

9 What Can Bayesian Statistics Do For Archaeological Predictive Modelling?

Andrew Millard¹

9.1 Introduction

The BBO Baseline Report sets out the rationale for developing archaeological predictive modelling². This paper focuses on commentary on the statistical aspects of the proposals and to developing some more detailed proposals about the potential for Bayesian statistical modelling within archaeological predictive models. In the world of predictive modelling for archaeological heritage management it seems to me that we must deal with four *sine quae non*:

- The First Law of Geography³: 'Everything is related to everything else, but near things are more related than distant things';
- What might be called the First Law of Statistics: always inspect your data before analysing them;
- All archaeological samples are biased;
- Our *understanding* of the variables and processes which influence where sites are located is very poor.

Only the last of these might change, and whatever modelling approach we adopt we must be cognisant of all of them.

9.2 Bayesian statistics

Bayesian statistics differs from Classical statistics in allowing the explicit incorporation of subjective prior beliefs into our statistical analysis. The philosophical basis for this cannot be explored here⁴; the mathematical basis is Bayes' Theorem, which may be stated in various ways:

prior knowledge	+	new information	→	revised knowledge
prior belief	×	likelihood	∝	posterior belief
$p(\text{parameters})$	×	$\frac{p(\text{data} \text{parameters})}{p(\text{data})}$	=	$p(\text{parameters} \text{data})$

where $p(\cdot)$ represents the probability of something taking a particular value and $p(\cdot | \cdot)$ represents the probability of something given the truth or value of something else.

If appropriate mathematical forms are chosen for the prior probability and the likelihood, then the posterior probability takes the same form as the prior, and it is possible to simply and directly update our knowledge. This happy situation is known as conjugacy. Although Bayes' theorem was published in the 18th century Bayesian approaches were a minor part of statistical analysis until recently, as most real-world problems cannot be expressed adequately in a conjugate form and calculation of posteriors in non-conjugate problems requires complex, multidimensional integrations.

¹ Department of Archaeology, University of Durham. South Road, Durham DH1 3LE, UK. The author can be reached at A.R.Millard@durham.ac.uk.

² This volume, chapter 3.

³ Tobler 1970.

⁴ But see, for example, Steel 2001.

In the last couple of decades Bayesian statistical analysis has undergone an enormous amount of development as new computational techniques have opened up areas which were previously inaccessible. The main driver has been the development of Markov-Chain Monte Carlo (MCMC) techniques for simulating from a posterior probability distribution. Fundamentally the technique is an algorithm for repeatedly drawing possible values of the variables involved, using random numbers to move from the current set of values to a new set chosen according to the probability distribution. It is not necessary to be able to write down a full mathematical expression for the posterior probability distribution: knowledge of prior probabilities and the likelihood suffices. It can be demonstrated that this leads to a set of samples which are a very good estimate of the required distribution, and they can be plotted as a histogram or summarised as a mean and standard deviation etc. A general overview of the current methods at a reasonably accessible level for the mathematically literate is given by Congdon⁵, but this does not fully cover Bayesian spatial statistics for which more specialist volumes⁶ should be consulted. There is a wide range of methods with applications to continuous spatial data, and to spatially discrete data aggregated at the level of irregularly shaped administrative units such as counties or electoral wards (particularly for the health literature). There are both correlative and explanatory methods, with a complexity well beyond any archaeological spatial model I have seen.

9.3 Why adopt a Bayesian approach to predictive modelling?

... it is impossible to separate opinions (prior beliefs), data and decisions/actions. In the 'classical' approach, our opinions influence our procedures in all sorts of subtle and little-understood ways, for example in choosing the significance level of a hypothesis test. It's better to be as explicit as we can about our prior beliefs, and let the theory take care of how they interact with data to produce posterior beliefs, rather than to let them lurk at the backs of our minds and cloud a supposedly 'objective' belief. This way the Bayesian approach can be more than just a nice piece of mathematics.

(Orton 2003)

The Baseline Report sets out a primary reason for using Bayesian approaches, that is incorporation of expert prior knowledge in a formal and transparent way into the predictive models, thus making them more rigorous and of higher quality⁷. However, there are a number of additional advantages of Bayesian models that address other problems and questions raised in the Baseline Report. A distinction is drawn in the Baseline Report between models with probabilistic and possibilistic outputs⁸. A Bayesian model will be of the former type, which in my opinion has some advantages. Not only can the output be used straightforwardly in testing the model with new data, but the quantitative nature of the statements resulting from such a model can feed into further analyses. For example, in a project covering a large area, statements of the probability of finding sites or particular types of site can be used directly to estimate the likely number of sites and to prepare budgets. However, on smaller development projects, probability statements are less likely to be of direct use and it is more likely that some sort of preliminary investigation ('assessment' in the English Heritage jargon) will be needed. In this case the methodology developed by Orton and Nicholson⁹ can use a probability statement to design an assessment to ensure with a specific confidence level that, if no archaeological remains are found, there are none there. A probabilistic statement thus has the desirable property of being able to directly feed into an algorithm for designing fieldwork strategies for assessment or mitigation.

Another feature of Bayesian models is that they can handle missing observations in straightforward manner, providing that some prior probabilities for the obser-

5 Congdon 2001.

6 Such as Elliott et al. 2000.

7 This volume, pp. 56, 64.

8 This volume, pp. 30-1.

9 Orton 2000a; 2000b; 2003.

vations have been specified. This is likely to be advantageous in modelling and predicting from multivariate archaeological survey data, as one missing value at a survey point does not have to lead to the loss of all the information from other observables at that point, as would happen in many classical statistical analyses (for example, with PCA a full set of measurements for every survey point would be required).

A final advantage is that hierarchical models can be constructed which allow the aggregation of differing but similar classes of samples in a single analysis without assuming that they are identical. This could allow models to be built which account for chronological and functional subsets of sites in the landscape, whilst achieving the sample sizes necessary to achieve useably small posterior confidence ranges.

9.4 Some comments on the past use of statistics in predictive models

Inductive models¹⁰ tend not to have any prior weighting of variables, although we do have some idea before the analysis which are likely to be most important, and indeed the selection of variables is in some cases a judgement of relevance. The weights are derived 'solely' from the training data. Deductive models¹¹ may err in the other direction: there is a prior specification of the weights of the variables in the prediction, but no quantitative method (and sometimes not even a formal method) for updating the weightings in the light of the data. The model is a rigid statement of prior beliefs. It would be much better to combine the two approaches in a system that allows us to specify weightings according to our prior knowledge and expert judgement based on other regions, or other types of sites, and then modify them using observed data for the study area. This is what a Bayesian model can do in one of several ways. One possibility is to assign prior weights explicitly, as in the deductive approach, but adding some statement of uncertainty so that their probability distributions can be modified using the data. An alternative, which is easier to specify, but less flexible to apply, is to use our prior knowledge to rank the variables in order of importance but without specific weights. Training data can then be used to determine the weights subject to maintaining their rank order.

Almost without exception the predictive modelling studies I have seen use a logistic regression to relate multiple environmental and other variables to the probability of site occurrence. I presume that this is because of the ready availability of logistic regression in GIS programs. In fact logistic regression is one of a class of regressions known as *quantal response models* which link a set of variables of any data types via a linear equation to a probability value. In general the equations take the form:

$$\text{link}(p) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n$$

with weights β_i attached to variables x_i . The link function may be any suitable function which maps the zero to one range of p , the probability of site occurrence, to the linear predictor range of $-\infty$ to $+\infty$; it describes the probability distribution of the linear predictor. As well as a logistic distribution using the log-odds link function [$\text{logit}(p) = \ln(p/(1-p))$], two other commonly used distributions are the normal distribution with a probit link function (inverse cumulative normal distribution) and the extreme value distribution with the complementary log-log link function [$\text{cloglog}(p) = \ln(-\ln(1-p))$]. The choice of link function affects the results of the regression in terms of the weights (in an inductive model), model predictions and goodness-of-fit. Hence the use of logistic regression is not as neutral as the Baseline Report, following Gibbon¹², implies when it says that it makes 'no assumptions about the form of the data'. Whilst this is true, assumptions are made about the probability distribution of

10 E.g., Warren & Asch 2000.

11 E.g., Dalla Bona 2000.

12 Gibbon 2002.

the linear predictor derived from the data, and, importantly, about the independence of the variables. More sophisticated quantal response models modify the linear predictor to allow for 'interaction terms' of the form $\beta_{ij}x_i x_j$, or explicitly allow for the covariance of variables via covariance of the β_i . Archaeological predictive models (Bayesian or Classical) could benefit from these modifications, to allow, for example, for the interaction of slope and soil type in preventing/enabling settlement.

Using a linear predictor also has other limitations, which necessitate invoking the first law of statistics and examining the data before it is used. A continuous or interval variable with a bimodal distribution will fit badly in a linear model with a unimodal link function. Further, some of our data may be circular in form, notably aspect. If aspect is measured as an angle from north and a north aspect is favoured for settlement then the distribution will be bimodal at small and large angles (close to 0° and 360°). Circular data might be adequately modelled via a transformation, e.g. using $\sin(\theta)$ and $\cos(\theta)$ as predictors, but it would be even better to drop the linear equation and use circular statistics such as a von Mises distribution.

In other cases the assumption that, all other things being equal, the probability of a site is symmetrically distributed with respect to a variable does not hold. For example, how does the probability of settlement depend on depth to the water table? It is clearly zero for large negative depths (i.e. deeply submerged areas), it might be small for small positive and negative depths representing areas where pile dwellings are possible, and then as dry ground is reached it increases to a constant value regardless of whether the water table depth is 1m or 100m. This is poorly described by a probability function which is not zero at any depth, peaks at a particular depth and then declines as large depths are reached.

Two previous papers have considered the possibilities of a Bayesian approach to archaeological predictive modelling. Van Dalen uses a simple application of Bayes theorem to modify the predictions of an 'at random' location model using a set of variables¹³. His method arrives at zero probability for many cells because it fails to account for the stochastic nature of the sample, and thus the possibility that sites are present on a particular soil type but have not been observed due to the (small) sample size. The method as applied also only works for variables of categorical and ordinal scales, but could be combined with a quantal response model for interval and ratio variables. Van Dalen's other proposal is that it is not always the immediate qualities of the site location, but also those of nearby locations which influence its position. He attempts to account for this by a form of kernel density estimate creating 'buffers' around sites in a geometric model. Kernel density estimates could equally be applied in his Bayesian approach¹⁴.

Verhagen proposes to use a beta distribution to model prior expert opinion on the proportion of sites within the study area falling in a particular land unit¹⁵. This has the advantage of being conjugate to a binomial likelihood based on the number of sites inside and outside the unit, so that the calculation of a posterior probability distribution in the form of a revised beta distribution is very easy. However, although there are multiple land units containing between them a certain number of observed sites, Verhagen treats them separately. The posterior beta distributions derived for the land units do not take account of the multivariate nature of the problem. If one draws values of the proportions from the entire set of beta distributions then there is no guarantee that they will have the required sum of unity. A better approach would be to model the numbers of sites in the set of land units as a multinomial distribution, where the conjugate prior is a Dirichlet distribution. The same prior parameters as Verhagen derived for the beta distributions can be combined as the set of parameters of the Dirichlet, and the marginal posteriors remain the same.

13 Van Dalen 2000.

14 Compare the work of Lucy et al. 2002.

15 Verhagen 2001.

9.5 On autocorrelation and covariance in predictive models

As repeatedly noted in the Baseline Report, archaeological predictive models have consistently failed to account for spatial autocorrelation in the predictor variables and thus observations are assumed to be independent, covariance is ignored, and the predictive power of the model is overestimated. When this is done, locations similar to known site locations are assigned probabilities of being sites which are too high and locations similar to known non-sites are assigned probabilities which are too low. The nature of the problem is well illustrated by the methods and data of Warren and Asch¹⁶. In their survey area they report that there are 59 identified sites, but in their grid of 5473 cells to be modelled as site or non-site, there are 265 site cells. Each site therefore occupies ~4.5 cells, but this appears not to be fully accounted for in the predictive modelling. Although the training set is chosen with cluster sampling to avoid the problem of cells from a given site occurring in both training and test datasets, no account is taken of the fact that each training site contributes several data points. The possibilities for improving the predictive power of models will be limited until spatial autocorrelation is taken into account.

If time is to be considered within a predictive model either as archaeological periods or more precisely as dates, then temporal autocorrelation might need to be considered as well, as sites tend to persist at one location. Accounting for this in a model with high temporal resolution will introduce a good deal more complexity. In constructing a model we need to find a balance between the level of detail needed for planning purposes and the complexity of producing the model.

When environmental variables are mapped it is not usually with the aim of recording them at archaeological sites. For archaeological purposes it is therefore often necessary to use interpolated variables (e.g. in a digital elevation model constructed from irregularly spaced survey points) as part of the predictive process. The process of interpolation assumes the existence of autocorrelation and uses it to interpolate, hence interpolated data is inherently autocorrelated. This autocorrelation is (or should be) deducible from the interpolation algorithm. Further transformations of interpolated data will introduce further complications in the structure of the autocorrelation, and if they require the use of more than one interpolated point (e.g. in going from elevation to slope) then additional autocorrelation is introduced. If we can, we should account for these autocorrelations in predictive modelling.

The Baseline Report suggests¹⁷ that to deal with spatial autocorrelation we could use techniques such as PCA to 'de-correlate' our set of variables, at the expense of real-world interpretability. It seems to me that we should embrace the autocorrelation and use it via explicitly spatial statistics. We are better off modelling our autocorrelation than trying to get around it with a generic tool whose assumptions may not match the properties of our data, and which transforms the data into a form to which we cannot apply our understanding of past processes. For example, we have sufficient knowledge of medieval settlement patterns to know that if a village occurs at a particular location, then there is unlikely to be another village within 2 or 3 km, and that area will likely have been occupied by fields. As distance from the known village increases from 3 km, there is an area with an increased chance of finding another village, but as distance increases to tens of kilometres, the location of the known village becomes less relevant to the probability of finding another village. This pattern of anti-correlation, correlation and declining correlation with distance would be difficult to capture in any generic tool. We might however attempt to represent it by kriging with an appropriate semi-variogram (see the section on 'modelling autocorrelation' below).

One interesting aspect of spatial correlation is raised in the Baseline Report in the guise of *complex types*¹⁸, which are essentially statements of the spatial

¹⁶ Warren & Asch 2000.

¹⁷ This volume, p. 67.

¹⁸ This volume, pp. 35-6.

correlation of different archaeological components. Their definition is complicated, hence the suggested need for an expert system, but in terms of a predictive model the resulting statements may be translated into probability statements which can be combined with statistical analysis of the location to predict the likelihood of a site occurring, and perhaps surviving. A simple example could be that a find of a quern-stone implies a high probability of the presence of a domestic site, but this statement should be qualified by the other characteristics of the location.

9.6 A conceptual model for the processes of archaeological site location and discovery

In order to have an effective predictive model we need to make sure that the underlying ideas are as good as we can make them, clearly articulated and open to discussion and improvement. In this vein I present a possible process model for the location, preservation and discovery of archaeological sites (figure 1). Each of the steps in the process must be modelled if we are to use our biased archaeological samples and incomplete knowledge of the factors determining site location to make inferences about the buried archaeological record. In order to make inferences several components must be in place:

- 1 We must have a list of relevant variables which is at least fairly complete; the list given in figure 1 is intended to be illustrative and far from exhaustive
- 2 We must model how these variables influence site location. The current default model, which I have critiqued above, is a logistic regression with linear predictor. However for some variables we may be able to do better, at least with representing the general form of the dependence.

The other components have received less attention from archaeological predictive modellers, but seem to me to be equally key to what we want to do:

- 3 We need to model the taphonomy and survival of sites. This depends on some of the same variables as in the list of (1) but certainly includes additional variables, including later human activities. For example, soil type and topographic variables might satisfactorily predict the location of a Neolithic long-house, but its survival will depend on the location (in three dimensions) of post-medieval cellar-digging and 5th century BC coastal erosion.
- 4 If we had perfect or even good models at this stage then this purely deductive approach might suffice, but in reality we need to calibrate, adapt and test our models using observations of the presence and absence of archaeological remains. These observations may in some cases be truly representative or random samples derived from dedicated surveys, but in a heavily populated and researched country like the Netherlands, the majority of the data will be non-representative in a variety of ways. If we can get some handle on the observational biases, then we should include an observation component in our model to relate the 'preserved sites' to the 'observed sites'.

In practical terms for archaeological heritage management we may not need to be very sophisticated in our modelling of the relationships of points (2), (3) and (4), as is shown by the fact that predictive models in the past have been reasonably successful with simple linear predictor functions. The form of the relationships is the province of enquiry for explanatory models, and whilst correlative models may draw on the understanding from such enquiry they may get away with forms of relationship which are workable rather than realistic¹⁹. That said, the more realism we can build into our models the less we will run into various problems of bias and error in the results.

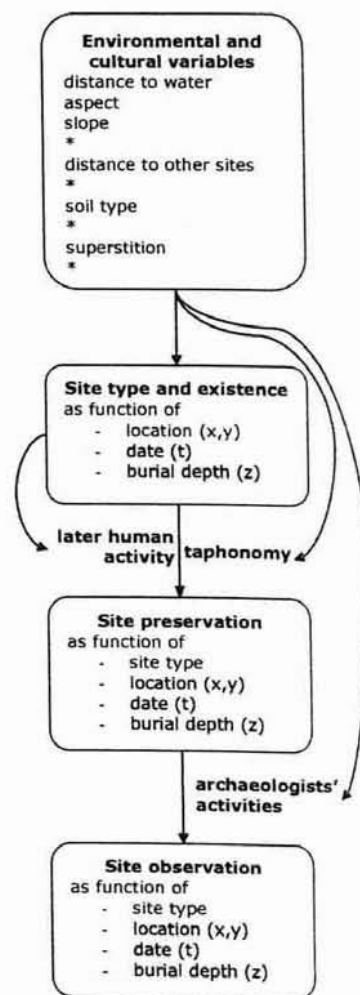


Fig. 1 Process model for the creation, preservation and observation of sites.

9.7 A Bayesian model for the processes of archaeological site location and discovery

What might a Bayesian statistical version of the above conceptual model look like? I will expound this with a highly simplified model, and indicative equations only, even so it will appear fairly complicated. In each of the subsections that follow, the first part expounds the ideas behind the model and may suffice as an explanation for non-mathematical readers. The notation is explained in table 1. It is easiest to consider the processes in reverse order, followed by aspects of covariance and autocorrelation. At each stage in the modelling the use of a Bayesian hierarchical model allows us to write equations which assume that we have perfect knowledge from the next stage, but ultimately the model includes the uncertainties from each stage.

Table 1 Summary of notation in section 9.7.

Variable	Meaning
t	time or date in the past
x, y, z	coordinates in three-dimensional space
Discovery	
$D(x,y,z)$	the discovery of a site at a particular place, xyz , ($D=1$) or failure to discover a site ($D=0$)
$S(x,y,z,t)$	the survival of a site at a particular place, xyz , from a particular time, t , ($S=1$) or non-survival of a site that once existed ($S=0$)
T	site type
U	modern land use
ϕ	mean probability of finding a site given that it survives
α_U	weighting for land-use U used to modify ϕ to give the actual probability of finding a surviving site
Survival	
$E(x,y,z,t)$	the existence of a site at a particular place, xyz , at a particular time, t , ($E=1$) or absence of a site ($E=0$)
P_R	probability that a site has survived natural destructive processes
P_C	probability that a site has survived destructive human activities
R	a collection of variables related to natural destructive processes
C	a collection of variables related to destructive human activities
β_R	coefficients in a logistic regression relating R to p_R which may depend on site type and date
Site type and location	
p_V	probability of a site being created at a specific place and time given environmental variables
p_z	probability of a site being created at a given vertical position (a complex function of location and date)
V	the set of environmental variables that determine site location
β_V	coefficients in a logistic regression relating V to p_V which may depend on site type and date
Modelling the parameters of the process	
β_i	the coefficient for a specific environmental variable for a particular site type and date
β	the collection of all β_i s for a particular site type and date
ψ_i	the mean value through time of β_i for a specific environmental variable and particular site type
ψ	the collection of all ψ_i s for a particular site type
M_i	the mean value over all site types of ψ_i for a specific environmental variable
M	the collection of all M_i s
σ_β	the standard deviation of the values of β_i for an environmental variable and site type (constant across all such type-variable pairs)
σ_ψ	the standard deviation of the values of ψ_i for an environmental variable (the same for all such variables)
Σ_β	covariance matrix for values of β_i (subsumes values of σ_β)
Σ_ψ	covariance matrix for values of ψ_i (subsumes values of σ_ψ)
Autocorrelation	
$\epsilon_R(x,y)$	deviation from mean when predicting p_R from R at a particular location
$\epsilon_V(x,y)$	deviation from mean when predicting p_V from V at a particular location
ϵ	the collection of values ϵ_R and ϵ_V across all locations of interest
0	vector of zeroes representing the mean values of ϵ
τ	covariance matrix expressing the autocorrelation of ϵ in space

Archaeological site discovery and recording

The process that we ought to know most about is how archaeologists discover and record sites that have survived from the past. In fact we seem to have little knowledge of this, probably because we (and our predecessors) have not bothered to think carefully and record *how* we do that part of archaeological practice which is discovering sites. As we have a large amount of data already available to us, it is worthwhile investigating how we got it and thus how we can use it. We know that different areas have been prospected to greater or lesser degrees, and can model this. A very simplified model might say that archaeologists have worked in most areas of the Netherlands, but because prospecting is easier in open countryside than in wooded or urban areas, the probabilities of a site having been found when it survived in open country, woodland and urban areas differ. We do not know the absolute values of these probabilities but prior estimates of the ratios of these parameters could be derived from experts using the methods outlined by Verhagen²⁰. Other factors that might be similarly modelled include soil type, and the site's period and type. Ultimately we write model equations for the chance of a site's discovery, relating this to information about its survival and type, and the modern land use. More formally we have an equation for $p(D(x,y,z) | S, T, U)$, (where D=site discovery, S=site survival, T=site type, U=modern land use, etc) and prior probabilities for the parameters in the equation. For the example above, we might have

$$p(D(x,y,z) | S(x,y,z,t), T, U, \phi, \alpha_U) = \phi \times \alpha_U$$

where: ϕ is the mean probability of finding a site, which we might guess as being small but not too small and write a prior of $\phi \sim \text{Beta}(1,3)$; and α_U is a weighting depending on land use with a prior estimate given by the appropriate term from the vector α given by $\alpha \sim \text{Dirichlet}(a)$, with the values of a derived from expert opinion. Where some areas have been subject to sampling by systematic survey, ϕ and α should be modelled separately for those areas.

The survival of sites

Again this is an under-researched area. We have some basic ideas about destructive processes that remove sites, such as cellar digging, gravel extraction, ploughing and erosion. Several EU funded projects have investigated the effect on various archaeological materials of soil and other environmental factors²¹, and their data might be useful in establishing the variables relevant to damage by slow, natural processes. As these processes are so poorly understood we probably have to fall back on modelling their effects via a quantal response model. We end up with an equation for the probability of site survival expressed in terms of information about the existence of the site in the past, the variables that relate to natural destructive processes and human activity on the site since it was created.

We seek to calculate $p(S(x,y,z,t) | E, T, \mathbf{R}, \mathbf{C})$ where E is the existence of a site in the past, \mathbf{R} are the variables relating to natural destructive processes and \mathbf{C} are variables relating to human activity and hence include $E(T, t')$ for later periods than t . If we approximate natural processes and later human activities as independent, then:

$$p(S(x,y,z,t) | E, T, \mathbf{R}, \mathbf{C}, \beta_R, \epsilon_R) = p_R \times p_C$$

$$\text{with } \text{logit}(p_R) = \beta_R(T,t)^T \mathbf{R} + \epsilon_R(x,y).$$

²⁰ Verhagen 2001.

²¹ *The most recent being bone: Kars & Kars 2002; Kars 1998.*

The coefficients β_R in the logistic regression of R are given as a function of site type and date, as this will alter the effects of the natural processes. The modelling of these parameters is most effectively done in a hierarchical manner, for which see below. The probability relating to later human activity, p_C , will be more complex in expression as activities which give rise to sites will be modelled in the next stage, but activities such as ploughing will not. The error term $\epsilon_R(x,y)$ is discussed below as it incorporates structure accounting for spatial autocorrelation.

The location and type of sites

This is the part of the model whose structure is of interest to explanatory modellers. The possibilities here are to adopt some sort of explicit model with specified relationships between the variables and the probability of a site, or a non-parametric relationship such as a quantal response model. Land evaluation is suggested in the Baseline Report as a possible explicit method for modelling land-use on the basis of ecological and socioeconomic data and confronting these with archaeological data to predict activity areas²². Land evaluation has never actually been developed sufficiently for predictive modelling, though if it could be made to work it would be very useful for giving realistic (rather than arbitrary) mathematical forms to the variable-site link in a process model. For the moment I simply offer the common logistic quantal response model, but with the addition of hierarchy:

$$p(E(x,y,z,t), T | V, t, z, \beta_V, \epsilon_V) = p_V \times p_Z$$

$$\text{with } \text{logit}(p_V) = \beta_V(T,t)^T V + \epsilon_V(x,y).$$

The depth at which a site is found depends on where it is and the period of the site, so p_Z is a complex function of t and (x,y) . Modelling of this will depend on the palaeo-geographical modelling for the Holocene areas of the Netherlands. Note, however, that only a probabilistic statement is required, which allows for areas where our knowledge is uncertain. The prior distribution of t could reflect expectations of the different relative numbers of sites created at different items/periods. The error term $\epsilon_V(x,y)$ is discussed below as it incorporates structure accounting for spatial autocorrelation.

Summary of model equations

A summary equation for posterior probabilities of the parameters in the full model where we have site/non-site observations is:

$$\begin{aligned} p(\beta_V, \beta_R, \epsilon_R, \epsilon_V, \phi, \alpha, V, R, C | D(x, y, z, t), T, U) \propto \\ p(D(x,y,z) | S(x,y,z,t), \phi, \alpha, U) \times p(\phi) \times p(\alpha) \\ \times p(S(x,y,z,t) | E, T, \beta_R, \epsilon_R, R, C) \times p(\beta_R) \times p(\epsilon_R(x,y)) \\ \times p(E(x,y,z,t), T | t, z, \beta_V, \epsilon_V, V) \times p(t) \times p(z) \times p(\beta_V) \\ \times p(\epsilon_V(x,y)) \times p(V(x,y)) \times p(R(x,y)) \times p(C(x,y)) \end{aligned}$$

The final line allows for our sets of variables influencing site location and survival to be incompletely known for a location: we then need to give some prior probabilities for the various possible values.

²² This volume, pp. 62-4.

Finally the real interest in a predictive model is not the discovered sites but the surviving sites, so that we can maximise our site discovery rate. We use the posterior estimates of the parameters derived above to predict site survival at a new location (x',y') where we have no observations, integrating out the parameters relating to site discovery:

$$\begin{aligned}
 p(S(x',y',z',t'), T', t' | \mathbf{V}, \mathbf{R}, \mathbf{C}, \beta_V, \beta_R, \epsilon_R, \epsilon_V) = \\
 \int_{\phi, \alpha} p(S(x',y',z',t') | E, T', \mathbf{R}, \mathbf{C}, \beta_R) \\
 \times p(E(x',y',z',t'), T' | \mathbf{V}, t, z, \beta_V, \epsilon_V) \\
 \times p(z, \mathbf{V}, \mathbf{R}, \mathbf{C}, \mathbf{U}, \beta_V, \beta_R, \epsilon_R, \epsilon_V, \phi, \alpha | D(x,y,z,t), T) d\phi d\alpha
 \end{aligned}$$

Hierarchical model for the parameters of the process

It is in the modelling of the parameters of the process which relate the observable variables to the chances of site survival ($\beta_V, \beta_R, \phi, \alpha$) that a Bayesian approach has one of its major advantages. This approach allows us to use prior knowledge to increase the precision of our posterior, as has been discussed above and elsewhere²³. It also allows the modelling of the many possible combinations of variables, site types and periods in more complex ways than either treating them all as independent (there is rarely enough data to do this), or lumping all sites and periods together (which we know is unrealistic). This alternative is a hierarchical model which recognises that there is a close relationship between the factors influencing sites of the same type but different periods, whilst allowing for some differences between them. It considers that there is a *distribution* of parameters, for sites of a particular type but different periods, so that information from one period gives us some (but less) information about other periods.

We need to specify prior probability distributions which express our prior beliefs about similarities between periods and site types respectively. If we decided that the similarities are primarily within periods and then secondarily between periods this hierarchy would be reversed. More complex hierarchies can also be devised, for example, if we believe that for the post-medieval period the use of new methods of land reclamation makes the distribution of sites utterly different to preceding periods, then the relevant sub-set of parameters can be modelled with a separate prior. Similarly we might decide that the location of 'religious' sites is influenced by our variables in a different way to settlement sites, in which case $\psi_i(\text{religious})$ has a separate prior. However the number of special cases should not be multiplied too far, or we lose the beneficial effects of the hierarchy. A simple and often adopted approach is to model the distribution of values of the coefficients as independent and normally distributed, so that

$$\beta_i(T, t) \sim \text{Normal}(\psi_i(T), \sigma_\beta)$$

where values of i represent the various possible environmental variables. We may further believe that all types of sites are related in their parameters, so that

$$\psi_i(T) \sim \text{Normal}(M_\psi, \sigma_\psi)$$

with prior probability distributions for M_ψ, σ_β and σ_ψ .

The modelling of these parameters can also allow us to account for temporal correlation between variables. Instead of treating the parameters as independent, we can model the whole set of parameters as drawn from a multivariate distribution with an explicit statement of covariance:

23 Verhagen 2001; Orton 2000a, 2000b.

$$\beta(T, t) \sim \text{MVNormal}(\psi(T), \Sigma_\beta)$$

$$\psi(T) \sim \text{MVNormal}(\mathbf{M}, \Sigma_\psi)$$

where now our prior distributions for Σ_β and Σ_ψ include terms for the covariance of variables. Our prior for \mathbf{M} could have further structure linking the mean values of parameters as well as their variances.

The elicitation of prior information on this hierarchical structure could become complicated. However, it might suffice to obtain information on the relative strengths of the relationships between the influence of variables across site types and periods, and then introduce a scaling factor with a very broad prior distribution to represent our lack of knowledge of the absolute values.

Object-oriented GIS²⁴ would appear to offer a similar view of hierarchy in certain circumstances, but OOGIS seems to be little developed and not to offer anything in the way of spatial statistics which acknowledge its hierarchical view of objects.

Modelling autocorrelation

I have argued above that site locations are spatially autocorrelated. To model this we have to allow that the random errors $\varepsilon_R(x,y)$ and $\varepsilon_V(x,y)$ in our regression equations above are not independent at all points, but are more similar to the values at nearby points. This is done by specifying a relationship between their values at different points. Two popular methods (of the many available) are:

- kriging, which assumes that the covariance between any two points j and k in continuous space depends on a function (the semi-variogram, γ) of the distance (h_{jk}) between them. A variety of models are possible depending on the choice of semi-variogram function. We then write

$$\varepsilon \sim \text{MVNormal}(\mathbf{0}, \tau), \quad \text{with } \tau_{jj} = \sigma^2 \quad \text{and } \tau_{jk} = \sigma^2 - 2\gamma(h_{jk})$$

- simultaneous autoregressive methods for discrete areas, which specify the co-variance between areas in terms of a spatial correlation, ρ , and a weighting matrix, W . Variety in these models derives from the possibilities for defining the weights, which may be based on distances, shared boundary lengths or simply whether areas are neighbours or not. The equations are rather more complicated than for kriging:

$$\varepsilon \sim \text{MVNormal}(\mathbf{0}, \tau), \quad \text{with } \tau = \sigma^2(I - \rho W)^{-1}(I - \rho W^T)^{-1}$$

where I is the identity matrix.

I am unclear about which expression of spatial covariance would be more suitable for archaeological predictive models. Kriging certainly has simpler calculations, and might suffice for a first approximation of spatial autocorrelation.

Model implementation

The model outlined here is fairly complex statistically and would need specialist research to ensure that the necessary MCMC computations could be carried out efficiently. It might be that simplifications could be made to the discovery sub-model (e.g., reducing it to a series of beta distributions) and site survival could be rolled into the same modelling equations as site creation. In doing

this we would need to recognise that pragmatic simplifications are being made; they would be the target for the next round of model improvement.

9.8 Testing models

Statistical modellers recognise that all models are imperfect representations of the real world and are therefore subject to the problem of mis-specification to a greater or lesser degree. The mis-specification is allowed for, together with random variation that cannot be modelled, by the variance terms included in the equations for the error terms. Part of the aim of model criticism, testing and refinement is to reduce the size of these terms. This may be done in simple models by examining the difference between predictions and observations to see if there is correlation of the residuals with some variable, which may or may not have featured in the prediction.

More complex models are tested and refined with a variety of procedures. How well they fit to the data may be expressed using some measure of fit, and models that fit too badly may be rejected (analogous to the use of p -values in classical statistics). Such measures of fit may be used simply for rejection of bad models, but they may also be used to help in model criticism and alteration in developing a model, or in comparing models. Data may be used to test the model directly as well, either by withholding some data from the 'training' data for the model, or by collecting new data and seeing how well it is predicted. Withheld data is of course available within the context of the IKAW revision, but new data will be collected in the future though not specifically for model testing. Some new data might be particularly useful for testing aspects of the model. For example, resurvey by the same method of an area will provide a test of the discovery model and lead to refinement of its parameters. With an explicit discovery model in place we can incorporate new information in a more rigorous way even though it is not collected with the intention of model testing. Future samples for testing need not be acquired by random or even representative sampling. The requirement becomes simply that our specification of how they were acquired is much more careful than it has been in the past.

Model comparisons depend on measures of fit, but may be useful in a variety of ways. They can be used within an overarching model to compare what happens when sub-models are altered, for example, removing a hierarchical description, or simplifying some part of the model. In these circumstances they may indicate that a model is more complex than is necessary or than is justified by the nature of the data. Model comparisons are also useful for comparing different prior estimates of the parameters, if we believe that the differences between different prior estimates are not reconcilable or if we do not know what mathematical form to give to the uncertainty in those estimates.

9.9 Looking forwards

Statistics, whether Bayesian or classical, is just a tool. The 'big' questions are the archaeo-political ones. By separating out and creating a sound methodology for dealing with the technical issues, we can focus attention on these 'big' issues, such as: What is a 'significant archaeological remain'? What is an acceptable risk of its unrecorded destruction? How do these answers vary according the location, period and nature of the remains?

(Orton 2003)

References

- Congdon, P. 2001: *Bayesian statistical modelling*. Chichester: Wiley.
- van Dalen, J. 2000: Probability modeling: A Bayesian and a geometric example, in M. Gillings, D. Mattingly, & J. van Dalen (eds), *Geographical Information Systems and Landscape Archaeology* (The Archaeology of Mediterranean Landscapes 3). Oxford: Oxbow, 117-24.
- Dalla Bona, L. 2000: Protecting cultural resources through forest management planning in Ontario using archaeological predictive modeling, in: K. Wescott & R. Brandon (eds), *Practical applications of GIS for archaeologists: a predictive modeling toolkit*. London: Taylor & Francis, 73-99.
- Elliott, P., J. Wakefield, N. Best, & D. Briggs (eds) 2000: *Spatial epidemiology: methods and applications*. Oxford: Oxford University Press.
- Gibbon, G. 2002: Archaeological predictive modeling: an overview, in G. Hudak, E. Hobbs, A.S. Brooks, C. Sersland, & C. Phillips (eds), *A predictive model of precontact archaeological site location for the State of Minnesota*. St Paul (MN): Minnesota Department of Transportation, <http://www.mnmodel.dot.state.mn.us/TOC.html>.
- Kars, E., & H. Kars (eds) 2002: *The degradation of bone as an indicator for the deterioration of the European archaeological property*. Amersfoort: Rijksdienst voor het Oudheidkundig Bodemonderzoek.
- Kars, H. 1998: Preserving our in situ archaeological heritage: a challenge to the geochemical engineer, *Journal of Geochemical Exploration* 62, 139-47.
- Lucy, D., R.G. Aykroyd, & A.M. Pollard 2002: Nonparametric calibration for age estimation, *Applied Statistics* 51, 183-96.
- Orton, C. 2000a: A Bayesian approach to a problem of archaeological site evaluation, in K. Lockyear, T. Sly, & V. Mihailescu-Birliba (eds), *Computer Applications and Quantitative Methods in Archaeology CAA 96* (British Archaeological Reports International Series 845). Iasi and Oxford: Editura Demiurg and Archaeopress, 1-7.
- Orton, C. 2000b: *Sampling in archaeology*. Cambridge: Cambridge University Press.
- Orton, C. 2003: *The Fourth Umpire: Risk in archaeology*. Inaugural Lecture, University College London, <http://www.ucl.ac.uk/archaeology/special/orton-inaugural-2003/index.htm> (Accessed 9 May 2003).
- Steel, D. 2001: Bayesian statistics in radiocarbon calibration, *Philosophy of Science* 68, S153-64.
- Tobler, W. 1970: Review of J Forrester 'Urban Dynamic', *Geographical Analysis* 2, 198-9.
- Tschan, A. 1999: An introduction to object-oriented GIS in archaeology, in J. Barceló, I. Briz, & A. Vila (eds), *New techniques for old times - CAA98. Computer Applications and Quantitative Methods in Archaeology* (British Archaeological Reports International Series 757). Oxford: Archaeopress, 303-16.

Verhagen, P. 2001: *Quantifying the qualified: the use of multi-criteria methods and Bayesian statistics for the development of archaeological predictive models*. Paper presented at: GIS and archaeological predictive modelling: larger scale approaches to establish a baseline for site location models, Argonne, IL 21-24 March 2001.

Warren, R., & D. Asch 2000: A predictive model of archaeological site location in the Eastern Prairie Peninsula, in K. Wescott & R. Brandon (eds), *Practical applications of GIS for archaeologists: a predictive modeling toolkit*. London: Taylor & Francis, 5-32.