Methods in Ecology and Evolution

Methods in Ecology and Evolution 2016 doi: 10.1111/2041-210X.12617

Designing occupancy studies when false-positive detections occur

Matthew J. Clement*,†

United States Geological Survey, Patuxent Wildlife Research Center, 12100 Beech Forest Road, Laurel, MD 20708, USA

Summary

1. Recently, estimators have been developed to estimate occupancy probabilities when false-positive detections occur during presence–absence surveys. Some of these estimators combine different types of survey data to improve estimates of occupancy. With these estimators, there is a trade-off between the number of sample units surveyed, and the number and type of surveys at each sample unit. Guidance on efficient design of studies when false positives occur is unavailable.

2. For a range of scenarios, I identified survey designs that minimized the mean square error of the estimate of occupancy. I considered an approach that uses one survey method and two observation states and an approach that uses two survey methods. For each approach, I used numerical methods to identify optimal survey designs when model assumptions were met and parameter values were correctly anticipated, when parameter values were not correctly anticipated and when the assumption of no unmodelled detection heterogeneity was violated.

3. Under the approach with two observation states, false-positive detections increased the number of recommended surveys, relative to standard occupancy models. If parameter values could not be anticipated, pessimism about detection probabilities avoided poor designs. Detection heterogeneity could require more or fewer repeat surveys, depending on parameter values. If model assumptions were met, the approach with two survey methods was inefficient. However, with poor anticipation of parameter values, with detection heterogeneity or with removal sampling schemes, combining two survey methods could improve estimates of occupancy.

4. Ignoring false positives can yield biased parameter estimates, yet false positives greatly complicate the design of occupancy studies. Specific guidance for major types of false-positive occupancy models, and for two assumption violations common in field data, can conserve survey resources. This guidance can be used to design efficient monitoring programmes and studies of species occurrence, species distribution or habitat selection, when false positives occur during surveys.

Key-words: detection heterogeneity, detection probability, false positives, misidentification, occupancy models, presence–absence, study design

Introduction

The proportion of sample units occupied by a target species has long been a state variable of interest for ecological questions related to species distribution modelling, meta-population ecology and community ecology (Bailey, MacKenzie & Nichols 2014). However, surveys of organisms typically include false negatives, which bias estimates of occupancy (Gu & Swihart 2004). MacKenzie *et al.* (2002) and Tyre *et al.* (2003) developed unbiased estimators of occupancy probability in the face of incomplete detection by analysing repeated surveys of sample units. However, these estimators are biased even by the low (e.g. 1–3%) false-positive probabilities that have been documented in experimental aural surveys of birds and amphibians (Alldredge *et al.* 2008; McClintock *et al.* 2010; Miller *et al.* 2012).

Recognizing the problem of false positives, Royle & Link (2006) developed an estimator of occupancy that allowed for false-positive detections. They parameterized detection probability as a 2-component finite mixture, with a true-positive detection probability, p_{11}, at occupied sample units and a false detection probability, p_{10}, at unoccupied sample units. With the constraint that $p_{11} > p_{10}$, the parameters are identifiable. However, model results are ambiguous because the same model can also be interpreted as a heterogeneous detection model with no false positives (Royle & Link 2006). Subsequently, Miller *et al.* (2011) introduced false-positive occupancy estimators that allow one to incorporate additional survey information. With this approach, a surveyor may record two observation states: uncertain detections and certain detections. For example, an animal detected at close range may generate a certain detection, while detection of a distant animal may generate an uncertain detection. The model

*Correspondence author. E-mail: mclement@gmail.com
†Present address: Research Branch, Arizona Game and Fish Department, 5000 W Carefree Hwy, Phoenix, AZ 85086, USA

assumes a certain detection cannot be a false positive, while an uncertain detection might be. Non-detections are always uncertain, in that they might be due to true absences or false negatives. Alternatively, Miller *et al.* (2011) proposed that surveyors might use two survey methods, one of which generates uncertain detections, while the other only generates certain detections (see Ferguson, Conroy & Hepinstall-Cymerman 2015 for a Bayesian formulation). For example, an acoustic survey might yield uncertain detections (false positives possible) while a capture survey might generate only certain detections (no false positives).

These new modelling methods raise questions about how studies should be designed (Clement *et al.* 2014). In particular, there is a trade-off between the number of sample units surveyed, and the number and type of surveys at each sample unit. Guidance exists for designing standard occupancy models (MacKenzie & Royle 2005; Bailey *et al.* 2007; Guillera-Arroita, Ridout & Morgan 2010; Guillera-Arroita & Lahoz-Monfort 2012, in press), but this guidance is unlikely to be appropriate for false-positive models. Under a standard occupancy model (MacKenzie *et al.* 2002), designing a study requires selecting an appropriate number of repeat surveys, K, given an anticipated probability of occupancy, ψ, and probability of true-positive detection, p_{11} (MacKenzie & Royle 2005). Under a two-observation-state model (Miller *et al.* 2011), the optimal number of repeat surveys depends on two additional parameters, p_{10} and b, the probability that a true-positive detection will be classified as certain. Under a two-method false-positive model (Miller *et al.* 2011), sampling design requires selecting an appropriate number of repeat surveys, T and S, for both the uncertain and the certain survey methods, respectively. In addition to ψ, survey design is affected by the probability of true-positive detection for the certain method, r_{11}, and the uncertain method, p_{11}, as well as the false-positive probability for the uncertain method, p_{10}.

Given the increasing use of false-positive occupancy models, guidance on effective study design would be useful for practitioners. Therefore, I investigated the optimal allocation of effort for false-positive occupancy models. For both the two-observation-state model and the two-method model, I considered how variation in different parameter values and the relative cost of different survey methods affected the optimal number of repeat surveys to perform. I also investigated how deviations from model assumptions affected optimal design in an effort to find robust designs. In particular, I considered how optimal design was affected by (i) errors in pre-survey expectations of parameter values and (ii) unmodelled detection heterogeneity. Based on these analyses, I make recommendations on how to allocate survey effort when designing false-positive occupancy studies.

Materials and methods

My goal was to identify optimal survey designs for false-positive occupancy studies under various scenarios. In this study, survey design consists of selecting the number and type of surveys to perform in each sample unit. Given a fixed budget, the number of repeat surveys then determines the number of surveyed sample units. I follow MacKenzie & Royle (2005) in defining the optimal design as the design that produces the most accurate estimate of ψ, given a fixed budget, although the accuracy of other parameter estimates could be considered as well (Guillera-Arroita, Ridout & Morgan 2010). First, I describe my general approach to investigating optimal survey design, and then I apply it to both two-observation-state models and two-method models.

MacKenzie & Royle (2005) used analytic methods to calculate asymptotic variances for different study designs. Although such closed-form solutions are convenient, they do not hold when the model is misspecified, which, unavoidably, is the case in many studies. Instead, I used numerical methods to estimate variances and mean square errors. More specifically, for each study design and scenario of interest (described below), I used expected values to generate encounter histories, each occurring at the expected frequency. I then used maximum-likelihood estimation to estimate parameters and their variances on the logit scale. Finally, I used the delta method to obtain variances on the probability scale (Williams, Nichols & Conroy 2002: 736). I completed all analyses with Program R (version 3.1.2; R Core Team 2014).

I considered the most accurate design to be the one with the lowest mean square error (variance + bias^2) for ψ on the probability scale. For correctly specified models, estimates were unbiased, and the most accurate design corresponded to the most precise design (i.e. the one generating the lowest variance in ψ). For correctly specified standard occupancy models, this approach yielded results identical to MacKenzie & Royle (2005). For misspecified models (described below), estimates were biased, so the mean square error and the optimal design were affected by bias as well as variance.

Initially, I created scenarios that met all model assumptions and identified optimal designs associated with different parameter values. Preliminary investigations of two-method models indicated that in contrast to standard occupancy models, slight differences in parameter values produced very different optimal survey designs. These results indicated that if the parameter values anticipated during the design stage were incorrect, the selected design could be very different from the true optimal design. I expect that errors in anticipating parameter values are ubiquitous in field studies, and therefore, hedging strategies protecting against such errors could benefit ecologists. A useful hedging strategy would select a survey design that ensures adequate parameter accuracy in the face of pre-survey parameter uncertainty. To identify good hedging strategies, I estimated the variance for several similar scenarios with different optimal designs. I considered the design with the lowest sum of variances across the set of scenarios to be the best strategy in the face of uncertainty.

I also investigated optimal designs for cases when model assumptions were not met. In particular, unmodelled site-specific detection heterogeneity is a common problem in ecological estimation with the potential to severely bias results (Link 2003; Royle 2006; Miller *et al.* 2015). For example, differences among sites in habitat or animal density that affect detection can bias occupancy estimates unless properly modelled. Furthermore, distinguishing between detection heterogeneity and heterogeneity due to false positives may be impossible (Royle & Link 2006). To better understand how unmodelled site-specific detection heterogeneity would affect optimal design, I associated a standard normal covariate with survey sites and made detection parameters linear functions (on the logit scale) of the covariate. By setting the coefficient on this covariate to one, I generated a high level of heterogeneity in detection across sites. For example, with an average detection probability of 0·40, 90% of sites would have a detection probability between

0·11 and 0·78. With an average detection probability of 0·01, the same sites would have detection probabilities between 0·002 and 0·050. I introduced heterogeneity to one or multiple detection parameters and calculated the effect on optimal design.

For my analyses, I set the budget to 400 total surveys. For correctly specified models, project budget does not affect the number of repeat surveys because estimates are unbiased (MacKenzie & Royle 2005). In contrast, for misspecified models, the budget affects survey effort and therefore the relative size of bias and variance, and therefore, optimal design. The various scenarios I investigated are illustrative, not exhaustive. Therefore, I also provide Program R code to assist the interested reader in investigating additional scenarios in Appendices S1 and S2 (Supporting Information).

TWO-OBSERVATION-STATE MODELS

Both occupancy, ψ, and true detection probability, p_{11}, are known to affect the optimal number of repeat surveys (MacKenzie & Royle 2005). I expected that false-positive probability, p_{10}, and the probability that a true detection will be classified as certain, b, would also affect optimal survey design. Therefore, I identified the optimal survey design when $\psi = \{0{\cdot}2, 0{\cdot}5, 0{\cdot}8\}$, $p_{11} = \{0{\cdot}2, 0{\cdot}4, 0{\cdot}6, 0{\cdot}8\}$, $p_{10} = \{0{\cdot}01, 0{\cdot}05, 0{\cdot}10\}$ and $b = \{0{\cdot}0, 0{\cdot}2, 0{\cdot}5, 0{\cdot}8\}$, yielding 144 combinations of parameter values. Note that when $b = 0$, the two-observation-state model reduces to the Royle & Link (2006) model, while $b = 1$ corresponds to a standard occupancy model (MacKenzie *et al.* 2002).

To identify good hedging strategies in case it is not possible to accurately predict parameter values prior to surveys, I estimated model accuracy for several different scenarios. In particular, I set $\psi = 0{\cdot}5$, $p_{10} = 0{\cdot}05$, $b = 0{\cdot}2$ and estimated variance for $p_{11} = \{0{\cdot}2, 0{\cdot}4, 0{\cdot}6\}$. I then identified the design with the lowest sum of variances. I also investigated designs robust to simultaneous poor anticipation of the true values of p_{11}, p_{10} and b. To do this, I set $p_{11} = \{0{\cdot}2, 0{\cdot}4\}$, $p_{10} = \{0{\cdot}05, 0{\cdot}10\}$ and $b = \{0{\cdot}2, 0{\cdot}5\}$, yielding eight scenarios. Again, I summed variances across scenarios to identify robust designs.

I also investigated the effect of detection heterogeneity on optimal survey design. I introduced detection heterogeneity through a site covariate that affected (i) only p_{11}, (ii) only p_{10}, (iii) only b or (iv) all three detection parameters. I considered the same 144 combinations of parameter values considered previously and identified the optimal survey design for each scenario.

TWO-METHOD MODELS

For the two-method models, I considered how five sampling schemes were affected by occupancy, ψ, the probability of true-positive detection for the certain method, r_{11}, and the uncertain method, p_{11}, the false-positive probability for the uncertain method, p_{10}, and the number of surveys for the uncertain and certain methods, T and S. The first sampling scheme was a standard approach in which T and S were fixed across all sites. For parameters to be identifiable, T or S (or both) must be ≥ 2. I allowed the special cases of $T = 0$ (standard occupancy) and $S = 0$ (Royle–Link model). The other four schemes were variations on a removal design in which surveys ceased when certain criteria were met (Guillera-Arroita & Lahoz-Monfort in press). Under the removal schemes, the number of surveys could be fixed, so that they were equal at all surveyed sites, or varying so that surveys continued until a detection occurred, or a maximum survey number was reached. These two options yielded four schemes: (i) a fixed-fixed scheme in which T uncertain surveys were performed at all sites, and then S certain surveys were performed only at sites with uncertain detections; (ii) a fixed-varying scheme in which T uncertain surveys were performed at all sites, and then, only at sites with uncertain detections, certain surveys were performed, ceasing when a detection occurred, or the maximum of S surveys was reached; (iii) a varying-fixed scheme in which uncertain surveys were performed at all sites, ceasing when a detection occurred or the maximum of T surveys was reached, and then S certain surveys were performed only at sites with uncertain detections; and (iv) a varying-varying scheme in which uncertain surveys were performed at all sites, ceasing when a detection occurred or the maximum of T surveys was reached, and then, only at sites with uncertain detections, certain surveys were performed, ceasing when a detection occurred, or the maximum of S surveys was reached. For completeness, I also considered a design consisting of 0 uncertain surveys and a varying number of certain surveys performed at all sites. I report the results of this design as a special case of scheme (iv). By discontinuing surveys after a detection, the removal schemes conserved resources for other sample units, but estimation might be less accurate with fewer detections at each site. Therefore, I simulated surveys using all five sampling schemes and compared the accuracy of estimates from each. I identified the optimal survey design when $\psi = \{0{\cdot}2, 0{\cdot}5, 0{\cdot}8\}$, $p_{11} = \{0{\cdot}2, 0{\cdot}4, 0{\cdot}6, 0{\cdot}8\}$, $p_{10} = \{0{\cdot}01, 0{\cdot}05, 0{\cdot}10\}$ and $r_{11} = \{0{\cdot}2\}$, yielding 36 combinations of parameter values for each sampling scheme.

I also considered the relative cost of the two survey types, so that certain surveys might cost the same as, or five times more than, uncertain surveys.

To identify good hedging strategies in the likely case of pre-survey uncertainty about parameter values, I estimated estimator accuracy for several different scenarios. Specifically, I set $\psi = 0{\cdot}5$, $r_{11} = 0{\cdot}2$, $p_{10} = 0{\cdot}05$ and estimated variance for $p_{11} = \{0{\cdot}3, 0{\cdot}45, 0{\cdot}6\}$. I selected $p_{11} = 0{\cdot}45$ because the optimal survey design changed radically at this value, and the other values are equidistant from this value. I also investigated the effect of simultaneous uncertainty about the true value of both p_{11} and r_{11} on optimal design. I allowed p_{11} to vary as before and set $r_{11} = \{0{\cdot}1, 0{\cdot}2, 0{\cdot}3\}$, yielding nine scenarios. I identified robust designs by summing variances across scenarios.

I also investigated the effect of detection heterogeneity on optimal survey design for the standard and the varying-varying schemes with equal costs for certain and uncertain survey methods with a total budget of 400 surveys. I introduced detection heterogeneity through a site covariate that affected (i) only p_{11}, (ii) only p_{10}, (iii) only r_{11} or (iv) all three detection parameters. I considered the same 36 combinations of parameter values considered previously and identified the optimal survey design for each scenario.

Results

TWO-OBSERVATION-STATE MODELS

When $b = 0$ under the two-observation-state model (i.e. the Royle–Link model), the optimal number of repeat surveys at each sample unit increased with ψ and p_{10}, and decreased with increases in p_{11} (Table 1). As b increased, the optimal number of repeat surveys decreased towards the recommendations associated with a standard occupancy model (MacKenzie & Royle 2005).

Under pre-survey uncertainty about parameter values, an effective hedging strategy was to increase repeat surveys at the cost of fewer survey sites. For example, if $\psi = 0{\cdot}5$, $p_{10} = 0{\cdot}05$ and $b = 0{\cdot}2$, then the optimal number of repeat surveys per sample unit for $p_{11} = \{0{\cdot}2, 0{\cdot}4, 0{\cdot}6\}$ was 28, 10 and 5. If these

Table 1. Optimal number of surveys for the two-observation-state occupancy model for different parameter values. ψ is probability of occupancy, p_{11} is probability of true detection, p_{10} is probability of false detection, and b is the probability that a true detection will be classified as certain. Note that $b = 0$ is equivalent to the Royle–Link (2006) model

		$b = 0$ ψ			$b = 0{\cdot}2$ ψ			$b = 0{\cdot}5$ ψ			$b = 0{\cdot}8$ ψ		
p_{11}	p_{10}	0·2	0·5	0·8	0·2	0·5	0·8	0·2	0·5	0·8	0·2	0·5	0·8
0·2	0·01	17	20	25	15	18	23	12	15	20	9	11	16
	0·05	41	40	47	26	28	37	15	18	25	9	11	16
	0·10	118	109	126	35	42	58	15	18	26	9	11	16
0·4	0·01	8	9	11	7	8	10	6	7	9	4	5	7
	0·05	11	11	13	9	10	12	6	7	10	5	5	7
	0·10	18	17	20	12	13	16	7	8	11	5	5	7
0·6	0·01	5	5	6	5	5	6	4	4	5	3	3	4
	0·05	5	6	7	5	5	6	4	4	6	3	3	4
	0·10	8	8	9	6	6	8	4	5	6	3	3	4
0·8	0·01	3	4	4	3	3	4	3	3	4	2	2	3
	0·05	3	4	4	3	3	4	3	3	4	2	2	3
	0·10	5	4	4	3	4	4	3	3	4	2	2	3

Table 2. Optimal number of surveys, given a budget of 400 surveys, for the two-observation-state occupancy model, when there is heterogeneity in the true detection probability, p_{11}. ψ is probability of occupancy, p_{10} is probability of false detection, and b is the probability that a true detection will be classified as certain

		$b = 0$ ψ			$b = 0{\cdot}2$ ψ			$b = 0{\cdot}5$ ψ			$b = 0{\cdot}8$ ψ		
p_{11}	p_{10}	0·2	0·5	0·8	0·2	0·5	0·8	0·2	0·5	0·8	0·2	0·5	0·8
0·2	0·01	10	27	30	11	23	51	9	18	34	7	15	25
	0·05	19	32	42	13	29	58	10	19	35	7	15	25
	0·10	21	39	44	15	29	59	10	18	34	7	15	25
0·4	0·01	7	14	18	7	13	26	5	9	17	4	8	13
	0·05	8	17	21	7	15	30	5	10	18	4	8	13
	0·10	12	22	27	8	17	33	6	10	18	4	8	13
0·6	0·01	5	7	13	5	7	12	4	6	9	3	4	7
	0·05	4	8	13	4	8	16	4	6	10	3	4	7
	0·10	6	12	16	5	9	18	4	6	11	3	4	7
0·8	0·01	4	5	6	3	4	6	3	3	5	2	2	4
	0·05	3	4	7	3	4	7	3	3	5	2	2	4
	0·10	5	6	8	3	4	8	3	3	5	2	2	4

scenarios were equally likely, the optimal hedging strategy was to perform 21 repeat surveys, which was higher than the median (10) or mean (14·3) surveys for the three individual scenarios. When I considered uncertainty in all three detection parameters, so $p_{11} = \{0{\cdot}2,\ 0{\cdot}4\}$, $p_{10} = \{0{\cdot}05,\ 0{\cdot}10\}$ and $b = \{0{\cdot}2,\ 0{\cdot}5\}$, then the optimal numbers of surveys for the eight scenarios were 42, 28, 18, 18, 13, 10, 8 or 7 repeat surveys. If these scenarios were equally likely, the optimal hedging strategy was 22 surveys per sample unit, higher than the median (15·5) or mean (18) of the individual scenarios.

When I introduced heterogeneity in p_{11}, the optimal number of repeat surveys decreased when $\psi = 0{\cdot}2$ and increased when $\psi = 0{\cdot}8$ (Table 2). If $\psi = 0{\cdot}5$, the effect also depended on the value of p_{10} and b. Heterogeneity in other detection parameters could also increase or decrease the optimal number of repeat surveys, depending on the combination of parameter values (Tables S1–S3).

TWO-METHOD MODELS

Under a standard survey scheme (fixed number of uncertain and certain surveys), the optimal survey design consisted of either all certain surveys or all uncertain surveys, depending on parameter values (Table 3). If p_{11} was low and/or p_{10} was sufficiently high, the optimal design was to only use the certain method, with S equal to the recommendations in MacKenzie & Royle (2005). If p_{11} increased or p_{10} decreased, the optimal design switched to using only the uncertain method, with T equal to my recommendations for the Royle–Link model (Table 1). If only the certain method was recommended, the number of repeat surveys increased with ψ. If only the uncertain method was recommended, the number of surveys increased with ψ and p_{10}, and decreased with p_{11}.

Optimal survey designs were more diverse across the four removal schemes (Table 4). Under numerous scenarios, the

Table 3. Optimal number of surveys for two-method occupancy models, using the standard design. ψ is probability of occupancy, p_{11} is probability of detection with uncertain method, p_{10} is probability of false detection with uncertain method, and r_{11}, the probability of detection with certain method, is set to 0·2 in all calculations. In the table, the numbers before and after the dash indicate the number of uncertain and certain surveys, respectively

		ψ		
p_{11}	p_{10}	0·2	0·5	0·8
0·2	0·01	0–7	0–9	0–13
	0·05	0–7	0–9	0–13
	0·10	0–7	0–9	0–13
0·4	0·01	8–0	9–0	11–0
	0·05	0–7	0–9	0–13
	0·10	0–7	0–9	0–13
0·6	0·01	5–0	5–0	6–0
	0·05	5–0	6–0	7–0
	0·10	0–7	8–0	9–0
0·8	0·01	3–0	4–0	4–0
	0·05	3–0	4–0	4–0
	0·10	5–0	4–0	4–0

optimal survey design included a mix of both survey methods. Only the varying-varying scheme included survey designs restricted to the certain method because the other three schemes required at least one uncertain survey. Patterns were apparent among the survey designs, but violated by exceptions. For example, the optimal number of surveys was generally low for high values of p_{11}, but an exception can be noted under the varying-fixed scheme when $\psi = 0{\cdot}8$, $r_{11} = 0{\cdot}2$, $p_{11} = 0{\cdot}6$, $p_{10} = 0{\cdot}1$. In general, as ψ increased, accuracy could be improved by switching towards uncertain surveys, but an exception is apparent under the fixed-varying scheme when $r_{11} = 0{\cdot}2$, $p_{11} = 0{\cdot}6$, $p_{10} = 0{\cdot}1$. The most accurate of the five sampling schemes depended on the parameter values (Table 5). When $\psi = 0{\cdot}2$, low values of p_{11} were best addressed with a standard scheme, while high values of p_{11} recommended a fixed-varying scheme. When $\psi = 0{\cdot}5$, the varying-varying scheme produced the most accurate estimates of ψ, unless both p_{11} and p_{10} were high, and then the standard scheme was most accurate. If $\psi = 0{\cdot}8$, the varying-varying scheme was best, although sometimes recommendations were identical to the varying-fixed scheme.

Unsurprisingly, increasing the cost of certain surveys shifted the optimal survey design towards uncertain surveys (Table 6). A mixture of two survey types was never recommended. Similarly, under the different removal schemes, the recommended number of uncertain surveys increased, relative to the number of certain surveys (Table 7). The most accurate of the five

Table 5. The sampling scheme with the lowest standard error on ψ, when the costs of certain and uncertain surveys are equal. Schemes include S = standard, FF = fixed-fixed, FV = fixed-varying, VF = varying-fixed and VV = varying-varying, described in text. ψ is probability of occupancy, p_{11} is probability of detection with uncertain method, p_{10} is probability of false detection with uncertain method, and r_{11}, the probability of detection with the certain method, is set to 0·2 in all calculations

		ψ		
p_{11}	p_{10}	0·2	0·5	0·8
0·2	0·01	S	VV	VV
	0·05	S	VV	VV
	0·10	S	VV	VV
0·4	0·01	FV	VV	VV = VF
	0·05	FV	VV	VV
	0·10	S	VV	VV
0·6	0·01	FV	VV	VV = VF
	0·05	FV	VV	VV = VF
	0·10	FV	VV	VV
0·8	0·01	FV	VV = VF	VV = VF
	0·05	FV	S = FF = FV	VV = VF
	0·10	FV	S = FF = FV	VV = VF

Table 4. Optimal number of surveys for two-method occupancy models, using different removal schemes. ψ is probability of occupancy, p_{11} is probability of detection with uncertain method, p_{10} is probability of false detection with uncertain method, and r_{11}, the probability of detection with certain method, is set to 0·2 in all calculations. In the table, the numbers before and after the dash indicate the number of uncertain and certain surveys, respectively

		Fixed-fixed ψ			Fixed-varying ψ			Varying-fixed ψ			Varying-varying ψ		
p_{11}	p_{10}	0·2	0·5	0·8	0·2	0·5	0·8	0·2	0·5	0·8	0·2	0·5	0·8
0·2	0·01	8–11	12–8	18–7	8–17	10–19	14–24	12–13	17–11	37–4	0–11	0–13	0–19
	0·05	8–11	10–11	15–13	8–14	10–17	13–22	13–11	18–12	25–14	0–11	0–13	0–19
	0·10	8–11	10–12	13–15	8–14	9–16	12–22	15–10	20–11	26–14	0–11	0–13	0–19
0·4	0·01	5–7	9–0	11–0	4–16	5–17	7–21	6–11	9–7	19–0	5–18	6–21	19–0
	0·05	4–9	6–7	13–0	4–13	5–15	7–20	6–10	8–10	38–0	5–14	6–17	0–19
	0·10	4–9	5–9	8–10	4–13	5–15	6–20	6–10	8–10	12–12	0–11	0–13	0–19
0·6	0·01	3–6	5–0	6–0	3–14	5–0	6–0	4–9	8–0	10–0	3–19	4–21	10–0
	0·05	3–7	6–0	7–0	3–11	3–14	7–0	3–9	5–7	14–0	3–14	4–17	14–0
	0·10	3–8	8–0	9–0	3–12	3–14	4–19	3–9	5–9	68–0	3–13	4–16	0–19
0·8	0·01	3–1	4–0	4–0	2–14	4–0	4–0	2–8	5–0	6–0	2–19	5–0	6–0
	0·05	3–1	4–0	4–0	2–10	4–0	4–0	2–9	6–0	7–0	2–15	2–18	7–0
	0·10	3–2	4–0	4–0	2–9	4–0	4–0	2–9	6–1	10–0	2–13	2–16	10–0

Table 6. Optimal number of surveys for two-method occupancy models, using the standard design, assuming certain surveys cost five times uncertain surveys. ψ is probability of occupancy, p_{11} is probability of detection with uncertain method, p_{10} is probability of false detection with uncertain method, and r_{11}, the probability of detection with certain method, is set to 0·2 in all calculations. In the table, the numbers before and after the dash indicate the number of uncertain and certain surveys, respectively

		ψ		
p_{11}	p_{10}	0·2	0·5	0·8
0·2	0·01	17–0	20–0	25–0
	0·05	0–7	0–9	0–13
	0·10	0–7	0–9	0–13
0·4	0·01	8–0	9–0	11–0
	0·05	11–0	11–0	13–0
	0·10	18–0	17–0	20–0
0·6	0·01	5–0	5–0	6–0
	0·05	5–0	6–0	7–0
	0·10	8–0	8–0	9–0
0·8	0·01	3–0	4–0	4–0
	0·05	3–0	4–0	4–0
	0·10	5–0	4–0	4–0

sampling schemes depended on the parameter values (Table 8). When $\psi = 0{\cdot}2$ or 0·5, the best strategy was typically to use a fixed number of uncertain surveys, meaning that the standard, fixed-fixed and fixed-varying schemes all made the same recommendation, unless p_{11} was very low. If $\psi = 0{\cdot}8$, the best strategy was generally to use a varying number of uncertain surveys, so that the varying-fixed and varying-varying schemes made the same recommendation.

I also sought strategies that would be robust to errors in pre-survey expectations about parameter values. Under a standard design, if $\psi = 0{\cdot}5$, $r_{11} = 0{\cdot}2$ and $p_{10} = 0{\cdot}05$, then $p_{11} = 0{\cdot}45$ was noteworthy because at or below this value, the optimal strategy was to perform nine certain surveys, while above this value, the optimal strategy was to perform nine uncertain surveys (Fig. 1). Although survey design was very different, accuracy of ψ estimates was very similar. When p_{11} was lower than anticipated, at 0·3, relying on uncertain surveys greatly increased variance in ψ (Fig. 2a). When p_{11} was higher than anticipated, at 0·6, uncertain surveys outperformed certain surveys, but the difference was modest (Fig. 2b). As a result, if there was pre-survey uncertainty about the true value of p_{11}, the best hedging strategy was to assume a low value for p_{11}, which favoured certain surveys. Similarly, if there was uncertainty about the true value of r_{11}, the best hedging strategy was to assume a low value and favour uncertain surveys. When there was simultaneous uncertainty about both p_{11} and r_{11}, each individual scenario considered recommended using only one survey method or the other. However, when I summed variances across the nine scenarios, the best hedging strategy was to use seven uncertain and five certain surveys (Fig. 3).

I also identified the optimal design for the standard scheme when detection heterogeneity occurred in p_{11}. When p_{11} was high and heterogeneous, using uncertain surveys improved the variance, but worsened bias. By using a combination of both survey methods, it was possible to moderate bias and variance and improve accuracy. Accordingly, the optimal design frequently included a mix of both survey types (Table 9). Unexpectedly, there were additional cases where p_{11} was low, and yet the optimal design included both survey types, such as $\psi = 0{\cdot}5, p_{11} = 0{\cdot}2, p_{10} = 0{\cdot}01, r_{11} = 0{\cdot}2$. In these cases, p_{11} was overestimated and p_{10} was underestimated such that the estimate of ψ was relatively accurate. Heterogeneity in other detection parameters also affected the optimal survey designs under the standard scheme (Table S4) and the varying-varying scheme (Table S5).

Discussion

Standard occupancy models generate biased estimates of occupancy if false positives occur during surveys (McClintock *et al.* 2010; Miller *et al.* 2011). Accordingly,

Table 7. Optimal number of surveys for two-method occupancy models, using the removal design, assuming certain surveys cost five times uncertain surveys. ψ is probability of occupancy, p_{11} is probability of detection with uncertain method, p_{10} is probability of false detection with uncertain method, and r_{11}, the probability of detection with certain method, is set to 0·2 in all calculations. In the table, the numbers before and after the dash indicate the number of uncertain and certain surveys, respectively

		Fixed-fixed ψ			Fixed-varying ψ			Varying-fixed ψ			Varying-varying ψ		
p_{11}	p_{10}	0·2	0·5	0·8	0·2	0·5	0·8	0·2	0·5	0·8	0·2	0·5	0·8
0·2	0·01	17–0	20–0	25–0	9–11	20–0	25–0	13–9	36–1	50–0	12–14	15–17	50–0
	0·05	9–8	40–0	47–0	8–12	11–13	16–18	13–8	18–9	26–12	12–12	15–14	0–19
	0·10	9–8	12–9	17–13	9–12	11–14	15–19	15–8	19–9	25–13	0–11	0–13	0–19
0·4	0·01	8–0	9–0	11–0	8–0	9–0	11–0	7–7	15–0	19–0	6–15	15–0	19–0
	0·05	11–0	11–0	13–0	11–0	11–0	13–0	6–8	25–0	38–0	5–12	25–0	38–0
	0·10	18–0	17–0	20–0	18–0	17–0	20–0	6–8	8–8	83–0	5–12	7–14	9–19
0·6	0·01	5–0	5–0	6–0	5–0	5–0	6–0	8–0	8–0	10–0	3–14	8–0	10–0
	0·05	5–0	6–0	7–0	5–0	6–0	7–0	4–7	11–0	14–0	3–12	11–0	14–0
	0·10	8–0	8–0	9–0	8–0	8–0	9–0	3–7	17–0	68–0	3–12	17–0	68–0
0·8	0·01	3–0	4–0	4–0	3–0	4–0	4–0	5–0	5–0	6–0	5–0	5–0	6–0
	0·05	3–0	4–0	4–0	3–0	4–0	4–0	6–0	6–0	7–0	6–0	6–0	7–0
	0·10	5–0	4–0	4–0	5–0	4–0	4–0	5–1	7–0	10–0	2–12	7–0	10–0

Table 8. The sampling scheme with the lowest standard error on ψ, when the cost of certain surveys is five times uncertain surveys. Schemes include S = standard, FF = fixed-fixed, FV = fixed-varying, VF = varying-fixed and VV = varying-varying, described in text. ψ is probability of occupancy, p_{11} is probability of detection with uncertain method, p_{10} is probability of false detection with uncertain method, and r_{11}, the probability of detection with certain method, is set to 0·2 in all calculations

		ψ		
p_{11}	p_{10}	0·2	0·5	0·8
0·2	0·01	FV	S = FF = FV	VV = VF
	0·05	FV	VV	VV
	0·10	S	VV	VV
0·4	0·01	S = FF = FV	S = FF = FV	VV = VF
	0·05	S = FF = FV	S = FF = FV	VV = VF
	0·10	S = FF = FV	S = FF = FV	S = FF = FV
0·6	0·01	S = FF = FV	S = FF = FV	VV = VF
	0·05	S = FF = FV	S = FF = FV	VV = VF
	0·10	S = FF = FV	S = FF = FV	VV = VF
0·8	0·01	S = FF = FV	VV = VF	VV = VF
	0·05	S = FF = FV	S = FF = FV	VV = VF
	0·10	S = FF = FV	S = FF = FV	VV = VF

recommendations for designing occupancy surveys have emphasized the need to adopt protocols intended to eliminate false positives (MacKenzie *et al.* 2006). Adopting strict criteria for detections can reduce false positives, but also lower detection probabilities, reducing accuracy of occupancy estimates (Miller *et al.* 2011). However, in some cases it may be difficult to significantly reduce false-positive probabilities through stricter protocols alone (Miller *et al.* 2012). Alternatively, more reliable but expensive survey methods may be adopted, reducing the number of samples that can be collected and therefore the precision of occupancy estimates. The advent of models that can estimate occupancy in the presence of false positives (Royle & Link 2006; Miller *et al.* 2011; Chambert, Miller & Nichols 2015) enables surveyors to consider additional survey methods that previously would have been rejected. However, the availability of more choices also raises new questions about survey design.

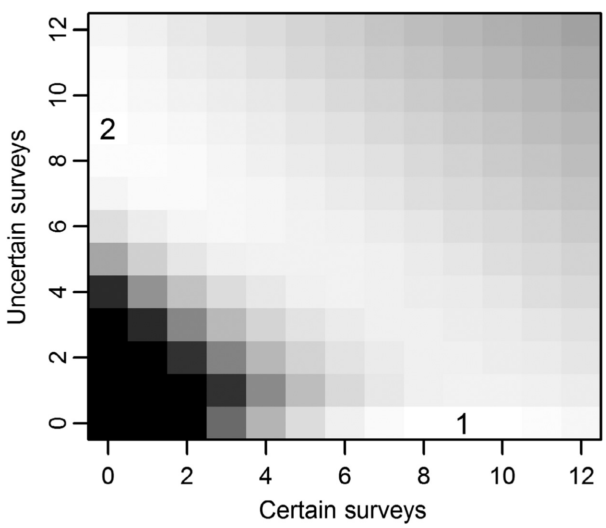

Fig. 1. Variance of estimates of ψ under several survey designs when ψ = 0·5, r_{11} = 0·2, p_{11} = 0·45, p_{10} = 0·05. Darker colours indicate higher variance. Square 1 is the optimal design. Square 2 is optimal when p_{11} = 0·46.

TWO-OBSERVATION-STATE MODELS

The optimal design for two-observation-state models requires more repeat surveys than a standard occupancy model in which the assumption of no false-positive detections is met. At the same time, a well-designed two-observation-state study requires fewer repeat surveys than a Royle–Link false-positive model, demonstrating that the ability to classify a subset of detections as certain can improve efficiency when false positives occur. If available survey methods allow one to achieve a high (>0·6) value for p_{11} or b, the gains in efficiency over the Royle–Link approach are large, while avoiding the bias in a standard occupancy model when false positives occur. However, if both p_{11} and b are low, accurate estimates of ψ require extensive surveys.

I expect that in most cases, it will be difficult to anticipate detection probabilities prior to conducting surveys. If p_{11} or b is high, survey design is relatively robust to such errors in prognostication. However, if p_{11} and b are low, poor prognostication can lead to poor survey design. In this case, I found that pessimism about detection probabilities, and therefore use of additional repeat surveys, helped avoid poor outcomes.

If the data include unmodelled detection heterogeneity, additional repeat surveys reduce bias, but also worsen precision (due to fewer sites being surveyed), so that the best survey strategy is not immediately obvious. I found that heterogeneity in p_{11} caused occupied sites with low p_{11} to generate detection histories that mimic false positives. This problem was greater when ψ was high, so additional repeat surveys improved accuracy, while fewer repeat surveys were appropriate when ψ was low. Of course, the preferable strategy would be to identify and model sources of heterogeneity. For example, if detection probability varies across sites with vegetation density, using vegetation metrics as model covariates will reduce heterogeneity and bias. Accordingly, optimal design will revert from the values in Table 2 to the values in Table 1.

TWO-METHOD MODELS

The two-method model was introduced to improve occupancy estimation when false positives occur. However, if all model assumptions were met and parameter values were correctly anticipated, then under a standard scheme that uses a fixed number of repeat surveys at each site, I found that it was always best to use a single survey method. Thus, optimal design should follow the Chesterfield Principle: 'Whatever is worth doing at all, is worth doing well', (Chesterfield 1774). Stated simply, dividing effort between two methods yielded poorer estimates of detection probability for both survey types, resulting in less precise occupancy estimates. As expected, high values for p_{11} and low values for p_{10} favoured the uncertain

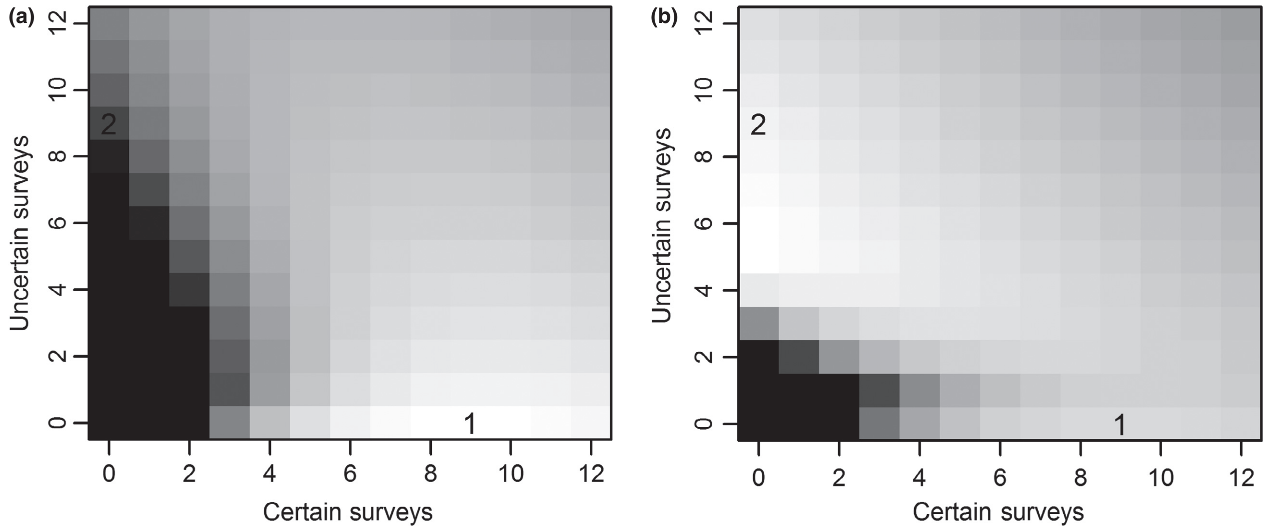

Fig. 2. Variance of estimates of ψ under several survey designs when ψ = 0·5, r_{11} = 0·2, p_{10} = 0·05. Darker colours indicate higher variance. Square 1 is the optimal design when p_{11} = 0·45 and square 2 is optimal when p_{11} = 0·46. (a) Scenario with p_{11} = 0·3. (b) Scenario with p_{11} = 0·6.

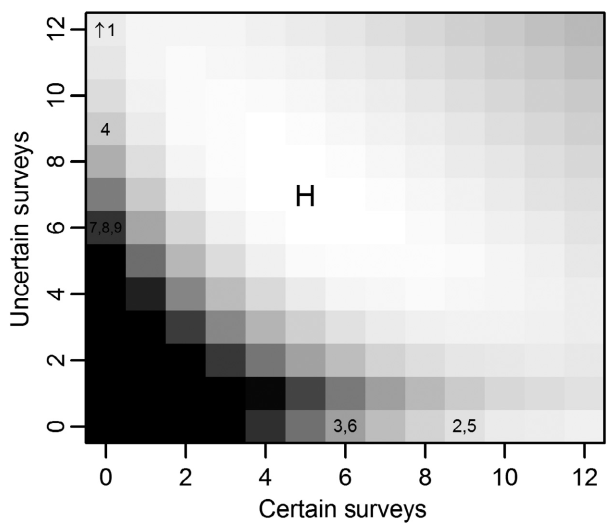

Fig. 3. Sum of variance in estimates of ψ under several survey designs when ψ = 0·5, p_{10} = 0·05, p_{11} = 0·3, 0·45 or 0·6, and r_{11} = 0·1, 0·2 or 0·3. Darker colours indicate higher variance. Numbers indicate optimal designs for individual scenarios, while the large *H* indicates the best hedging strategy.

Table 9. Optimal number of surveys for two-method occupancy model, given a budget of 400 surveys, using the standard design, with heterogeneity in p_{11}. ψ is probability of occupancy, p_{11} is probability of detection with uncertain method, p_{10} is probability of false detection with uncertain method, and r_{11}, the probability of detection with certain method, is set to 0·2 in all calculations. In the table, the numbers before and after the dash indicate the number of uncertain and certain surveys, respectively

		ψ		
p_{11}	p_{10}	0·2	0·5	0·8
0·2	0·01	0–7	3–3	4–7
	0·05	0–7	0–9	0–13
	0·10	0–7	0–9	0–13
0·4	0·01	2–1	2–2	3–6
	0·05	0–7	2–1	2–4
	0·10	0–7	0–9	2–2
0·6	0·01	2–1	2–1	2–4
	0·05	4–1	2–1	2–2
	0·10	5–2	4–4	2–1
0·8	0·01	3–0	2–1	2–1
	0·05	3–0	4–1	2–1
	0·10	3–1	3–2	2–1

method, and my analysis provides some guidance about what constitutes 'high' and 'low' (Table 1). However, it might be difficult to correctly anticipate parameter values, and in this case, using both survey methods can be a good hedging strategy. As detection parameters change, estimator precision responds nonlinearly, so misplaced optimism about detection probabilities may be particularly costly (Fig. 2). In this case, putting similar effort into each survey method will not produce the most precise estimate of occupancy (which will be found in one of the numbered blocks in Fig. 3), but it provides good insurance against a highly imprecise estimate. Accordingly, in a study of sufficient duration, a potential strategy would be to use multiple survey methods initially, and then to phase out all but one method after learning about the relative performance of each survey method (Guillera-Arroita, Ridout & Morgan 2014).

For a wide variety of parameter values, the varying-varying scheme (uncertain surveys stop after a detection, followed by certain surveys that stop after a detection) generated the most precise estimates of occupancy, similar to the results of MacKenzie & Royle (2005). However, there are at least two reasons to prefer a standard scheme (a fixed number of uncertain and certain surveys). In a varying scheme, surveys stop after a single detection, so detection histories (for each method) will contain at most one detection. If detection probabilities are constant through time and space, as in my simulations, these detection histories are sufficient to obtain unbiased estimates of occupancy. However, if detection probabilities vary, then space, time and mixture parameters become confounded. A standard scheme, where multiple detections can be observed at a single site, is better suited to disentangling these processes. Given that detection often varies by time or location, researchers may benefit from selecting a standard scheme due to the additional protection against detection heterogeneity and bias (Guillera-Arroita & Lahoz-Monfort in press). Secondly, the varying schemes imply that uncertain surveys occur before certain surveys, but this may not always be desirable. Sequential surveys are appealing because more expensive certain surveys can be limited to promising sites identified by the uncertain survey. However, if detection probabilities vary through time, it might be more appropriate to intersperse the two survey methods. And in some cases, it may be cost-effective to perform surveys simultaneously, such as simultaneous visual and aural surveys. I also note that if model assumptions are met, using only one survey method is more efficient than a two-phase survey (Table 3).

When the data included unmodelled heterogeneity in p_{11}, estimates of ψ could be improved by using both survey methods. Often this approach worked because the uncertain method improved precision, while the certain method reduced bias. However, a review of the results indicated that some of the 'optimal' designs achieved a low variance in ψ by misestimating all of the detection parameters, with p_{10} typically near 0. While these designs produced accurate estimates of ψ, I recommend caution before using this strategy. It would be preferable to select or modify survey methods to reduce heterogeneity to the extent possible or collect information on covariates that can be used to explain heterogeneity in the model. Nonetheless, my results suggest that a combination of survey methods may improve occupancy estimates if detection heterogeneity persists.

Because false-positive occupancy models are relatively new, little guidance exists for study design (Clement *et al.* 2014). Miller *et al.* (2015) compared a false-positive occupancy analysis of frog surveys using an uncertain method against an analysis that incorporated additional data from a certain-detection survey and found the second approach yielded a lower mean absolute error. Their work highlights how making use of additional available survey data and how data with certain detections can improve estimate accuracy. However, when allocating new survey effort, I found that if all model assumptions are met under a standard scheme, then just one survey method should be selected. But if model assumptions are not met, then multiple survey methods could be warranted.

I also note that the two-method approach does not fully resolve the ambiguity present in the Royle–Link model. For example, we might interpret a set of parameters, ψ, r_{11}, p_{11} and p_{10} as indicating that detection probabilities for the two methods are r_{11} and p_{11} at ψ occupied sites, with false-positive probabilities of 0 and p_{10} at $1 - \psi$ unoccupied sites. However, we could equally interpret this result as true-positive detection heterogeneity so that all sites are occupied, and ψ is a mixture parameter indicating the share of sites with high detection probability (r_{11} and p_{11}), while 1-ψ indicates the share of sites with low detection probability (0 and p_{10}). To distinguish between these interpretations, we must resort to *a priori* plausibility, rather than mathematical certainty, as with the Royle–Link approach. Furthermore, I note that there is no particular requirement that the second method generates no false positives. Instead, we can allow $r_{10} > 0$ and estimate this parameter, just as we do p_{10}. Such an approach may raise further ambiguity about how to interpret results, but it may be justified. For example, acoustic surveys for bats may be plagued by false-positive detections (Clement *et al.* 2014), but capture surveys can also yield misidentifications (Weller *et al.* 2007). Given that low false-positive probabilities can bias occupancy estimates, estimating false-positive probabilities for both methods might be appropriate in some circumstances.

My results are based on asymptotic approximations to variances, which hold when sample sizes are large. Smaller sample sizes will yield greater variances and possibly bias as well. By providing a lower bound on variances, my results highlight general issues of effort allocation, and the results may be accurate for those with large sample sizes. However, ecological studies often rely on small sample sizes, so a thorough approach to study design would include simulation studies to obtain variance estimates corresponding to the planned survey effort (Bailey *et al.* 2007; Guillera-Arroita, Ridout & Morgan 2010). One simulation study of standard occupancy models indicated that when sample sizes were small, the optimal number of repeat surveys was greater than when sample sizes were large (Guillera-Arroita, Ridout & Morgan 2010).

Additional complications could be considered. For example, the assumption that detections classified as certain are never false positives might be violated (Ferguson, Conroy & Hepinstall-Cymerman 2015). One could investigate the effect of this violation on optimal design or the effect of more conservative or liberal protocols for classifying detections as certain on the accuracy of model estimates (Miller *et al.* 2015). Alternatively, more cost structures could be considered. For example, the initial survey of a site might be more expensive than subsequent surveys. In this case, increasing repeat surveys would presumably improve accuracy (MacKenzie & Royle 2005). Similarly, the marginal cost of using a second survey method might be negligible, as when conducting simultaneous visual and aural surveys. The number of questions I investigated was necessarily limited, but the R functions in Appendices S1 and S2 may be useful in exploring additional design questions.

Acknowledgements

I thank Dave Miller and Jim Nichols for sharing their deep knowledge of occupancy models and two reviewers for providing constructive comments on the manuscript. Any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Data accessibility

Results are based on calculations, not data. Calculations can be recreated with computer code in Appendices S1 and S2.

References

Alldredge, M.W., Pacifici, K., Simons, T.R. & Pollock, K.H. (2008) A novel field evaluation of the effectiveness of distance and independent observer sampling to estimate aural avian detection probabilities. *Journal of Applied Ecology*, **45**, 1349–1356.

Bailey, L.L., MacKenzie, D.I. & Nichols, J.D. (2014) Advances and applications of occupancy models. *Methods in Ecology and Evolution*, **5**, 1269–1279.

Bailey, L.L., Hines, J.E., Nichols, J.D. & MacKenzie, D.I. (2007) Sampling design trade-offs in occupancy studies with imperfect detection: examples and software. *Ecological Applications*, **17**, 281–290.

Chambert, T., Miller, D.A.W. & Nichols, J.D. (2015) Modeling false positive detections in species occurrence data under different study designs. *Ecology*, **96**, 332–339.

Chesterfield, P.D.S. (1774) *Letters to His Son*. J. Dodsley, London, UK.

Clement, M.J., Rodhouse, T.J., Ormsbee, P.C., Szewczak, J.M. & Nichols, J.D. (2014) Accounting for false-positive acoustic detections of bats using occupancy models. *Journal of Applied Ecology*, **51**, 1460–1467.

Ferguson, P.F.B., Conroy, M.J. & Hepinstall-Cymerman, J. (2015) Occupancy models for data with false positive and false negative errors and heterogeneity across sites and surveys. *Methods in Ecology and Evolution*, **6**, 1395–1406.

Gu, W. & Swihart, R.K. (2004) Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation*, **116**, 195–203.

Guillera-Arroita, G. & Lahoz-Monfort, J.J. (2012) Designing studies to detect differences in species occupancy: power analysis under imperfect detection. *Methods in Ecology and Evolution*, **3**, 860–869.

Guillera-Arroita, G. & Lahoz-Monfort, J.J. (in press) Species occupancy estimation and imperfect detection: shall surveys continue after the first detection. *Advances in Statistical Analysis*.

Guillera-Arroita, G., Ridout, M.S. & Morgan, B.J. (2010) Design of occupancy studies with imperfect detection. *Methods in Ecology and Evolution*, **1**, 131–139.

Guillera-Arroita, G., Ridout, M.S. & Morgan, B.J. (2014) Two-stage Bayesian study design for species occupancy estimation. *Journal of Agricultural, Biological, and Environmental Statistics*, **19**, 278–291.

Link, W.A. (2003) Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics*, **59**, 1123–1130.

MacKenzie, D.I. & Royle, J.A. (2005) Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*, **42**, 1105–1114.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Andrew Royle, J. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.

MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006) *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Academic Press Elsevier, London, UK.

McClintock, B.T., Bailey, L.L., Pollock, K.H. & Simons, T.R. (2010) Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. *Ecology*, **91**, 2446–2454.

Miller, D.A., Nichols, J.D., McClintock, B.T., Grant, E.H.C., Bailey, L.L. & Weir, L.A. (2011) Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification. *Ecology*, **92**, 1422–1428.

Miller, D.A.W., Weir, L.A., McClintock, B.T., Grant, E.H.C., Bailey, L.L. & Simons, T.R. (2012) Experimental investigation of false positive errors in auditory species occurrence surveys. *Ecological Applications*, **22**, 1665–1674.

Miller, D.A.W., Bailey, L.L., Grant, E.H.C., McClintock, B.T., Weir, L.A. & Simons, T.R. (2015) Performance of species occurrence estimators when basic assumptions are not met: a test using field data where true occupancy status is known. *Methods in Ecology and Evolution*, **6**, 557–565.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: http://www.R-project.org/.

Royle, J.A. (2006) Site occupancy models with heterogeneous detection probabilities. *Biometrics*, **62**, 97–102.

Royle, J.A. & Link, W.A. (2006) Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, **87**, 835–841.

Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K. & Possingham, H.P. (2003) Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications*, **13**, 1790–1801.

Weller, T.J., Scott, S.A., Rodhouse, T.J., Ormsbee, P.C. & Zinck, J.M. (2007) Field identification of the cryptic vespertilionid bats, *Myotis lucifugus* and *M. yumanensis*. *Acta Chiropterologica*, **9**, 133–147.

Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of Animal Populations*. Academic Press Elsevier, San Diego, CA, USA.

Received 22 January 2016; accepted 5 July 2016
Handling Editor: Nigel Yoccoz

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. R functions used by R code in Appendix S2.

Appendix S2. R code to recreate analysis presented in this paper.

Table S1. Optimal number of surveys for the two-observation-state method for different parameter values, when there is heterogeneity in the false detection probability, p_{10}.

Table S2. Optimal number of surveys for the two-observation-state method for different parameter values, when there is heterogeneity in the certainty around true detections, b.

Table S3. Optimal number of surveys for the two-observation-state method for different parameter values, when there is heterogeneity in all detection probabilities.

Table S4. Optimal number of surveys, using two-method models and a standard survey scheme, with heterogeneity in different detection probabilities.

Table S5. Optimal number of surveys, using two-method models and a varying-varying survey scheme, with heterogeneity in different detection probabilities.