special issue:
Advances in tools for bird population monitoring in New Zealand

Three population estimation methods compared for a known South Island robin population in Fiordland, New Zealand

Terry C. Greene* and Moira A. Pryde

Department of Conservation, PO Box 11089, Christchurch 8443, New Zealand
*Author for correspondence (Email: tgreene@doc.govt.nz)

Published on-line: 30 July 2012

Abstract: We evaluated the accuracy and precision of three population estimation methods (mark–resight, distance sampling and five-minute bird counts) for two populations of South Island robin (*Petroica australis australis*) of known size in the Eglinton Valley, Fiordland, over 5 years (March and August, 2005–2009). The performance of these population estimators was compared to known robin abundance derived from simultaneous territory mapping of individually marked birds. Mark–resight methods performed well with Bowden's estimator generating accurate and precise population size estimates and trends very similar to those obtained from territory mapping. Distance sampling estimates displayed significant positive bias and poor precision even though we could identify the general population trends derived from territory mapping at Knobs Flat and Walker Creek. Five-minute bird counts (and associated generalised linear mixed models) performed well when the assumption of constant detectability was met, and poorly when it was not. Such failures prevented robust inference and confounded longer-term trend analyses. As robins are attracted towards stationary observers, we recommend that they be counted from line transects rather than points. Whenever monitoring objectives demand accurate and unbiased estimates of population abundance, the monitoring methods used should explicitly account for incomplete detectability wherever possible.

Keywords: abundance; assumptions; detectability; Eglinton Valley; *Petroica australis australis*, trend

Introduction

Conservation managers require reliable tools to systematically detect, record and report on changes in species status and trends at a variety of spatial and temporal scales. Without effective inventory and monitoring programmes, it is difficult to evaluate the relative success of conservation management outcomes (Thompson et al. 1998; Bibby et al. 2000; Sutherland 2006; Elphick 2008). Prior to commencing any monitoring programme, it is therefore essential that the biases and limitations of the survey method(s) being considered are recognised, understood and, if practical, minimised (Nichols et al. 2009).

Absolute measures of bird population abundance and density (i.e. a true census) are usually impossible or difficult and costly to obtain (Thompson et al. 1998; Sutherland 2006). As a result, inferences are usually based on counts by observers standing at points or moving along line transects (Nichols et al. 2009). The majority of these counts (around 95%; Rosenstock et al. 2002) are usually treated as indices of abundance with spatial or temporal comparisons between them, assumed to have constant probability of detection and to represent a relatively constant (albeit unknown) proportion of the sampled population.

Although the assumption of constant detectability (or proportionality) needed for indices may be somewhat overstated (large-scale changes in abundance are likely to be detected even if relatively small changes in detectability occur; Johnson 2008; Nichols et al. 2009), larger failures of assumptions can obscure significant changes or trends in density or abundance (Thompson et al. 1998; Norvell et al. 2003). Despite this risk, indices remain attractive to managers, as data collection is simpler (they do not require supplementary data such as distance to birds), they are cheap to run in the field, and relatively easy to compute.

There has been considerable recent debate, much of it highly critical, over the assumptions and application of indices (Diefenbach et al. 2003; Ellingson & Lukacs 2003; Norvell et al. 2003; Buckland 2006). The debate has coincided with the proliferation of newer estimation methods (Buckland et al. 2001, 2004; Kissling & Garton 2006; Mackenzie et al. 2006; Dawson & Efford 2009). Although these alternatives usually explicitly address concerns over variable detectability, they also come with inherent weaknesses (often in the form of restrictive assumptions) along with requirements for increasingly complex field designs and analyses, and associated increased costs (Broekema & Overdyck 2012). Cost-efficient

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

methods that are capable of providing quality data sufficient to detect important changes in number and trend are, therefore, essential (Macleod et al. 2012).

It is unlikely that a single solution exists that will provide the best estimation option in all situations (Efford & Dawson 2009; Nichols et al. 2009). Weaknesses inherent in all estimation approaches dictate that all survey methodologies need to be customised to some degree. Study objectives, species to be monitored (number, behaviours, rarity, etc.), habitat types and topography are all important determinants of method selection and survey design (Buckland et al. 2001; Norvell et al. 2003; McClintock et al. 2006; Dawson & Efford 2009; Nichols et al. 2009).

Objective field tests of the various survey methodologies used to infer the density and abundance of bird populations are clearly required. Such tests would be extremely useful in determining (1) which of the methodologies suit particular objectives for a given set of conditions or species, (2) the precision of the estimator for a given species or group of species, (3) the optimum design and allocation of effort required to draw reasonable inferences, and (4) cost.

Although several studies have compared methods of estimating bird abundance (e.g. Hamel 1984; Casagrande & Beissinger 1997; Buckland 2006; Kissling & Garton 2006; Morgan et al. 2012; Spurr et al. 2012), we are aware of few studies where such comparisons have been conducted on populations of known size and structure (Gill 1980; Tarvin et al. 1998; Nelson & Fancy 1999; Buckland 2006) and none where more than one estimation method have been used.

In the Eglinton Valley on New Zealand's South Island, two sub-populations of South Island robins (*Petroica australis australis*), a widespread, territorial and conspicuous forest passerine, were individually marked as part of a long-term demographic study (C.F.J. O'Donnell, Department of Conservation, pers. comm.). This enabled territories to be mapped and provided a basis against which the performance, precision, and potential sources of bias of other sampling methodologies could be assessed. In this paper we present population estimates and trends for robins based on territory mapping, mark–resight estimators, distance sampling, and five-minute counts. We then use the results to compare their accuracy and precision, robustness to assumption violations, and practical application in the field.

Methods

Study area

The study was conducted in the Eglinton Valley, Fiordland, South Island, New Zealand (44°58' S, 168°01' E) (Fig. 1). The valley, c. 250–500 m above sea level, is of glacial origin, steep sided with a flat floor 0.5–1.5 km wide. Annual rainfall ranges from c. 1200 mm near the valley mouth to >5000 mm at its headwaters. Partly modified grassland covers much of the valley floor. Terraces, outwash fans and steep valley walls are covered with temperate beech (*Nothofagus* spp.) forest to the treeline at 1000–1200 m above sea level. Near the valley floor the forest is dominated by red beech (*Nothofagus fusca*) and silver beech (*N. menziesii*) with mountain beech (*N. solandri* var. *cliffortioides*) becoming increasingly common with increasing altitude. The understorey is generally open with few plants other than scattered broadleaf (*Griselinia littoralis*), small-leaved coprosmas (*Coprosma* spp.) and a ground cover of mosses (O'Donnell 2000).

Robins were monitored at two c. 100-ha sites within the Eglinton Valley: Knobs Flat and Walker Creek (Fig. 1). The Knobs Flat site, in the middle of the valley, is situated on a gently sloping alluvial fan with silver beech and a very open understorey dominating the periphery. With increasing elevation and distance from the margins, the forest becomes taller and is dominated by red beech trees with a much denser understorey. The Walker Creek site is further (c. 14 km) down the valley, and drier. Although the forest composition of both sites is similar, the forest at Walker Creek is generally taller. The topography is also more rugged, with a series of steep-sided alluvial terraces bisected by gullies formed by small streams.

Survey methods

Territory mapping, mark–resight, distance sampling, and five-minute bird counts were used to survey the robin population on each study site. All four survey methods made use of established, randomly placed grids (50 × 50 m) at both sites. For territory mapping and mark–resight survey methods this grid helped ensure that spatial coverage of each site was as even as possible. For those methods based on point sampling (distance sampling and five-minute bird counts), data were recorded at grid points spaced at 150-m intervals. At Knobs Flat, data were collected at 41 points and at Walker Creek at 43 points (Figs 1 & 2). Surveys for robins were conducted in August (pre-breeding) and March (post-breeding) of 2005–2009.

Territory mapping

Territory mapping was used as a benchmark for assessing the performance of the three other survey methods (Buckland 2006). During the period between 2005 and 2009, the majority of robins inhabiting the two study sites were captured in cage traps in spring (September & October) and autumn (March) and banded with unique leg-band combinations (one numbered metal band and three plastic colour bands), as part of a wider investigation into productivity and survival in response to pest control. Sightings of these marked birds were mapped throughout each field season (August–March), using information from formal resighting surveys and incidental observations of birds located and followed during the intervening period. An updated territory map was constructed for each site at monthly intervals (Fig. 2). Using this information, the actual number of robins (including any unmarked birds) at each site could be calculated (i.e. a true census) at the time bird counts were conducted (over a 7–10 day period) and compared directly with population estimates derived from the other survey methods.

Mark–resight modelling

Mark–resight data were collected for South Island robins from the two study sites. Even coverage probabilities (Nichols et al. 2009) were obtained by walking slowly (to reduce noise and attraction of robins toward observers) and systematically through each grid recording all marked and unmarked robins. Resighting surveys were conducted during the morning (at least 1 h after sunrise until 1300 hours), when robin activity was at its highest. Surveys were only attempted in good weather conditions, with no rain or significant wind. Knobs Flat could be surveyed by one person in a morning whereas Walker Creek was usually surveyed by two people per morning because of the more complex topography.

The free software Program NOREMARK (White 1996a) is a relatively simple DOS-based analysis program that computes

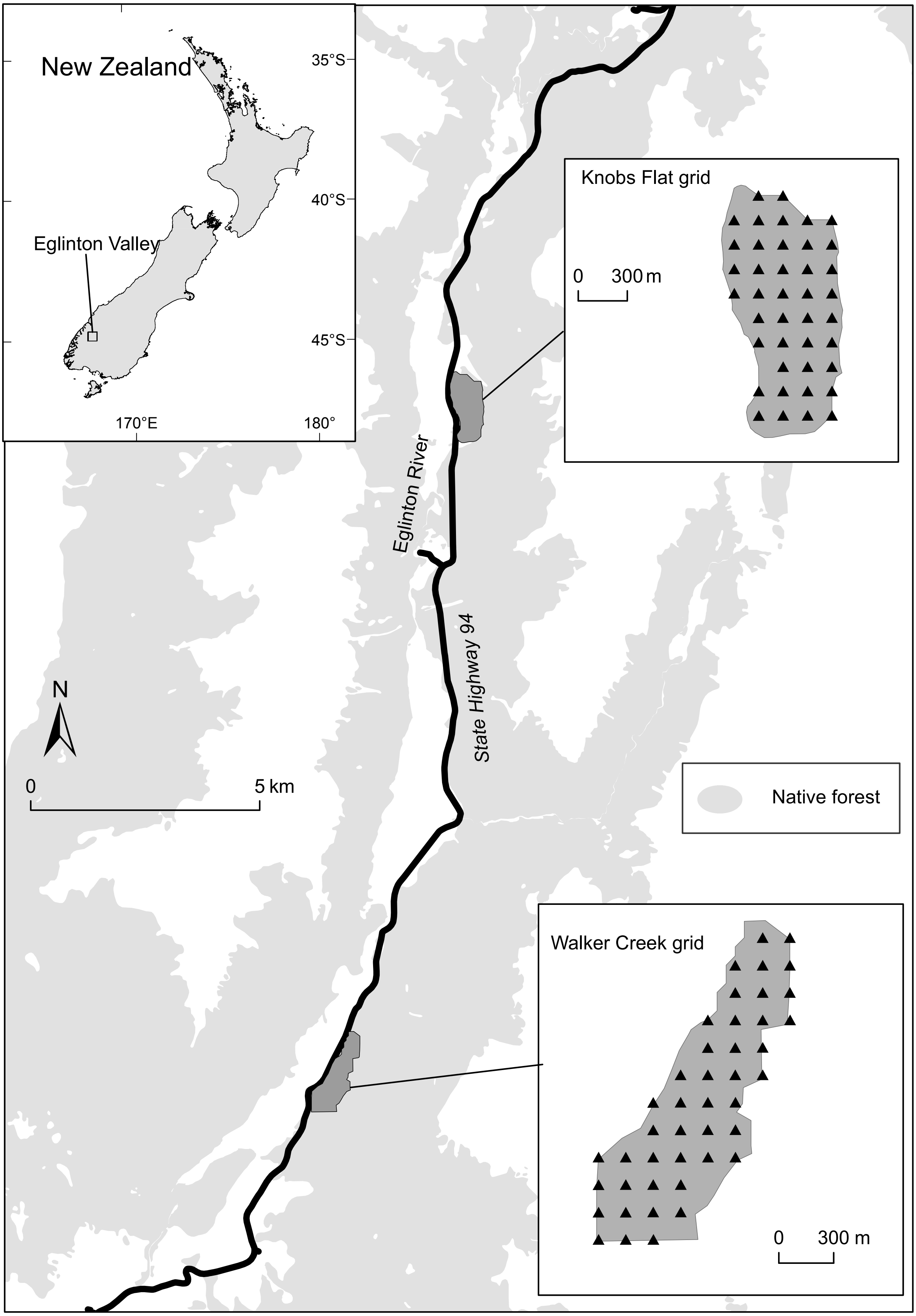

Figure 1. Location of study areas and South Island robin (*Petroica a. australis*) sampling points at Knobs Flat and Walker Creek, Eglinton Valley, Fiordland, New Zealand.

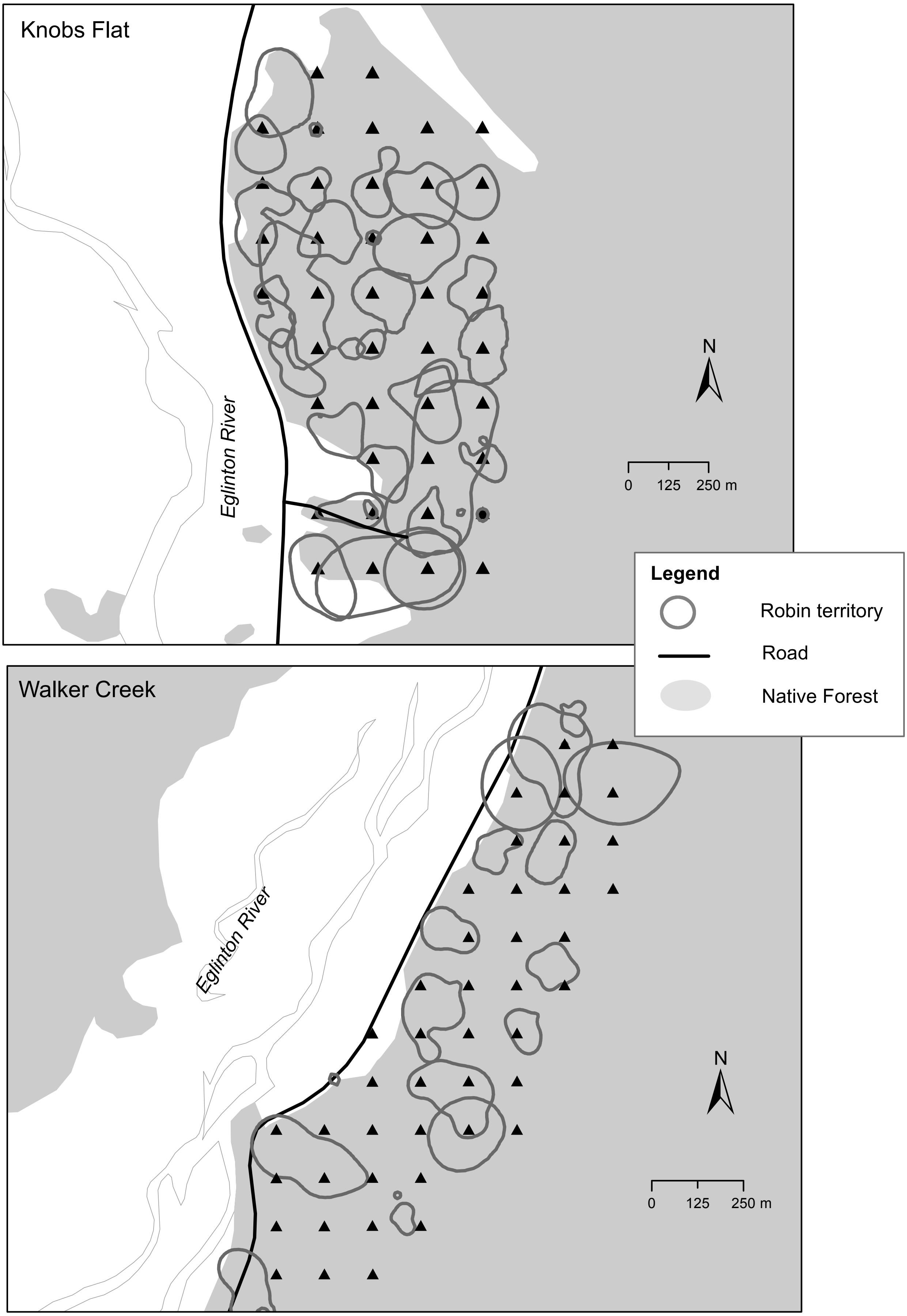

Figure 2. Maps of Knobs Flat and Walker Creek study sites (Eglinton Valley, Fiordland, New Zealand) showing typical territories for South Island robins (*Petroica a. australis*).

abundance estimates ($\hat{N}$) from the number of marked animals seen within a population over multiple resighting surveys. We used the simulation routines provided in this program to investigate various aspects of survey design in order to guide the field effort. These simulations showed that if there were (a) at least seven sighting occasions within each survey period (March and August for both study sites), (b) 50% of the expected proportion of the population was marked, and (c) the probability of resighting was between 40 and 50%, provided that we maintained the high number of banded birds, we would achieve precise estimates (total confidence interval length <20% of estimated abundance).

Distance sampling

South Island robins (along with tomtits *Petroica macrocephala macrocephala*, grey warblers *Gerygone igata*, bellbirds *Anthornis melanura*, and chaffinches *Fringilla coelebs*) were counted from designated fixed points using conventional distance-sampling methods (Marsden 1999; Nelson & Fancy 1999; Buckland 2006; Greene et al. 2010). Each point was visited twice during each of the 10 survey periods between March 2005 and August 2009. To maximise detectability of robins, counts were commenced no earlier than 1 h after sunrise and completed prior to 1300 hours and only conducted in reasonable weather (no significant rain or wind). Visits to all sample points within each study site took two observers a minimum of two half-days to complete.

As robins are known to be attracted to observers (particularly those that are stationary; Heather & Robertson 2005), it was thought likely that birds would move toward the observer before they were detected. Thus, smaller detection distances would be more common than expected, resulting in upward bias in density estimates. Observers were therefore instructed to approach each point quietly, scanning to detect any robins at their initial locations (particularly those close to the point) at the start of each count period. Birds were only recorded if they were detected within 50 m of a point (birds beyond this distance were ignored) to minimise undetected movement of robins.

Robins were located throughout a 2-min period after which the horizontal radial distance from the birds' location when first detected to the observation point was measured to the nearest metre using a laser rangefinder (Bushnell Yardage Pro 500$^{\mathrm{TM}}$). Distances less than the minimum focal distance of the rangefinder (usually <10 m) were estimated visually with the help of pieces of coloured plastic tape tied to vegetation at 5-m and 10-m intervals around each observation point. For those birds that were not clearly seen or only heard, measurements were made to vegetation at equivalent distances to the bird's estimated location. When an initial location could not be determined with confidence, or where it was obvious that a robin had moved prior to detection (e.g. sudden appearances of robins at observers' feet), the bird was not recorded.

Five-minute bird counts – an index of relative abundance

All birds within the Knobs Flat and Walker Creek study sites (including SI robin) were also surveyed by five-minute bird counts, using the methods outlined by Dawson and Bull (1975). The number and species of birds seen or heard (>90% of detections) within a 5-min period were counted by observers standing at the same points used for distance-sampling surveys. Counts were unbounded in terms of the area being surveyed and no attempt was made to adjust for the detectability of birds encountered. Each point at both study sites was visited four times during the 10 survey periods between March 2005 and August 2009. Two of these counts commenced immediately following the 2-min-count intervals used for distance sampling (i.e. total of 7 min spent at point). Every effort was made to ensure that timing of counts (from at least 1 h after sunrise to 1300 hours) and conditions under which they were counted (minimal rain and wind) were similar across all survey periods. Two observers took a minimum of four half-days to complete these counts.

Analysis

Mark–resight estimators

Data were analysed using the Program NOREMARK (White 1996a, b). Several mark–resight models are provided by the program. Two models were found to be suitable; the Joint Hypergeometric Maximum Likelihood Estimator (JHE) (Bartmann et al. 1987; Neal 1990; Neal et al. 1993) and Bowden's estimator (Bowden & Kufeld 1995).

The JHE model requires that several assumptions are met: (1) there is geographic and demographic closure; (2) no marks are lost; (3) animals are always correctly identified; (4) the probability of capture and recapture is the same for all animals; (5) the probability of sighting all animals is the same within a sampling occasion; (6) each animal is seen only once within a sampling occasion (Neal et al. 1993; White & Shenk 2001). Bowden's estimator relaxes four of these assumptions by allowing temporary movement out of the study site, variation in resighting probabilities, sampling with replacement, and not requiring all animals to be correctly identified (White & Shenk 2001). Although the JHE model often appears to have greater precision, it is overly precise when there is heterogeneity in sighting probabilities and, therefore, often performs poorly when estimating confidence interval coverage compared with the Bowden's estimator (McClintock et al. 2006). Because of these more relaxed assumptions, along with the behaviours exhibited by robins themselves (i.e. heterogeneous resighting probabilities), we chose Bowden's estimator to compute mark–resight abundance estimates (White 1996a).

Distance sampling

Distance data were examined and analysed using the free software Distance version 6.0 (Thomas et al. 2010). Observed differences in the vegetation composition, structure, and general topography between each study site meant that detection functions for robins at each site were likely to differ. Data were therefore analysed independently for each study site and each season (i.e. March and August) (Buckland et al. 2001). As distances to birds were recorded to the nearest metre, distances were left ungrouped rather than being aggregated into intervals. Although little estimation efficiency is lost by grouping data prior to analysis (and is recommended where there is evidence of movement of birds in response to observers prior to detection), choice of the width of distance intervals is often more critical for point counts than for line transects, and can result in relatively large variability in density estimates (Buckland et al. 2001).

To increase sample size and estimate precision, data from all surveys at a site were pooled (Buckland et al. 2001). Histograms of distance measurements were constructed and global detection functions were calculated for each site (Knobs Flat and Walker Creek) and each season (March and August).

Using these global detection functions, data were post-stratified by year and histograms of radial distance measurements constructed. A selection of robust models and appropriate expansion functions recommended by Buckland et al. (2001) were then fitted. Model fit was assessed using Akaike's Information Criterion (AIC), Goodness of Fit (GoF) and Q-Q plots (Buckland et al. 2001, 2004; Burnham & Anderson 2002). Good model fit was usually attained using uniform or half-normal models with varying numbers of adjustment terms. Data truncation within the 50 m limit imposed on field observations did little to improve estimate precision (average probability of detection at 50 m was less than 0.1) and was therefore considered unnecessary (Buckland et al. 2001). Where analysis highlighted competing models (ΔAIC < 2), selection uncertainty was addressed using a model-averaging procedure (Burnham & Anderson 2002).

Five-minute bird counts

Simple indices of abundance were calculated by aggregating the mean number of robins counted at each sample point for each site (i.e. the mean of the mean number counted at each point for each site) and computing appropriate 95% confidence intervals. Repeated visits to points within and between survey periods suggested temporal correlation of counts was likely to be an issue. To account for this potential correlation and address some of the factors influencing detection probability, the data were modelled using generalised linear mixed models (Fox 2008; Nichols et al. 2009). A number of weather and site covariates from each count could then be incorporated, including information on wind, temperature, precipitation, minutes of sunshine, cloud cover, environmental noise (e.g. stream and traffic) and observers.

Comparison of estimates

To investigate trends we calculated and plotted the population estimates and associated confidence intervals for each monitoring method over the sampling period. Territory mapping results were used as the standard against which each monitoring method was compared using logged linear models and their R-squared values. To assess their similarity we also plotted the mean five-minute bird count estimates against the fitted values derived from the generalised linear mixed models (Table 2).

Results

Territory mapping

Up to 46 individual robins (males and females) were identified in March and 42 in August at the Knobs Flat study site. Lower numbers of robins were present at Walker Creek with up to 40 identified in March and 29 in August (Table 1; Figs 3a & 4a). At Knobs Flat the trend was one of initial increase from 2005 counts (to a maximum of 46 in March and 42 in August 2006) followed by a sharp decline in numbers (minimum of 20 in March and 12 in August 2008) followed by a slight recovery in the last year (2009) for both August and March survey periods. At Walker Creek the trend was slightly different with relatively stable numbers for the first three years, with the exception of the small increase seen in March 2007 (to a maximum of 40 cf. 25 in August 2006), declining sharply to a low in 2008 (minimum of 16 in March and 15 in August) then, like Knobs Flat, followed by a slight recovery in the last year for both March and August survey periods. Male and female robins were equally observable provided August surveys were completed prior to commencement of nesting.

Table 1. Robin abundance estimates and 95% confidence intervals (CI) using territory mapping, mark–resight estimation, and distance sampling for Knobs Flat and Walker Creek, Eglinton Valley, Fiordland, New Zealand.

Site	Season	Year	Territory mapping (N)	Mark–resight ($\hat{N}$) (±95% CI)	Distance ($\hat{N}$) (±95% CI)
Knobs Flat	March	2005	34	38 (33–45)	175 (95–383)
		2006	46	48 (44–53)	213 (116–433)
		2007	30	25 (24–27)	157 (75–369)
		2008	20	17 (17–17)	123 (41–300)
		2009	23	20 (18–22)	174 (58–365)
	August	2005	36	32 (28–36)	64 (40–103)
		2006	42	34 (28–43)	81 (54–122)
		2007	20	17 (16–19)	50 (31–83)
		2008	12	12 (12–14)	45 (26–78)
		2009	22	19 (16–23)	55 (33–91)
Walker Creek	March	2005	34	39 (33–45)	327 (87–662)
		2006	32	36 (32–40)	261 (87–529)
		2007	40	35 (31–39)	198 (63–417)
		2008	16	15 (13–17)	111 (21–636)
		2009	32	27 (26–29)	276 (68–606)
	August	2005	29	31 (26–36)	61 (24–109)
		2006	25	15 (12–19)	38 (13–75)
		2007	27	29 (23–35)	24 (6–55)
		2008	15	14 (13–14)	21 (6–48)
		2009	25	18 (16–19)	47 (16–91)

Table 2. Correlation (r^2) between known numbers of robins (N – from territory mapping) and estimates derived from mark–resight estimation, distance sampling, five-minute bird counts (5MBC) and generalised linear mixed models (GLMM).

	Knobs Flat March	Walker Creek March	Knobs Flat August	Walker Creek August
N ~ Mark–resight	0.96	0.83	0.98	0.52
N ~ Distance	0.72	0.58	0.90	0.46
N ~ 5-min bird count	0.76	0.86	0.002	0.74
N ~ GLMM	0.36	0.84	0.01	0.84
5MBC ~ GLMM	0.71	0.98	0.98	0.96

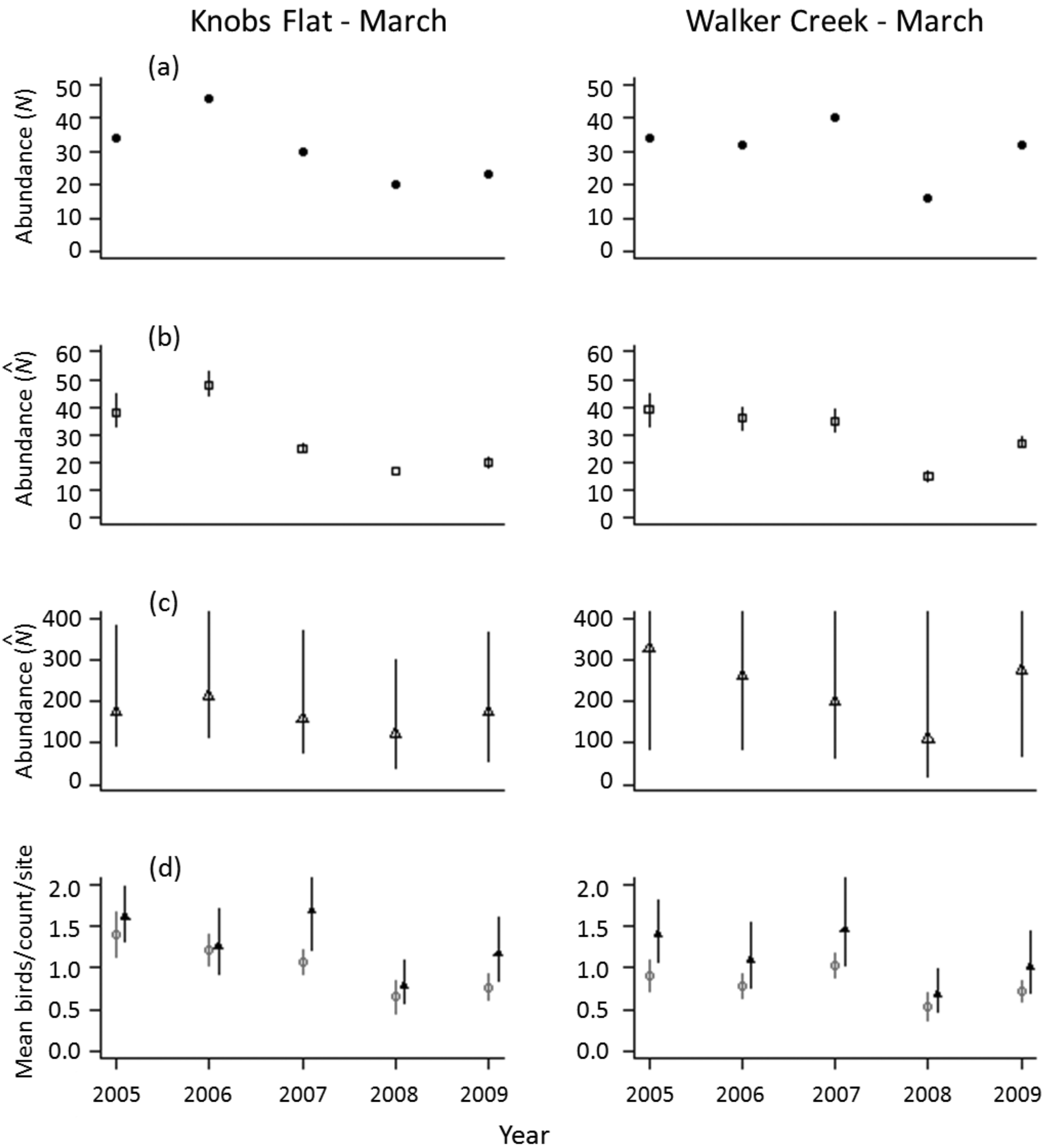

Figure 3. South Island robin (*Petroica a. australis*) population estimates for March surveys (2005–2009) at Knobs Flat and Walker Creek (Eglinton Valley, Fiordland, New Zealand), using (a) territory mapping, (b) mark–resight estimation, (c) distance sampling and (d) five-minute bird counts (open circles) and generalised linear mixed models (triangles).

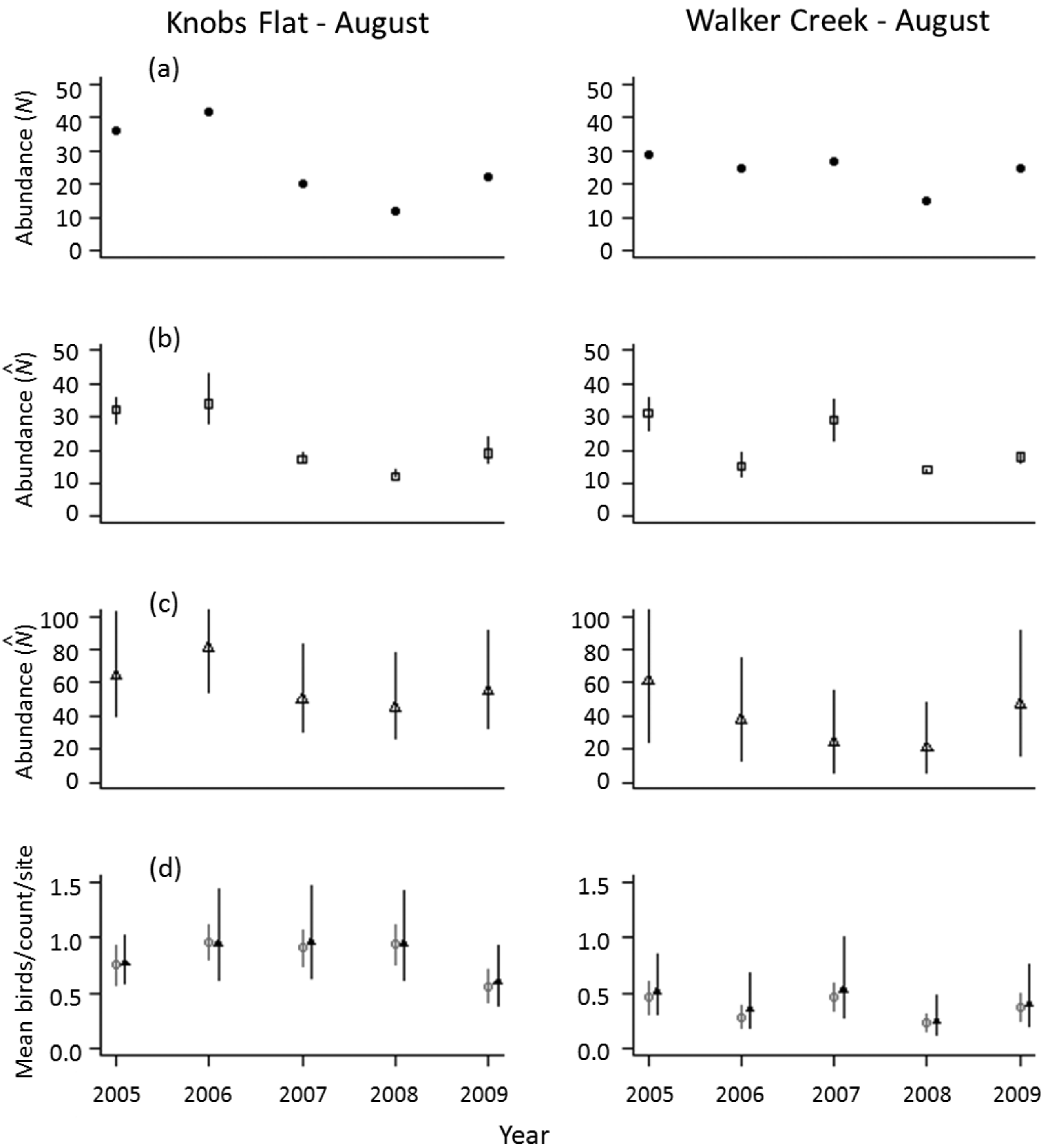

Figure 4. South Island robin (*Petroica a. australis*) population estimates for August surveys (2005–2009) at Knobs Flat and Walker Creek (Eglinton Valley, Fiordland, New Zealand), using (a) territory mapping, (b) mark–resight estimation, (c) distance sampling, and (d) five-minute bird counts (open circles) and generalised linear mixed models (triangles).

Mark–resight estimates

With the exception of surveys at Walker Creek in August, population estimates (N) and general population trends derived from mark–resight surveys using Bowden's estimator for both study sites were very similar to and highly correlated with those derived from territory mapping (Table 1; Figs 3b & 4b). At Knobs Flat, robin abundance was underestimated by 11–19% on seven occasions and overestimated by 4–10.5% on two occasions. At Walker Creek, robin abundance was underestimated by 6–40% for six of the surveys and overestimated by 6.5–12.8% for four of the surveys. Despite the relatively small variation in estimated accuracy, 9 of the 20 calculated 95% confidence intervals only narrowly failed to include the known number of robins.

Distance sampling

Robins were either seen or heard at 25% of points visited at Knobs Flat and 15% of points at Walker Creek for all surveys. During any one survey period, robins were observed at 32–76% of points at Knobs Flat and 14–47% of points at Walker Creek.

Distances to a total of 246 robins at Knobs Flat and 132 robins at Walker Creek were recorded. Of these, 83 clusters (33%) of more than one bird (range 2–4) were recorded at Knobs Flat and 31 clusters (23%) (range 2–3) at Walker Creek. Modal cluster size was 1.0 for both sites. Mean cluster size for Knobs Flat was 1.4 (95% CI = 1.32–1.48) and for Walker Creek 1.3 (95% CI = 1.20–1.40). The impact of clusters on abundance estimation was therefore likely to be small (Greene et al. 2010) and was subsequently ignored during analysis.

Relatively few distance measurements to individuals or clusters of robins were recorded for each survey period at either Knobs Flat (range 13–36) or Walker Creek (range 6–28). Pooling detections for each study site (assuming detectability at each site remains the same over time) and applying a global detection function and post-stratifying by year provided a partial solution to the lack of data for some years, but extremely low sample sizes in some years (particularly at Walker Creek) seriously compromise the precision of abundance estimates (Buckland et al. 2001).

Half-normal or uniform models were found to fit data well when pooled across March and August survey periods for both study sites. Up to three competing models were highlighted for each analysis.

Abundance estimates and corresponding 95% confidence intervals (model-averaged where appropriate) are graphed in Fig. 3c for surveys conducted in March (post-breeding) and Fig. 4c for surveys conducted in August (pre-breeding) for both study sites. Robin abundance estimates at both Knobs Flat and Walker Creek exhibit highly significant positive bias when compared with actual numbers. Similarly, confidence intervals are extremely wide and estimate-precision is therefore poor. Highly inflated abundance estimates and large confidence intervals are particularly pronounced for the March survey periods. Despite these inaccuracies, the general relationship between abundance estimates derived from territory mapping and distance sampling, particularly for Knobs Flat, is quite strong (Table 2; Figs 3c & 4c). Although this relationship is not as marked for Walker Creek (where robins were less commonly encountered), distance sampling was able to identify the general trend of decline in abundance to a low in 2008 followed by a subsequent increase in 2009.

Five-minute bird counts

During the August survey periods, robins were detected at least once at 81–95% of points at Knobs Flat and 58–79% of points at Walker Creek. In March, robins were detected at least once at 76–98% of points at Knobs Flat and 81–100% of points at Walker Creek. The mean number of robins recorded in the August survey period ranged from 0.56 to 0.97 at Knobs Flat and 0.23 to 0.46 at Walker Creek. In the March survey period the mean number of birds ranged from 0.65 to 1.40 at Knobs Flat and 0.53 to 1.03 at Walker Creek. The fitted values ranged from 0.59 to 0.96 at Knobs Flat and 0.24 to 0.53 at Walker Creek in the August survey, and, from 0.79 to 1.61 at Knobs Flat and 0.68 to 1.40 at Walker Creek in March. Mean values have smaller confidence intervals than the fitted values (Figs 3d & 4d) but this does not reflect accuracy. The mean counts are calculated as the mean of the point mean and assume that all points are independent. The fitted values take advantage of the data having been collected from the same sample units over time, can incorporate variables such as weather, and usually lead to larger confidence intervals.

Five-minute bird counts and territory mapping generally showed very similar trends for surveys of Knobs Flat and Walker Creek during March survey periods. The exception to this occurred at Knobs Flat in March 2006 where five-minute bird counts and the fitted values failed to detect the 26% increase in robin numbers between 2005 and 2006. However, both methods did detect significant decreases in robin numbers in 2008 followed by subsequent increases in 2009. At Walker Creek, territory mapping and five-minute bird counts showed very similar trends in robin numbers (Table 2; Figs 3 & 4) with population declines readily observable in 2006 and 2008. The fitted values also show a very similar pattern despite the increase in variability following the inclusion of relevant covariates. In contrast, surveys of Knobs Flat in August show quite different results. Five-minute bird counts and the fitted values suggest a relatively stable robin population with a sharp decline between 2008 and 2009. Territory mapping, on the other hand, showed that robin numbers increased slightly between August 2005 and 2006, declined steeply to 2008, then began increasing again over the following year. Relative abundances derived from five-minute bird counts and repeated measures modelling show little correlation with these trends (Table 2; Figs 3 & 4). Five-minute bird counts and the fitted values generally showed a strong relationship as expected ($R^2 > 0.90$), except for the March counts at Knobs Flat where the inclusion of the variables increased the fitted value for 2007 (Table 2; Fig. 3d).

Discussion

Although abundance estimates derived from territory mapping of marked birds were the most accurate (Thompson et al. 1998), they are heavily dependent on a high proportion of the population being marked, the accurate identification of these marked individuals, and the assumption that all unmarked individuals can be identified. Although we were able to meet these assumptions, the time and effort required to maintain and map sufficient marked birds is a luxury not normally possible at other sites or for other species (Thompson et al. 1998; Sutherland 2006).

Despite the inherent expense and relative inefficiency of territory mapping as a general tool for landscape-scale monitoring, robin abundance was accurately assessed without the need for estimates of precision. Abundance estimates were reliable and less vulnerable to environmental and behavioural variations and restrictive assumptions compared with other methods (Sutherland 2006). We are confident, therefore, that our results provided an accurate baseline measurement of known population size from which robust comparisons of the accuracy and precision of robin abundance could be made.

With few exceptions our mark–resight estimates for Knobs Flat and Walker Creek were very similar to those derived from territory mapping. The 95% CIs were usually small and either overlapped the actual number of birds present or came within one or two birds of doing so. The large proportion of marked birds relative to unmarked birds and the relatively large number of resighting occasions (7) at each study site were critical to achieving this accuracy and precision and are key components for any mark–resight study design (McClintock et al. 2006; Sutherland 2006). Bowden's estimator, with its more relaxed assumptions, further improved estimate accuracy. Despite these obvious advantages, the relationship between actual robin numbers as determined by territory mapping and those derived from mark–resight data was not uniformly consistent. Although estimates of precision for the August surveys at

Walker Creek appeared high, the relationship between known numbers and the mark–resight estimates was poor ($R^2 = 0.52$). This seems almost entirely attributable to robins breeding earlier than expected in August 2006. Nesting female robins were not available to be counted during resighting surveys, distance sampling, or five-minute bird counts and the population estimate simply reflected the number of male birds within the study site. The actual number of robins present in the area could only be determined following more intensive territory mapping. The importance placed on these sources of variation will depend on their frequency of occurrence relative to the period over which population trends are being measured. For this reason care should always be taken when interpreting point estimates for any one year.

In contrast, using distance sampling methods from points to monitor robin populations resulted in overestimates of population abundance and poor precision. Clearly our attempts to limit bias and prevent violations of distance sampling assumptions by modifying our sampling regime were of limited success. Violations of the assumption that birds are detected at their initial locations appeared (as predicted) particularly problematic despite rejection of observations for which initial locations were known to be uncertain. This approach also drastically reduced the available sample size and undoubtedly contributed to the poor precision of calculated estimates.

The scale of the reported bias appears to depend largely on the time of year counts are conducted. March abundance estimates generated considerable overestimates compared with those derived from August counts. This is presumably related to both the seasonal increase of robins at each study site (with juveniles present post-breeding), their increased attraction to observers following feeding with mealworms during autumn (March) capture periods, and changes in the types and rates of calls. Considering the scale of the observed bias it was therefore surprising that the abundance estimates from distance sampling were so well correlated with those from territory mapping for Knobs Flat irrespective of time of year. It is not clear why the estimates of robin abundance at Walker Creek were less strongly correlated and precise, but it is likely that combinations of behavioural and site factors (e.g. female robins breeding earlier than expected in August 2006, and thus unavailable during counts, as well as differences in forest structure and topography) played a role in some years. Despite these problems, distance sampling did generate abundance estimates for Walker Creek that reflected the general decline in robin numbers to a minimum in 2008 and subsequent increase in 2009.

Buckland (2006) suggests that the standard implementation of point-based distance sampling (i.e. recording distance estimates to all birds detected within a fixed time period) can often yield abundance estimates with considerable bias as birds move about the plot area. Variations on the method, including the closely related 'snapshot' of bird locations (Buckland et al. 2001) and cue counting (Buckland 2006), do not assume that individual birds are fixed at a single location over the count period and therefore avoid this potential source of overestimation. Despite the apparent advantages of these alternatives, only the use of cue counting seems likely to improve the accuracy and precision of robin abundance estimates derived from point counts. The use of the snapshot method, even using very short detection periods (e.g. the 2-min interval used here), does not solve the issue of significant and often undetected movement of robins toward observers. Cue-based methods will only work well for cues (e.g. calls) that are short, well defined, easily detected and for which cue production rates can be reliably estimated (Buckland 2006). Robin call rates can vary tremendously among individuals depending on season, sex and the demographic structure of the population being monitored. It may be possible, however, to estimate cue rates for specific call types (e.g. downscales regularly produced by male and less commonly by female SI robins between January and July; Powlesland 1983) within a survey period, in which case cue-counting from points may be a viable alternative, at least for territory holding adult males.

If the habitats and topography being traversed are relatively undemanding, a far easier way of increasing the accuracy and precision of abundance estimates from distance sampling might be to count robins from line transects rather than points (Buckland et al. 2001; Efford & Dawson 2009). Results from trial monitoring, of North Island robins (*Petroica australis longipes*), using distance sampling from line transects suggest that movement of birds toward observers is much reduced if observers are moving rather than stationary, and noticeable improvements in the precision of estimates of robin abundance can be made (O. Overdyck, Department of Conservation, pers. comm.). Other species that are often attracted to observers (e.g. tomtits *Petroica macrocephala*, grey warblers *Gerygone igata*, and bellbirds *Anthornis melanura*) may also benefit from this approach.

Comparison of the trends derived from territory mapping and five-minute bird counts provided rather mixed results. For Walker Creek the relationship between abundance estimates from the two methods was strong for both the March and August survey periods. In contrast, at Knobs Flat there was little correspondence between the known number of birds and the index. Of particular note is the complete failure of five-minute bird counts to detect the 71% decline in robin numbers at Knobs Flat between August 2006 and August 2008 and the subsequent increase in 2009. This decline was caused by a rapid increase in predator numbers at Knobs Flat during 2007, following a large beech seeding event, resulting in significant robin predation (C.F.J. O'Donnell, pers. comm.). As females are more vulnerable to predation when nesting (O'Donnell 1996), a male-dominated population resulted. Single males are known to call and move about more frequently as they look for a mate (Powlesland 1983) and it seems likely that five-minute bird counts were only detecting this behavioural change (i.e. the increased calling by males) rather than any real increase in numbers.

The sex ratio of the robin population at Knobs Flat in August 2009 remained skewed toward males. Despite this, estimates derived from five-minute bird counts suggested a population decline between 2008 and 2009 while the number of birds counted by territory mapping increased. In this instance, cold weather at the time of the August 2009 survey period delayed the onset of breeding activity for a month. The reduction in the production and intensity of full song, despite the strongly male-biased population, was reflected in the relative abundance of robins measured by five-minute bird counts.

Although Gill (1980) found a linear relationship between indices derived from five-minute bird counts and the actual density of South Island robins over a 17-month period, we had less success. Our inability to measure robin population trends accurately at Knobs Flat compared with Walker Creek appeared to be directly influenced by disproportionate predation of females at Knobs Flat, where there was no rodent control, and the subsequent changes in male robin detectability (increased calling). Appropriate selection of environmental covariates can correct for factors influencing detection probability in some circumstances, but did little to improve accurate measurement of trends for either of our sites. Where predators simultaneously influence the abundance and detection probability of robins, indices of relative abundance such as five-minute bird counts are not likely to be adequate for monitoring robin population trends over the medium to long term.

Prediction of when these conditions might occur is obviously problematic, especially without detailed knowledge of pest numbers and the demographic structure, timing of breeding seasons, and the behaviour of robins. Our results indicate that five-minute bird counts conducted in March provide reasonable estimates of robin population density and abundance, despite being somewhat inflated by the presence of juvenile birds. However, survival of robins (particularly juveniles) over the following winter and inter-annual variations in productivity are not taken into account, which could potentially increase the precision of population estimates. On the other hand, if robins are counted in August then there is considerable potential for behavioural changes (predation induced or otherwise) to affect five-minute bird counts that either mask changes in abundance or produce spurious trends (Gibbs & Wenny 1993). Thus, it is extremely important that the assumptions and inherent biases of any index of abundance are understood and results interpreted with caution, particularly where the demographic parameters for the species being monitored are poorly understood.

Cost

Territory mapping of robins in the Eglinton Valley was the most costly but the most accurate monitoring method applied in this study. The effort required effectively limited monitoring to two relatively small sites. If the objective was to monitor populations over large-landscape-scale areas, a less intensive monitoring regime with reduced accuracy and precision would be required (Sutherland 2006).

Although mark–resight surveys require a considerable initial investment (i.e. catching and individually marking a significant part of the population), for relatively small South Island robin populations, very accurate and precise estimates of population size and trend can be made. Once established, such surveys can be maintained by suitably qualified technical staff. The cost-efficiency of mark–resight methods is dependent on the proportion of the target population that can be captured and marked, the number of sighting occasions that can be conducted in any one survey period and the resighting probability of marked individuals. This will be difficult for many bird species, particularly those that are rare, sparsely distributed, or difficult to identify accurately once marked. Any reduction in the proportion of individuals marked or the numbers of sampling occasions within each survey period will greatly reduce the accuracy and precision of population estimates that can be made.

One of the great advantages of distance sampling is that it can yield estimates of abundance adjusted for detectability over large representative areas with only modest resource requirements (Buckland 2006). As long as critical assumptions can be met and enough data of sufficient accuracy can be collected (often difficult in a field situation; Alldredge et al. 2007; Efford & Dawson 2009), legitimate comparisons can be made across sites and times as well as between species (Buckland et al. 2001; Diefenbach et al. 2003; Sutherland 2006). Adoption of line-transect survey methods rather than point-based distance sampling would, at least for robin populations where conditions are suitable, improve the cost-efficiency and reliability of the data collected. Such a change may increase the number of robins encountered and the area surveyed, and reduce the observed positive bias.

Five-minute bird counts required less field and analytic effort than the other methods and as such are inherently attractive to managers seeking cost-effective methods to monitor forest bird populations on a landscape scale. Although we were able to demonstrate a reasonable relationship between the relative abundance of robins and their actual abundance, any increase in the precision of our index would require large increases in sample sizes (Dawson 1981). If this requires considerable additional effort and expense then managers would be better off using more formal estimation methods that adjust for incomplete detectability or accepting that only relatively large changes (40–50%) in population size are likely to be detected. Similarly, the failure of five-minute bird counts to detect the predation-mediated declines in robin numbers at Knobs Flat clearly demonstrates that indices can and do fail to meet the assumption of constant proportionality and can easily be invalidated by unanticipated future events (Thompson et al. 1998). Unless alternative sources of information (e.g. demographic data, predator densities, etc.) are available, spurious results derived from indices could have far more serious consequences (i.e. undetected declines) than anticipated.

Indices can provide useful information about the relative abundance and spatial distribution of species and they may be the only viable survey option for multi-species avian community monitoring programmes. Nevertheless, their use should be carefully considered and results interpreted, weighing the relative importance of cost (and practicality) against inferential strength (Thompson et al. 1998; Williams et al. 2002; Norvell et al. 2003).

Recommendations

Monitoring objectives, estimator assumptions, the desired level of estimate accuracy and precision, and the characteristics of the species of interest should all influence the choice of an appropriate monitoring method. Although robins were in many ways ideal candidates for this comparison (they are territorial, relatively common and easily identifiable with obvious song and calls, as well as being easy to capture and mark), no monitoring method will work well where robins are routinely fed or for other species with different behavioural characteristics. This is likely to be particularly true for those species that are cryptic, sparsely distributed, hard to capture and mark, or rare, and for which sample sizes are likely to be small.

Some general recommendations are, however, possible. If the objectives of a monitoring programme demand particularly accurate and precise estimates of abundance (e.g. impact of pest control on vulnerable indicator species, status and trends of threatened species, etc.), the use of resource-intensive monitoring methods such as territory mapping and mark–resight estimators may be justified. Although these methods can usually only be applied to bird populations in relatively small areas, and generally in a research context (unless there is an easily maintained pre-existing marked population), the use of individually marked birds allows monitoring to be expanded to cover complementary demographic parameters such as survival, productivity, and rates of population change. Although less intensive monitoring methods are clearly attractive, our data show there are significant risks in relying on indices of relative abundance to measure population trends. This is particularly so over longer time-frames when the already tenuous assumption of constant detectability is likely to be invalidated by unanticipated events such as shifts in behaviour. Alternatives that yield abundance estimates adjusted for differential detectability (e.g. distance sampling) are potentially much more robust and should be carefully considered. Providing that the critical assumptions of distance sampling can be met in the field, the sample size is reasonably high, and the survey methodology can be tailored to the species of interest (or small groups of species that share similar behaviours or detection probabilities), the relatively small increase in expense and effort for more reliable abundance estimates seems justified. There will be occasions, however, when methods such as distance sampling will not perform well (e.g. this study), further emphasising the need to tailor survey methods (e.g. observation radius and length of count period) to the monitoring situation and objectives. In these instances, indices of abundance derived from point counts are still useful and may prove more robust (Efford & Dawson 2009). Such counts can be used advisedly to interpret spatial and temporal variation providing (a) sample size is high, (b) there is reasonable likelihood that detection probabilities are similar, or the main sources of variation in detection probability can be identified, measured and modelled as covariates (e.g. in a repeated-measures framework), and (c) there is a clear understanding that only relatively large differences in abundance are likely to be detected (Thompson et al 1998; Sutherland 2006; Nichols et al. 2009). If detection probability cannot be estimated from the count data then it is important to standardise the sampling design to control for factors influencing detectability or to collect data on sampling conditions that can be used to model abundance estimates (e.g. weather variables, observers, etc.). This approach is often extremely difficult as there are usually too many factors to account for (Thompson et al. 1998), particularly for surveys in which multiple species of variable detectability are being monitored simultaneously.

Our data, along with extensive evidence from elsewhere, suggest that reliance on unadjusted indices of relative abundance (e.g. five-minute bird counts) for inferential monitoring in some situations is likely to be misplaced. If unbiased and reliable counts are required and constant detectability cannot be reasonably assumed, we strongly recommend that monitoring methods that explicitly account for incomplete detectability or sources of variation in detection probabilities be used wherever possible.

Acknowledgements

We thank Greg Coates, Peter Dilks, Richard Earl, Brice Ebert, Hannah Edmonds, Richard Ewings, Lynette Hartley, Rod Hay, Danilo Hegg, Iris Broekema, Paul Jacques, Sue Lake, Colin O'Donnell, Georgina Pickerell and Emma Williams for assistance in the field. We also thank Derek Brown for GIS assistance and Brenda Greene, Rod Hay, Ian Jamieson and an anonymous reviewer for helpful comments on earlier drafts.

References

Alldredge 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.

Bartmann RM, White GC, Carpenter LH, Garrott RA 1987. Aerial mark-recapture estimates of confined mule deer in pinyon-juniper woodland. Journal of Wildlife Management 51: 41–46.

Bibby CJ, Burgess ND, Hill DA, Mustoe S 2000. Bird census techniques. 2nd edn. London, Academic Press. 302 p.

Bowden DC, Kufeld RC 1995. Generalized mark-sight population size estimation applied to Colorado moose. Journal of Wildlife Management 59: 840–851.

Broekema I, Overdyck O 2012. Distance sampling to estimate densities of four native forest bird species during multi-species surveys. New Zealand Journal of Ecology 36: 353-364.

Buckland ST 2006. Point-transect surveys for songbirds: robust methodologies. The Auk 123: 345–357.

Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L 2001. Introduction to distance sampling: estimating abundance of biological populations. 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 University Press. 416 p.

Burnham KP, Anderson DR 2002. Model selection and multimodel inference: a practical information-theoretic approach. 2nd edn. New York, Springer. 488 p.

Casagrande DG, Beissinger SR 1997. Evaluation of four methods for estimating parrot population size. Condor 99: 445–457.

Dawson DG 1981. Experimental design when counting birds. In: Ralph CJ, Scott JM eds Estimating numbers of terrestrial birds. Studies in Avian Biology 6. Pp. 392–398.

Dawson DG, Bull PC 1975. Counting birds in New Zealand forests. Notornis 22: 101–109.

Dawson DK, Efford MG 2009. Bird population density estimated from acoustic signals. Journal of Applied Ecology 46: 1201–1209.

Diefenbach DR, Brauning DW, Mattice JA 2003. Variability in grassland bird counts related to observer differences and species detection rates. Auk 120: 1168–1179.

Efford MG, Dawson DK 2009. Effect of distance-related heterogeneity on population size estimates from point counts. Auk 126: 100–111.

Ellingson AR, Lukacs PM 2003. Improving methods for regional landbird monitoring: a reply to Hutto and Young. Wildlife Society Bulletin 31: 896–902.

Elphick CS 2008. How you count counts: the importance of methods research in applied ecology. Journal of Applied Ecology 45: 1313–1320.

Fox J 2008. Applied regression analysis and generalized linear models. 2nd edn. Thousand Oaks, CA, Sage.

Gibbs JP, Wenny DG 1993. Song output as a population estimator: effect of male pairing status. Journal of Field Ornithology 64: 316–322.

Gill BJ 1980. Abundance, feeding, and morphology of passerine birds at Kowhai Bush, Kaikoura, New Zealand. New Zealand Journal of Zoology 7: 235–246.

Greene TC, Jones A, Dennis G, Sachtleben T 2010. Distance sampling to determine kaka (*Nestor meridionalis septentrionalis*) density within Waipapa Ecological Area, Pureora. New Zealand Journal of Ecology 34: 297–305.

Hamel PB 1984. Comparison of variable-circular plot and spot-map censusing methods in temperate deciduous forest. Ornis Scandinavica 15: 266–274.

Heather BD, Robertson HA 2005. The field guide to the birds of New Zealand. Rev. edn. Auckland, Viking. 440 p.

Johnson DH 2008. In defence of indices: the case of bird surveys. Journal of Wildlife Management 72: 857–868.

Kissling ML, Garton EO 2006. Estimating detection probability and density from point-count surveys: a combination of distance and double-observer sampling. The Auk 123: 735–732.

Marsden SJ 1999. Estimation of parrot and hornbill densities using a point count distance sampling method. Ibis 141: 327–390.

Mackenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey LA, Hines JE 2006. Occupancy modeling and estimation. San Diego, CA, Academic Press. 324 p.

MacLeod CJ, Greene T, MacKenzie DI, Allen RB 2012. Monitoring widespread and common bird species on New Zealand's conservation lands: a pilot study. New Zealand Journal of Ecology 36: 300-311.

McClintock BT, White GC, Burnham KP 2006. A robust design mark–resight abundance estimator allowing heterogeneity in resighting probabilities. Journal of Agricultural, Biological and Environmental Statistics 11: 231–248.

McClintock BT, White GC, Burnham KP, Pryde MA 2008. A generalized mixed effects model of abundance for mark–resight data when sampling is without replacement. In: Thomson DL, Cooch EG, Conroy MJ eds Modeling demographic processes in marked populations. Environmental and Ecological Statistics 3. New York, Springer. Pp. 271–289.

McClintock BT, White GC, Antolin MF, Tripp DW 2009. Estimating abundance using mark–resight when sampling is with replacement or the number of marked individuals is unknown. Biometrics 65: 237–246.

Morgan DKJ, Waas JR, Innes J, Arnold G 2012. Native bird abundence after Australian Magpie (*Gymnorhina tibicen*) removal from localised areas of high resource availability. New Zealand Journal of Ecology 36: 333-339.

Neal AK 1990. Evaluation of mark–resight population estimates using simulations and field data from mountain sheep. Unpublished MS thesis, Colorado State University, Fort Collins, CO.

Neal AK, White GC, Gill RB, Reed DF, Olterman JH 1993. Evaluation of mark–resight model assumptions for estimating mountain sheep numbers. Journal of Wildlife Management 57: 436–450.

Nelson JT, Fancy SG 1999. A test of the variable circular-plot method where exact density of a bird population is known. Pacific Conservation Biology 5: 139–143.

Nichols JD, Thomas L, Conn PB 2009. Inferences about landbird abundance from count data: Recent advances and future directions. In: Thomson DL, Cooch EG, Conroy MJ eds Modeling demographic processes in marked populations. Environmental and Ecological Statistics 3. New York, Springer. Pp. 201–235.

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. Predators and the decline of New Zealand forest birds: an introduction to the hole-nesting bird and predator programme. New Zealand Journal of Zoology 23: 213–219.

O'Donnell CFJ 2000. Influence of season, habitat, temperature and invertebrate availability on nocturnal activity of the New Zealand long-tailed bat (*Chalinolobus tuberculatus*). New Zealand Journal of Zoology 27: 207–221.

Powlesland RG 1983. Seasonal and diurnal variation in vocal behaviour of the South Island robin. New Zealand Journal of Zoology 10: 225–232.

Rosenstock SS, Anderson DR, Giesen KM, Leukering T, Carter MF 2002. Landbird counting techniques: current practices and an alternative. Auk 119: 46–53.

Spurr EB, Borkin KM, Drew KW 2012. Line-transect distance sampling compared with fixed-width strip-transect counts for assessing tomtit *(Petroica macrocephala)* population trends. New Zealand Journal of Ecology 36: 365-370.

Sutherland WJ ed. 2006. Ecological census techniques: a handbook. 2nd edn. Cambridge University Press. 432 p.

Tarvin KA, Garvin MC, Jawor JM, Dayer KA 1998. A field evaluation of techniques used to estimate density of Blue Jays. Journal of Field Ornithology 69: 209–222.

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, White GC, Gowan C 1998. Monitoring vertebrate populations. San Diego, CA, Academic Press. 365 p.

White GC 1996a. Program NOREMARK software reference manual. Fort Collins, CO, Department of Fishery and Wildlife Biology, Colorado State University. http://warnercnr.colostate.edu/~gwhite/noremark

White GC 1996b. NOREMARK: Population estimation from mark-resighting surveys. Wildlife Society Bulletin 24: 50–52.

White GC, Shenk TM 2001. Population estimation with radio-marked animals. In: Millspaugh JJ, Marzluff JM eds Radio tracking and animal populations. San Diego, CA, Academic Press. Pp. 329–350.

Williams BK, Nichols JD, Conroy MJ 2002. Analysis and management of animal populations: modelling, estimation, and decision making. San Diego, CA, Academic Press. 817 p.