STEPHEN T. BUCKLAND (*) – NICOLE H. AUGUSTIN (**)
VERENA M. TRENKEL(***)(****)
DAVID A. ELSTON (***) – DAVID L. BORCHERS (*)

Simulated inference, with applications to wildlife population assessment

INTRODUCTION

We review some of the ways in which computer-intensive "simulated" inference has impacted wildlife population assessment. Each of our examples illustrates a different extension of the bootstrap, to show how the simple concept of resampling the data leads to powerful tools for data analysis. In the first example, we show how to quantify model selection uncertainty. In the second, the problem of accounting for uncertainty in the specification of a population dynamics model is addressed, using weighted bootstrap samples. The last two examples show that we need not assume that observations are independently and identically distributed to bootstrap them. We use a model for spatial distribution of wildlife as an example of bootstrapping when the observations are not independently distributed. Bootstrapping of observations that are not identically distributed is illustrated using overdispersed counts from a fisheries survey.

(*) School of Mathematical and Computational Sciences, University of St Andrews, The North Haugh, St Andrews - Fife KY16 9SS, Scotland
(**) Biomathematics and Statistics Scotland, Scottish Crops Research Institute - Invergowrie, Dundee, Scotland
(***) Biomathematics and Statistics Scotland, Macaulay Land Use Research Institute - Craigiebuckler, Aberdeen, Scotland
(****) Current address: IFREMER, Centre de Nantes, Nantes, France

Computational power has freed statisticians of the constraints imposed by the (non-)availability of analytic results. In academia, the impact has been most dramatic in Bayesian statistics, in which the Gibbs sampler and other Markov chain Monte Carlo methods have allowed analysts to implement models that are unmanageable by purely analytic methods. More important to most applied statisticians is the development of the bootstrap and related resampling methods.

Classical statistical inference depends largely on the existence of analytic results that allow us to draw inferences from an observed sample about a population of interest. Statistical inference based on computer-intensive methods removes that dependence. Thus we can choose the models we believe are most appropriate for the data, not those for which we have analytic results. We can relax assumptions as far as the data allow, and are no longer subject to the vagaries of what is mathematically possible, and what has historically been derived. We term such computer-intensive statistical inference "simulated inference". This term is intended to refer both to frequentist and to Bayesian methods. In a frequentist framework, we consider the observed sample to be one of many (usually an infinity of) possible samples, and develop inference according to the variability we would expect if we observed a large number of these samples. Simulated inference allows us to generate this large number, from which we extract summary information, and hence draw inferences by a much more direct route than classical methods allow. In a Bayesian framework, we condition on the observed sample, but generate different realizations of the set of parameters by simulation, from which we draw inference.

Undergraduate teaching programmes in statistics have been slow to respond to change. Adopting a frequentist framework, it is now easy to generate resamples from data by a variety of methods, and to treat these resamples as a subset of all possible samples. Inference is then merely an exercise in extracting appropriate summary information. This is more intuitive and less restrictive than relying on the Central Limit Theorem, or deriving Student's t distribution, yet classical inference remains the cornerstone of our teaching. Simulated inference, which is more directly rooted in the history of statistical inference, but whose development was delayed until appropriate computational tools became available, is too often treated as an afterthought, or ignored altogether. Contrast this with practice. In many fields, the bootstrap is now the method of preference for quantifying precision.

2. THE BOOTSTRAP AS A TOOL FOR QUANTIFYING MODEL SELECTION UNCERTAINTY

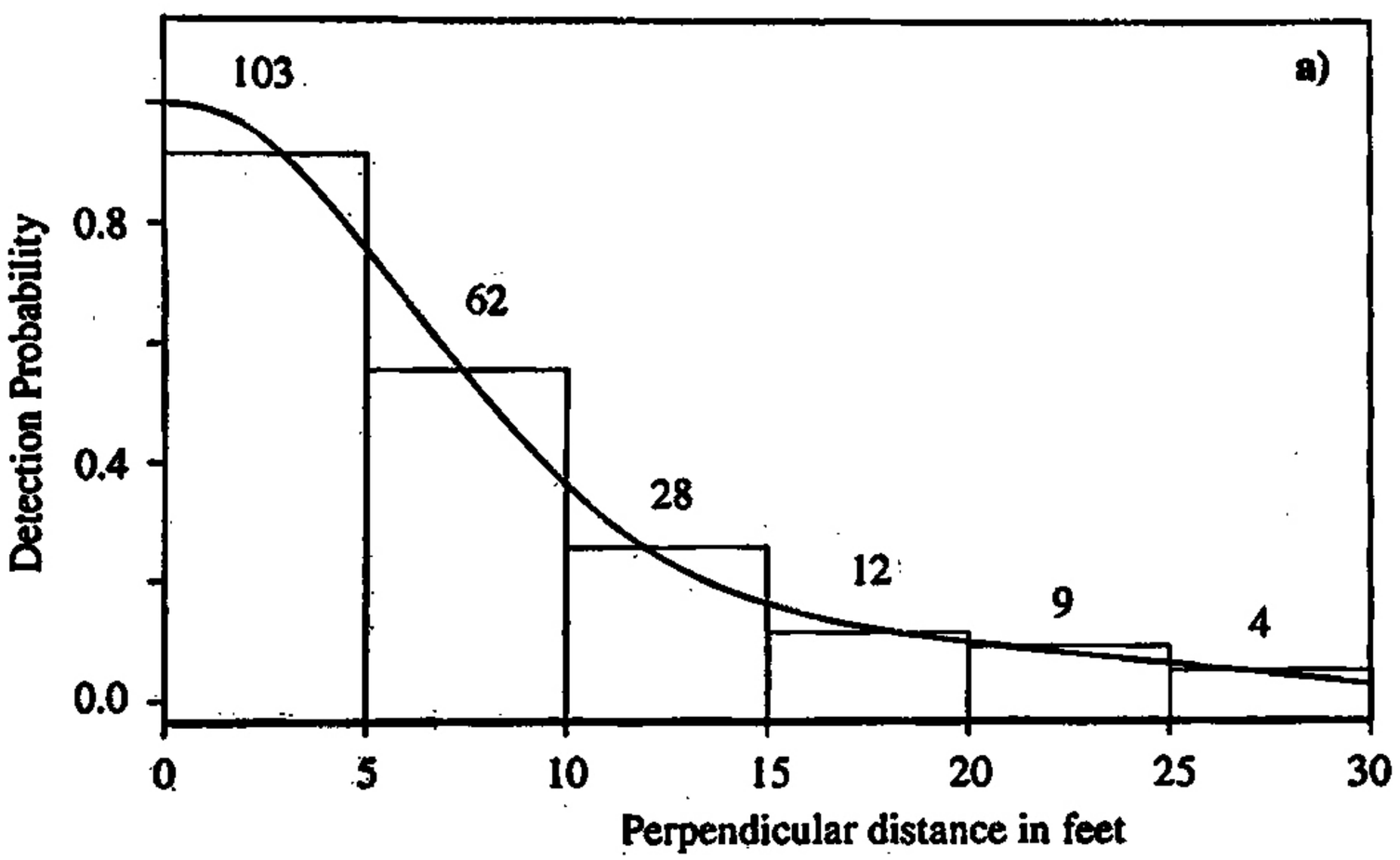
Although model selection is widely recognised as central to good inference, paradoxically, it has seldom been integrated fully into inference. For example, there are many methods in multiple regression for identifying an appropriate subset of covariates. Having identified them, subsequent inference is usually conditional on the selected model; that is, we assume that the model is correct. It is more defensible to recognise the uncertainty in model selection when quantifying the precision of an estimator. Under this philosophy, "model mis-specification bias" is not bias at all, but merely a component of the variance. The difficulty in incorporating model selection uncertainty into inference can be circumvented using the bootstrap. The model selection procedure is applied independently in each bootstrap resample, and inference is based on the resulting bootstrap estimates (Buckland, 1982; Hjorth, 1994). We illustrate the approach with a line transect example.

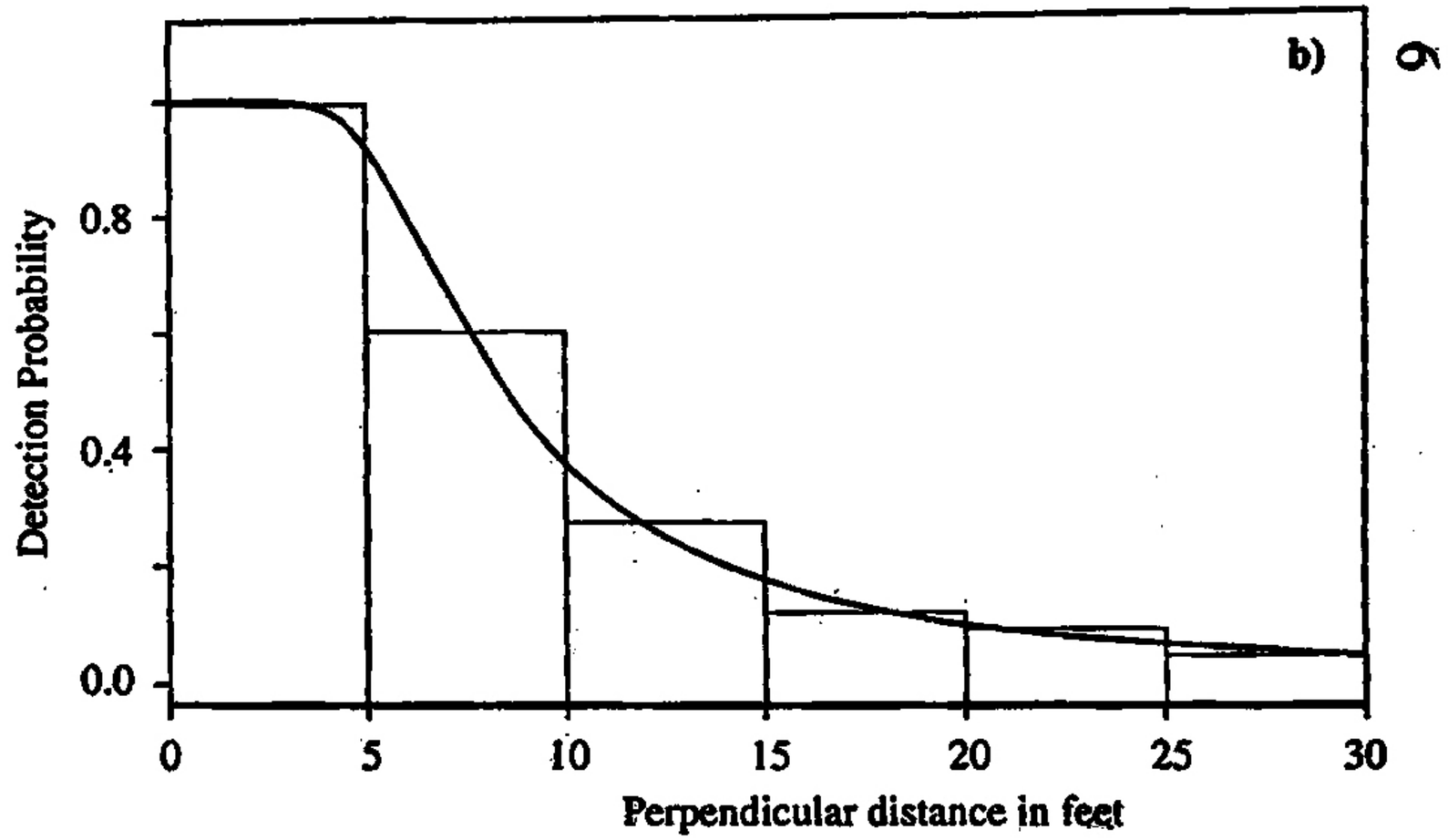
In line transect sampling, an observer walks along a line, and records the perpendicular distance from the line of each animal detected. The usual estimate of animal density D is a function of $f(0)$, the fitted probability density function of perpendicular distances, evaluated at zero distance:

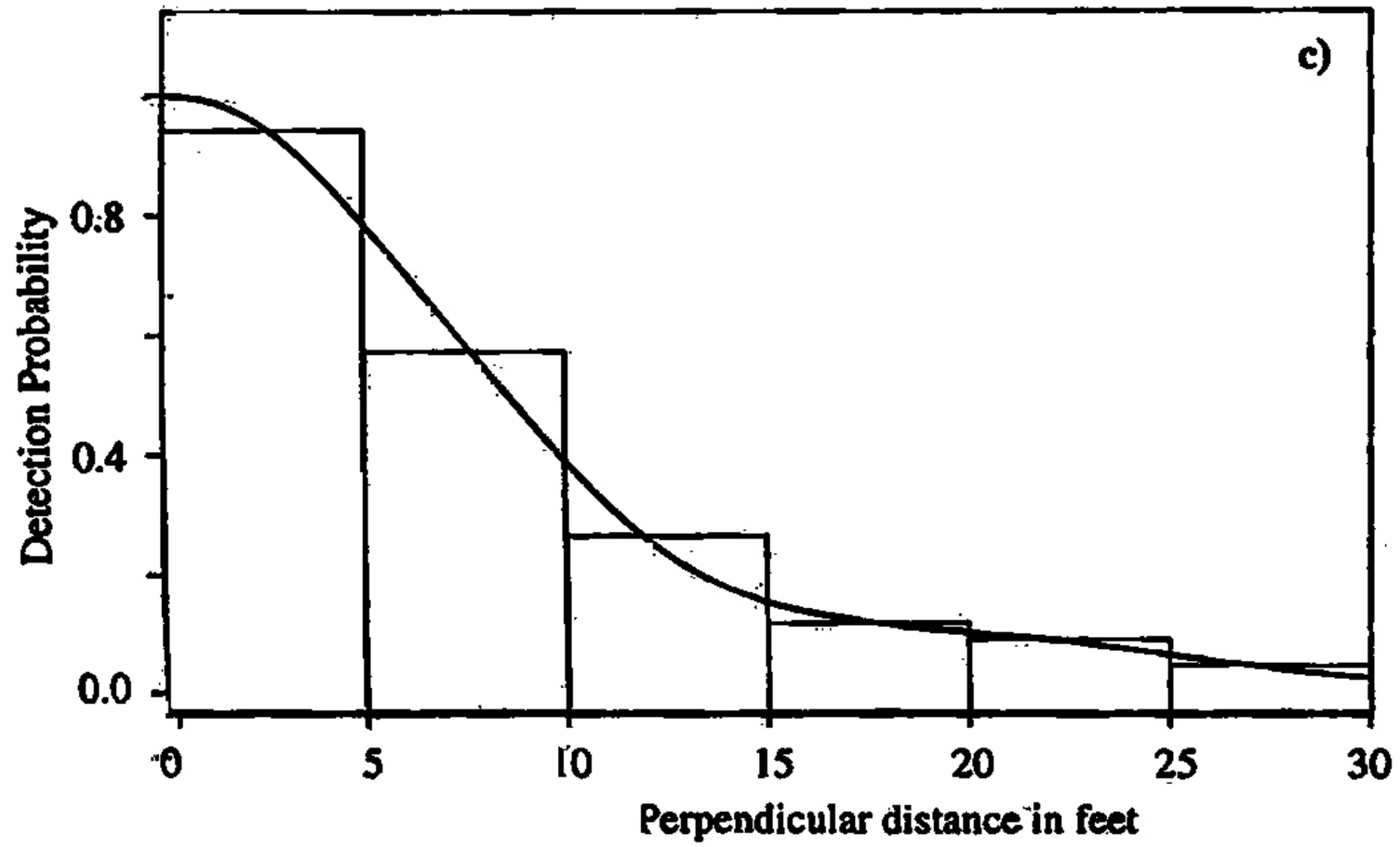
$$\hat{D} = \frac{n \cdot \hat{f}(0)}{2L} \tag{1}$$

where n is the number of animals detected, and L is the total length of transect travelled (Buckland *et al.*, 1993).

We use a ruffed grouse dataset taken from Gates (1979), in which $n = 218$, as an example of the effect of incorporating model selection uncertainty in inference. Analyses were carried out in DISTANCE (Laake *et al.*, 1993), which allows the user to approximate the probability density function by a parametric "key" function, together with a choice of adjustment terms, for when the key function alone fails to provide an adequate fit (Buckland, 1992). Model selection was restricted to four models here: a half-normal key and cosine adjustment terms; a hazard-rate key and simple polynomial adjustment terms; the Fourier series model (uniform key and cosine adjustment terms); and a negative exponential key and simple polynomial adjustment terms. Plots of the data, together with the fitted models, are shown in figure 1.







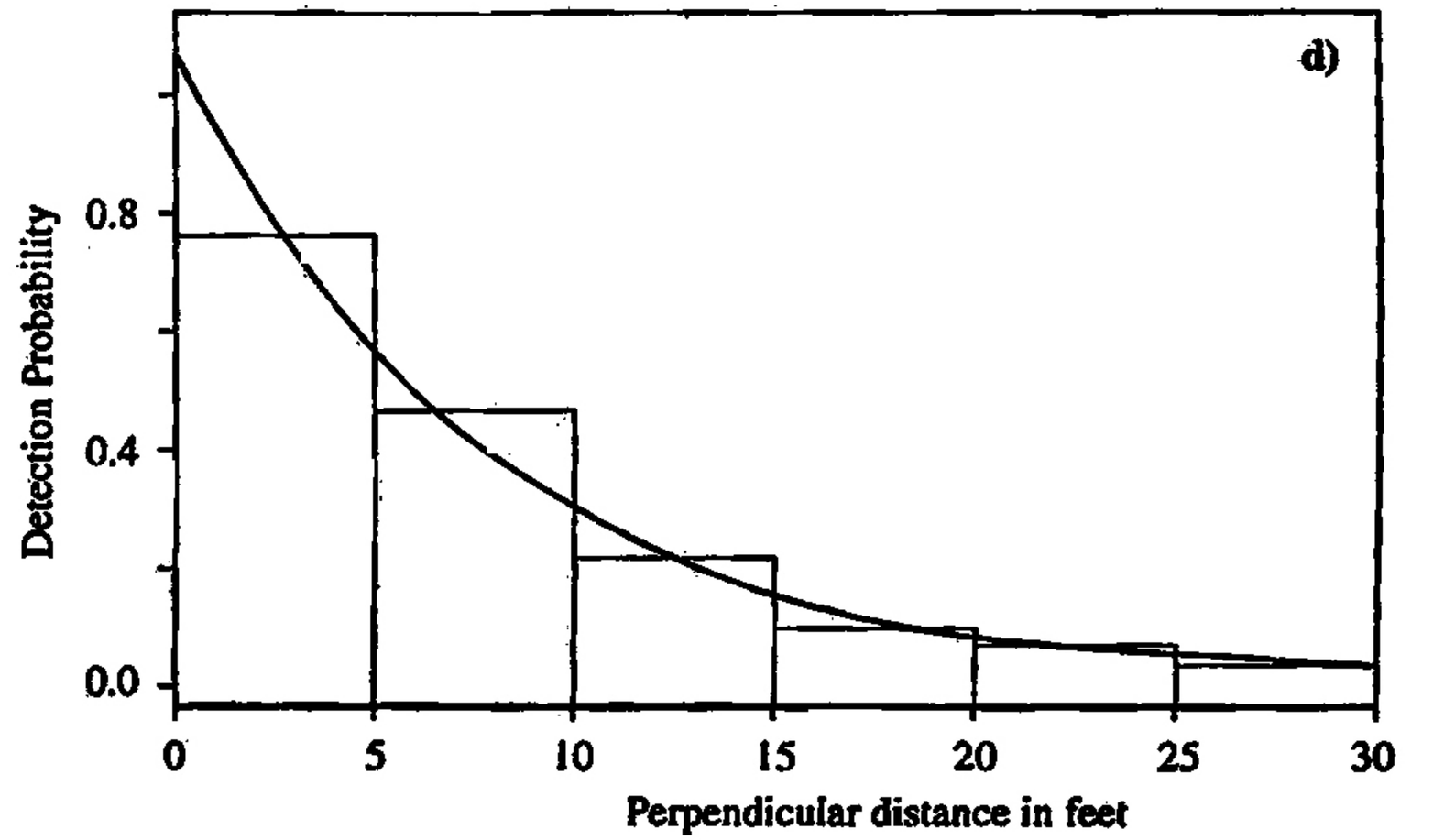


Fig. 1. Fit of (a) the half-normal, (b) the hazard-rate, (c) the Fourier series and (d) the negative exponential models to the ruffed grouse perpendicular distance data. No adjustment terms were selected in the cases of the hazard-rate and negative exponential models, whereas one cosine term adjustment was made to the half-normal model, and three cosine terms were selected for the Fourier series model.

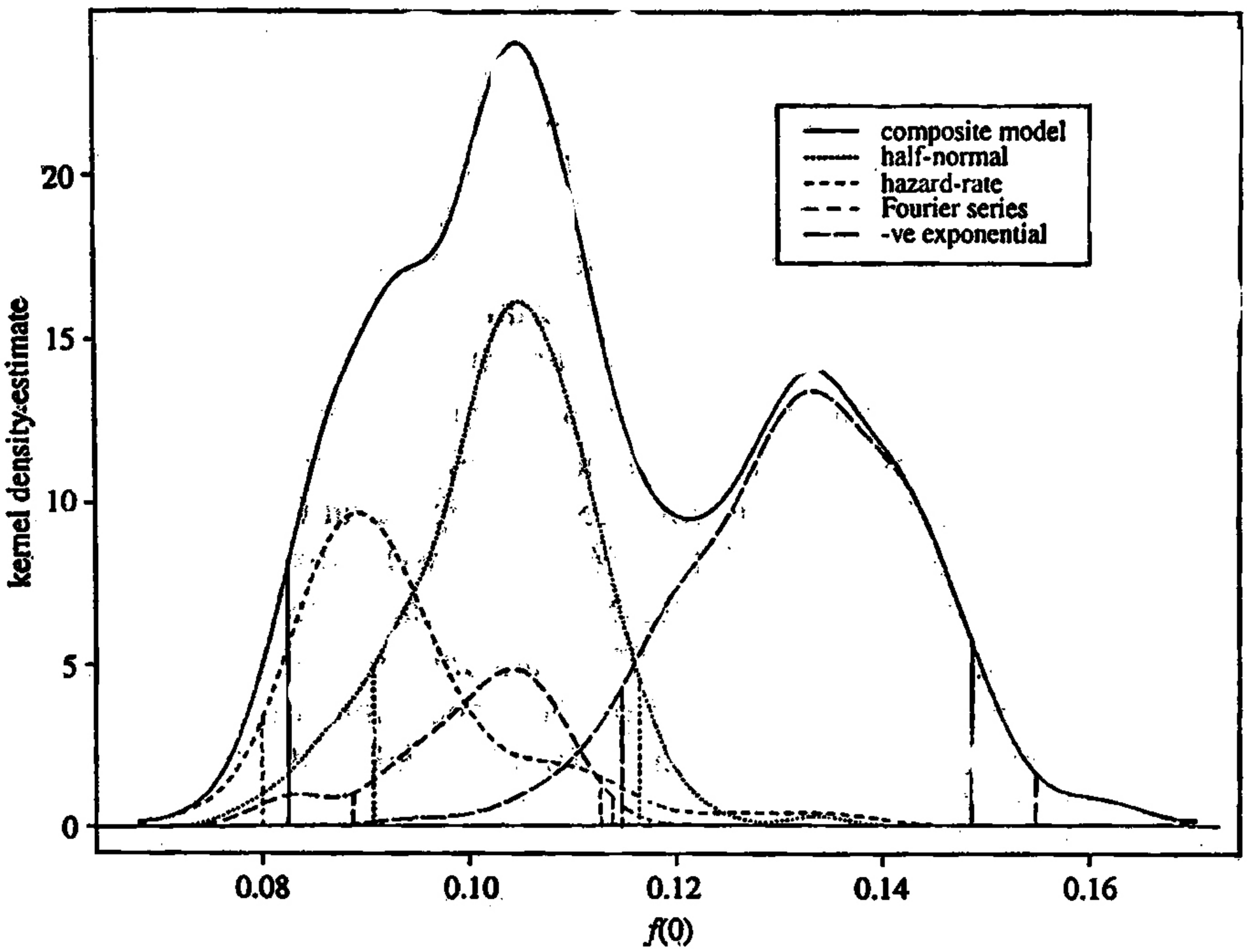


Fig. 2. The distribution of bootstrap estimates under each model from figure 1. The curves were fitted using the kernel algorithm of Silverman (1982). Each curve was scaled so that it integrates to the proportion of times the corresponding model was selected when analysing bootstrap resamples. Thus the sum of the curves (the "composite model") estimates the probability density function of the estimate $\hat{f}(0)$. The vertical lines indicate 95% percentile confidence limits for $f(0)$ under the respective models.

Usually in line transect sampling, the nonparametric bootstrap is implemented by resampling from the transects, rather than individual detections. However, for estimating the precision of $\hat{f}(0)$, resampling from individual detections works well, and we adopt that strategy here, as we only have the data pooled across lines. Each model was fitted both to the original data and to each of $b = 1000$ resamples, and Akaike's Information Criterion (AIC) was used to select between the four models. In figure 2, we show the distribution of bootstrap estimates of $f(0)$ under each model, using kernel density estimation to smooth each distribution. Each curve is scaled so that the area under it is equal to the proportion of times the corresponding model was selected in the bootstrap resamples. Also shown is the sum of these four curves; the area under this composite curve is unity. For each curve, the endpoints of a 95% percentile confidence interval, calculated as described by Buckland (1984), are shown. It can be seen that the

lower confidence limit under the negative exponential model lies above the upper limit for the hazard-rate model. The two remaining models yield confidence intervals intermediate between these, although the lower limit under the negative exponential model is close to the upper limit under all three of the other models. The limits corresponding to the composite curve better reflect uncertainty in the true value $f(0)$ than do those from any of the individual curves.

The average of the 1000 bootstrap estimates of $f(0)$ is 0.1120. The estimate under the negative exponential model, selected by AIC applied to the original sample, is considerably higher, at 0.1333. The 95% percentile confidence interval for $f(0)$ obtained from the full set of 1000 bootstrap estimates is (0.0830, 0.1486), and the bootstrap standard error is $\text{se}\{\hat{f}(0)\} = 0.0194$. The data fail to rule out any of the possible models, and the bootstrap confidence interval consequently spans the range of estimates for $f(0)$ obtained under the different models.

3. Simulated inference using a weighted bootstrap

Suppose we have a model from which we can readily simulate data, if the model parameters are specified, but for which we have no reliable means of estimating the parameters, given data. The bootstrap cannot be applied in the normal way, because we have no means of fitting the model to the bootstrap resamples. Instead, we can specify prior distributions for the model parameters, and generate bootstrap resamples from these distributions. The priors should be sufficiently diffuse that the posterior distributions are insensitive to the choice of prior distributions. Having generated a set of input parameters from the prior distributions, predictions are derived (deterministic models) or simulated (stochastic models) from the model determined by those parameter values. A weight is defined that relates the real data to the output parameters. This weight is the likelihood of the data, given the model and the model parameters.

The above process is repeated a large number of times. Thus for simulations in which the model parameters match closely the "true" model, the weight associated with that bootstrap replicate is likely to be high; when they do not, the weight is small. Inference proceeds either by resampling the parameter sets with probability proportional to the weights (e.g. Smith and Gelfand, 1992), so that the resulting parameter

sets are a sample from the joint posterior distribution and provide an empirical estimate of that distribution, or by extracting the values of the predictions from each resample, and attaching their respective weights; when estimating the true parameters and the corresponding precision of the estimates. If data are collected sequentially, then the weighted resampling method may be applied sequentially, with the posterior distribution from one step providing the prior distribution for the next, Gordon *et al.* (1993) called this approach the recursive Bayesian filter, based on work by Smith and Gelfand (1992). Liu and Chen (1998) give an overview of recent developments of the general class of methods which they refer to as sequential importance sampling.

We illustrate the method using a population dynamics model. Given count and cull data on a population of red deer (*Cervus elaphus*), we wish to draw samples from the posterior distributions of the parameters, to allow us to project the population forwards and examine the impact of different culling strategies. Exploitation and natural mortality are strong functions of both age and sex, so that a large number of parameters is required to model populations adequately.

The model parameters are initial population structure and proportion of hinds with calves; parameters that allow the natural mortality rate to be a function of age, sex and population density; and parameters that allow fecundity to be a function of age, status (milk hind or yeld hind, corresponding with whether or not the hind currently has a calf), body weight and population density. The data are counts (which may or may not be conducted annually), tallied separately for adult males (stags), adult females (hinds) and calves; and annual cull data, tallied by the same three categories. The initial conditions are given by the first count. For each model parameter, a value is generated by simulation from its prior distribution. Having generated a set of values for the model parameters, the likelihood of the data given those parameters is evaluated, and the population is projected forwards to yield predictions of population size and structure. If the population dynamics model is deterministic, the predictions are calculated by applying a Leslie matrix model. A stochastic model is readily accommodated, by simulating the fate of each animal in the population individually. The process is repeated a large number of times, and the posterior distribution of any predicted variable is estimated by weighting the set of values obtained from the simulations, or by resampling those sets with probability proportional to the weights, where the weights are determined from the likelihood. More details are given by Trenkel *et al.* (1996).

To implement this simulated inference approach, we need to be able to assign weights to each simulated population. In addition, if there are contending population dynamics models, we require weights to assign to them. Consider first how to calculate a weight for simulation i. By defining an appropriate state space model, we can define and evaluate the likelihood of the counts of population size, given our population dynamics model. The process of simulating from the prior distributions of the parameters ensures appropriate prior weighting, so that we may interpret the ratio of likelihoods, L_i/L_j, for two simulations i and j as the posterior odds ratio for the respective simulations. Thus the weight assigned to simulation i is simply

$$w_i = \frac{L_i}{\sum L_j} \tag{2}$$

where summation is over all simulations. Suppose the simulation is advanced a year at a time. For any year in which we have a count, we can update the likelihood of all data collected up to that time, given our model. Thus the weight in year y is

$$w_{i,y} = \frac{w_{i,y-1} L_{iy}}{\sum_j w_{j,y-1} L_{jy}} \tag{3}$$

where L_i is the contribution of the data in year y to the likelihood corresponding to simulation i. Note that the weights only change in years for which there are counts. Because values for the parameters are simulated from their prior distributions, $w_{i\,1} = 1/b$ for all i, where b is the total number of simulations.

Within each simulation, if there are contending population dynamics models, the same approach might be applied to assign weights to each model. However, if the number of parameters differs between models, a penalized likelihood, such as Akaike's Information Criterion, might be used (Buckland, 1995; Buckland *et al.*, 1997). We do not pursue this here.

Each simulation provides estimates of population size and structure by year. If we resample at each step with probabilities proportional to the weights, then the parameter sets generated at the final step provide an empirical estimate of the joint posterior distribution, and appropriate percentiles from the empirical distribution for each parameter provide endpoints for an interval estimate for that parameter. If the initial parameters sets are carried through, together with their

associated weights, an average across simulations of the parameter estimates of interest, weighted by the final weights for the simulations, provides the estimate of a given parameter, and a weighted sample variance yields the corresponding variance estimate. A modification of the bootstrap percentile method of setting intervals can be used, in which the bootstrap estimates of a parameter are ranked from smallest to largest, and the weights associated with the ranked estimates summed to determine the percentile points in the distribution of simulated estimates. For example, to estimate the lower 95% limit, the location of the 2.5 percentile is found by summing the weights of the smallest estimates until the sum is equal to 2.5% of the sum of all weights, b. The limit will be the estimate associated with the weight that first takes the running sum above 2.5% of b.

There will be a tendency for most of the simulated populations to have small weights after several counts, so that inference is dominated by a small proportion of the populations originally generated. Thus, the number of populations b generated from the priors should be large, perhaps around one million. Alternatively, if we adopt the sequential weighted resampling approach, and perturb the resampled parameter values slightly at each step using the smoothed bootstrap, we generate new simulated populations after each year for which there was a count. In this case, only a relatively small number of populations, perhaps a few thousand, need be generated at each step.

The above approach is similar to the policy of simulating from a grid of starting values for the input parameters, which has been adopted by many researchers and wildlife managers. That approach is dogged with difficulties. The outcomes are often implausible, and no formal inference is possible, because the grid of starting values takes no account of their relative likelihoods; if an unlikely combination of starting values is adopted, it is unsurprising if the outcome is implausible. Simulated inference avoids the difficulty by assigning distributions to the input parameters that reflect uncertainty, and by evaluating the likelihood associated with the data. In common with all Bayesian methodologies, the approach raises the question of how to specify the prior distributions.

We show in figure 3 results from a population of red deer in western Scotland. Counts and cull data were available on this population for the period 1989 to 1997. Estimates of input parameters were found from the literature, and prior distributions were centred on these estimates. The spread of the priors was chosen to reflect both uncer-

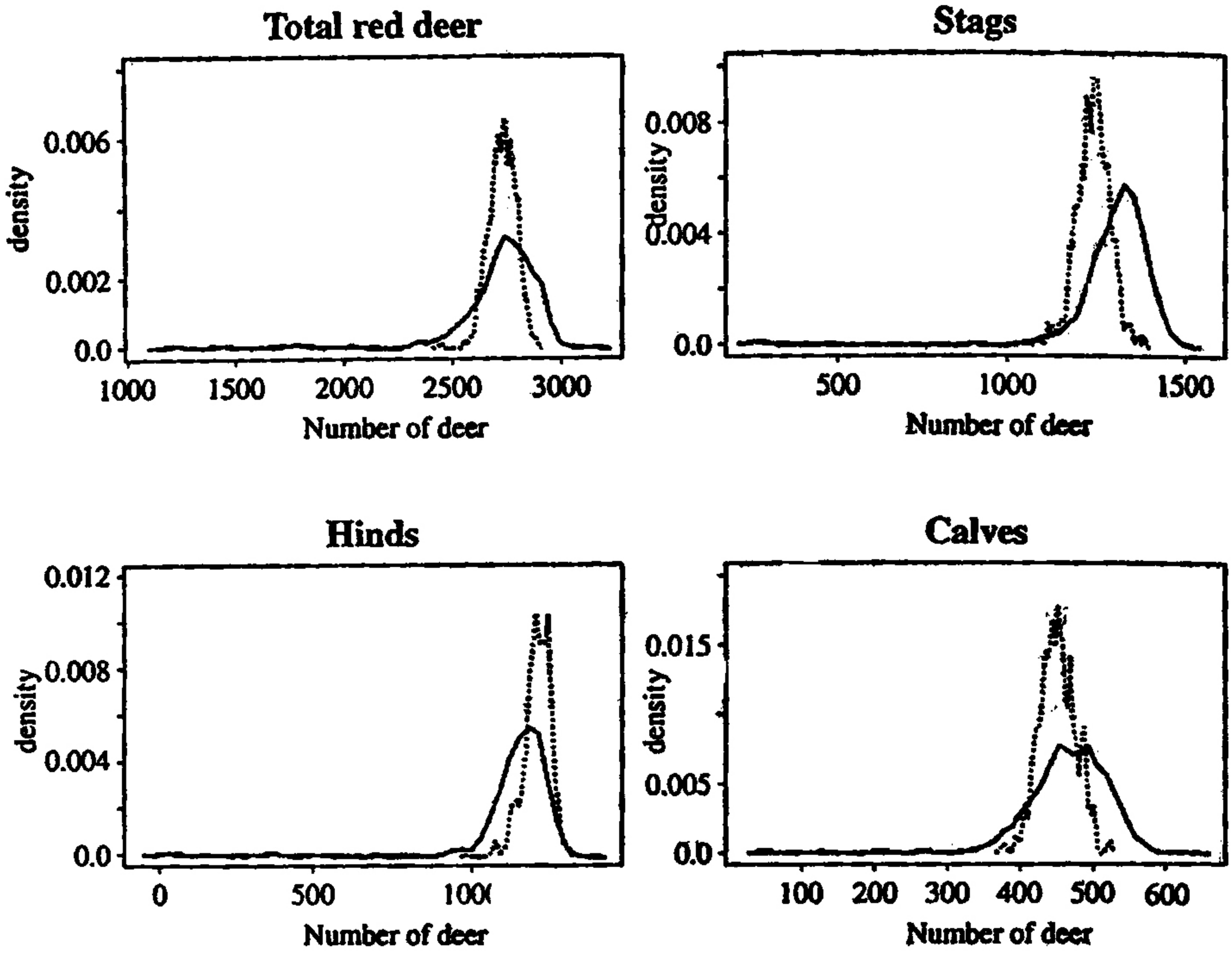


Fig. 3. Prior and posterior distributions for the predicted size of the stag, hind and calf components of a population of red deer, and for the overall total, in spring 2000.

tainty in estimation in the earlier studies and uncertainty arising from applying parameter estimates from populations of red deer elsewhere in Scotland to the study population. Figure 3 shows the implicit prior distributions for the size of the stag, hind and calf populations and of the total population for the year 2000. These priors are derived from the priors specified on the input parameters. Also shown are the corresponding posterior distributions.

4. Bootstrapping correlated spatial data

The standard bootstrap assumes that observations are independently distributed. There has been research interest in defining "moving blocks" (Künsch, 1989), on the principle that observations sufficiently far apart in space or time can be assumed independent. However, this approach seems unsatisfactory as the first few observations in one block are unlikely to be independent of the last few in the previous

block. A more appealing strategy is to incorporate dependence into the fitted model, and then use the parametric bootstrap to generate simulated data with the desired correlation structure. We illustrate this approach by bootstrapping from an autologistic regression model fitted to presence/absence data on red deer. This example is summarized from Augustin *et al.* (1996, 1998).

The primary objective of the modelling was to estimate the geographic distribution of wildlife species from incomplete field surveys of a sample of grid squares. The autologistic regression model was chosen because it allowed us to predict presence/absence in a square as a function of habitat and other spatial covariates, and of an autocovariate that is a function of the responses in neighbouring squares. The Gibbs sampler was used to predict presence/absence in the unsurveyed squares.

Let y_i be the response ($y_i = 0, 1$) and x_{ki} be the value of covariate k for square i, $k = 1, \ldots, m-1$. Further, let $p_i = \Pr(Y_i = 1 | \{Y_j = y_j,\ j \neq i\}, \{x_{ki},\ k = 1, \ldots, m-1\})$.

Then the model is defined by

$$\log \frac{p_i}{1 - p_i} = \alpha + \sum_{k=1}^{m} \beta_k x_{ki} \tag{4}$$

where x_{mi} represents the proportion of neighbouring squares that is occupied.

The algorithm used for fitting the model is implemented as follows. First, we fit an ordinary logistic model to data from the surveyed squares. Next we predict the response in unsurveyed squares by generating $\hat{y}_i$ from $\hat{p}_i$ stochastically: $\hat{y}_i \sim \text{binary}(\hat{p}_i)$. We then use the predicted responses to calculate the autocovariate for every square. The autologistic model is then fitted to data from the surveyed squares. New $\hat{p}_i$ are obtained, and new $\hat{y}_i$ generated from these $\hat{p}_i$. The Gibbs sampler is then applied as described by Augustin *et al.* (1996, 1998).

The process appeared to converge after about $T = 20$ steps, but as the stochastic variation was large, convergence was difficult to judge. Thus the algorithm was modified by replacing the predicted $\hat{y}_i$ by $\hat{p}_i$ when calculating the autocovariate term. This increased the speed of the algorithm, as a stochastic realization of the distribution was only required in the final iteration of the Gibbs sampler, and stochastic variation was reduced substantially, with clear convergence occurring within $T = 10$ steps. The results presented below used this modification of the algorithm.

Having fitted the autologistic model, it is straightforward to extract the estimated probabilities of occupation by square, and to generate stochastically new presence/absence data from these estimated probabilities. A bootstrap replicate is then obtained by selecting a new random sample of squares, and the above algorithm is applied to it. The whole process is repeated for a large number of bootstrap replicates.

Note that the Gibbs sampler allows iterative generation of simulated data for unsampled squares in a way that preserves spatial correlation. This allows the autocovariate term to be evaluated, and so methods can be applied that would otherwise be restricted to the case in which data are available for all squares. If generation of presence/absence data is always restricted to the unsurveyed squares, and the predictions are supplemented with the observed responses from surveyed squares, then this is analogous to using a finite population correction to estimate the variance of the estimated proportion of occupied squares in sampling theory. We implement this restriction in the analyses presented here.

Figure 4 shows the observed distribution by 1km square of red deer in north-east Scotland. A 20% sample of squares was selected from these data, to assess how well the method could recreate the (known) true distribution. Analysis of deviance for the autologistic model with various definitions for neighbouring squares suggested that a "clique" of side seven was optimal. Thus the autocovariate for square i was the proportion of occupied squares amongst the 48 neighbouring squares in the clique of size 7×7 with square i at its centre.

Figure 5 shows stochastic realisations from the predicted probabilities of presence for the above model and for a simple logistic regression model. It is clear that the autologistic model has captured the degree of clustering exhibited in figure 4 better than the logistic model. The bootstrap was used on both models to quantify the precision of the estimated number of occupied squares, given the 20% sample. Under the autologistic model, this number was estimated to be 237 with standard error 59; under the simple logistic model, the estimated number of occupied squares was 241 with standard error 25. The true number was 190. Results from simple random sampling agreed closely with those under the logistic model, but with a slightly higher standard error. The autologistic model gives a substantially higher standard error because it models the population as if it were a random realization from some super-population, whereas the simple

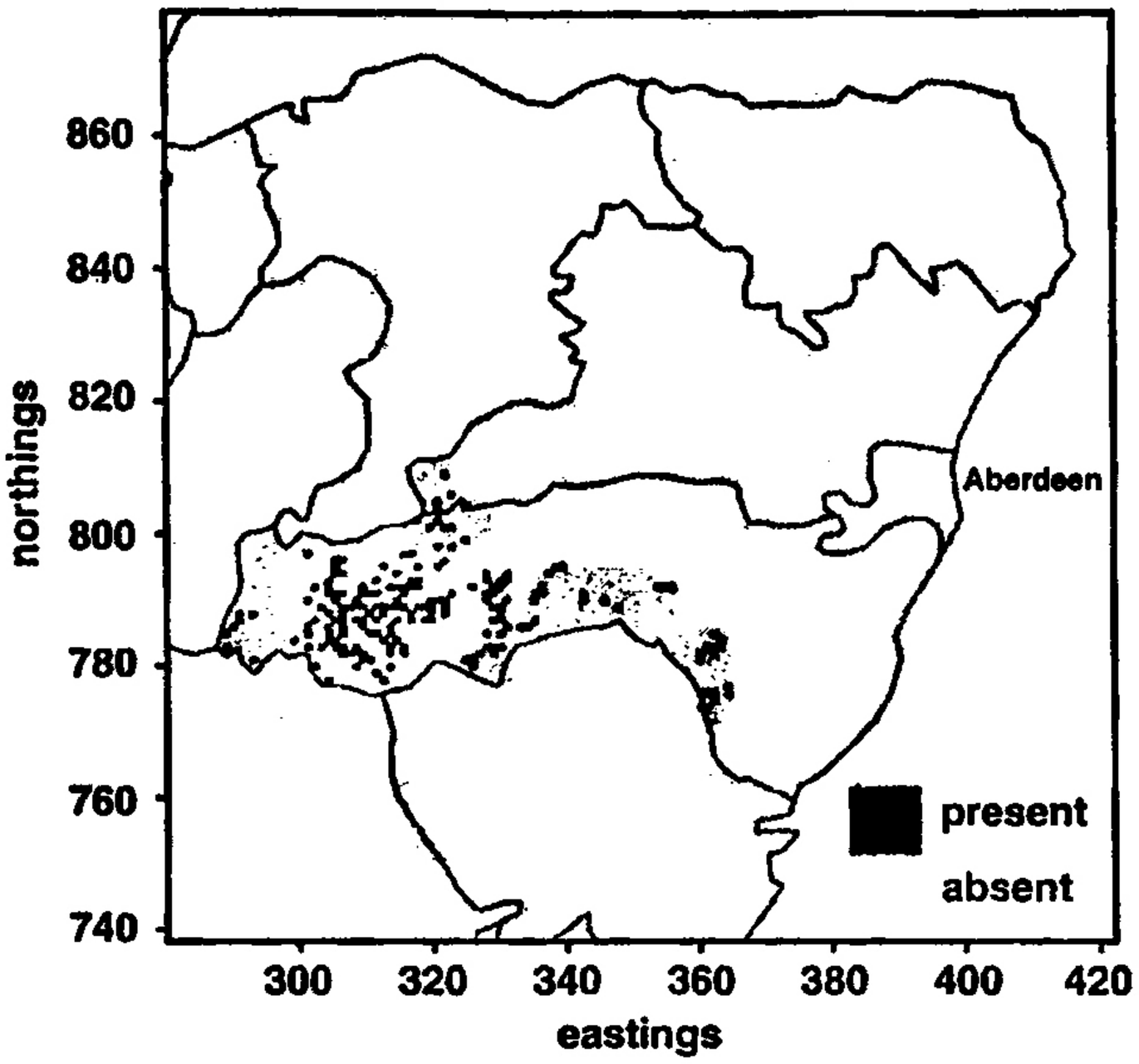


Fig. 4. The observed distribution by 1km square of red deer in north-east Scotland.

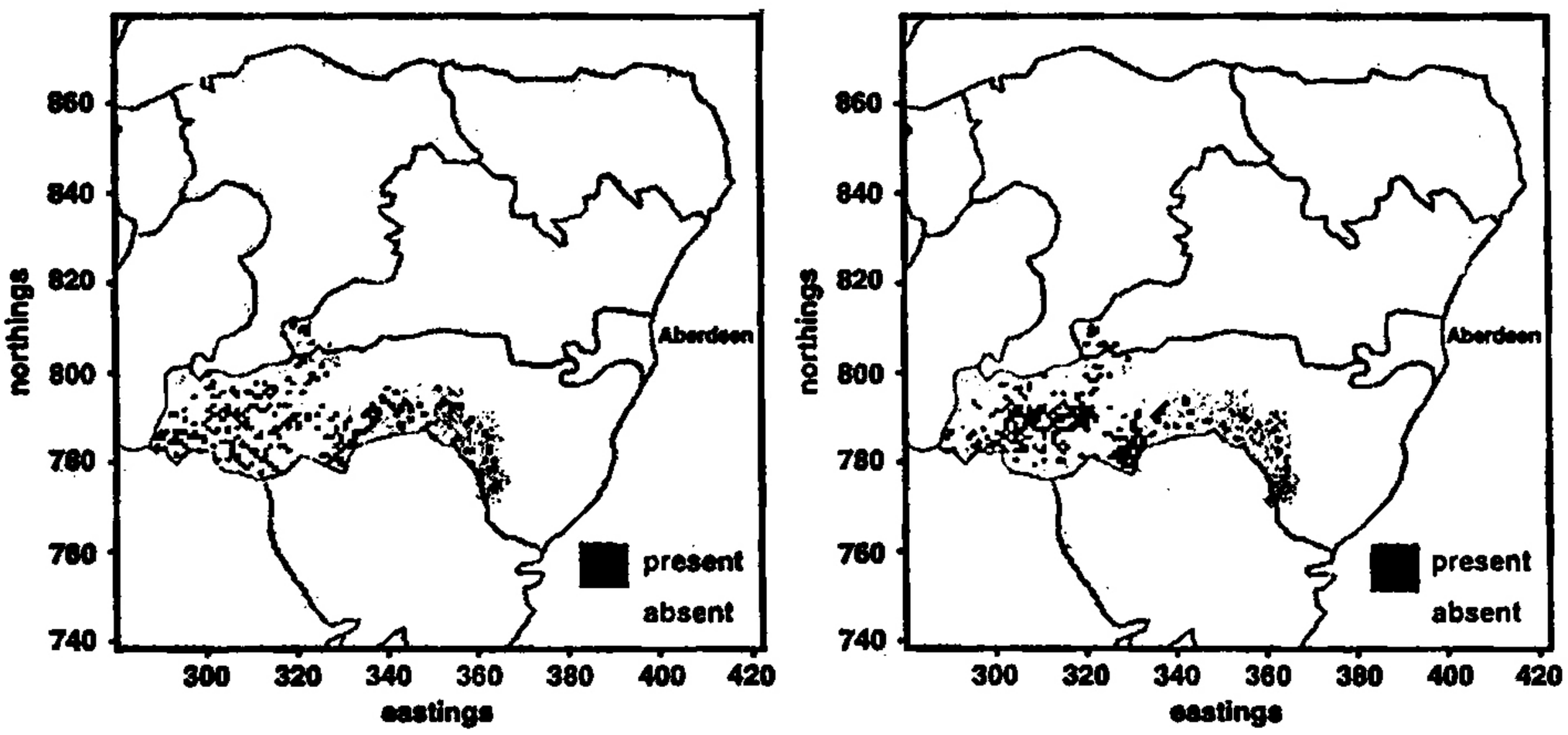


Fig. 5. A stochastic realization generated from predicted probabilities of presence for a simple logistic regression model (left) and an autologistic regression model (right).

logistic model merely quantifies the number of squares occupied at the time of the survey. The latter is thus better for quantifying population size, or other global parameters, whereas the former is able to model local distribution more effectively.

5. Bootstrapping overdispersed counts

Western mackerel and horse mackerel stocks support the largest international fishery in European waters. Spawning stock biomass estimates are essential to sound management of the stocks. In recent years, biomass estimates have been obtained from both the annual egg production method and the daily egg production method. Generalized additive models (GAMs) allow spatial variation in egg density to be modelled, leading to improved precision. We summarize here how an extension of the bootstrap due to Bravington (1993) was used to quantify variances in egg abundance estimates from the daily method, using GAMs, by Borchers *et al.* (1997).

The data from an egg survey comprise counts of eggs in sampled water, in which the volume of water sampled varies. In the mackerel and horse mackerel surveys, the counts ranged from zero to several hundred, so could not be assumed to be identically distributed, ruling out simple resampling from the data. Further, the counts were overdispersed, so that simple parametric bootstrapping from a fitted Poisson model was also ruled out.

The methods used to model the mackerel data were as follows. (The horse mackerel data were modelled slightly differently; see Borchers *et al.*, 1997.) A GAM with log link, overdispersed Poisson error structure, and an offset parameter representing the volume of water sampled (with a correction for bottom depth where this was less than the normal maximum sampling depth) was found to be adequate. The eggs were substantially overdispersed relative to a Poisson distribution, with estimated dispersion parameter of 9.8. However, this model proved superior to one in which the cv was assumed to be constant. Smoothing splines of four covariates were chosen for inclusion in the model selection process. They were latitude, longitude, bottom depth and distance from the 200m contour. Three first order interactions were also found to be significant: linear interactions between latitude and each of the other covariates.

Egg abundance was estimated by integrating over the fitted surface. A parametric bootstrap was used to estimate the variance as follows. In each of 200 bootstrap replicates, a sample was generated, the selected model refitted, and egg abundance estimated by integrating under the newly fitted surface. The variance of these bootstrap abundance estimates is our variance estimate. Generating a bootstrap sample requires the generation of overdispersed Poisson random variables at each sam-

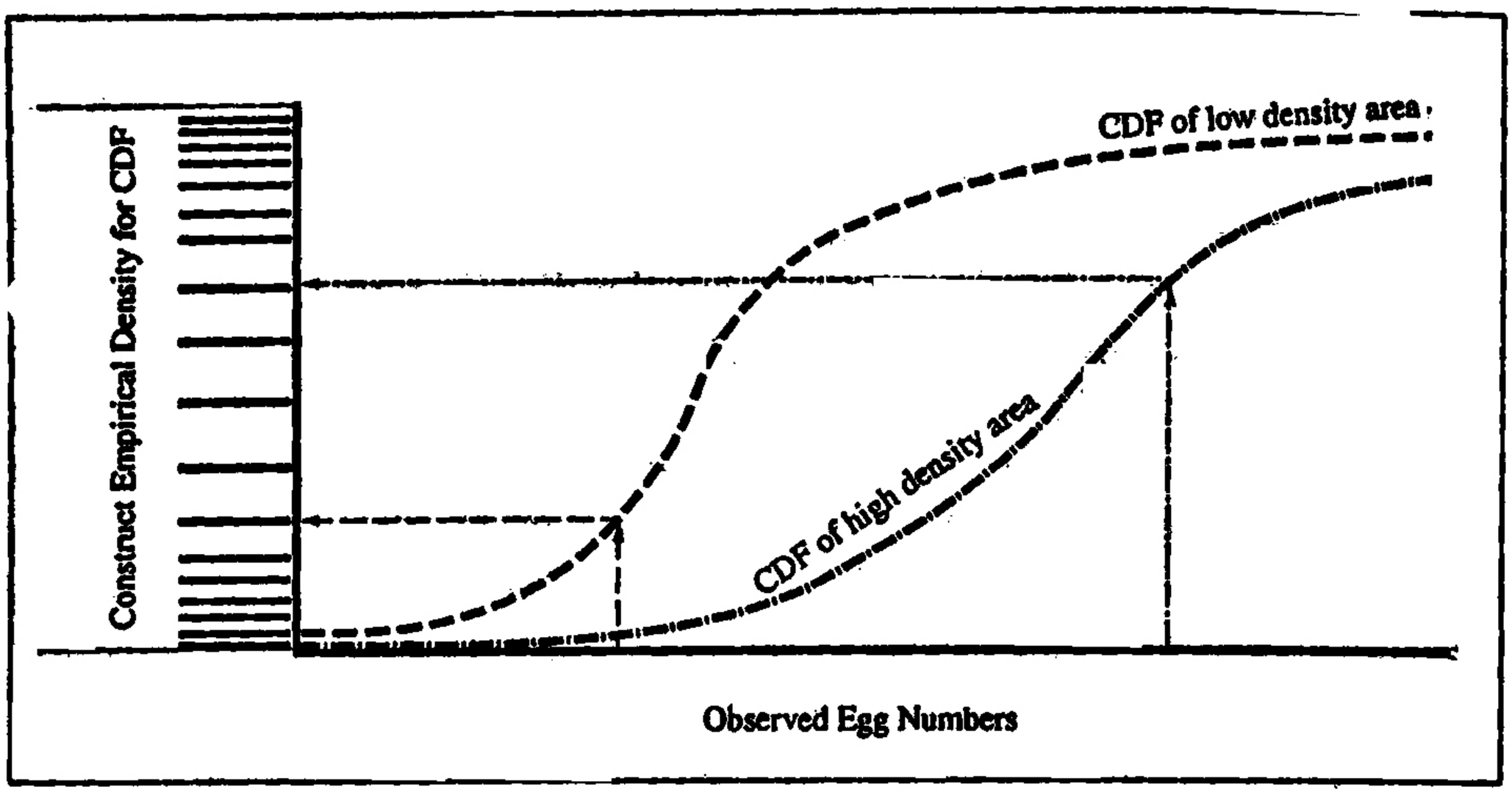
ple point, with mean equal to the modelled abundance at that point. This was achieved using the method of Bravington (1993), which we now describe.

We obtain the predicted value $\hat{y}_i$ from our model, corresponding to observation y_i, and hence estimate that the distribution of y_i is Poisson with mean $\hat{y}_i$. This determines the estimated cumulative distribution function (cdf) of the observation Y_i, which we evaluate at $Y_i = y_i$, to give u_i. In the absence of overdispersion, the resulting values are approximately uniformly distributed on (0,1). (The approximation arises because we estimate $E(y_i)$ by $\hat{y}_i$.) We now resample the u_i, reassign the resampled u's to the predicted values, and back-transform to generate the bootstrap sample of observations. In the presence of overdispersion, the u_i's are not approximately uniform, but instead are more likely to be close to zero or one. The above procedure preserves the overdispersion in the bootstrap resamples.

The procedure is illustrated in figure 6. Figure 6(a) illustrates the process of constructing the empirical distribution of u_i values from the sampled egg numbers using the fitted model. (Two realizations of the fitted model are represented in the figure, the dashed line for a location of low egg density and the dot-dashed line for a location of high egg density.) Overdispersion is illustrated by the clustering of the u_i values of the sampled numbers towards 0 and 1. Figure 6(b) illustrates the resampling process in which the u_i values are randomly permuted among the sampled locations and back-transformed to egg numbers, using the fitted model evaluated at the appropriate location.

The following slight modification to the bootstrap procedure is required when the response y_i is discrete. Instead of using $u_i = \widehat{F}(y_i)$ (where $\widehat{F}(y_i)$ is the estimated cdf at y_i), we choose u_i at random from the interval $(\widehat{F}(y_i - 1), \widehat{F}(y_i))$. Bravington (1993) discusses the modification in more detail.

In our example, the coefficient of variation (cv) of egg abundance was reduced from 10% using stratified sampling to just 3.4% using GAMs, leading to a reduction in the cv of the biomass estimate from 13% to less than 9% (table 1). The variance of the egg abundance estimate contributed 60% of the variance of the biomass estimate when stratified sampling was used, and just 16% when GAMs were used. Similar improvements were achieved for horse mackerel.



b)

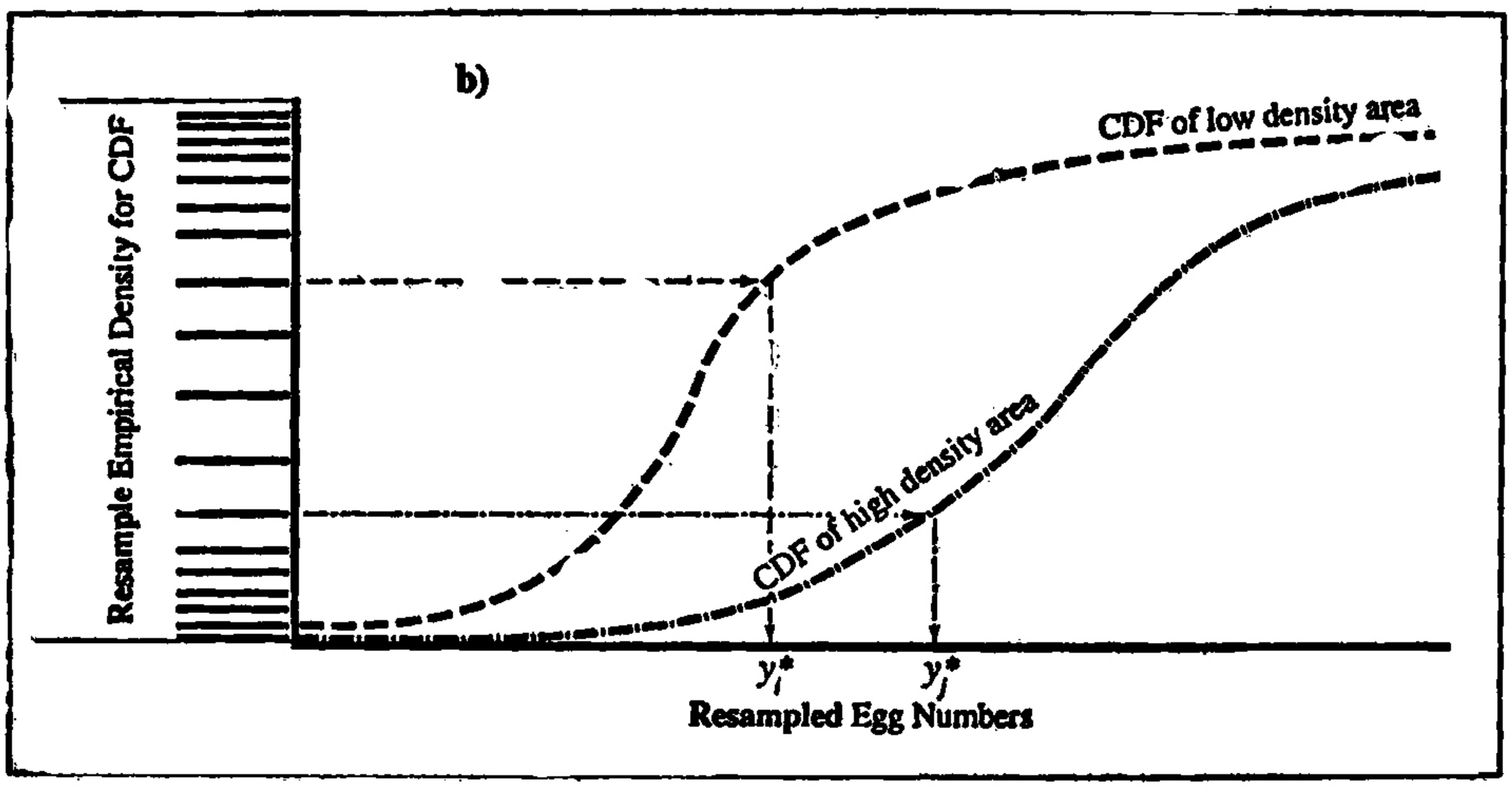


Fig. 6. Diagrammatic representation of the resampling algorithm for generating overdispersed Poisson variates.
Fig. 6(a): Step I. Construction of the empirical cumulative distribution function (cdf). Each observation y is transformed into an "observed" cdf value using the Poisson distribution with mean equal to the GAM fitted value at that point. (Overdispersion results in clustering of the "observed" cdf values towards 0 and 1, as indicated on the vertical axis.) The transformation is shown for two observations only.
Fig. 6(b): Step II. Construction of a new sample from the the empirical cdf. By permuting the "observed" cdf values, resampled cdf values are generated for each of the sampled points. Each of these is then back-transformed using the Poisson distribution with mean equal to the GAM fitted value at that point to yield resampled y's. The back-transformation is shown for the same two observations as in (a) and for the case in which permutation resulted in swapping the two associated cdf values.

TABLE 1: *Estimates by stratum of egg abundance ($\widehat{P}_0$), fecundity ($\widehat{F}$), spawning fraction ($\widehat{S}$) and spawning stock biomass ($\widehat{B}$) of western mackerel. Estimates are shown for both the GAM estimation method and the previously used stratified method. Standard errors are shown in round brackets, and percent cv's in square brackets.*

Stratum	GAM $\widehat{P}_0$ ($\times 10^{12}$)	Stratified $\widehat{P}_0$ ($\times 10^{12}$)	$\widehat{F}$ (eggs/g)	$\widehat{S}$	GAM $\widehat{B}$ ($g \times 10^{11}$)	Stratified $\widehat{B}$ ($g \times 10^{11}$)
Northern	1.20 (0.18) [14.9]	1.20 (0.56) [46.2]	28.7 (2.7) [9.5]	0.467 (0.085) [18.1]	1.79 (0.45) [25.3]	1.80 (0.91) [50.5]
Middle	7.42 (0.34) [4.6]	9.19 (1.54) [16.7]	47.4 (2.2) [4.6]	0.573 (0.055) [9.7]	5.46 (0.64) [11.7]	6.86 (1.36) [19.9]
Southern	10.30 (0.50) [4.9]	13.17 (1.69) [12.9]	49.8 (4.4) [8.7]	0.480 (0.040) [8.4]	8.62 (1.12) [13.0]	11.03 (1.95) [17.7]
Total	18.92 (0.64) [3.4]	23.56 (2.35) [10.0]			15.87 (1.37) [8.6]	19.69 (2.55) [12.9]

6. DISCUSSION

The above examples illustrate the bootstrap as a tool for incorporating model selection uncertainty into inference; a weighted bootstrap, for the case that not all bootstrap replicates are equally probable; a bootstrap method for when observations are not independently distributed; and a bootstrap method for when observations are not identically distributed. Other important extensions of the bootstrap have received little attention to date. For example, if estimation is carried out in two or more stages, analytic methods of variance estimation typically quantify precision at one stage, conditional on any previous stages. By contrast, bootstrap resamples can be generated at every modelling stage, so that a single replicate comprises a sequence of resampling and modelling steps. It thus reflects the uncertainty propagated from one stage to another. Buckland and Elston (1993) gave such a case, for modelling changes in the spatial distribution of wildlife over time. Their first stage was to model habitat suitability of sites for a species using logistic regression. In the second stage, proba-

bility of occupation of a site at one time point was modelled as a function both of these habitat suitabilities and of distance from sites occupied at a previous time point. The precision of predicted distribution maps was quantified using the bootstrap. In a single replicate, the parametric bootstrap was used to generate a resample from the fitted logistic regression model; a new logistic regression was fitted to the resample; the model for spread, fitted by maximum likelihood to the real data, was then used to simulate a "bootstrap" distribution for the second time point; the model for spread was then refitted, using the bootstrap suitabilities and distribution; and finally predictions of future distribution from that model were obtained. Buckland, Elston and Beaney (1996) used the proportion of such bootstrap replicates in which a species became extinct in a given time period as an estimate of the probability of extinction for scarce species.

Another extension of the bootstrap that merits more investigation is the multi-level bootstrap. For example in line transect sampling, it is often unclear whether the sampling unit is more correctly the individual detection (usually assumed for the variance of estimated effective width of search, or of mean cluster size when animals occur in clusters) or the line (usually assumed for the variance of estimated encounter rate). A solution is to resample lines, then resample detections within lines, which is thus a two-level bootstrap. If all lines are traversed by each of a number of observers, then a third level might be added: resample observers, lines within observers, and detections within lines. Program DISTANCE (Laake *et al.*, 1993) allows such analyses.

REFERENCES

AUGUSTIN, N. H., MUGGLESTONE, M. A. and BUCKLAND, S. T. (1996) An autologistic model for the spatial distribution of wildlife, *J. App. Ecol.*, 33, 339-347.

AUGUSTIN, N. H., MUGGLESTONE, M. A. and BUCKLAND, S. T. (1998) The role of simulation in modeling spatially correlated data, *Environmetrics*, 9, 175-196.

BORCHERS, D. L., BUCKLAND, S. T., PRIEDE, I. G. and AHMADI, S. (1997) Improving the precision of the daily egg production method using generalized additive models, *Can. J. Fish. Aquat. Sci.*, 54, 2727-2742.

BRAVINGTON, M. (1993) *The effects of acidification on the population dynamics of brown trout in Norway*, Ph.D. thesis, Imperial College, London. 330 pp.

BUCKLAND, S. T. (1982) A note on the Fourier series model for analysing line transect data, *Biometrics*, 38, 469-477.

BUCKLAND, S. T. (1984) Monte Carlo confidence intervals, *Biometrics*, 40, 811-817.

BUCKLAND, S. T. (1992) Fitting density functions using polynomials, *Applied Statistics*, 41, 63-76.

BUCKLAND, S. T. (1995) Comment on "Inference from a deterministic population dynamics model for bowhead whales" by RAFTERY A. E., GIVENS G. H. and ZEH. J. E., *J. Amer. Statist. Soc.*, 90, 417- 420.

BUCKLAND, S. T., ANDERSON, D. R., BURNHAM, K. P. and LAAKE, J. L. (1993) *Distance Sampling: Estimating Abundance of Biological Populations*, Chapman and Hall, London. 446 pp.

BUCKLAND, S. T., BURNHAM, K. P. and AUGUSTIN, N. H. (1997) Model selection: an integral part of inference, *Biometrics*, 53, 603-618.

BUCKLAND, S. T. and ELSTON, D. A. (1993) Empirical models for the spatial distribution of wildlife, *Journal of Applied Ecology*, 30, 478-495.

BUCKLAND, S. T., ELSTON, D. A. and BEANEY, S. J. (1996) Predicting distributional change of birds in north-east Scotland, *Global Ecology and Biogeography Letters*, 5, 66-84.

BURNHAM, K. P. and ANDERSON, D. R. (1992) Data-based selection of an appropriate biological model: the key to modern data analysis. Pp 16-30, in *Wildlife 2001: Populations*, MCCULLOUGH D. R. and BARRETT R. H. (eds). Elsevier Science Publishers, London.

BURNHAM, K. P., ANDERSON, D. R. and WHITE, G. C. (1994) Evaluation of the Kullback-Leibler discrepancy for model selection in open population capture-recapture models, *Biom. J.*, 36, 299-315.

GATES, C. E. (1979) Line transect and related issues, In *Sampling Biological Populations*, CORMACK R. M., PATIL G. P. and ROBSON D. S. (eds), 71-154. International Cooperative Publishing House, Fairland, Maryland.

GORDON, N. J., SALMOND, D. J. and SMITH, A. F. M. (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings-F*, 140, 107-113.

HJORTH, J. S. U. (1994) *Computer Intensive Statistical Methods: Validation, Model Selection and Bootstrap*, Chapman and Hall, London.

KÜNSCH, H. R. (1989) The jackknife and the bootstrap for general stationary observations, *Ann. Statist.*, 17, 1217-1241.

LAAKE, J. L., BUCKLAND, S. T., ANDERSON, D. R. and BURNHAM, K. P. (1993) *DISTANCE User's Guide V2.0*, Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, CO. 72 pp.

LIU, J. S. and CHEN, R. (1998) Sequential Monte Carlo methods for dynamic systems, *J. Amer. Statist. Soc.*, 93, 1032-1044.

RAFTERY, A. E., GIVENS, G. H. and ZEH, J. E. (1995) Inference from a deterministic population dynamics model for bowhead whales (with discussion), *J. Amer. Statist. Soc.*, 90, 402-430.

SILVERMAN, B. W. (1982) Algorithm AS176: kernel density estimation using the fast Fourier transform, *Applied Statistics*, 31, 93-99.

SMITH, A. F. M. and GELFAND, A. E. (1992) Bayesian statistics without tears: a sampling-resampling perspective, *The American Statistician*, 46, 84-88.

TRENKEL, V. M., ELSTON, D. A. and BUCKLAND, S. T. (1996) *A simulation framework for re- estimation of parameters in a population model for application to a particular locality*, Proceedings in Computational Statistics: 12th Symposium, Barcelona, Spain. pp 477-481. Physica-Verlag.

Simulated inference, with applications to wildlife population assessment

SUMMARY

Large increases in computational power are leading to far-reaching changes in statistical inference. For statisticians, simulation used to be merely a tool for testing the validity of simple results from statistical inference when assumptions fail. Now, it is replacing traditional statistical inference. The practical use of simulated inference far surpasses the importance attached to it in undergraduate degree programmes. In this paper, we illustrate the power and flexibility of simulated inference, using examples from the field of wildlife population assessment. We restrict our attention to extensions of the bootstrap: incorporating model selection uncertainty into inference; use of a weighted bootstrap when there are insufficient data given the complexity of the model; bootstrapping spatial data in the presence of autocorrelation; and bootstrapping overdispersed counts.

Inferenza simulata con applicazioni alla valutazione della popolazione della fauna selvatica

RIASSUNTO

I grandi incrementi di potenza del calcolo automatico stanno portando a grandi cambiamenti nell'inferenza statistica. Nel passato la simulazione è stata utilizzata in statistica soltanto come uno strumento per verificare la validità di semplici risultati ottenuti con metodi inferenziali dai dati empirici, nel caso di non validità delle ipotesi di base dell'inferenza. Oggi i metodi dell'inferenza simulata sono largamente diffusi e spesso sostituiscono i metodi tradizionali. È interessante notare che ciò sta avvenendo anche nell'insegnamento della statistica. Nel presente articolo applichiamo il metodo dell'inferenza simulata a diversi casi di stima della popolazione animale di un territorio. L'attenzione è rivolta ai metodi bootstrap.

KEY WORDS

Bootstrap; Bootstrapping overdispersed counts; Bootstrapping spatial data; Model selection uncertainty; Simulated inference; Weighted bootstrap.

[Manuscript received June 1999; final version received November 1999.]