Received: 6 February 2018 | Accepted: 20 August 2018
DOI: 10.1111/2041-210X.13082

RESEARCH ARTICLE

Methods in Ecology and Evolution — British Ecological Society

Model selection with overdispersed distance sampling data

Eric J. Howe[1,2] | Stephen T. Buckland[1] | Marie-Lyne Després-Einspenner[3] | Hjalmar S. Kühl[3,4]

[1]Centre for Research into Ecological and Environmental Modelling, University of St Andrews, The Observatory, Fife, UK

[2]Wildlife Research and Monitoring Section, Ontario Ministry of Natural Resources and Forestry, DNA Building, Trent University, Peterborough, Ontario, Canada

[3]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

[4]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

Correspondence
Eric J. Howe
Email: ejh20@st-andrews.ac.uk

Handling Editor: Jason Matthiopoulos

Abstract

1. Distance sampling (DS) is a widely used framework for estimating animal abundance. DS models assume that observations of distances to animals are independent. Non-independent observations introduce overdispersion, causing model selection criteria such as AIC or AIC_c to favour overly complex models, with adverse effects on accuracy and precision.
2. We describe, and evaluate via simulation and with real data, estimators of an overdispersion factor ($\hat{c}$), and associated adjusted model selection criteria (QAIC) for use with overdispersed DS data. In other contexts, a single value of $\hat{c}$ is calculated from the "global" model, that is the most highly parameterised model in the candidate set, and used to calculate QAIC for all models in the set; the resulting QAIC values, and associated ΔQAIC values and QAIC weights, are comparable across the entire set. Candidate models of the DS detection function include models with different general forms (e.g. half-normal, hazard rate, uniform), so it may not be possible to identify a single global model. We therefore propose a two-step model selection procedure by which QAIC is used to select among models with the same general form, and then a goodness-of-fit statistic is used to select among models with different forms. A drawback of this approach is that QAIC values are not comparable across all models in the candidate set.
3. Relative to AIC, QAIC and the two-step model selection procedure avoided overfitting and improved the accuracy and precision of densities estimated from simulated data. When applied to six real datasets, adjusted criteria and procedures selected either the same model as AIC or a model that yielded a more accurate density estimate in five cases, and a model that yielded a less accurate estimate in one case.
4. Many DS surveys yield overdispersed data, including cue counting surveys of songbirds and cetaceans, surveys of social species including primates, and camera-trapping surveys. Methods that adjust for overdispersion during the model selection stage of DS analyses therefore address a conspicuous gap in the DS analytical framework as applied to species of conservation concern.

KEYWORDS
animal abundance, camera trapping, cue counting, distance sampling, model selection, overdispersion, QAIC

© 2018 The Authors. Methods in Ecology and Evolution © 2018 British Ecological Society

1 | INTRODUCTION

Distance sampling (DS) is an established framework for estimating animal abundance (Borchers, Buckland, & Zucchini, 2002; Buckland et al., 2001, 2004). It allows for imperfect detection by assuming detection probability is a function of the distance between objects (e.g. animals or their sign), and observers. Careful modelling of this function is required to obtain accurate abundance estimates (Buckland et al., 2001, 2004). Exploratory analyses, goodness-of-fit (GOF) testing, and model selection are therefore critical components of DS analyses (Buckland et al., 2001, 2004; Marques, Thomas, Fancy, & Buckland, 2007). GOF tests evaluate the null hypothesis that a model adequately fits the data; tests for continuous and binned DS data were described by Buckland et al. (2001). Rejection may indicate problems in the data or the structure of the model being tested, or violations of model assumptions. The purpose of model selection is the identification of a model or models that optimise the trade-off between bias and precision of the parameters estimated from a dataset, where the inclusion of more parameters reduces both bias and precision (Burnham & Anderson, 2002; Johnson & Omland, 2004). The remainder of this paper assumes readers are familiar with both distance sampling, and information-theoretic model selection, as described by Buckland et al. (2001, 2004), and Burnham and Anderson (2002).

Distance sampling methods assume that observations are independent (Buckland et al., 2001), but some DS surveys violate this assumption. For example, some animals travel in groups. Violation of the independence assumption can be avoided by treating the group as the unit of observation, measuring or estimating distances to the centre of detected groups, and estimating animal density as the product of group density and mean group size (Buckland et al., 2001). However, this is only effective if the size and central location of the group are measured accurately (Buckland, Plumptre, Thomas, & Rexstad, 2010; Buckland et al., 2001). When they cannot be, for example, because groups are widely spread or in motion, the recourse is to treat the individual as the unit of observation, and to record distances to all group members detected, in which case the data include non-independent observations. Furthermore, some animals, such as cetaceans that are often submerged, or songbirds that perch concealed in trees, are only available to be observed intermittently. However, if they give discrete cues of their presence and location, such as whale blows or bursts of birdsong, density of cues can be estimated using DS methods, and converted to estimates of animal density by dividing by the cue production rate (Buckland, 2006; Buckland et al., 2001). During cue counting surveys, distances to all cues are recorded, so the data may include observations of distances to multiple cues given by the same animal(s), which again violates the independence assumption (Buckland, 2006). Finally, Howe, Buckland, Després-Einspenner, and Kühl (2017) extended DS methods to accommodate data from camera traps (CTs). Distances to animals when first detected by CTs are expected to be positively biased, so authors recommended programming cameras to record video, or multiple still images, each time the sensor is triggered, and measuring distances to each detected animal multiple times at predetermined "snapshot moments" during an independent encounter with a CT. Authors acknowledged that these observations would not be independent of each other. Violations of the independence assumption do not bias point estimates of model parameters, but introduce overdispersion (Buckland et al., 2001).

When distance data are overdispersed: (a) GOF tests, and likelihood ratio tests (LRTs) to compare the fit of nested models, are invalid and prone to reject the null hypotheses that a model adequately fits the data (GOF tests), or that the simpler of two models provides a better fit than a more complex one (LRTs); (b) model-based analytic variances underestimate the actual uncertainty associated with the estimates, though empirical design-based estimators are robust (Fewster et al., 2009), and bootstrap estimators that resample points or transects are unaffected (Buckland, 1984); (c) model selection criteria that have not been adjusted for overdispersion favour overly complex models with more than the optimal number of parameters (Buckland et al., 2001, 2010; Burnham & Anderson, 2002; Cox & Snell, 1989). Akaike's Information Criterion (AIC; Akaike, 1973) is usually recommended for selecting among candidate models of the detection function (Buckland et al., 2004; Marques et al., 2007), however, if the data are overdispersed, AIC is likely to favour unnecessarily complex models (Buckland, 2006; Buckland et al., 2001, 2010). This additional complexity reduces precision, and can cause bias if it affects the slope of the detection function near the line or the point. Criteria adjusted to account for overdispersion have not been developed previously.

Detectability may vary in response to multiple factors other than distance. DS methods are pooling robust, so the total or average density estimated from the entire dataset will generally be unbiased even when variation in detectability is ignored (in the case of differences between distinct spatial subsets of the greater study area, sampling effort should be proportional to the areas of the subsets; Buckland et al., 2004). However, density estimates specific to different population strata among which detectability varies, which might be different species, treatments, habitat types, time periods, etc., are expected to be biased if estimated from a common detection function (Buckland et al., 2004; Marques et al., 2007). Observations within different strata can be analysed separately to avoid this bias, but this can reduce sample sizes to the point where densities of some strata may not be estimable, or estimates may be too imprecise to be useful. The multiple covariate approach to DS analysis improves efficiency by modelling variation in detectability using covariates (Buckland et al., 2004; Marques et al., 2007). It also casts decisions about how much stratification is necessary as a model selection problem, but in this case, the quality of inferences about strata-specific densities is affected by the reliability of the model selection criterion. When the independence assumption is suspected or known to have been violated, it has been recommended that analysts constrain the complexity of the detection function and the number of covariates to avoid overfitting (Buckland et al., 2004, 2010; Marques et al., 2007). However, limiting the candidate set to simple models may not be desirable if there are multiple potential covariates of the detection function. Model selection criteria unadjusted for overdispersion will tend to select models that subdivide the data more than necessary, with adverse effects on precision. Conversely, "underfitting", that is, failure to include significant sources of variation in the estimating model, would cause stratum-specific densities to be underestimated if true detection probabilities in that stratum tend to be lower, and vice versa. Adjusted criteria could underfit if they overcompensated for overdispersion (e.g. if the magnitude of overdispersion was overestimated).

Although explicitly modeling the sources of overdispersion would be ideal, this is not always possible or practical with real data (Burnham & Anderson, 2002; Cox & Snell, 1989; Lebreton, Burnham, Clobert, & Anderson, 1992). An approximation that is often sufficient in practice is to estimate a single, omnibus overdispersion factor ($\hat{c}$) from a χ^2 GOF test of the global model (i.e. the most highly parameterised or most general model) divided by its degrees of freedom (*df*), and to include $\hat{c}$ in the calculation of information criteria adjusted for overdispersion for all models in the candidate set (Burnham & Anderson, 2001, 2002; Cox & Snell, 1989; Lebreton et al., 1992; Liang & McCullah, 1993). The adjusted version of AIC (QAIC) is:

$$\text{QAIC} = -2\left\{\frac{\log L(\hat{\theta})}{\hat{c}}\right\} + 2K,$$

where log L is the log likelihood value, θ is a vector of maximum likelihood parameter estimates, and K is the number of parameters in the current model (Lebreton et al., 1992). Burnham and Anderson (2001, 2002) clarified that $\hat{c}$ should be included as one of the K parameters.

Given an estimator of c ($\hat{c}$), the same approach could be used to calculate QAIC for models of the DS detection function. However, candidate sets usually include models with different general forms (termed "key functions"; for example, half-normal, hazard rate, and uniform; Buckland et al., 2001) as well as different numbers of adjustment terms and covariate combinations (Buckland et al., 2004; Marques et al., 2007). Models with different key functions are not nested, hence it may not always be straightforward to identify a single "global" model from which to estimate $\hat{c}$. Below we propose and evaluate two estimators of $\hat{c}$, and a two-step model selection procedure that does not require that a single global model is identifiable, for use with overdispersed DS data.

2 | MATERIALS AND METHODS

2.1 | Model selection criteria and procedures

We suggest the χ^2 GOF statistic for binned distance data (Buckland et al., 2001, p. 71, eqn. 3.57) divided by its degrees of freedom as one estimator of c ($\hat{c}_1$). To allow for the possibility that multiple models may include the maximum number of parameters, and the fact that DS models have different general forms, we propose the following two-step model selection procedure. In step one we use QAIC to identify the best-supported model within each key function, and in step two we compare the GOF of the best-supported models with different key functions. More specifically, in step one, we obtain $\hat{c}_1$ from the most highly parameterised model within each key function (rather than from the most highly parameterised model overall), use those values of $\hat{c}_1$ to calculate QAIC for all models with the same key function, and use QAIC to identify the best-supported model within each key function. In this step, the same value of $\hat{c}_1$ is used to calculate QAIC for all models with the same key function, but different values of $\hat{c}_1$ are used to calculate QAIC for different key functions. In step two, we compare values of the χ^2 GOF statistic divided by its *df* across QAIC-selected models (one from each key function), and choose the model with the smallest value for estimation. If continuous distances are recorded in the field, distance observations will first need to be grouped into categories so that the GOF test for binned data can be performed. See Buckland et al. (2001) for advice regarding binning continuous observations.

The number of distance observations recorded per independent encounter between an animal and an observer provides an alternative measure of the magnitude of overdispersion ($\hat{c}_2$). $\hat{c}_2$ will often be calculable from the raw data, and will be the same for all models in the candidate set. In CT surveys of solitary animals, $\hat{c}_2$ would be the mean number of distance observations recorded during a single pass by an animal in front of a CT. In surveys of social animals employing human observers, $\hat{c}_2$ would be the mean number of detected animals per detected group, and in CT surveys of social animals $\hat{c}_2$ would be the mean number of distance observations recorded during an encounter between a group of animals and a CT. $\hat{c}_2$ could be used instead of multiple values of $\hat{c}_1$ to calculate QAIC values as in step one above. QAIC values would still be compared only within key functions, and the χ^2 GOF statistic divided by its *df* would still be used in step two to select among QAIC-selected models with different key functions. Hereafter, we will refer to QAIC calculated from $\hat{c}_1$ as QAIC_1, and from $\hat{c}_2$ as QAIC_2.

2.2 | Simulations

We conducted simulations where non-independent observations were all at the same distance so that we could evaluate performance where the true magnitude of overdispersion (c), and the true underlying model were known, but we would not expect this scenario to arise in practice. When non-independent observations during a single independent encounter are at different distances (e.g. to different members of a group, different cues from a moving animal, or as an animal moves past a CT), true c is unknown because the different distance observations contribute information about the shape of the detection function. We therefore also simulated camera-trapping (CT) surveys of moving animals where cameras recorded video and distance was recorded every 2 s as animals moved through the field of view. These simulations mimic real surveys where animals move and c is unknown. Furthermore, the distribution of observed distances differed from the expected distribution of independent detections (see Supplemental Material), so the true underlying model was also unknown.

For the simulations with known c, we sampled distances to animals within a circular point transect with radius 20 m, where the true density was 2.00/m^2. To generate independent DS data, we simulated detections via random trials where detection probability declined according to a half-normal function with scale parameter (σ) of 7. Each observation was arbitrarily assigned one of three levels of a spurious categorical covariate that had no effect on detectability, which we will refer to as "observer". We then replicated each observation five times to generate overdispersed data with $c = 6$. We fitted eight point transect DS models to each dataset, including the half normal model used to generate the data, and overparameterised models.

For the CT surveys, we simulated sampling of ungulates inhabiting old growth forests, recently logged forests, and previously logged but regrowing forests. Simulation parameters were based on Howe et al.'s (2017) survey of Maxwell's duikers, but were also selected to ensure that data were overdispersed, not sparse, and included multiple potential covariates of detectability. We assumed that the density of understorey vegetation increased immediately after logging and decreased gradually as forests regrew, such that food supply and therefore animal density was highest, but detection probability as a function of distance was lowest, in recently logged forests; we further assumed a larger difference in detection probability between old growth and logged forests than between recently logged and regrowing forests (Table 1).

We simulated movements of 10, 12 and 15 animals within 1 km^2 study areas in old growth, regrowing, and recently logged habitats respectively. Each animal started with a random initial location and heading, after which new locations were generated every 2 s for 12 hr. Step lengths were drawn from an exponential distribution with a rate parameter of 2, and turn angles were drawn from a normal distribution with *M* of 0 and *SD* of 0.05 radians. Animals that moved beyond the boundaries of the study areas reappeared on the opposite side of the same study area at the same heading. We simulated sampling at a grid of 36 CTs at 150 m spacing within each study area. We defined the zone of potential detection by a CT as a sector with a central angle of 0.733 radians and a radius of 25 m, and recorded distances between CTs and animal locations that fell within these sectors. We initially conducted random trials according to a half-normal function with σ as in Table 1 to determine whether animals were detected at each time step. However, we assumed that cameras were programmed to record video when triggered, so once an animal was detected we set the probability of subsequent detection to 1.0 for as long as the animal remained within the sector. Therefore, the observed distances were those recorded within the sector defined by the location and angle of view of the CT, at predetermined snapshot moments after initial detection, following Howe et al. (2017). Each animal travelled 10.7 to 11.0 km in a meandering path over 12 hr. Most step lengths were between 0 and 0.5 m, which ensured that animals would be observed multiple times, including at similar distances, during each independent encounter, and hence distance data would be severely overdispersed. Density remained constant, and the expected distribution of animal locations was uniform within each study area.

TABLE 1 Animal densities (D) and scale parameters (σ) of a half normal detection probability function in different habitat types used to generate simulated distance sampling data

Forest type	D	σ
Old growth	10	7.0
Regrowing	12	5.5
Recently logged	15	5.0
Mean	12.33	

We analysed data from all three habitat types simultaneously using multiple covariate distance sampling. Different habitat types were treated as different strata, with the potential to estimate a common detection function across all strata, or to model differences in detectability among strata using categorical covariates affecting the scale parameter of the detection function. We considered a habitat type covariate with two levels (old growth or logged), and one with three levels (old growth, regrowing, and recently logged). The 36 cameras in each study area were arbitrarily assigned to one of three different CT models (12 of each type). Detectability therefore varied among habitat types (Table 1) but not among camera trap models. Both habitat type and camera trap model were considered as potential covariates of the detection function; only one habitat type covariate was included in any model. We fitted 20 models with either the half-normal or the hazard rate key function, 0 or 1 cosine adjustment terms, and different covariate combinations to each dataset.

During both sets of simulations, distance data were binned into intervals prior to analysis. Howe et al. (2017), were confident of their assignments of duikers into 1 m intervals out to 8 m, but found it more difficult to estimate distances to this level of precision beyond 8 m. We similarly binned data into one-meter intervals out to 8 m, and at 10, 12, 15 and 20 m. In the case of the CT survey of moving animals, distance observations <1 m and >20 m were truncated. We conducted 500 replicate iterations, recording the number of estimated parameters, the log-likelihood value, the estimated density ($\hat{D}$) and associated empirical, design-based variances (Fewster et al., 2009), and the χ^2 GOF statistic and its *df* and *p*-value, from all models fit to each dataset. We selected among candidate models by comparing AIC values across all models fitted to the same dataset, and using both $QAIC_1$ and $QAIC_2$ following the two-step procedure described in the methods section. Simulations were performed using R software, version 3.3.2 (R Core Team, 2016).

2.3 | Applications with real data

We applied the same model selection criteria and procedures used in the simulations to real data from Maxwell's duikers in Taï National Park, Côte d'Ivoire, originally presented in Howe et al. (2017). We also reanalysed point count data from singing males of four species of songbirds sampled at Montrave Estate in Fife, Scotland, originally presented in Buckland (2006). The Montrave study area was small enough that densities of singing males were

TABLE 2 Number of times each detection function model fitted to simulated, overdispersed data from stationary animals was selected by AIC, and by each of $QAIC_1$ and $QAIC_2$ following the two-step procedure described in the methods section, of 500 replicate iterations. "Key" denotes the key function, either half-normal (hn) or hazard rate (hr); "Adj." denotes the number of adjustment terms

Models				Model selection criteria		
Key	Covariates	Adj.	Parameters	AIC	$QAIC_1$	$QAIC_2$
hn	None	0	1	14	248	238
hn	Observer	0	3	54	21	29
hn	None	1	2	9	32	39
hn	Observer	1	4	92	9	4
hr	None	0	2	1	30	29
hr	Observer	0	4	8	1	2
hr	None	1	3	57	120	127
hr	Observer	1	5	265	39	32

estimable by mapping their territories; these estimates were expected to have low bias, and served as benchmarks by which the accuracy of DS estimates were assessed (Buckland, 2006). Aware of the potential for overdispersion and therefore overfitting with cue count data, Buckland (2006) did not consider models with >2 parameters, and used a combination of AIC and plots of fitted probability density functions and detection functions to select among six models with different key functions and numbers of adjustment terms. We fitted a total of nine models to each dataset (uniform with 1, 2 or 3 cosine adjustment terms, half-normal with 0, 1, or 2 Hermite polynomial adjustment terms, and hazard rate with 0, 1, or 2 cosine adjustment terms) and used the two-step procedure with $QAIC_1$ to select among them. Truncation distances and cut points for the χ^2 GOF test followed Buckland (2006). $QAIC_2$ could not be calculated because $\hat{c}_2$ was unknown. We used diagnostic plots only to identify and exclude implausible models, such as cases where estimated detection probabilities exceeded 1.0, or fitted detection functions that were not monotonically nonincreasing.

3 | RESULTS

3.1 | Simulations

In simulations where c and the correct underlying model were known, mean sample sizes of distance observations in overdispersed datasets was 3,630. The χ^2 GOF test rejected the null hypothesis of adequate fit of the correct model for 492 of 500 datasets. $\hat{c}_1$ varied among iterations, but on average it estimated the true magnitude of overdispersion reasonably accurately (mean and median $\hat{c}_1$ from the data generating model were 6.16 and 5.73, respectively; true c was 6.0). AIC selected the most highly parameterised model most frequently, selected models with the spurious observer covariate for 71.4% of datasets, and selected the correct model for only 2.8% of datasets (Table 2). QAIC selected the correct model most frequently, followed by the hazard rate model with one adjustment and no covariates. $QAIC_1$ and $QAIC_2$ selected models with the spurious covariate for 14% and 13% of datasets respectively (Table 2). $\hat{D}$ from QAIC-selected models was both more accurate and more precise than $\hat{D}$ from AIC-selected models (Table 3).

TABLE 3 Medians of density estimates ($\hat{D}$) and of coefficients of variation (CV) of those estimates, from models fitted to simulated, overdispersed data from stationary animals, selected by AIC, and by $QAIC_1$ and $QAIC_2$ following the two-step procedure, across 500 iterations. True D was 2.00

	AIC	$QAIC_1$	$QAIC_2$
Median $\hat{D}$	1.89	2.00	2.00
Median CV($\hat{D}$)	0.054	0.029	0.028

In our simulated CT surveys of moving animals, where we assumed that, after initial detection, detection probability was 1.0 for as long as the animal remained in the field of view of the CT (as though CTs were programmed to record long bursts of still images or videos) observed distances included more observations at longer distances than where animals were detected via random trials at each time step (as though CTs were programmed to record a single image when triggered). The mode of the distribution was shifted right, and the number of observations at longer distances declined more slowly than under the data generating model (Supporting Information Figure S1). These differences arose because detected animals moving away from the CT continued to contribute observations at longer distances, where detection probability would otherwise be low. As a result, hazard rate models frequently provided a better fit than the half-normal model from which the random detections were simulated (Supporting Information Figure S1).

The χ^2 GOF test rejected the null hypothesis of adequate fit of 89% of the 10,000 models fitted. Sample sizes, and numbers of observations per independent encounter ($\hat{c}_2$), were slightly higher in old growth forests where detection probability as a function of distance was highest, although densities there were lowest (Table 4). $\hat{c}_1$ was generally lower, indicating less overdispersion, than $\hat{c}_2$ from a given dataset and model; $\hat{c}_1$ was also more variable among iterations than $\hat{c}_2$ (Table 4).

AIC again tended to select highly parameterised models. Density was not estimable from the AIC-minimising model in nine cases, and in 53 other cases, estimates were unrealistically high (>10 times the true density). $QAIC_1$ and $QAIC_2$ each selected models from which density was not estimable twice, and from which

TABLE 4 Mean (SD) sample sizes (*n*) of distance observations and numbers of observations per independent encounter ($\hat{c}_2$) from each habitat type and from data pooled across habitat types, and mean values of the χ^2 GOF statistic divided by its degrees of freedom ($\hat{c}_1$) from the most highly parameterised half-normal and hazard rate models fitted to the pooled datasets, across 500 iterations

	Old growth	Regrowing	Recently logged	Pooled data
n	919 (149)	730 (134)	787 (130)	2,437 (246)
$\hat{c}_1$ half-normal	–	–	–	6.70 (3.61)
$\hat{c}_1$ hazard rate	–	–	–	8.57 (7.74)
$\hat{c}_2$	16.8 (1.66)	14.8 (1.78)	14.2 (1.65)	15.3 (0.99)

density was severely overestimated 4 times; these problems were associated with the same six datasets. AIC favoured detection function models with more complex forms, selecting adjusted hazard rate models for 49% of datasets, and either unadjusted hazard rate or adjusted half normal models for another 45%, whereas QAIC selected unadjusted hazard rate models most frequently,

TABLE 5 The number of times, of 500 iterations, that each of the 20 candidate models was selected by each model selection criterion, and below this, the number of times each of four forms of the detection function, and each of three covariate effects, was included in the selected models. "Key" denotes the key function, either half-normal (hn) or hazard rate (hr); covariates were: Logging (2), with differences in detectability between logged and old growth forests, Logging (3), with differences among all habitat types, and camera trap model (CT), which did not affect detectability; "Adj." denotes the number of adjustment terms

Models				Model selection criteria		
Key	Covariates	Adj.	Parameters	AIC	$QAIC_1$	$QAIC_2$
hn	None	0	1	0	8	26
hn	Logging (2)	0	2	0	52	104
hn	Logging (3)	0	3	3	34	19
hn	Logging (2) + CT	0	4	9	32	7
hn	Logging (3) + CT	0	5	20	14	3
hn	None	1	2	0	0	0
hn	Logging (2)	1	3	3	37	28
hn	Logging (3)	1	4	6	22	5
hn	Logging (2) + CT	1	5	16	18	0
hn	Logging (3) + CT	1	6	71	13	0
hr	None	0	2	0	45	70
hr	Logging (2)	0	3	2	106	182
hr	Logging (3)	0	4	15	48	35
hr	Logging (2) + CT	0	5	31	31	12
hr	Logging (3) + CT	0	6	79	18	1
hr	None	1	3	0	0	0
hr	Logging (2)	1	4	4	11	8
hr	Logging (3)	1	5	28	2	0
hr	Logging (2) + CT	1	6	50	7	0
hr	Logging (3) + CT	1	7	163	2	0
Forms				**AIC**	**$QAIC_1$**	**$QAIC_2$**
hn (0 adjustment terms)				32	140	159
hn (1 adjustment term)				96	90	33
hr (0 adjustment terms)				127	248	300
hr (1 adjustment term)				245	22	8
Covariate effects				**AIC**	**$QAIC_1$**	**$QAIC_2$**
Logging (2-level)				115	294	341
Logging (3-level)				385	153	63
CT model				439	135	23

followed by unadjusted half normal models (Table 5). AIC always supported an effect of habitat type on detection probability, and supported the three-level habitat covariate for 77% of datasets (Table 5). $QAIC_1$ and $QAIC_2$ selected models with habitat type covariates for 89% and 81% of datasets, respectively, but tended to favour the two-level covariate (selected for 59% and 68% of datasets, respectively) over the three-level covariate (Table 5). Most (88% of) AIC-selected models, 27% of $QAIC_1$-selected models and 5% of $QAIC_2$-selected models included the spurious CT model covariate (Table 5). Model selection uncertainty across iterations was greatest with $QAIC_1$ (Table 5), which is not surprising given the variability in $\hat{c}_1$ across datasets (Table 4).

$QAIC_2$ and the two-step model selection procedure maximised both the accuracy and precision of $\hat{D}$ (Figure 1, Supporting Information Table S1). AIC-selected models yielded negatively biased $\hat{D}$ on average (Figure 1). AIC-selected models yielded the most accurate $\hat{D}$ only in recently logged forests (Figure 1). QAIC-selected models rarely included the three-level habitat covariate, and as a result, $\hat{D}$ in recently logged forests, and differences in $\hat{D}$ among habitat types, were underestimated (Figure 1, Supporting Information Table S1). However, QAIC-selected models yielded more accurate estimates of total density, and of density in regrowing and old growth forests (Figure 1). $QAIC_2$-selected models yielded the most precise density estimates, followed by $QAIC_1$-selected models (Supporting Information Table S1).

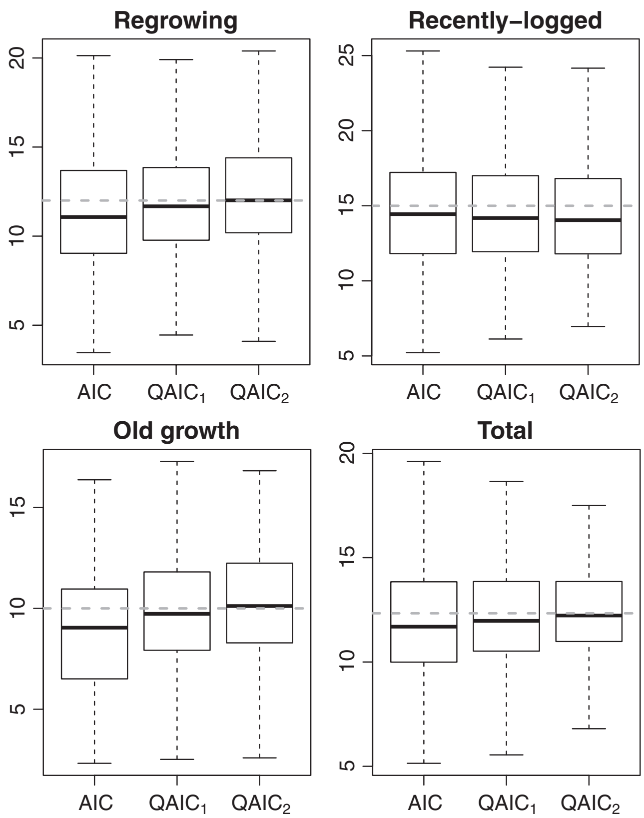

FIGURE 1 Animal densities (on y-axes) estimated from AIC-, $QAIC_1$- and $QAIC_2$-selected models fitted to simulated distance sampling data collected at camera traps in three different habitat types, and total density across all 3 habitat types. Dashed grey lines show true densities. Heavy black lines show medians across 438 and 494 AIC- and QAIC-selected models, respectively, from which density was estimable and the estimate of total density was within an order of magnitude of the true value. Whiskers extend 1.5 times the interquartile range from the boxes; outliers were excluded from the plots

3.2 | Applications with real data

The number of observations of Maxwell's duikers per independent encounter ($\hat{c}_2$) was 15.35 during the daytime, and 16.98 during times of peak activity. The χ^2 GOF statistic divided by its *df* ($\hat{c}_1$) from different models fitted to the daytime dataset ranged between 20 and 25, and from models fitted to the peak activity dataset ranged between 12 and 35 (Supporting Information Tables S2 and S4). Model selection criteria and procedures adjusted for overdispersion selected the same models as AIC for estimation from each dataset (the unadjusted hazard rate model, see Howe et al., 2017 and Supporting Information Tables S2–S5), so $\hat{D}$ was unaffected.

In our reanalysis of songbird data from Montrave Estate, $QAIC_1$ did not consistently outperform either AIC, or the combination of

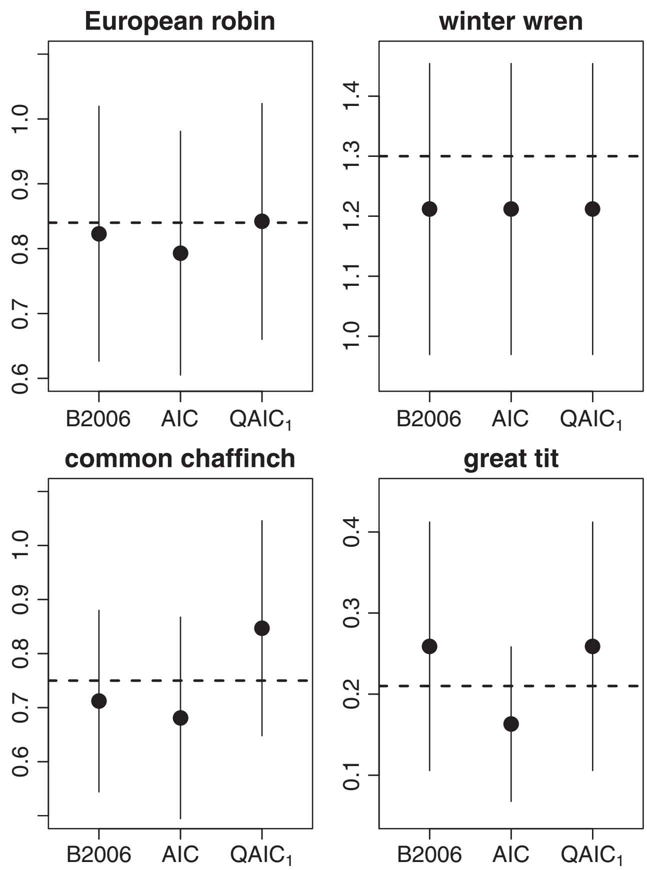

FIGURE 2 Densities of songbirds at Montrave Estate (on *y*-axes), estimated from models selected by Buckland (2006; "B2006" on *x*-axes), AIC-minimising models, and models selected by $QAIC_1$. Densities estimated by mapping territories, which were assumed to have low bias, are shown as dashed lines. Error bars show the point estimate ±1 *SE*

AIC, a constrained candidate model set, and reference to diagnostic plots employed by Buckland (2006). Model selection via $QAIC_1$ yielded a superior density estimate for European robins, the same estimate as Buckland (2006) for winter wrens and great tits, and an inferior estimate for common chaffinches (Figure 1). See the supplemental material for a detailed description of the results of our reanalysis including comparisons to models selected by AIC and Buckland (2006).

4 | DISCUSSION

Simulations with known *c* demonstrated that: (1) AIC was prone to overfitting, selecting unnecessarily complex models, (2) $\hat{c}_1$ was an accurate if variable estimator of the true magnitude of overdispersion and (3) QAIC and our two-step procedure outperformed AIC in that it selected the correct underlying model more frequently, and QAIC-selected models yielded more accurate and precise $\hat{D}$ than AIC-selected models.

Our simulations with animal movement were designed to be challenging from a model selection perspective, in that we sought criteria and procedures that would support small but real differences in detectability while excluding spurious effects from estimating models. AIC consistently supported models with adjustment terms although density was sometimes inestimable or drastically overestimated by these models, and models with a covariate that had no real effect on detectability. Associated inferences regarding both animal abundance and sources of variation in detectability were flawed. Models selected by QAIC and our two-step model selection procedure included fewer adjustment terms, were much less likely to include the spurious CT model covariate, and yielded more accurate and precise $\hat{D}$. Of the two proposed estimators of the magnitude of overdispersion, the mean number of observations per independent encounter ($\hat{c}_2$) was more consistent than the χ^2 GOF statistic divided by its degrees of freedom ($\hat{c}_1$). $QAIC_2$-selected models yielded the most accurate and precise density estimates on average.

Relative to AIC, QAIC more frequently supported models where detectability differed between old growth and logged forests, but not between recently logged and regrowing forests. This suggests that QAIC underfitted, selecting models with fewer parameters than the optimal model. This may not indicate that QAIC will underfit generally because the difference in detectability between recently logged and regrowing forests was slight. Furthermore, the effect of certain detection after initial detection on the distribution of observed distances would have obscured differences between study areas. Sources of variation in detectability that have small effect sizes may go undetected by any model selection criteria. Nevertheless, failure to detect and support this difference in our simulated data caused underestimation of density where detection probability was lowest. The difference in density between recently logged and other forest types was therefore underestimated; however, differences among all three habitat types were still apparent.

We applied model selection criteria and procedures "blindly", in that we always estimated $\hat{D}$ from the model that minimised AIC, or χ^2/df from the QAIC-minimising model within each key function. AIC might have performed better if we had followed established practices for DS analyses and multimodel inference, including exploratory analyses of relationships between distance observations and covariates, careful examination of fitted detection functions and associated parameter estimates, and consideration of ΔAIC values and AIC weights (Buckland et al., 2001, 2004; Burnham & Anderson, 2002; Marques et al., 2007). Therefore our results, where adjusting for overdispersion improved inferences from simulated data, but only improved inferences from real data in one of six cases, as well as Buckland et al.'s (2010) simulation results, suggest that adverse effects of overfitting by AIC may often be minor. Furthermore, our two-step approach to model selection using QAIC leads to the selection of a single model for estimation. QAIC values are not comparable between key functions, so metrics like ΔQAIC and QAIC weights cannot be used to compare relative support for models with different key functions, or to estimate detectability by model averaging across all models in a candidate set that includes different key functions.

We did not attempt to fit density surface models that allow researchers to assess support for covariation between density and spatially referenced covariates (Hedley & Buckland, 2004; Miller, Burt, Rextad, & Thomas, 2013). However, we note that such analyses can be performed either in two stages, where detectability is estimated during the first stage, and plot-specific counts or abundance estimates are modelled during the second stage, or by maximising a full likelihood model whereby parameters related to both detectability and local abundance are estimated simultaneously (Hedley & Buckland, 2004; Johnson, Laake, & Ver Hoef, 2010; Miller et al., 2013). If a two-stage approach to fitting density surface models is adopted, any model selection criteria or procedures, including those described here, could be used when estimating detectability. It is therefore possible to account for overdispersion in the distance data when estimating the detection function, and still fit density surface models to test for effects of spatial covariates on density. Johnson et al. (2010) proposed a one-stage, model-based approach for simultaneously estimating detectability and spatially variable abundance from DS data. They also evaluated the effectiveness of an overdispersion factor calculated from a χ^2 test performed on transect-specific counts for inflating model-based variances around abundance estimates to account for overdispersion introduced by fine-scale variation in local abundance. They found that variances were still underestimated except where there were many transects, and suggested the χ^2 GOF test for binned distance data divided by its degrees of freedom (our $\hat{c}_1$) as an alternative estimator. However, it is not clear to us how a statistic derived from the observed distances would quantify overdispersion induced by variation in local abundance, and we prefer to use this statistic to adjust for overdispersion only when modeling the detection function.

We analysed overdispersed data from simulated and real cue counting and CT surveys; however, model selection criteria adjusted for overdispersion could also be useful when social animals that travel in loosely clumped or moving groups are surveyed. Buckland et al. (2010) simulated line transect sampling of primate groups and found that treating the individual as the unit of observation and selecting among models of the detection function using AIC yielded more accurate and precise $\hat{D}$ than approaches that treated the group as the unit of observation, "despite obvious overfitting in some cases" (p. 835).

5 | SYNTHESIS AND RECOMMENDATIONS

We described novel approaches to estimating an overdispersion factor ($\hat{c}$), and QAIC-based procedures for selecting among models of the DS detection function when the assumption that observations are independent is violated. These novel methods improved inference from simulated data. However, we conducted limited simulations with severely overdispersed data, and reanalyses of real datasets did not unambiguously indicate that adjusting for overdispersion at the model selection stage improved inference. We therefore recommend additional research, but also that these criteria and procedures be considered as alternatives to AIC when the independence assumption is violated. They are most likely to be useful where: (a) overfitting by AIC is apparent (e.g. if AIC favours models that include both adjustment terms and covariates, multiple adjustment terms, or weak or imprecisely estimated covariate effects); (b) it is not practical or not desirable to constrain the candidate set to include only simple models (e.g. where there are multiple potential covariates of the detection function, or where models with unadjusted key functions do not fit the observed data well) or (c) where researchers wish to avoid subjectivity during model selection.

ACKNOWLEDGEMENTS

We thank the Robert Bosch Foundation, the Max Planck Society and the University of St Andrews for funding, the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique and the Ministère de l'Environnement et des Eaux et Forêts in Côte d'Ivoire for permission to conduct field research in Taï National Park, Dr. Roman Wittig for permitting data collection in the area of the Taï Chimpanzee Project, and Dr. Joeseph Nocera for informal discussions about model selection in the presence of overdispersion.

AUTHORS' CONTRIBUTIONS

E.J.H. and S.T.B. conceived the ideas, E.J.H. analysed the data and led the writing of the manuscript. M.-L.D.-E. helped to design and conducted the field study of Maxwell's duikers, H.S.K. designed the field study of Maxwell's duikers. All authors contributed critically to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

Data from Maxwell's duikers in Tai National Park are archived at the Dryad Data Repository https://doi.org/10.5061/dryad.b4c70 (Howe et al., 2017), as are the raw Montrave Estate cue count data (https://doi.org/10.5061/dryad.j2p2h3s), which are also available as Distance projects at https://synergy.st-andrews.ac.uk/ds-manda/#montrave-songbird-case-study and http://distancesampling.org/ (Howe, Buckland, Despres-Einspenner, & Kühl, 2018).

ORCID

Eric J. Howe http://orcid.org/0000-0002-4715-3958

REFERENCES

Akaike, H. (1973). Information Theory as an Extension of the Maximum Likelihood Principle. In B. N. Petrov, & F. Csaki (Eds.), *Second international symposium on information theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.

Borchers, D. L., Buckland, S. T., & Zucchini, W. (2002). *Estimating animal abundance: Closed populations*. New York, NY: Springer Science & Business Media. https://doi.org/10.1007/978-1-4471-3708-5

Buckland, S. T. (1984). Monte Carlo confidence intervals. *Biometrics*, *1*, 811–817. https://doi.org/10.2307/2530926

Buckland, S. T. (2006). Point transect surveys for songbirds: Robust methodologies. *The Auk*, *123*, 345–357. https://doi.org/10.1642/0004-8038(2006)123[345:PSFSRM]2.0.CO;2

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (2001). *Introduction to distance sampling: Estimating abundance of biological populations*. Oxford, UK: Oxford University Press.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (2004). *Advanced distance sampling: Estimating abundance of biological populations*. Oxford, UK: Oxford University Press.

Buckland, S. T., Plumptre, A. J., Thomas, L., & Rexstad, E. A. (2010). Design and analysis of line transect surveys for primates. *International Journal of Primatology*, *31*, 833–847. https://doi.org/10.1007/s10764-010-9431-5

Burnham, K. P., & Anderson, D. R. (2001). Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildlife Research*, *28*, 111–119. https://doi.org/10.1071/WR99107

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and inference: A practical information theoretic approach*, 2nd ed. New York, NY: Springer Science and Business Media.

Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data*, 2nd ed. New York, NY: Chapman and Hall.

Fewster, R. M., Buckland, S. T., Burnham, K. P., Borchers, D. L., Jupp, P. E., Laake, J. L., & Thomas, L. (2009). Estimating the encounter rate variance in distance sampling. *Biometrics*, *65*, 225–236. https://doi.org/10.1111/j.1541-0420.2008.01018.x

Hedley, S. L., & Buckland, S. T. (2004). Spatial models for line transect sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, *9*, 181–199. https://doi.org/10.1198/1085711043578

Howe, E. J., Buckland, S. T., Després-Einspenner, M. L., & Kühl, H. S. (2017). Distance sampling with camera traps.

Methods in Ecology and Evolution, *8*, 1558–1565. https://doi.org/10.1111/2041-210X.12790

Howe, E. J., Buckland, S. T., Despres-Einspenner, M.-L., & Kühl, H. S. (2018). Data from: Model selection with overdispersed distance sampling data. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.j2p2h3s

Johnson, D. S., Laake, J. L., & Ver Hoef, J. M. (2010). A model-based approach for making ecological inference from distance sampling data. *Biometric*, *126*, 310–318. https://doi.org/10.1111/j.1541-0420.2009.01265.x

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, *19*, 101–108. https://doi.org/10.1016/j.tree.2003.10.013

Lebreton, J.-D., Burnham, K. P., Clobert, J., & Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monogrographs*, *62*, 67–118. https://doi.org/10.2307/2937171

Liang, K.-Y., & McCullah, P. (1993). Case studies in binary dispersion. *Biometrics*, *49*, 623–630. https://doi.org/10.2307/2532575

Marques, T. A., Thomas, L., Fancy, S. G., & Buckland, S. T. (2007). Improving estimates of bird density using multiple-covariate distance sampling. *The Auk*, *124*, 1229–1243. https://doi.org/10.1642/0004-8038(2007)124[1229:IEOBDU]2.0.CO;2

Miller, D. L., Burt, M. L., Rextad, E. A., & Thomas, L. (2013). Spatial models for distance sampling data: Recent developments and future directions. *Methods in Ecology and Evolution*, *4*, 1001–1010. https://doi.org/10.1111/2041-210X.12105

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Howe EJ, Buckland ST, Després-Einspenner M-L, Kühl HS. Model selection with overdispersed distance sampling data. *Methods Ecol Evol*. 2018;00:1–10. https://doi.org/10.1111/2041-210X.13082