and 'other' were consistently associated with higher detection probability than level 'veg' (denoting woody vegetation; Table 3). 'Season' was most important for skylarks (Fig. 2), but although there was some tendency for higher detectability in the breeding season, there was no clear correlation. Thrush models showed lower detectability in 'post-breeding' than in the other seasons, although the importance of this covariate was low. The other species showed no trends (Table 3).

Table 3 also shows those correlations between sigma ratios and effect directions that appear relatively consistent over all farms and species. Globally, only the effects of habitat type ('other' corresponds to a higher or equal detectability than 'open' and to a higher detectability than 'vegetation') and of observer identity (main observer corresponds to a higher detectability than the other observers) seem to be applicable. Models without any added covariates were not very common for any species, and particularly rare for blackbirds (Fig. 2); thrushes had the highest incidence of such models, probably

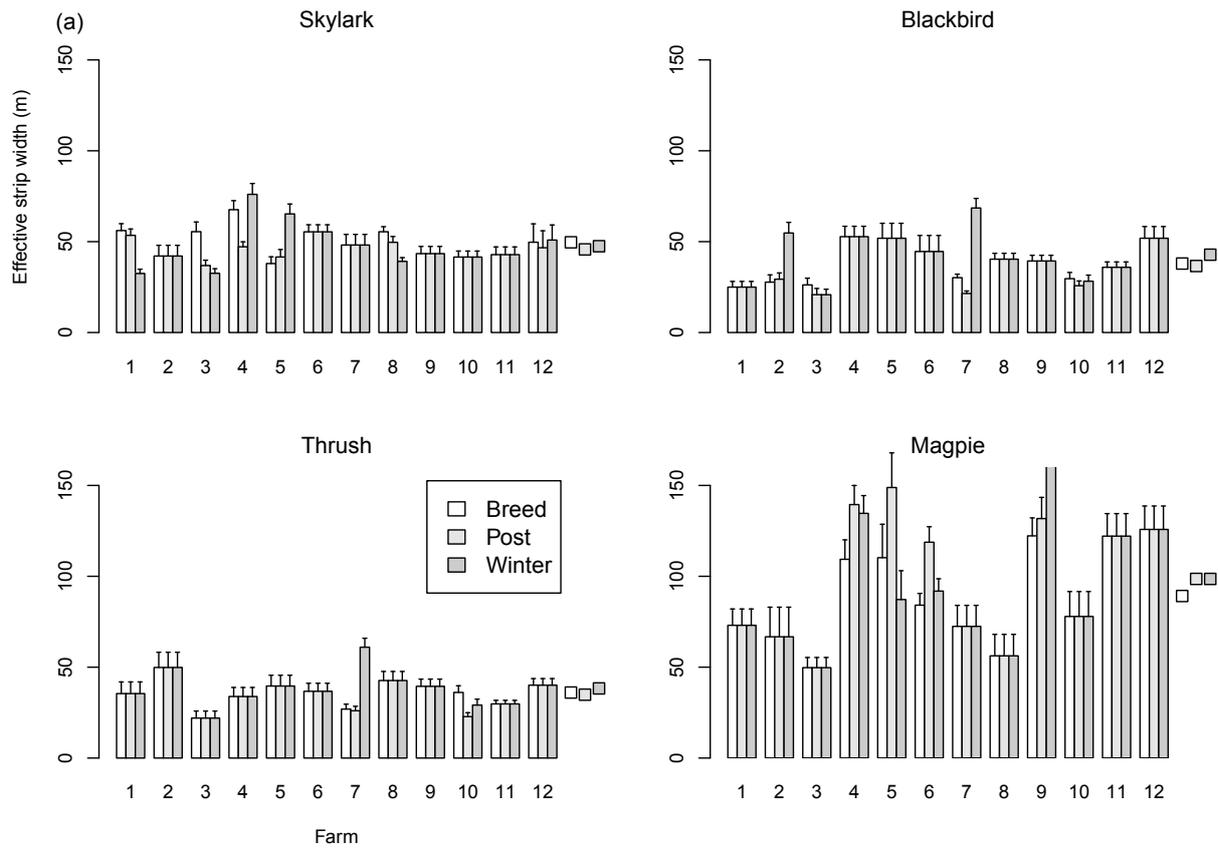


Figure 1. Seasonal effective strip widths (sESWs). (a) sESWs by species and farm (+ 1 SE). The averages over all farms per season and species are indicated by three symbols to the right of each plot. (b) sESWs averaged over all farms (± 1 SE).

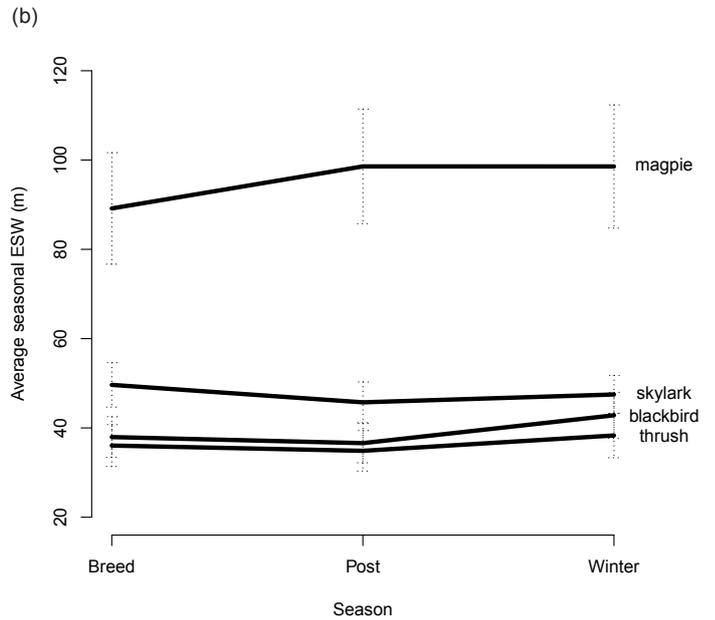


Figure 2. Cumulative covariate scores (sums of Akaike weights of models containing the covariate) per species, averaged over all farms, denoting the covariate's relative importance in defining well-fitting models for that species. 'None' denotes models with no covariates. Type size is graded in five levels from lowest value present (0.08) to highest (0.44). For individual scores see Appendix 2.

species	none	season	habitat	observer	heard_seen	wind	minutes
skylark	0.19	0.32	0.12	0.22	0.14	0.19	0.34
blackbird	0.11	0.25	0.40	0.17	0.21	0.31	0.44
thrush	0.22	0.14	0.08	0.34	0.20	0.19	0.23
magpie	0.15	0.15	0.19	0.23	0.21	0.31	0.33

Table 3. Consistent trends over all farms in covariate scale factors (sigmas) per species, denoting levels (factor covariates) or deviations from 0 (continuous covariates) that were associated with increased detection probability. See Table 2 for description of parameters.

Covariate	Sigma ratios over all farms & species	Sigma ratios over all farms within species			
	All species	Skylark	Blackbird	Thrush	Magpie
Season	-	Breed > post	-	Breed, winter > post	-
Habitat	Other ≥ open, vegetation	Open, other > vegetation	Open, other > vegetation	-	Other > open
Observer	Main ≥ other, main+other	Main, main+other > other	-	Main, other > main+other	Main > other, main+other
Heard/seen	-	-	Heard > seen	-	Heard > seen
Wind	-	-	Positive	-	Negative
Minutes	-	Positive	-	-	-

due to the often poor detection numbers, which did not allow the fitting of complex models.

Influence of farm-level drivers on detectability

Seasonal and farming system effects

Season-specific effective strip widths were constructed for about one-third of all species × farm × season combinations. Annual ESWs were used as seasonal values in the remainder. There were notable differences between individual farms, but these did not follow any readily discernible pattern of farm geographical location (Fig. 1a). Consistent interspecies differences were present for each species, with magpies having by far the largest and thrushes the smallest widths (Fig. 1a, b). However, there were few seasonal changes in detectability within each species or across all species (Fig. 1b).

Lack of systematic differences in detectability between either farm locations or sampling seasons was confirmed by an analysis of variance. No significant sESW differences between seasons or management systems were found for any species in linear mixed models (REML) fitted to the log₁₀-transformed estimates and incorporating panel, season, and panel within season as fixed (treatment) factors. Season was also used as a random (blocking) factor.

Dependence of effective strip width on farm-level habitat parameters

A general linear model linking sESWs to habitat of each species worked very well for magpies, acceptably well for blackbirds, but less well for skylarks (average $R^2 = 70.8, 28.9$ and 12.9 respectively). An F -test of the model's fit was still in an acceptable range ($P = 0.063$) for skylarks. No workable models could be found for thrushes (Table 4).

Detectability decreased strongly as woody vegetation cover increased. A significant decline in detectability in skylarks and blackbirds was connected only with increase in vegetation cover, whereas both vegetation cover and introduced vegetation cover predicted ESW in magpies. Magpie models also displayed a positive interaction between these two vegetation types, indicating a mutual reinforcement of these indices' negative effect on detection probability (Table 4).

We found no evidence of seasonal differences in detectability except perhaps in blackbirds. Although inclusion of 'season' yielded a borderline significant increase in the blackbirds' model fit ($P = 0.061$), levels breed/post/winter were not sufficiently different from each other to unequivocally indicate the effect of each level. Detectability was higher in winter than in the other two seasons. A highly significant interaction between the influence of vegetation cover and season

Table 4. Parameters with significant effects in the hierarchical general linear model (HGLM) regressing seasonal effective strip width on farm-level habitat parameters. For vegetation percentage measures, \\\ signifies a negative effect on effective strip width with increasing parameter magnitude. (*) signifies cases where a factor parameter contributed significantly to model fit, but no difference between parameter-level effects was present. Akaike weight-averaged R^2 and F -test probability (P) are derived from general linear models with the same data structure as the HGLMs, but no blocking factors.

Species	R^2	P	Parameters with significant effects in confidence set				
			Continuous vegetation	Panel	Season	Cohort	Interactions
Skylark	12.95	0.063	Vegetation cover \\\	-	-	-	-
Blackbird	28.93	0.002	Vegetation cover \\\	-	(*)	-	Vegetation cover × season: negative interaction
Thrush	-	-	-	-	-	-	-
Magpie	70.80	<0.001	Vegetation cover \\\, percent introduced \\\	-	-	-	Vegetation cover × percent introduced: positive interaction

in blackbirds points in the same direction, i.e. detectability on farms with little vegetation was higher in winter than either in the breeding season ($P = 0.116$) or post breeding ($P = 0.083$). No difference was found between the two annual cohorts in any species, showing that detectability did not differ between the two years that the survey spanned.

Farming system explained none of the variation in sESW, just as no evidence of effects on detectability differences were found in our earlier panel \times season REML. This agrees with an assessment of the distribution of the apparent primary detectability driver (woody vegetation cover) in the panels – a test by analysis of variance showed no significant differences between panels for any of the three vegetation percentage measures (vegetation cover $P = 0.632$, introduced cover $P = 0.861$, percent introduced $P = 0.715$).

Discussion

Reliability of distance modelling

The reliability of the results of any study employing a complex sampling methodology depends on meeting the method's basic assumptions. Distance sampling is based on several fundamental assumptions that need to be satisfied by the data to avoid the introduction of bias into the abundance estimates obtained (Buckland et al. 2001; Table 5). The condition of complete detection of all birds on the transect line itself was relatively easy to meet in the mostly open farmland habitats. The main cause of missed detections on the line in land bird surveys is an obscuring overhead canopy, which was only present in a small percentage of transects. A random and independent distribution of birds in the surveyed area is a condition for the reliable estimation of the results' precision, and may be jeopardised if transects tend to follow gradients of bird density, and if birds occur in clusters. We avoided this bias by placing transects randomly, and by modelling the presence of flocks within Distance (Buckland et al. 2001). Further, it is assumed that animals do not move in response to the observer before they are detected. This is difficult to assure in the field as some species tend to detect the observer

early and retreat to a comfortable distance, which can bias the distribution of detection distances. A similar effect results from violations of the fourth fundamental assumption, that distance measurements are free of systematic error; this is often not the case if observers tend to round values (particularly angles). Both of these problems were present in some of our data, but could be effectively addressed during modelling by allocating detections into intervals to smooth rounding artefacts, and by employing functions with a constrained shape (shoulder) close to the origin where biases can have the greatest impact on estimation (Buckland et al. 2001). We conclude that the basic assumptions of distance sampling were satisfactorily met in this survey and that, in conjunction with the intensive model screening approach employed, the conditions for reliable estimation of detectability were present.

Impacts of covariates and farm-level variables on detectability

Of the parameters available as covariates for distance model building, the effects of observer identity and broad habitat type showed the greatest consistency across all species, even though the identities of all observers beside the main one had to be combined for model fitting. The main observer was consistently associated with high or the highest detectability. It might have been expected that the double-observer transects (main+other), with approximately double the detection capability, would result in higher detectability. However, this was not the case, probably because only a few of these were performed (3–4% of all transects) and because they were used for instructing new observers. Observer identity was particularly important in thrush models, probably because some part-time observers (taking part in the survey for periods of no more than three circuits each) had difficulty in distinguishing between thrushes of either sex and female blackbirds, and between singing male thrushes and blackbirds. The long-term observers rapidly gained confidence to identify thrushes. Failure to find a model for thrush ESW with any significant fit for any of the three main vegetation measures (Table 4) was also probably due to the comparatively low reliability of thrush data. Several studies have shown that novice observers will generally tend

Table 5. Assumptions of line transect distance sampling and their handling in this survey. Based on Borchers et al. (2002).

Assumption	Effect of violation	Assumption met in this study?
All animals on the line are detected (i.e. $g(0) = 1$)	Density estimate is negatively biased in proportion to actual $g(0)$; e.g. $g(0) = 0.8$ results in an estimate that is 80% of true density	Yes: only a small percentage of transects was in locations where detection on the line might be missed (e.g. high canopy)
Animals do not move before detection	Bias negligible if movement is random. If movement is in response to observer, negative or positive bias can result	Evasive movement frequently present, but addressed in the same manner as above (grouping data into intervals, use of models with shoulder)
Measurements are exact	Method is robust to random errors. Systematic errors like habitual rounding or over/underestimation will result in bias	Rounding of angles frequently present, but addressed in the same manner as above (grouping data, models with shoulder)
Animals are randomly and independently distributed	Biases confidence intervals of estimates	Transect placed randomly in regard to possible density gradients. Clustered populations present, but addressed in modelling (clusters treated as individual detections and cluster size recorded separately)

to miss more detections than experienced observers and have lower accuracy in identification, negatively biasing the survey outcome, but usually quickly gain the same level of competence (Sauer et al. 1994; McLaren & Cadman 1999; Jiguet 2009; but see Kendall et al. 1996). This underlines the importance of providing new observers with sufficient training to minimise this source of bias. Inferences from thrush detections were likely less reliable in this survey than for the other species.

'Open' and 'other' habitats tended to be associated with higher detectability than areas covered by woody vegetation, the values for 'other' tending to be higher again than for open. Lower detectability in vegetated areas was an expected outcome due to shortened lines of sight, concealment of birds, and dampening of calls. As the 'other' category is comprised largely of sown or ploughed paddocks, this illustrates the effect of birds congregating in these places to forage, enabling frequent detection over long distances. Vegetation had less effect and showed no directional influence for thrushes, perhaps because of behavioural differences. Blackbirds tend to congregate around food sources (often ploughed fields) for feeding during their autumn moult, while thrushes become secretive and keep to vegetation (Heather & Robertson 2000). Conversely, the relatively weak fit obtainable for skylark ESW models probably reflects genuinely low dependence of detectability on farm-level vegetation patterns for this species. Most skylarks were detected at medium range for birds in flight or displaying above open areas, limiting the obscuring effect of any vegetation present on the farm. Dependence of ESW on vegetation cover was pronounced in blackbirds, which might be found in vegetation as well as in open areas. It has been shown that in songbird surveys, the reliability of distance estimations from purely auditory cues is likely to be poor when both the observer and the bird are situated within vegetation (Alldredge et al. 2007). While detection distance would have been curtailed when the observer himself was situated within bush or tree patches, this negative effect was likely offset to some extent by aural detections of singing or calling blackbirds in vegetation when the observer was in the open. Both kinds of observations would be associated with vegetation, but only the former would result in reduced detection distances and accounted for less than 4% of total sightings. The indication of a seasonal reduction of this effect in winter likely relates to the winter congregation behaviour around food sources in open areas.

Increasing percentage cover of both total vegetation and introduced vegetation were associated with strong decreases in detectability in magpies. It is likely that the large detection and 'flushing' distances for magpies (c. 155 m, more than twice as large as the average detection distances for the other species) makes them very susceptible to blockage of line of sight, particularly by shelterbelts. The effects of such blockage are not counteracted by magpies calling from such woody vegetation in the way observed for blackbirds. Detectability of skylarks was much higher later in the day. This does not relate to song activity, which has been shown to be about constant throughout the day (Kayser 1999), but might be related to feeding and resting patterns. Increased wind speed correlated with increased detectability in blackbirds, but decreased detectability in magpies. Wind noise would be expected to decrease detections by sound, and birds might undertake fewer or less sustained flights in high winds, so the increased detectability of blackbirds in higher wind speeds was unexpected. While not very strong (c. 36% sigma deviation from neutral), there is no obvious explanation.

The distinction between 'heard only' and 'seen' was a moderately important explanatory variable in all species. However, the direction of its effect varied on different farms for skylark and thrush, whereas places of increased detectability of magpies and blackbird were associated with more of the birds being just heard rather than seen (or seen and heard). In the latter species, distant detections while perched in vegetation across open terrain might explain the effect because calls/singing are probably easier to notice over large distances than are visuals. While this would also be true for displaying skylarks, these birds were never concealed in vegetation – a confirmed skylark detection invariably meant visual detection – leading to lowest importance of this covariate among species. Our grouping of detections into 'heard only' and 'heard and/or seen' as used in this study is non-standard and performed markedly less well than the more common grouping of 'heard first' and 'seen first', in which form the parameter has frequently proven to be among the most important explanatory variables of detectability (Marques et al. 2007; Johnson 2008; Newson et al. 2005, 2008; MacLeod et al. 2012). This common form seems preferable as it results in groupings that can be more profitably mapped to frequently encountered categories of detectability.

Season affected detectability in skylarks and, to a lesser degree, blackbirds, although no consistent directionality of this effect was apparent for blackbirds. Skylark displaying behaviour led to a clear increase in detectability during the breeding season, when males mainly sing and hover 50–100 m above the ground (Heather & Robertson 2000). Apparently displaying was more important for detectability than was their flocking behaviour in winter (Cramp 1977) because most skylark detections were first detected by hearing their singing. Inconspicuous behaviour during moult (Heather & Robertson 2000) was presumably the reason for lower thrush detectability post-breeding, in the few cases where this covariate was selected for the species. For blackbirds and magpies, the order of the three seasons tended to even out over all farms.

Reliability of inference from covariates

The highest selection percentage for any single covariate within a species was 44% (minutes from sunrise in blackbirds); the average selection percentage was 23% across all covariates and species. There was thus little dominance of individual covariates even within species. In several species–farm sets, a variety of covariates could be included to achieve similar model fits. Also, the majority of covariates showed both negative and positive influence on detectability with increasing magnitude (continuous parameters) or permutations in the order of level influence (factor parameters) between farms within a species. Despite this low consistency, just 17% of total Akaike weight over all farms was made up of models without any covariates at all, so they almost invariably added to model fits, even under the parsimony constraints imposed by AIC selection.

These observations raise the question of how well-founded are inferences based on the frequency and effects of these covariates. When modelling complex ecological relationships, there is a danger of finding effects that are merely artefacts of the particular dataset (Anderson et al. 2001). First, small sample sizes relative to the number of parameters being estimated might result in over-fitting of incidental characteristics, and (particularly with iterative fitting processes like the one used by program Distance) instability of models, both promoting the overestimation of set artefacts into significant characteristics. The use of parsimony criteria like AIC is a method to prevent

inflation with parameters. Our further screening (see Methods) was designed to remove models with obvious over-fitting of incidental features or empty over-parameterisation ('pretending variables'; Burnham & Anderson 1998). The recommended minimum sample size for distance analysis (c. 60 detections; Buckland et al. 2001) was not met in nine out of 48 species-by-farm cases, five of them in thrushes. While this will have weakened the inferences made for that species in particular, similar effects would not be expected across all other species.

Second, the 'all possible subsets' method of model selection constitutes a mild form of data dredging that potentially seeks out complex parameter combinations that happen to coincidentally fit a particular dataset very well (Anderson et al. 2001). Nevertheless the method is probably more robust than its alternatives. For example, fitting only models designed from previous knowledge of plausible parameter-to-detectability relationships runs the risk of completely missing unconsidered explanatory structures. Stepwise selection runs the same risk of special fitting, and will bypass alternative branches of models that may have explanatory power of their own (Burnham & Anderson 1998). Therefore, although our interpretation of covariate effects can only be preliminary, we recommend the investigated covariates (time of year, time of day, habitat type, observer identity, heard/seen, and wind speed) for recording in the field and use in modelling. Further parameters that were not recorded in this study might be considered – these include data derived from terrain models, such as steepness or landform restrictions to lines of sight, and specific interactions between covariates such as wind speed and vegetation cover.

However, there is no need in distance sampling to individually account for every possible source of variation in detectability. This is ensured by the property of pooling robustness (Buckland et al. 2004), which allows for largely unbiased estimates from pooled data collected under different detectabilities. If interest is mainly in the generation of unbiased pooled density estimates, this property will often make it unnecessary to employ covariates. The use of covariates, on the other hand, generally provides a marked benefit to estimate precision (Marques & Buckland 2003), and may be needed when estimating differentiating segments of the population. For example, it has been shown that in species with pronounced detectability differences between genders, densities will tend to be underestimated in distance sampling if sex is not taken into account (Newson et al. 2008).

Implications for design of bird monitoring methods

Coverage of farm areas was not perfectly uniform, as comparatively few transects were performed in areas where woody vegetation was very heavy – firstly because no GPS coordinates (intended for further spatial analysis of bird distributions) could be collected here, and secondly because it was found that close-to-zero detections resulted from transects through very difficult terrain because the observer was preoccupied. Although this may have led to a weakening of the explanatory power of the farm-level vegetation cover measures, this effect was probably less severe with our four focal species than it would have been with native species that show a much higher habitat reliance on large native bush patches. While conclusions drawn from these models cannot automatically be extended to such species, the three species that could be fully modelled cover a species that is at home on open pastures and does not associate with woody vegetation at all (skylark), a species that forages mostly in the open but roosts and nests in vegetation (magpie), and a species that

spends a substantial amount of time in or near bush and tree patches (blackbird). This suggests that our findings can be generalised to some extent.

Overall, the gross percentage of general woody vegetation on farms emerged as the dominant driver of detectability, while vegetation composition (native or introduced) had little effect. As vegetation patches on the surveyed farms varied from coniferous shelterbelts to native bush, this suggests that bird detectability is similar in different woody vegetation types, even though woody vegetation generally obscures birds from detection, and the detectability of birds in vegetation tends to be noticeably different from that in open habitats. Had we simply used bird counts without accounting for detectability, or not built measures of habitat into the Distance analysis to enable variability in detection functions, the true extent of the importance of woody vegetation for biodiversity (Blackwell et al. 2008; Meadows et al. 2008; Moller et al. 2008) would have been obscured. The bias when comparing bird abundance (and perhaps also diversity) in landscapes with differing degrees of woody vegetation patches and open areas will mislead inferences. Therefore we caution that abundance estimation by relative indices (e.g. five-minute bird counts) will only be robust within large and uniformly wooded patches, or within large uniformly open ground. Seasonal differences did not show up as drivers of ESW differences, but season was a significant distance sampling covariate for skylarks in particular. Small sample sizes may have greatly reduced power to detect such effects in other species; it would therefore be unwise to ignore season as affecting detectability when choosing bird monitoring methods.

None of our four focal species showed any differences in detection probability between certified organic, integrated management, and conventional farms. For research programmes that investigate the ecosystem effects of different approaches to farm management, e.g. the ARGOS group, this finding is of interest since it points at the possibility of using simpler and highly efficient index count methods to compare populations among study sites. Nevertheless such an approach would crucially rely on standardising counts with respect to the other detectability drivers noted above. While index count methods are generally attractive for solving questions of population dynamics over time in comparable habitats, if such standardisation can reliably be implemented, this may be a more difficult task than is often realised by researchers. When possible drivers of detectability have to be dealt with by controlling for them at the survey design stage, any missed might only become apparent, and probably still remain unidentified, once different batches of estimates were compared (Weller 2012). This might in the worst case invalidate a season's work and necessitate a redesign of the sampling regime. In a modelling-based approach like Distance sampling, drivers can be tested and accommodated as needed, allowing great flexibility at the analysis stage as long as the necessary information has been collected in the field. Once the basic increase in effort for the more complex method has been committed to, the cost of recording additional parameters is typically very low, enabling the surveyor to cover many possible bases at little extra investment. Such a regime can also easily be simplified later on by removing ineffective parameters from the protocol. In most cases it will be easier to simplify a flexible method than to try to upgrade an inherently limited one. We recommend that in addition to the more obvious considerations of surveying effort and estimate quality (Weller 2012), researchers should be aware of this trade-off in analysis resilience when designing a survey.

Acknowledgements

Florian Weller wishes to thank the University of Otago for granting a postgraduate scholarship that enabled this study. We are grateful to Eric Rexstad and an anonymous reviewer for their helpful comments on the draft of this paper. ARGOS project work was funded by the Foundation for Research, Science and Technology (Contract No. AGRB0301) with financial assistance from the Certified Organic Kiwifruit Producers Association, Fonterra, Merino New Zealand Inc., a meat packing company, Te Rūmanga o Ngāi Tahu, and ZESPRI Innovation Company.

References

- Agrawal A 2005. *Environmentality: technologies of government and the making of subjects*. Durham, NC, Duke University Press. 344 p.
- Akaike H 1973. Information theory and an extension of the maximum likelihood principle. In: Petran BN, Csáki F eds *International Symposium on Information Theory*. 2nd edn. Budapest, Hungary, Akadèmiai Kiadó. Pp. 267–281.
- Allredge MW, Simons TR, Pollock KH 2007. A field evaluation of distance measurement error in auditory avian point count surveys. *Journal of Wildlife Management* 71: 2759–2766.
- Anderson DR, Burnham KP, Gould WR, Cherry S 2001. Concerns about finding effects that are actually spurious. *Wildlife Society Bulletin* 29: 311–316.
- Blackwell G, Fukuda Y, Maegli T, MacLeod CJ 2008. Room for everyone? Refugia and native biodiversity in New Zealand's agricultural landscapes. *New Zealand Journal of Agricultural Research* 51: 473–476.
- Blackwell G, Lucock D, Moller H, Hill R, Manhire J, Emanuelsson E 2011. Abundance and diversity of herbaceous weeds in sheep/beef pastures, South Island, New Zealand. *New Zealand Journal of Agricultural Research* 54: 53–69.
- Borchers DL, Buckland ST, Zucchini W 2002. *Estimating animal abundance: closed populations*. New York, Springer. 314 p.
- Bryk AS, Raudenbush SW 1992. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA, Sage.
- Buckland ST 2006. Point-transect surveys for songbirds: robust methodologies. *The Auk* 123: 345–357.
- Buckland ST, Goudie IBJ, Borchers DL 2000. Wildlife population assessment: Past developments and future directions. *Biometrics* 56: 1–12.
- Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L 2001. *Introduction to distance sampling: estimating abundance of biological populations*. Oxford, Oxford University Press. 432 p.
- Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L 2004. *Advanced distance sampling: Estimating abundance of biological populations*. Oxford, Oxford University Press. 416 p.
- Burnham KP, Anderson DR 1998. *Model selection and inference: a practical information-theoretic approach*. 1st edn. New York, Springer. 353 p.
- Coleman G, Moller H, Bengé J, MacLeod C 2009. Could fantails provide a marketing edge for New Zealand kiwifruit? *Kiwifruit Journal* July/August 2009: 18–23.
- Cramp S eds 1977. *Handbook of the birds of Europe and the Middle East and North Africa: The birds of the Western Palearctic*. Oxford, Oxford University Press.
- Dawson DG, Bull PC 1975. Counting birds in New Zealand forests. *Notornis* 22: 101–109.
- Heather B, Robertson H 2000. *The field guide to the birds of New Zealand*. 2nd edn. Auckland, Penguin Books. 440 p.
- Hooson S, Jamieson IG 2003. The distribution and current status of New Zealand Saddleback *Philesturnus carunculatus*. *Bird Conservation International* 13: 79–95.
- Jiguet F 2009. Method learning caused a first-time observer effect in a newly started breeding bird survey. *Bird Study* 56: 253–258.
- Johnson DH 2008. In defense of indices: The case of bird surveys. *Journal of Wildlife Management* 72: 857–868.
- Kayser B 1999. Diurnal and seasonal variation in song activity: Results from a Danish survey covering 20 species. *Dansk Ornitologisk Forenings Tidsskrift* 93: 91–103.
- Kendall WL, Peterjohn BG, Sauer JR 1996. First-time observer effects in the North American Breeding Bird Survey. *The Auk* 113: 823–829.
- Lawes Agricultural Trust 2006. *GenStat Version 9.1*. Harpenden, UK, Lawes Agricultural Trust, Rothamsted Experimental Station.
- MacLeod CJ, Blackwell G, Moller H, Innes J, Powlesland R 2008. The forgotten 60%: bird ecology and management in New Zealand's agricultural landscape. *New Zealand Journal of Ecology* 32: 240–255.
- MacLeod CJ, Blackwell G, Weller F, Moller H 2012. Designing a bird monitoring scheme for New Zealand's agricultural sectors. *New Zealand Journal of Ecology* 36: 312–323.
- Marques FFC, Buckland ST 2003. Incorporating covariates into standard line transect analyses. *Biometrics* 59: 924–935.
- Marques TA, Thomas L, Fancy SG, Buckland ST 2007. Improving estimates of bird density using multiple-covariate distance sampling. *The Auk* 124: 1229–1243.
- McLaren MA, Cadman MD 1999. Can novice volunteers provide credible data for bird surveys requiring song identification? *Journal of Field Ornithology* 70: 481–490.
- Meadows S, Gradwohl M, Moller H, Rosin C, MacLeod CJ, Weller F, Blackwell G, Perley C 2008. Pathways for integration of biodiversity conservation into New Zealand's agricultural production. *New Zealand Journal of Agricultural Research* 51: 467–471.
- Moller H, Wearing A, Pearson A, Perley C, Steven D, Blackwell G, Reid J, Johnson M 2005. Environmental monitoring and research for improved resilience on ARGOS farms. ARGOS Working Papers 6. Online at http://www.argos.org.nz/working_papers.shtml
- Moller H, MacLeod CJ, Haggerty J, Rosin C, Blackwell G, Perley C, Meadows S, Weller F, Gradwohl M 2008. Intensification of New Zealand agriculture: implications for biodiversity. *New Zealand Journal of Agricultural Research* 51: 253–263.
- Morrison ML, Block WM, Strickland MD, Kendall WL 2001. *Wildlife study design*. New York, Springer. 210 p.
- Newson SE, Woodburn RJW, Noble DG, Baillie SR, Gregory RD 2005. Evaluating the Breeding Bird Survey for producing national population size and density estimates. *Bird Study* 52: 42–54.
- Newson SE, Evans KL, Noble DG, Greenwood JJD, Gaston KJ 2008. Use of distance sampling to improve estimates of national population sizes for common and widespread breeding birds in the UK. *Journal of Applied Ecology*

- 45: 1330–1338.
- Norton DA 1998. The myth of reserves and the future of nature conservation in New Zealand. *Ecological Society Newsletter* 89: 8–9.
- Norvell RE, Howe FP, Parrish JR 2003. A seven-year comparison of relative-abundance and distance-sampling methods. *The Auk* 120: 1013–1028.
- O'Donnell CFJ 1996. Monitoring mohua (yellowhead) populations in the South Island, New Zealand, 1983-93. *New Zealand Journal of Zoology* 23: 221–228.
- Perley C, Moller H, Hamilton WJ, Hutcheson J 2001. Towards safeguarding New Zealand's agricultural biodiversity: Research gaps, priorities and potential case studies. *Ecosystems Consultants Report* 23. 230 p.
- Powlesland RG, Wills DE, August ACL, August CK 2003. Effects of a 1080 operation on kaka and kereru survival and nesting success, Whirinaki Forest Park. *New Zealand Journal of Ecology* 27: 125–137.
- Quinn GP, Keough MJ eds 2002. *Experimental design and data analysis for biologists*. 1st edn. Cambridge, Cambridge University Press. 537 p.
- Sauer JR, Peterjohn BG, Link WA 1994. Observer differences in the North American Breeding Bird Survey. *The Auk* 111: 50–62.
- Thomas L, Buckland ST, Rexstad EA, Laake JL, Strindberg S, Hedley SL, Bishop JRB, Marques TA, Burnham KP 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology* 47: 5–14.
- Thompson WL 2002. Towards reliable bird surveys: Accounting for individuals present but not detected. *The Auk* 119: 18–25.
- Weller F 2012. A comparison of different approaches to monitoring bird density on New Zealand sheep and beef farms. *New Zealand Journal of Ecology* 36: 382-390.