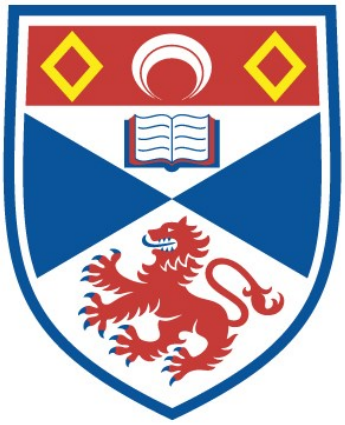
SPATIAL AND SPATIO-TEMPORAL MODELS WITH APPLICATIONS IN VEGETATION DYNAMICS AND WILDLIFE POPULATION ESTIMATION

Nicole Helene Augustin

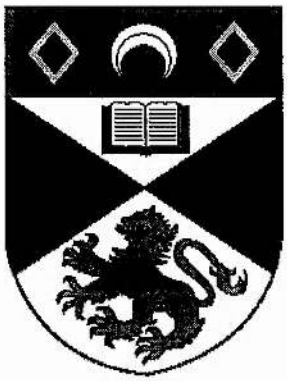
A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1999

Spatial and spatio-temporal models with applications in vegetation dynamics and wildlife population estimation

Nicole Helene Augustin



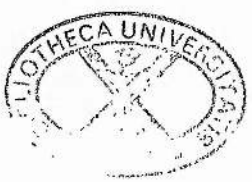
Thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

in the School of Mathematical and Computational Sciences,

UNIVERSITY OF ST ANDREWS.

August, 1999.



TL D412

Declarations

1. I, Nicole Helene Augustin, hereby certify that this thesis, which is approximately 43,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

   date 31.8.99 signature of candidate .....

2. I was admitted as a research student in February 1994 and as a candidate for the degree of PhD in February 1995; the higher study for which this is a record was carried out in the University of St Andrews between 1994 and 1999.

   date 31.8.99 signature of candidate .....

3. I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

   date 31/8/99 signature of supervisor .....

4. In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

   date 31.8.99 signature of candidate .....

Abstract

This thesis applies spatial and spatio-temporal modelling to two broad areas of environmental statistics: wildlife abundance estimation and vegetation dynamics.

The first methodology considered is *spatial modelling for estimating global characteristics* through predicting the value of a response variable at new locations. The approach is based on generalized additive models and illustrated using spatio-temporal fisheries survey data. The method incorporates historical data to overcome shortcomings in the survey design. The GAM-based method substantially improves the precision of estimates over a traditional estimation method and is also useful in explaining complex space-time trends using environmental variables.

The second methodology addressed is *spatial modelling for the description of the underlying process.* Its objectives lie in exploring local properties, such as autocorrelation. Auto-models (Markov Random Fields) are used for modelling discrete data. Autocorrelation is estimated directly from the response, as a fixed effect, through the specification of a conditional probability of each observation, given its neighbouring values. The auto-Poisson model for counts has traditionally been restricted to the modelling of negative autocorrelation. This restriction is overcome by right truncating the Poisson distribution. Further modifications of this model are also investigated. Parameter estimation methods for this truncated auto-Poisson model are then compared via a simulation study. The method with accompanying model selection and validation techniques is illustrated for the auto-Poisson and auto-negative binomial model using seed and mite counts. An example of modelling the presence and absence of deer illustrating the auto-logistic model for binary data is also presented.

Finally, methodology for *spatio-temporal modelling of the underlying process* is considered. The use of transition models for modelling change of semi-natural vegetation in Scotland is investigated. The transition model is extended to incorporate spatial effects and it is shown that estimates of transition probabilities for Markov models can be improved.

Acknowledgements

I would like to thank the following people. Firstly, my supervisors Stephen Buckland and James McNicol for their encouragement, patience and help. Moira Mugglestone, my former supervisor, for her encouragement. My colleague Maria Durbán Reguera for her help and many stimulating discussions. My friends and colleagues at the Statistical Ecology Group in the Maths department at St Andrews University and at Biomathematics and Statistics Scotland (especially the Dundee group) for many helpful discussions. Also, the IT staff at the Scottish Crop Research Institute for invaluable computing support, especially Philip Smith for proof reading. My colleagues from the various countries involved in the mackerel project for collaboration. Gladys Wright from the Scottish Crop Research Institute for making the seed count data available. Roger Cummins, Donald French, David Elston and Howard Grubb for inspiring discussions and comments on the work presented in chapter 6. Finally, many thanks to my family and Sam for their encouragement and support.

This work was partly funded by the Scottish Executive, Rural Affairs Department. The research on spatio-temporal modelling of mackerel eggs presented in chapter 2 was supported by the Commission of the European Communities Directorate General for Fisheries (DG XIV) project No. 94/107. The financial support is greatly appreciated.

Contents

Chapter 1

General introduction

In environmental studies there are two broad areas, apart from epidemiology and toxicology, for which spatial and spatio-temporal modelling are relevant, namely wildlife abundance estimation and modelling vegetation dynamics. This thesis describes applications from both areas. Spatial modelling might be carried out for at least two fundamentally different purposes, leading to different conceptual approaches. The first approach can be described as *spatial modelling for estimating global characteristics* through predicting the value of a response variable at new locations. We illustrate this with an example from fisheries abundance estimation. The second approach is more concerned with *spatial modelling for the description of the underlying process.*

Besides abundance estimation there are many instances where the first approach is appropriate, e.g. using data from a few sample points of a forest to estimate the mean tree health status throughout the forest, or using disease incidence data in epidemiology to estimate total incidence for an entire region. As a by-product predicted maps of tree health status throughout the forest or maps of expected incidence for the region are obtained. In addition such spatial modelling aids exploration of the relationships between spatially referenced response variables and explanatory variables.

For the first approach of predicting the spatial distribution, empirical modelling, where parameters do not necessarily have a meaningful interpretation, is sufficient. We use generalised additive models (GAMs) (Hastie and Tibshirani, 1990) for prediction and assume that observations are independent. We present an application of fish egg population esti-

mation in Atlantic mackerel (*Scomber scombrus*) (chapter 2). At the onset of the thesis we believed that autocorrelation could be eliminated by using very flexible models, such as GAMs. In our application in chapter 2 this turned out not to be the case and therefore the approach of modelling the spatio-temporal mackerel egg abundance data for stock assessment is not a statistically rigorous spatio-temporal model; it is merely a step in the development of robust methods for modelling the spatio-temporal trend in abundance.

The objectives of the second approach, spatial modelling for the description of the underlying process, lie in exploring the interactions between neighbouring responses, which can be competitive or co-operative. This also involves estimating the nature and extent of spatial interaction. There are a number of different methods involving explicit forms of the spatial auto-correlation structure in the modelling process. Even in the case of spatial independence, these types of models can obtain a better fit than models which do not account for autocorrelation, because they may capture the influence of spatial explanatory variables which could not be measured because of cost or other practical reasons. Hence these models are well suited for prediction of spatial distributions. Many of the existing approaches for this task have been developed for use with continuous data. These methods are often inappropriate for use with counts, presence/absence indicators, or categorical data which commonly arise in environmental studies. Here we have concentrated on models for discrete data. This 'process model approach' is presented in chapters 3 to 5, where auto-models (Besag, 1972) for discrete data are investigated. In chapter 6 we investigate the use of transition/Markov models (Agresti, 1990; Diggle et al., 1994) for modelling change of semi-natural vegetation in Scotland. This method aims to improve estimates of transition probabilities for Markov models, often used in ecological simulation studies.

In chapters 3 to 6 we emphasise suitable models for describing vegetation dynamics processes, though applications presented are also from wildlife abundance estimation. A literature review on vegetation dynamics models reveals that most of the models are mathematical; they can be either stochastic or deterministic but are usually not directly based on data. They describe the dynamics of a process, with the aim of investigating different scenarios. Examples of such models applied to vegetation dynamics include cellular automaton models (Green, 1989; Jeltsch et al., 1996), models using differential equations (Tongeren and Prentice, 1986), neighbourhood population dynamic models (Pacala, 1986),

spatial patch dynamic models (Wu and Levin, 1994) and Markov models (Lerzman, 1995; Scanlan, 1994; Childress et al., 1998). In some of the above approaches statistical methods are only used for calibration purposes, that is to estimate the model input parameters. There is broad scope for improvement of 'ecological process models' by incorporating data and we hope that chapters 3 to 6 have contributed to the improvement of ecological process models.

Chapter 2

Modelling spatial correlation using locational covariates

2.1 Introduction

In this chapter we present an approach for modelling spatial data which assumes that values of the response variable at different locations are independent. This approach is often used in wildlife distribution mapping. Regression methods in the form of generalised linear models (GLMs) or generalised additive models (GAMs) are applied under the assumption that values of the response variable at different locations are independent (Walker, 1990; Osborne and Tigar, 1992; Buckland and Elston, 1993). In these examples, counts or presence/absence indicators for a particular plant or animal species, which are obtained by dividing a region of interest into a regular grid and visiting some or all of the squares in the grid, are regressed on spatially referenced covariates. The covariates are typically measurements of physical and habitat-related characteristics drawn from a geographical information system (GIS). If covariate data are available for all the squares in a grid then a model fitted to the response in surveyed squares can be used to obtain predictions for all squares, even those which were not surveyed. A complete map of the spatial distribution of a species can then be produced. If abundance or density is estimated in surveyed squares, the map can also show densities throughout the area. Integration under the density surface allows abundance to be estimated within any area of interest. However, these models only explore relationships between response and covariates: they do not help in

estimating effects of spatial autocorrelation. If there is no intrinsic spatial autocorrelation in the data present, these models are sufficient. Inclusion of locational covariates such as eastings and northings (or longitude and latitude) can possibly eliminate most of the spatial correlation in the residuals, especially when generalized additive models (GAMs) are applied. These covariates may act as proxies for unknown covariates that are causally related to the response.

In section 2.2 we describe GAMs and present an application of GAMs for modelling spatio-temporal abundance data. In this example GAMs are used to estimate the spatial distribution of the density of mackerel eggs in the North Atlantic. The aim is to improve the precision of the total biomass of mackerel which is estimated indirectly from the total annual egg production (annual egg production method). This method requires multiple egg surveys during the spawning period, so that the total production of eggs can be estimated. Previously no spatial information was used to obtain this estimate. Here egg density is modelled as a function of spatial covariates and time, exploiting the fact that each egg survey is not carried out instantaneously. Since there is not full spatial coverage of the survey area throughout the spawning season, confounding of time and space presents a major challenge in these data. We show how, by constraining the model in time and space using historical stock boundary data, we can interpolate and extrapolate to time points and locations where no data are available. Using the GAM we can model complex trends in density with respect to space, time, and other explanatory covariates. In addition the changing spatial distribution of mackerel eggs can be monitored in time. This application is also published in Augustin et al. (1998a).

In section 2.3 we discuss the dangers of overfitting the data when such flexible models as GAMs are used and show how resampling methods for simulating data allow model selection uncertainty to be incorporated into estimates of precision. This work is published in Augustin et al. (1998b).

2.2 Spatio-temporal modelling for the Annual Egg Production Method of Stock Assessment using Generalized Additive Models (GAMs)

2.2.1 Introduction to the use of GAMs in modelling fisheries data

Generalized additive models (GAMs) provide a powerful and flexible statistical tool for modelling spatio-temporal distributions of animal abundance from survey data. They have previously been applied to relate fish survey or abundance data to locational and environmental covariates, e.g. in Swartzman et al. (1992, 1994, 1995); Swartzman (1997); Welch et al. (1995) and Maravelias and Reid (1997). Wood and Horwood (1995) relate sole egg abundance data to locational variables only using thin plate splines. All of these applications have the emphasis on applying GAMs to explore relationships between abundance and environmental and/or locational variables; they do not attempt to produce reliable abundance estimates. Swartzman et al. (1992) give abundance estimates for Bering sea groundfish survey data, but find substantial downward bias in their estimates compared with estimates using design-based methods. Deriso et al. (1996) estimate an abundance index for Pacific sardine using GAMs, but only use relative abundance estimates. Borchers et al. (1997b) go one step further and use GAMs to model egg density as a function of locational and environmental variables to obtain egg abundance estimates for the daily egg production method (DEPM). Here we want to improve upon the traditional annual egg production method (AEPM) by modelling egg density survey data as a smooth function of space, time and oceanographic variables and thus improve precision of egg abundance estimates. The AEPM and DEPM estimate total fish biomass indirectly from an estimate of the total annual egg production and daily egg production respectively (see for details Gunderson, 1993; Anon., 1994). In contrast to commercial catch-based methods, they provide (like acoustic surveys) a fishery independent estimate of stock size. The AEPM requires egg plankton surveys of the spawning area at multiple time points through the spawning season, so that egg production can be estimated at each time point of the spawning season. In the DEPM, egg production is only estimated at a single point in time. While a single survey with sufficient spatial coverage, ideally at a single point in time, is adequate for the application of GAMs to the DEPM, spatio-temporal modelling of egg abundance for the AEPM requires adequate coverage of the survey area in both

space and time. Logistically, it is impossible to cover the whole survey area instantaneously at any given time, and effects of space and time will always be confounded to some degree. We present GAM-based methods for spatio-temporal modelling which are more robust to the effects of this sort of confounding by incorporating historial data for spawning boundaries. We illustrate the methods using the 1995 egg survey data of the western stock of Atlantic mackerel (*Scomber scombrus* Linnaeus). When modelling egg production the random variable observed is egg count. Although we are interested in egg production, egg count as a response variable in the model is a natural choice and there are several distributions which can be used for such a response variable. The conversion of counts into egg production estimates can be handled by specifying an offset comprising the conversion factors required. Swartzman et al. (1992) use counts in their GAM as the response variable, but obtain estimates 30 to 50% lower than estimates from design based methods. Borchers et al. (1997b) use counts as response in their GAM for the daily egg production method (DEPM) and their abundance estimate for Atlantic mackerel is 20% lower than the estimate obtained from the traditional method. We illustrate how such a count model in the presence of sub-sampling can lead to biased estimates from the GAM-based methods and how this bias can be avoided. Finally we compare the performance of the GAM-based AEPM with the traditional AEPM.

2.2.2 Methods for fish egg abundance estimation

The traditional egg production estimation method (AEPM)

In the traditional AEPM, egg surveys are carried out over a number of periods through the spawning season and estimates for each period are assigned to the mid point of the period. The egg production curve is estimated by plotting these estimates against time and joining the estimates, either by straight lines or by using a nonlinear function such as cubic smoothing splines (Gunderson, 1993). Total annual egg production (E_a) is estimated by integrating under the curve.

The area of spawning is usually divided into strata and random samples are taken in each stratum and in each survey period. Samples are taken by lowering a plankton net down to 200m depth. Then the net is towed diagonally through the water-column over several nautical miles before it is retrieved (Figure 2.1). The multiple egg surveys carried

out throughout the spawning period are designed to provide full spatial coverage of the spawning area in each survey period. Mean daily egg production is estimated for each stratum and survey period, for example by estimating egg production per grid square, and summing these estimates over the grid squares in the stratum. The estimated total daily egg production by period is then obtained by summing over strata.

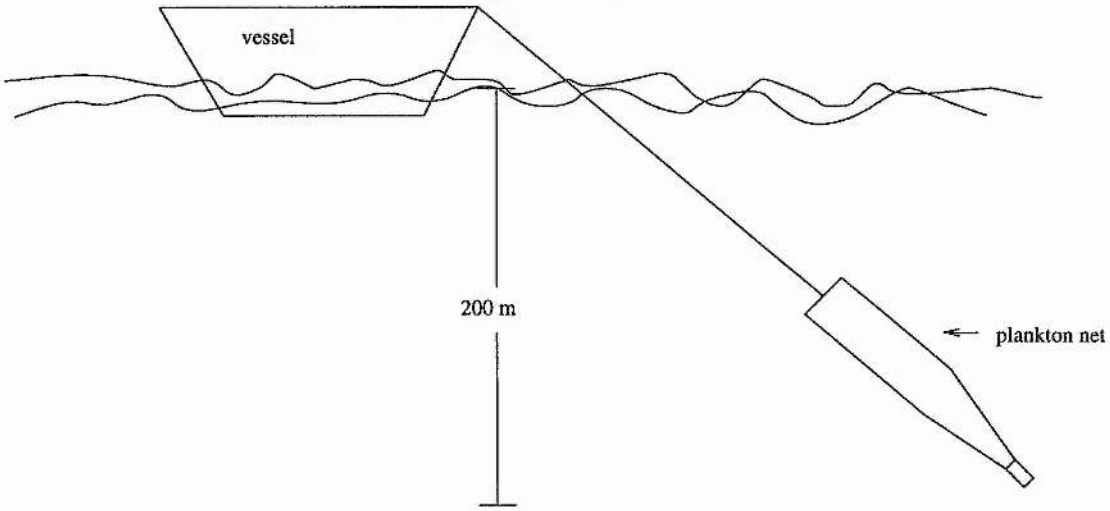


Figure 2.1: Schematic representation of the plankton net being towed through the 200 m water-column.

Variance estimation is based on the assumption that observed egg numbers are distributed with a constant coefficient of variation (cv) in space and time (Anon., 1996). When the egg production curve is assumed to be piecewise linear, as is the case with the AEPM as applied to Atlantic mackerel, the estimated variance for total annual egg production (E_a) neglects the component of variance due to estimating the shape of the egg production curve.

The GAM-based AEPM

Although the traditional method treats data from each survey period as if they came from a single point in time (the survey period midpoint), sampling within each period is spread over time. Here we treat time as continuous and model egg density as a function of location, oceanographic variables and time, using GAMs. GAMs are an extension of generalized linear models (GLMs) (McCullagh and Nelder, 1989). Both accommodate a

variety of distributions for the response, but unlike GLMs, GAMs allow flexible nonlinear effects of the covariates on the response to be estimated from the data. GAMs have the following general form (Hastie and Tibshirani, 1990):

$$E(y) = g^{-1}\left(\beta_0 + \sum_k S_k(x_k)\right)$$

The function $g(.)$ is the link function, which defines the relationship between the response and the linear predictor, $\beta_0 + \sum_k S_k(x_k)$. The response, y, is assumed to be distributed according to one of a wide family of statistical distributions. The term x_k is the value of the kth covariate, such as location or date, β_0 is an intercept term, and $S_k(.)$ is a one dimensional smoothing function for the kth covariate. The degree of smoothing performed by a smoother $S_k(.)$ is determined by the number of neighbouring points used for the smoothed estimates. We use cubic spline smoothers for $S_k(.)$, where the degrees of freedom (df) associated with the smoother control the amount of smoothing; the fewer the degrees of freedom, the greater the smoothing. For example, $df = 1$ corresponds to a linear effect of the associated covariate, as in a GLM. For count data a common choice is a Poisson distribution with a log link function. The choice of distribution is intrinsically determined by the assumed relationship between y and its variance. For example, the Poisson distribution assumes the variance is equal to the mean ($Var(y) = \mu$, where μ is the expectation of the response); an overdispersed Poisson distribution assumes the variance exceeds of the mean ($Var(y) = \sigma^2\mu$, where $\sigma^2 > 1$ is the dispersion parameter); and the negative binomial (NegBin) distribution assumes $Var(y) = \mu + \mu^2/k$. The negative binomial distribution allows for more flexibility in the variance of the response than the Poisson distribution through the term μ^2/k . The parameter k can be estimated from the data separately and kept fixed. It is also called the aggregation parameter. The greater k, the more similar the negative binomial distribution is to the Poisson distribution.

GAMs for egg density

Three possible types of response are: egg density, egg counts or egg presence/absence in a sample. The last of these is not illustrated here, but is useful in cases where zero eggs are observed over a large part of the survey area. Then egg density can be modelled in two stages, with one model for the presence/absence of eggs (stage one) and a second model for nonzero egg counts (stage two). This procedure is described by Borchers et al. (1997b).

When it comes to choosing the response in the GAM for modelling egg density it is useful to envisage the sampling process. The eggs obtained from a plankton sample are easily

counted, but to determine to which of the five identifiable stages each egg belongs is more time consuming. Therefore, subsamples for egg staging are typically taken in high density areas whereas in low density areas all sampled eggs might be staged. For different plankton samples there will be a similar number of eggs in the subsample, but the egg density in the sampled areas will vary considerably. If several vessels are involved in the survey, there might be other differences in the sampling process, e.g. longer hauls will result in higher egg counts, as will the use of a plankton sampler with a larger opening.

Traditionally the number of eggs counted (n_{ij}) at a given stage is converted into egg density ($density_{ij}$) per day per m^2 by multiplying by the following conversion factor ($conv_{ij}$):

$$conv_{ij} = \frac{ns_{ij}}{nsub_{ij}} * \frac{depth_{ij}}{volume_{ij}} * \frac{1}{duration_{ij}} * mort.adj.$$

where

- i is the index for location,
- j is the index for sampling day,
- ns represents the number of eggs of all stages in the sample,
- $nsub$ represents the number of eggs of all stages in the subsample,
- $depth$ is the depth of the water column sampled,
- $volume$ is the water volume sampled,
- $duration$ is the duration of the selected stage and
- $mort.adj.$ is an egg mortality adjustment

(see Gunderson (1993) for further details).

Density model

If we choose *density* as the response variable, the best choice of distribution for the response is not obvious. A normal distribution with an identity link can lead to negative estimates of density in some localities at some occasions and while the actual distribution of egg density is often appreciably skewed, the normal is symmetric. A Poisson, gamma or negative binomial distribution might therefore be more appropriate. Although the Poisson distribution is only appropriate for discrete counts, the variance function of the Poisson

distribution might fit the data best. The model can then be fitted via quasi-likelihood, which only requires the specification of the variance function (McCullagh and Nelder, 1989).

Our model for expected egg density of a given stage in the sample from location i and at day j with $j = minday, ..., maxday$, ($minday$ and $maxday$ are the first and last days with data available), is as follows:

$$E(density_{ij}) = \exp\left(\beta_0 + \sum_k S_k(x_{kij}) + \sum_{k=1}^{K}\sum_{l>k}^{K} S_{kl}(x_{kij}, x_{lij})\right) \tag{2.1}$$

where a log link function is used and

- exp(.) is the inverse of the log link function,
- x_{kij} represents the kth covariate at the ith location on day j,
- β_0 is the intercept parameter,
- $S_k(.)$ is the smoothing spline for the kth covariate and
- $S_{kl}(.)$ is the smoothing spline for the interaction term of the kth and lth covariate.

Given estimates $\hat{\beta}_0$, $\hat{S}_k(.)$ and $\hat{S}_{kl}(.)$ $(k = 1, ..., K; k < l \leq K)$, the expected egg density $\hat{E}(density_{ij})$ at a given point in space and time with covariates x_k (k=1,...,K) can be calculated.

Count model

In the absence of sub-sampling, the preferred choice is to model the number of eggs observed in the sample n_{ij} from location i at day j (y_{ij}; j=$minday$,...,$maxday$) using a GAM with an appropriate distribution (e.g. Poisson, gamma or negative binomial) and link function. However, in areas of high egg density the sample of eggs is typically sub-sampled to save time. Thus the volume of water sampled corresponding to a given count tends to be small when egg density is high. The assumed mean-variance relationship for counts does not reflect the mean-variance relationship of egg density. GAMs use the variance of the response (counts in this case) to construct weights in the iterative fitting procedure. When the error distribution is Poisson and the log link is used, the weights are equal to the mean response. This response is systematically reduced in areas of high

density by the sub-sampling procedure, so that too little weight is attached to data from such areas and this leads to bias.

By moving the term $conv_{ij}$ from $density_{ij} = n_{ij} * conv_{ij}$ in equation 2.1 to the right hand side we obtain the count model:

$$E(n_{ij}) = \exp\left(offset_{ij} + \beta_0 + \sum_k S_k(x_{kij}) + \sum_{k=1}^{K}\sum_{l>k}^{K} S_{kl}(x_{kij}.x_{lij})\right) \tag{2.2}$$

where $offset_{ij} = -\log(conv_{ij})$ and a log link is used.

Given estimates $\hat{\beta}_0$, $\hat{S}_k(.)$ and $\hat{S}_{kl}(.)$ $(k = 1, ..., K; k < l \leq K)$, the expected egg density at a point with covariates x_k $(k=1,...,K)$ is given by setting the *offset* term to zero:

$$\widehat{E(density_{ij})} = \exp\left(\hat{\beta}_0 + \sum_k \hat{S}_k(x_{kij}) + \sum_{k=1}^{K}\sum_{l>k}^{K} \hat{S}_{kl}(x_{kij}.x_{lij})\right)$$

Integration over space and time

Once a model has been selected, the GAM provides a smooth expected egg density surface which is integrated numerically over space, within the survey area, and time, within the spawning period, to provide an estimate of the total egg production in the survey area (Figure 2.2). Moreover, by integrating over the appropriate spatial and temporal limits, egg production can be estimated at any spatial and/or temporal resolution. In particular, the method can provide estimates of daily egg production at any given day in the spawning period over the whole spawning area, as well as an estimate of annual egg production (E_a).

In order to integrate the estimated egg production surface over space and time, the surface must be evaluated at points other than those at which samples were obtained. Explanatory variables which are not well defined except at sampled points and times (duration of haul and sampling depth, for example) present difficulties in this regard and should be excluded as candidate covariates. Time-dependent covariates are also only available at the sampled points and times. In order to use these covariates in integrating the egg production surface to obtain the egg production curve, they would have to be modelled as functions of space and time. This would add another dimension to the modelling problem, which we do not address.

Dealing with sparse survey coverage

Unless the spawning area is small relative to sampling resources available, surveys will

rarely cover the complete spawning area. If observed egg densities at the outer limits of the sampled area are greater than zero, spatial and temporal extrapolation into unsampled regions of the spawning area can lead to positive bias. This can be avoided by defining outer boundaries beyond which spawning is known not to occur and constraining the fitted models to be zero (or nearly zero) at those points. Often historical data help to define these outer boundaries of spawning. In most cases these spatial boundaries will vary over time. Inadequate coverage in the time dimension can be dealt with similarly. The fitted models can be constrained at the beginning and the end of spawning by specifying times before which spawning would not have begun and after which spawning will have ended. Artificial data points with zero egg density along the spatial boundaries, and scattered over the whole survey area at the temporal limits, are added to the data. The additional data points are referred to as "structural zeros" and are used in estimation, model selection and variance estimation. In model selection the contribution of structural zeros to the deviance is removed before comparing models.

Model selection

Model selection with GAMs involves choosing a distribution for the response, a link function, an appropriate set of covariates and appropriate degrees of freedom for smoothing the covariates. We used plots of deviance residuals standardized by their estimated variance versus fitted values to examine the appropriate link function and distribution used. To examine whether the spatial distribution was adequately modelled, we plotted averages of deviance residuals by grid square and month.

Variance estimation

The variance of the annual egg production estimator ($\hat{E}_a$) can be estimated using parametric bootstrap procedures. This involves generating b pseudo samples of the egg survey data using the fitted model, and refitting the GAM to each of these pseudo samples. For instance, for a model using the negative binomial distribution, a pseudo sample at location i and at day j is produced by sampling from $\text{NegBin}(\hat{\mu}_{ij}, k)$, where k is fixed and μ_{ij} is the fitted value. Integrating the predicted surface of each resample over space and time yields b bootstrap estimates for E_a. The variance of these bootstrap estimates is the estimate of the variance for the GAM estimate. Bootstrap percentile confidence intervals (Efron and Tibshirani, 1993) can also be constructed from the b bootstrap estimates. The bootstrap procedure can generate variance estimates for egg production estimates at any

spatial and/or temporal resolution.

Figure 2.2 gives an overview of the estimation procedure for the GAM-based AEPM.

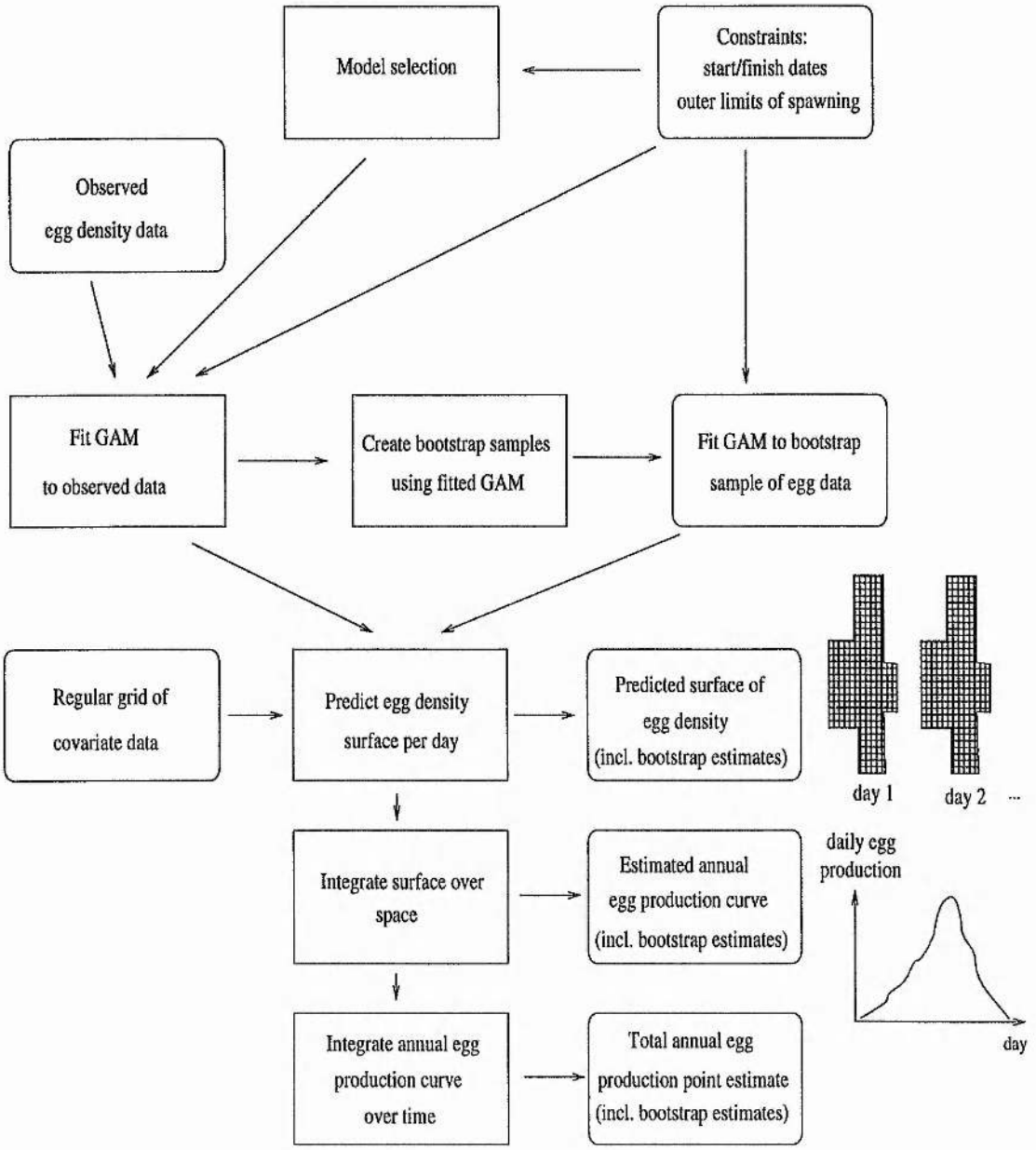


Figure 2.2: Schematic representation of the GAM egg production method. Boxes with rounded corners correspond to input data and output estimates, boxes with straight corners correspond to estimation procedures.

2.2.3 Example: Western Atlantic mackerel 1995 egg survey data

The International Council of the Exploration of the Seas (ICES) has carried out triennial mackerel and horse mackerel egg surveys since 1977. We illustrate the GAM method using the 1995 ICES survey data of the Atlantic "western" mackerel stock. This survey was conducted by England, Germany, Ireland, Netherlands, Norway, Portugal, Scotland and Spain. Plankton sampling was carried out by national versions of a Gulf III type sampler with the exception of Spain which used Bongo and Pairovet samplers (Anon., 1996). Although the stock assessments for the series of ICES surveys currently use the traditional AEPM, the GAM-based AEPM has been applied to data from three surveys (1989, 1992 and 1995) of the western area and from the 1995 survey of the southern area (Borchers et al., 1997a).

Figure 2.3 shows the Atlantic mackerel 1995 survey coverage over intervals of 15 days spanning the sampling period. Note that there is only partial spatial coverage of the survey area during any time period. For example in the beginning of the survey period in February samples are only taken in the south of the survey area. This confounding of sampling in space and time makes it difficult to distinguish the effects of time and space at the beginning of the survey. An adaptive sampling strategy in the east-west direction was used in this survey. This strategy ensured that outer boundaries of spawning were sampled even if these extended beyond the "standard" sampling area. In comparison to previous survey years, the 1995 survey had the most complete coverage in space and time.

The traditional AEPM applied to Atlantic mackerel

In the traditional AEPM as applied to Atlantic mackerel, egg density is estimated separately in each survey rectangle (0.5^o latitude by 0.5^o longitude) per period, using the mean observed egg density of stage 1 eggs in each rectangle as the estimate. The survey was divided into six periods: (1) 16/3 – 25/3, (2) 26/3 – 14/4, (3) 15/4 – 21/4, (4) 22/4 – 16/5, (5) 17/5 – 8/6, (6) 9/6 – 29/6, (7) 30/6 – 16/7. Egg mortality is ignored on the basis that it is negligible for the youngest egg stage, although ignoring mortality results in slight underestimation of biomass. For unsampled rectangles which are adjacent to at least two sampled rectangles and within the survey area, egg density is estimated by the arithmetic

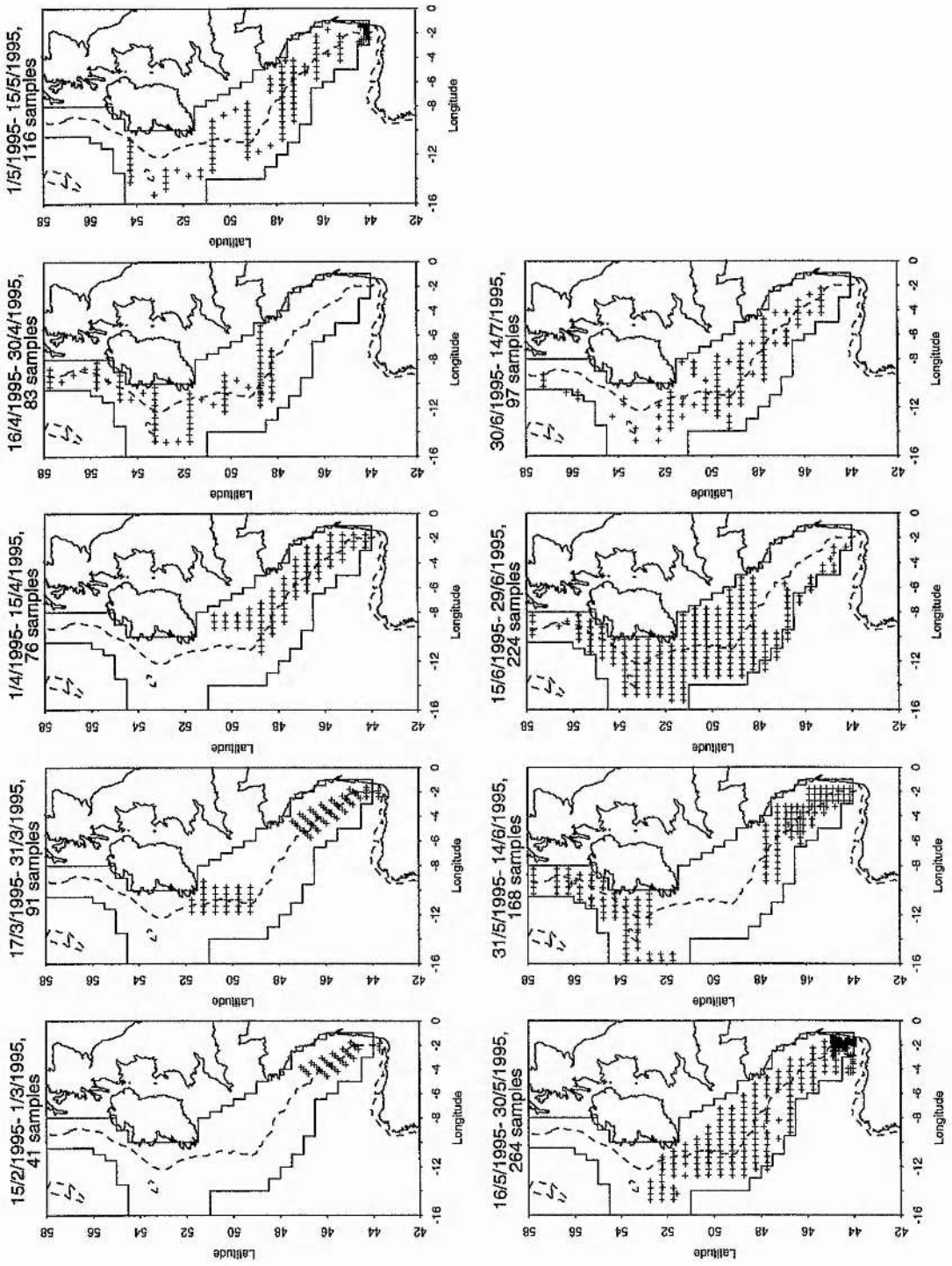


Figure 2.3: Locations of egg samples taken during the 1995 Atlantic mackerel (*Scomber scombrus*) survey. Samples are shown summarized in 15 day intervals. The 200m sea depth contour line is shown as a dotted lines. The survey area is shown by a continuous line.

mean of the densities in adjacent rectangles. Other unsampled rectangles within a period and within the survey area are assumed to have zero egg density. Estimates of egg density within each rectangle are converted to estimates of egg abundance by multiplying them by the surface area of the associated survey rectangle. Total egg production in a single period is estimated by the sum of the estimates of egg abundance in each rectangle, over all rectangles in the survey region.

Annual egg production is estimated by integrating under the piecewise linear egg production curve over the whole spawning period (Anon., 1996). Each knot in this piecewise curve corresponds to the egg production estimate at the period mid-point. In order to estimate the total egg production over the whole survey period, an initial and a final time point, at which there are assumed to be no eggs in the area, are chosen. These two time points define the limits of the spawning period. Data for estimating the initial time point (the last time before the survey at which there were no eggs) and the final time point (the first time after the survey at which there were no eggs) are scant as there are no directed surveys in these periods. Although all available data are used to decide where to locate the points in time, their date is necessarily somewhat arbitrary.

Start/finish dates and outer limits of spawning for the GAM-based AEPM

In 1995, sampling for mackerel seldom spanned the entire spawning area and did not span the entire temporal range of spawning. This is apparent from observed egg densities (Figure 2.4) at the outer limits of the sampled area and at the latest sample times. In both cases, egg densities were often substantially greater than zero. In order to avoid positive bias being introduced by spatial and temporal extrapolation into unsampled regions of the spawning area at any time, the outer boundaries of spawning were defined prior to fitting the GAMs, and the fitted models were constrained to be close to zero at these outer boundaries by using structural zeros, as described above. The spatial boundaries we used vary over time, in accordance with what are believed to be conservative estimates of the true outer boundary of spawning at various times through the spawning season (i.e. the true boundaries are likely to be within the assumed boundaries). The locations of these boundaries are based on data from all surveys of the area conducted up to 1995. Each of the five months between March and July was assigned a different spatial boundary defined by appropriately placed structural zeros (Figure 2.5). Except in July, no outer boundary

was specified at the southern end of the western survey area. Egg densities there were high, and counts were made south of the boundary, for stock assessment in the southern area. We included these counts, to improve the fit of the GAM at the boundary. Spatial boundaries were chosen as follows: Using a grid of 0.5^o latitude by 0.5^o longitude, and extracting by grid rectangle the maximum egg density observed between 1977 and 1995, the boundary was defined to be the first rectangle that lay at least 0.5^o longitude beyond the outermost rectangle with zero or near zero egg density. These rules were not always rigidly adhered to, but were adjusted slightly in the light of additional data from other sources.

A temporal limit at which it was assumed that spawning had not yet begun, and a temporal limit at which it was assumed spawning had ended, were defined and used similarly. The start date (*minday*) for spawning was assumed to be the 10th of February and the finish date was assumed to be the 31st of July (*maxday*). The same dates were used for the traditional AEPM.

Model Selection for the GAM-based AEPM

In order to simplify the model selection process, we adopted the approach taken by Borchers et al. (1997b) for model selection and considered only splines with either 4 degrees of freedom (*df=4*) or one degree of freedom (*df=1*) for covariates and their first order interactions. The covariates first entered the model with *df=4*, and backward stepwise elimination was used to select a set of covariates. Selection between smooths with *df=4* and smooths with *df=1* was performed in the next step. Finally, first order interactions of the previously selected covariates were first entered with *df=4*, again using backwards stepwise elimination for model selection, and selection between smooths with *df=4* and *df=1* was performed in the next step. Comparisons between models were made on the basis of approximate F-tests (Hastie and Tibshirani, 1990). The covariates (x_k) used in the GAMs are as follows.

date = date in days since the 1st of January,
lat = latitude in degrees,
lon = longitude in degrees,
cdist = closest distance to the 200m depth contour line in nautical miles (negative if

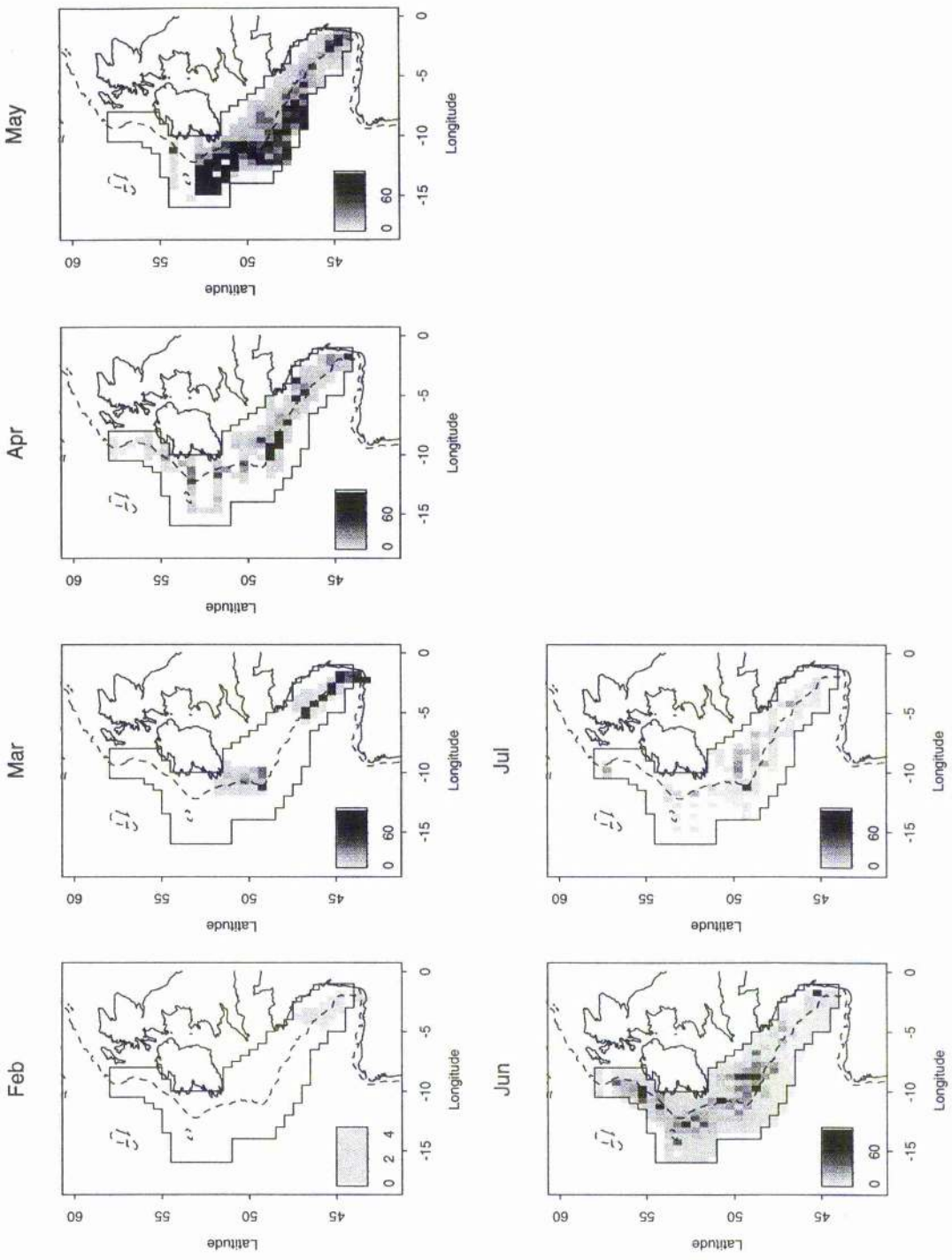


Figure 2.4: Observed mackerel egg densities (eggs m^2 day^{-1}) averaged within each half degree block in the survey area and per month. Unshaded half degree blocks were not sampled. The 200m contour line is shown as a dotted line. The survey area is shown with a continuous line. The same scale of shading is used for all plots.

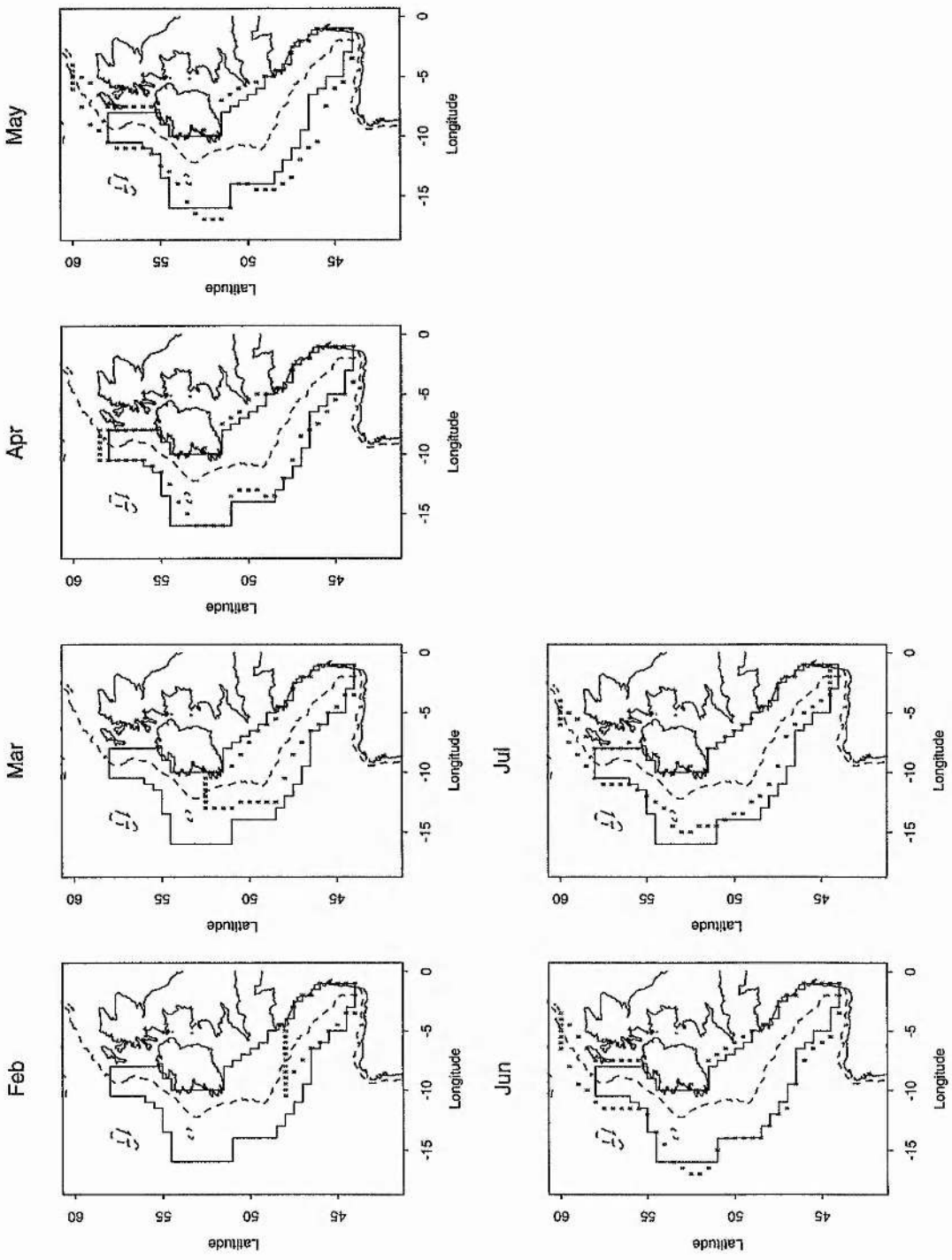


Figure 2.5: Assumed outer limits of spawning for Atlantic mackerel in the survey area. The shaded boxes indicate location of structural zeros. The survey area is shown with a continuous line. The 200m contour line is shown as a dotted line.

bottom depth $>200m$); the 200m contour line corresponds to the edge of the continental shelf,

gdist = distance along the 200m contour line in a north-south direction in nautical miles, and

log(depth) = logarithm of bottom depth in metres.

Substantial confounding between position, time and vessel precluded inclusion of a vessel factor. Sea surface temperature at a given location was found to be highly correlated with date. We chose to include date in preference to sea surface temperature as this avoids difficulties associated with predicting temperature at unsampled points. Aside from date, all covariates used in the model are constant with respect to time and all are known over the full range of the survey. The digitized bathometry data set (DBDB5) of the British Oceanographic Data Centre has been used to obtain bottom depths.

While models were selected using structural zeros to constrain the fitted model to be at or close to zero at the spatial and temporal boundaries of spawning, the contribution from these structural zeros to the deviance was removed before comparing models.

Results

GAM for egg density

Because subsamples were taken in areas of high density, we fitted egg density using equation (1), with a negative binomial distribution and a log link function. The selected model for the spatial distribution of egg density includes the following covariates selected with smooths of *df=4:s(date), s(cdist), s(gdist), s(log(depth))*, together with all smoothed 1st order interactions between the covariates except *cdist*gdist* and *cdist*log(depth)*. The residual deviance for this model is 1826 on 1587 df, with an estimated dispersion parameter of 1.1, which indicates that the model fits fairly well. Approximate F-tests indicate that *date* is the most significant covariate, followed by *log(depth)*. Deviance residuals plotted versus fitted value appear to be fairly random. Average deviance residuals plotted in space by grid square and per month (Figure 2.6) suggest that there may be some remaining unmodelled spatial correlation. Calculating the correlograms (not shown here) on these averages of deviance residuals confirms this. The autocorrelation of deviance residuals in time, ignoring the spatial dimension, show strong correlation too (Figure 2.7). In March

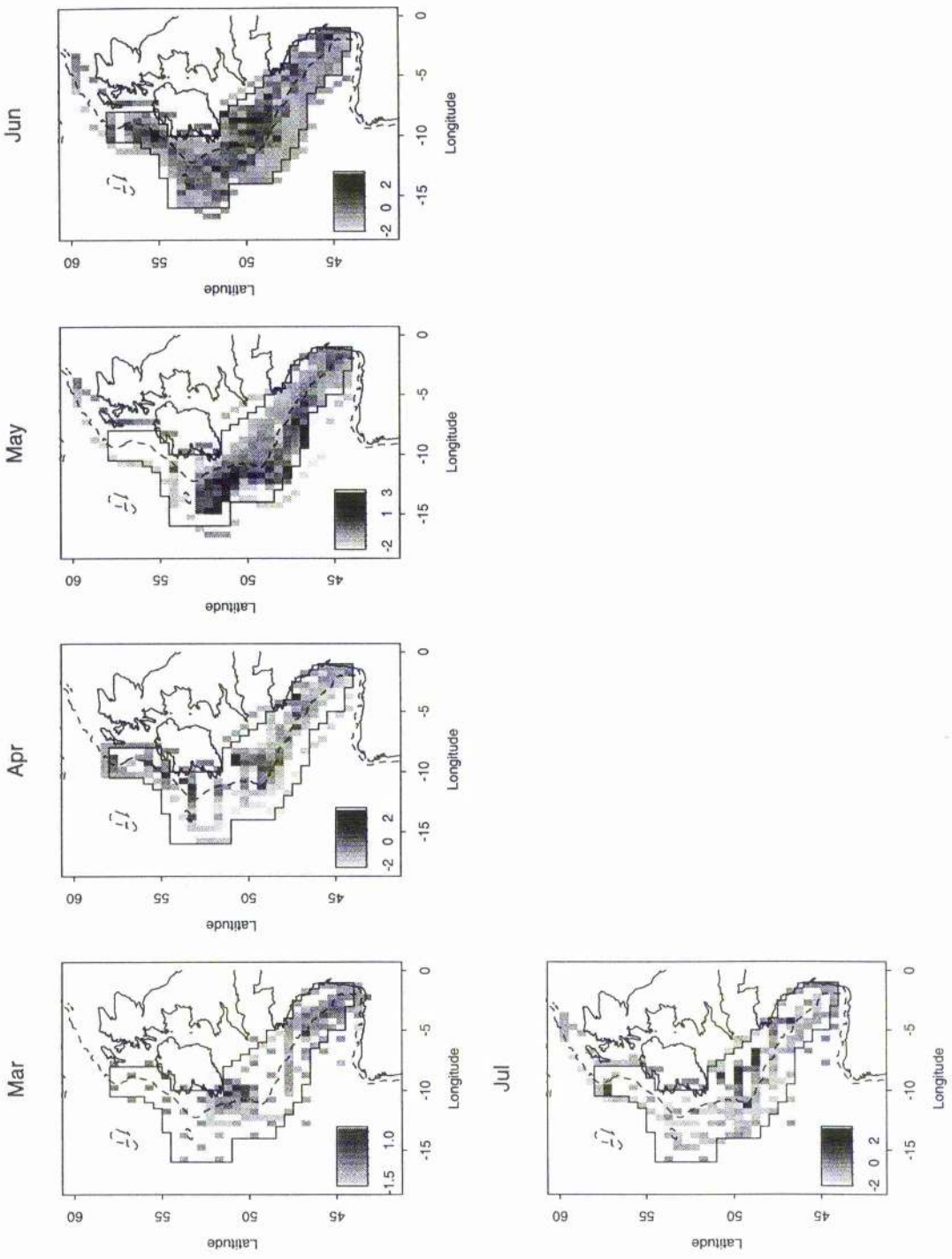


Figure 2.6: The spatial distribution of standardized deviance residuals summarized by months for the *density model* (response is density) with *df=4* fitted to the 1995 Atlantic mackerel survey data. The 200m contour line is shown as a dotted line. The survey area is shown with a continuous line. The same scale of shading is used for all plots.

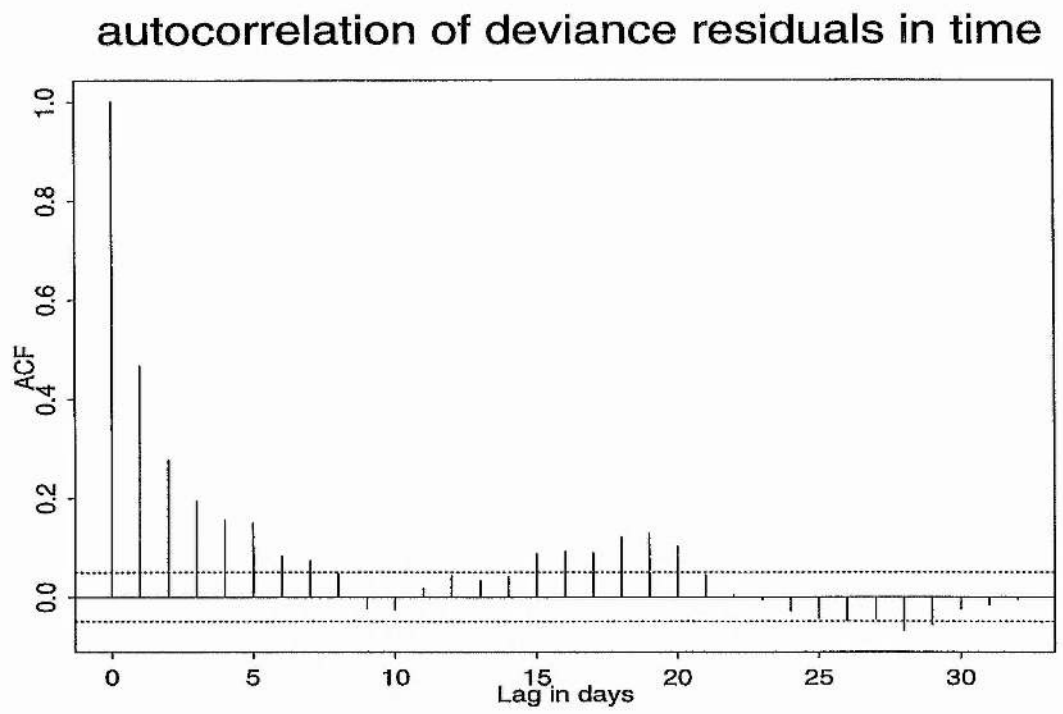


Figure 2.7: The autocorrelation function at different time lags (in days). The dotted lines correspond to approximate 95% confidence interval bands about the line of zero correlation.

observed average mackerel densities are highest in the south of the survey area along the 200m contour line, while in May high densities are further north and to the west of the 200m contour line (Figure 2.4). A comparison of the observed distribution of mackerel eggs (Figure 2.4) and the predicted spatial egg densities (Figure 2.8) indicates that the model has captured the south-north shift in spawning along the 200m contour line, as well as the peak in abundance and westward shift during May. The egg production curve (Figure 2.9) suggests a slightly lower peak than the curve from the traditional method, but overall it agrees fairly well. The plots using the centred covariates of the effects of the partial smooth functions, i.e. the effect of date given the inclusion of the other covariates, are given in Figure 2.10. The smooth corresponding to date indicates the peak of spawning in the middle of the survey period. The effect of the smooth function fitted for distance from the 200m contour line (Figure 2.10a) shows a peak west of the 200m metre contour line. The total annual egg production estimate from the density model is 1.49 x 10^{15}(standard error (s.e.) = 0.10; coefficient of variation (cv) = 6.9) in comparison to the traditional method estimate of 1.48 x 10^{15} (s.e. = 0.17, cv = 11.5) (Table 2.1). The

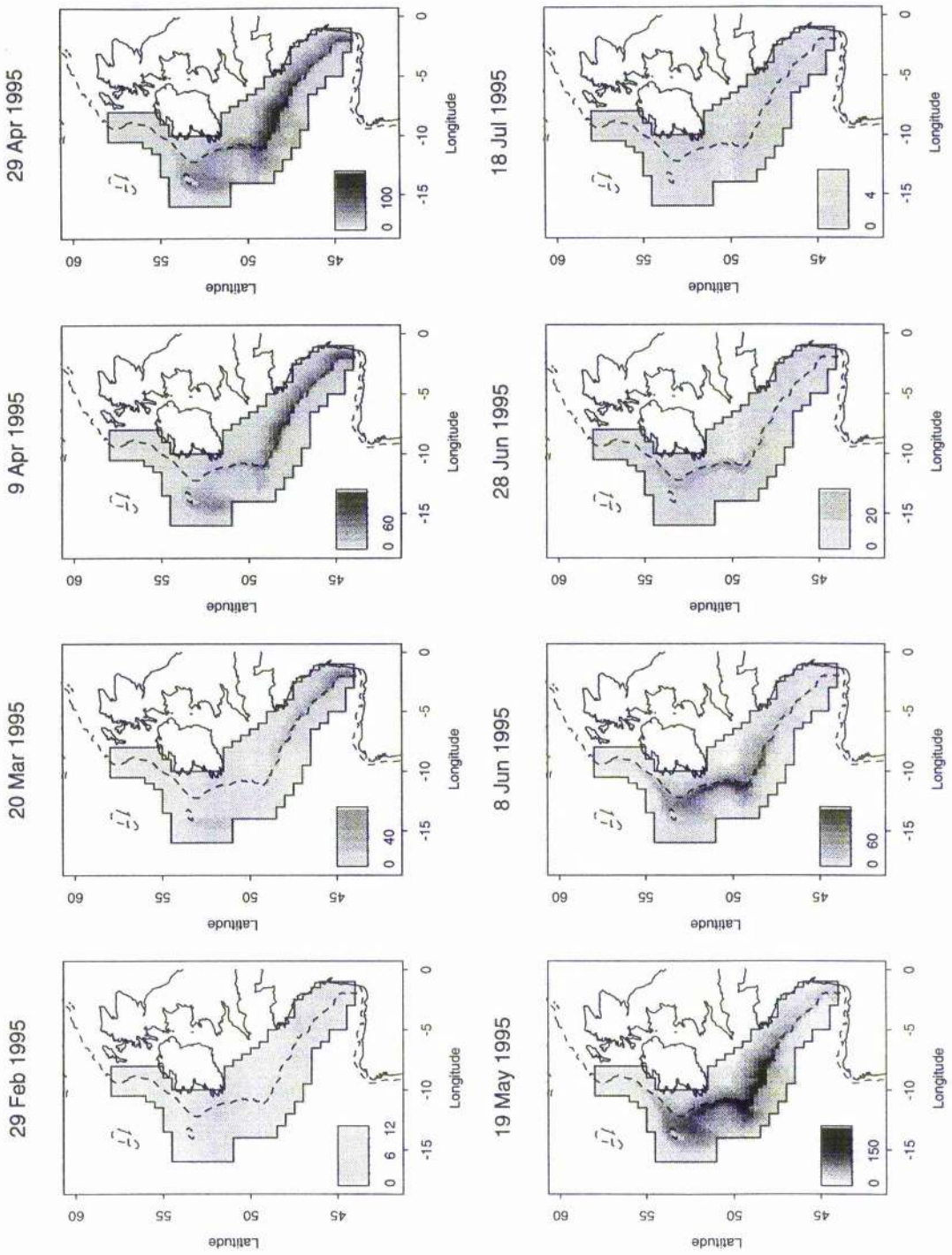


Figure 2.8: Spatial distribution of estimated Atlantic mackerel in 1995 egg density (eggs m^2 day^{-1}) using the *density model* at selected dates in 20 day intervals. The 200m contour line is shown as a dotted line. The survey area is shown with a continuous line. The same scale of shading is used for all plots.

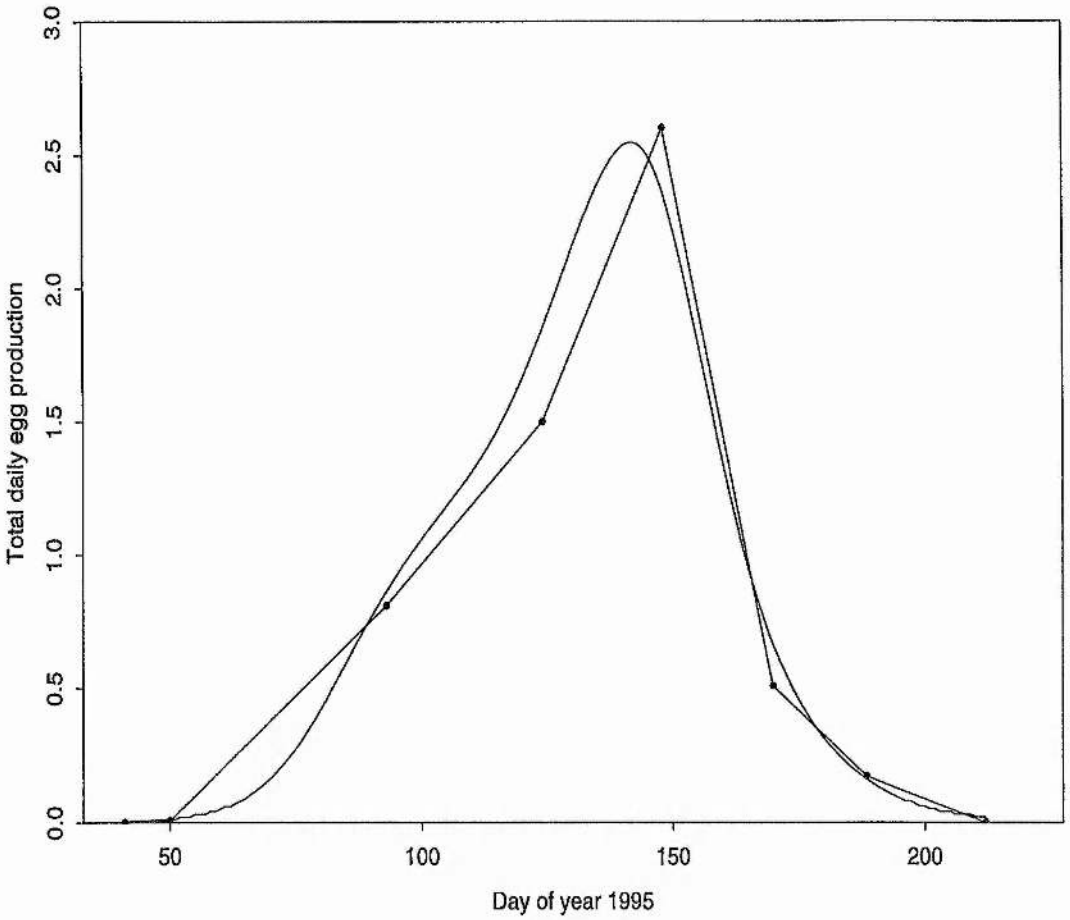


Figure 2.9: Estimated egg production curves for Atlantic mackerel in 1995 using the *density model* (smooth curve) compared with the corresponding curve from the traditional method (piecewise linear curve).

difference between estimates is small in comparison to the estimated standard error of the estimates. The cv of the GAM method estimate is just over half the cv of the traditional method. This increase in precision is not unexpected given the substantial spatial and temporal trend in observed egg densities within the survey region. The GAM method is able to model this trend relatively parsimoniously using smooths, whereas the traditional method models spatial trend much less parsimoniously: it effectively models the spatial trend via a two-dimensional step function with a step (and hence a parameter) at every half degree block. In contrast to the above results, if we take egg count as the response, we obtain a much lower estimate. Egg densities of up to 800 were recorded, and these tended to have high conversion factors (Figure 2.11a), which means that volume of the

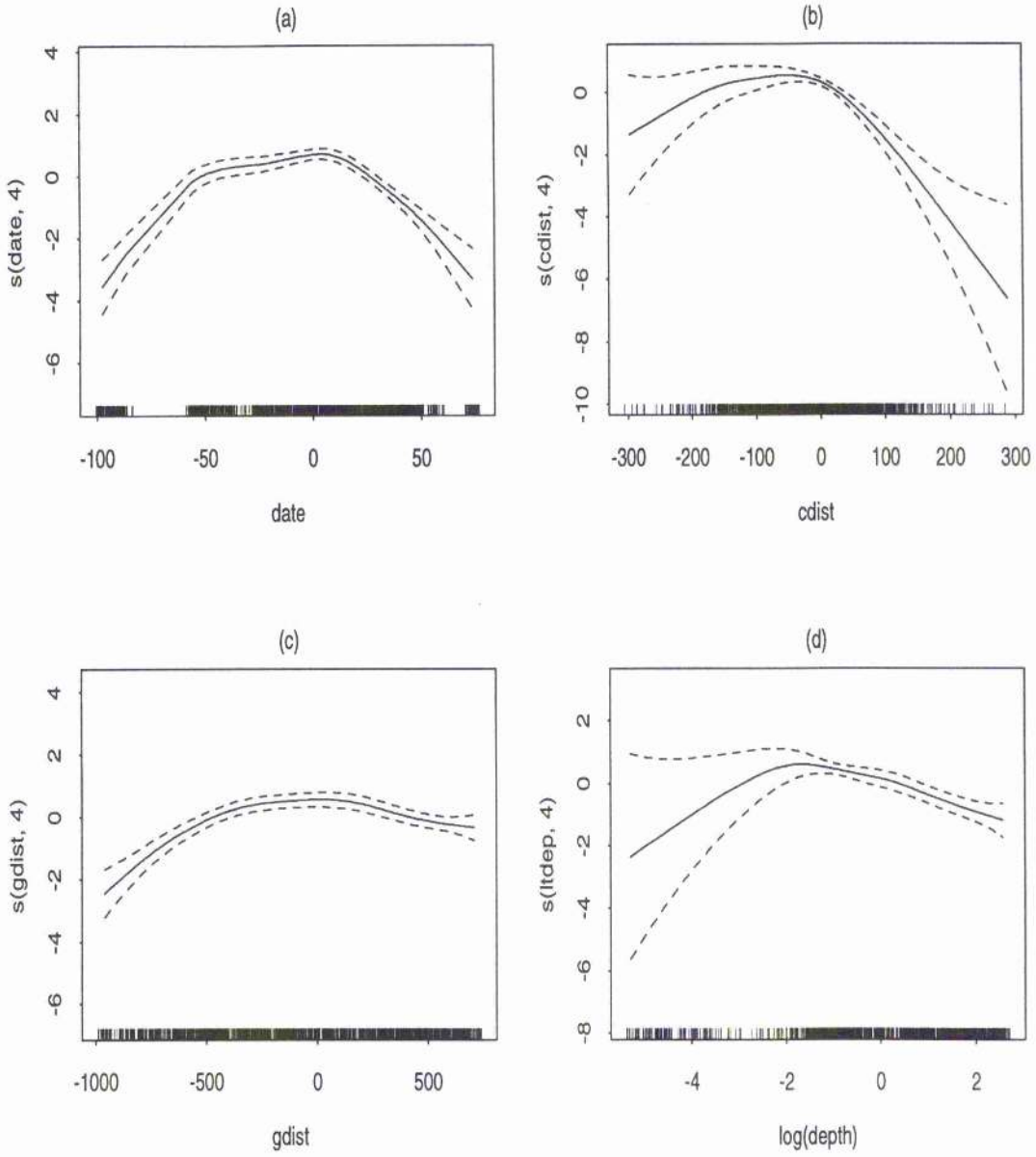


Figure 2.10: Estimated smooths of the covariates (a) *date* = date, (b) *cdist* = distance to contour (in nautical miles; negative if *depth* >*200*), (c) *gdist* = distance along the 200m contour lines (in nautical miles) and (d) *log(depth)* = log(bottom depth) fitted in the *density model* (response is density). The smooth indicates the partial effect of date on the linear predictor, i.e. the effect given that the other covariates are in the model. Zero on the vertical axis corresponds to no effect of the covariates fitted in the GAM. The covariates are centred, i.e. zero corresponds to the mean of the covariate. The dashed lines show the 95% confidence bands.

subsample taken was small. If the count model is fitted here, the fitting procedure gives these high densities weights that are lower than is appropriate for their actual variances (Figure 2.11b). The resulting egg production estimate for 1995 is 43% lower than that obtained from the density model.

2.2.4 Discussion and comparison of methods for the Annual Egg Production method of Stock Assessment

GAM-based AEPM requires sampling to span the spawning area adequately throughout the spawning period. In cases of less adequate survey designs, the success of the methods depends on the availability of historical data for constraining the spatial distribution. Incorporating historical knowledge of the spatio-temporal distribution of mackerel in the form of constraints, namely start/finish dates and spatial boundaries of spawning, we can extrapolate beyond the sampled region. This is illustrated in our example for the predicted mackerel egg densities early in the season in 1995 (29 February and 20 March in Figure 2.8). The GAM method has extrapolated for the whole survey area in that time, when there were few sampling points and these were concentrated towards the south along the 200m contour line (Figure 2.4). Sampling effort can be concentrated on time points and locations where high abundances are expected. The traditional AEPM is less sensitive to confounding of sampling in space and time within a single period, but it is equally important to determine the boundaries of the spawning area, and with the traditional method all extrapolation is essentially *ad hoc*, whereas with the GAM method it is model-based.

Table 2.1: Comparison of estimates of total egg production in 1995 ($\hat{E}_a$) from the GAM method using the density model and the traditional method.

	GAM method density model	Trad. method
$\hat{E}_a.10^{15}$	1.49	1.48
Standard error	0.10	0.17
Coeff. of variation (%)	6.9	11.5

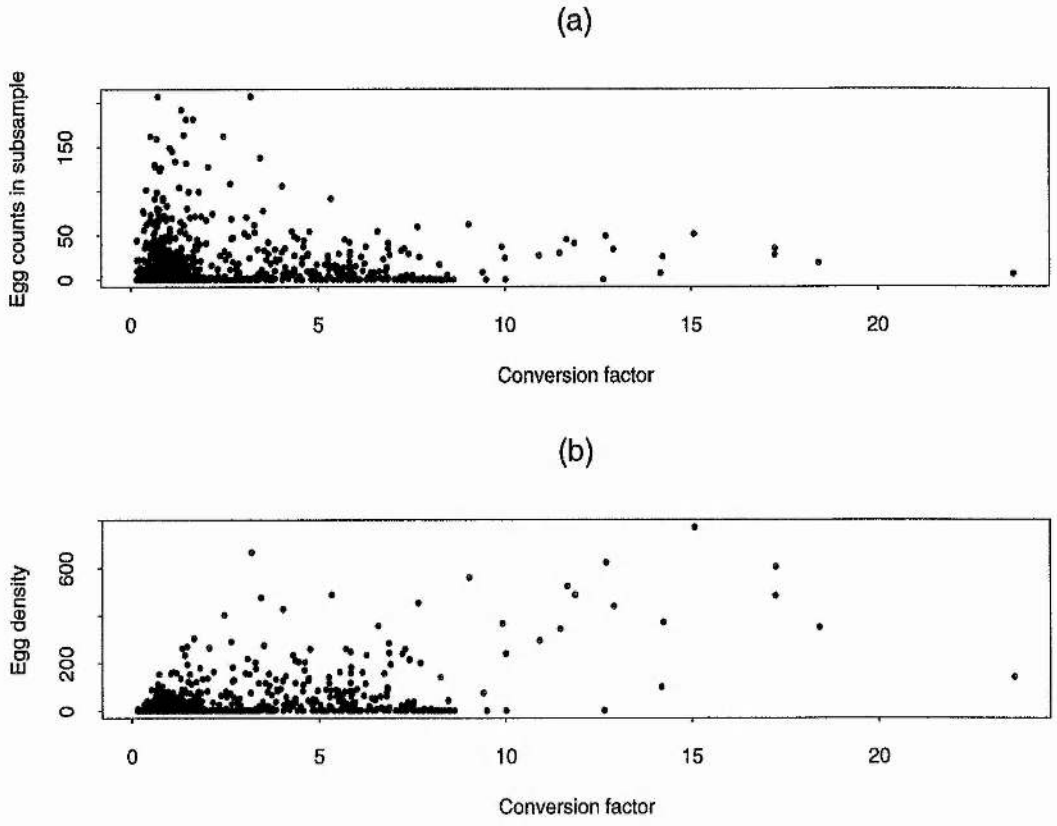


Figure 2.11: (a) Atlantic mackerel egg counts in subsample versus the conversion factor. (b) Atlantic mackerel egg density versus the conversion factor, which converts counts in the subsample to egg density.

The GAM-based AEPM in our example yields a more precise estimate and provides a variance estimate which contains all relevant components of variation, with the exception of the component of variation due to unmodelled autocorrelation in residuals, but the traditional method does not account for this either (see paragraph below for more details). In contrast to the GAM-based AEPM, the traditional method neglects the component of variance due to estimation of the shape of the egg production curve. Viewed in terms of survey effort required to achieve a given precision, the improved precision of the GAM method estimates would translate into a substantial reduction in survey costs. The survey design could be improved to better suit the GAM method, but the exact requirements and feasibility of optimal survey designs for the GAM-based AEPM deserves further investigation. In addition to improving precision, it is clear from our example that the GAM method provides information on the relationship between the response and covariates. Not surprisingly, the most important relationship is between date, which is a proxy for water temperature and other physical covariates, and egg density. This information on covariate response relationships could in principle be used to improve survey design.

As a by-product, the GAM-based AEPM is able to model complex trends in density with respect to space, time and other covariates, without *ad hoc* assumptions about the form of the trends. In addition, the method provides information on the nature of these trends with respect to a wide variety of covariates, and at a resolution which is likely to provide useful insights into the underlying mechanisms driving spawning distribution. In the traditional method only substantial changes in the distribution with time are visible, and only between a few time points.

A notable feature of the egg production curve estimated using the GAM method shown (Figure 2.9) is its relative insensitivity to the assumed start and end times for spawning (provided the times do not span too narrow a period). When egg production is low early in the season, the GAM estimated egg production curve remains low until substantial spawning is observed. When spawning starts early, the GAM curve is able to respond appropriately. The traditional method estimate, on the other hand, is completely unable to respond to what evidence there may be in the data for onset and ending of spawning; these limits are fixed by assumption, and changing them can have a substantial effect on the estimates, independently of the data.

Attractions of the traditional method are its simplicity and its supposedly assumption free

nature. While it is true that the point estimates of egg production for a given short period in time from the traditional method involve few assumptions, fairly strong assumptions are used in estimating the egg production curve as a function of time (start and end times for spawning and the piecewise linear shape of the curve are assumed without reference to the data), and *ad hoc* procedures are required to fill in unsampled rectangles. It is also true that the GAM method involves substantially more modelling than the traditional method, and like any powerful tool it could give misleading results if used inappropriately. As illustrated in the mackerel example, care needs to be taken when choosing the appropriate response variable in the model. An attraction of the GAM method is that it allows the data to determine the model. This is not true for the traditional method.

We considered two choices for response variable, egg count and egg density. When egg count is used, it is important to check whether the conversion factor depends on egg density. If, as here, it does, strong negative bias in the egg production estimate can be introduced. Similar bias has been observed by others, although the issue of bias correction has received little attention in the literature. It seems likely that the bias observed by Swartzman et al. (1992) was due to the bias described above. We advise in such circumstances to use density as a response.

If there is still unmodelled spatial correlation as in our example (see Figures 2.6 and 2.7) this could be due to the fact that the smoothed functions of the covariates cannot fully explain the similarities between egg densities in neighbouring survey rectangles, either because a potentially important covariate has not been measured or because of more direct influences such as social behaviour. With the survey design of our example, it is unlikely that the autocorrelation is due to social behaviour of the mackerel. The distances between sampling locations are greater than the distance over which a spawning school might typically travel. Also the plankton sampler is towed for several nautical miles and the egg counts are aggregated over space. It is most likely that the remaining autocorrelation in residuals is due to bad model fit, because the smoothers fitted are not flexible enough or important covariates such as information on currents and nutrients are absent. Not accounting for such autocorrelation results in underestimation of variance of the predicted surface and estimated abundance. It is likely that the same problems apply to the traditional method. There are several possible solutions. First, it might be possible to obtain covariate information on currents and nutrients at the appropriate resolution from other

data sources. Second, fitting more flexible smoothers, which are multi-dimensional in the time and space dimensions, might reduce the autocorrelation. Such multi-dimensional smoothers could allow for interaction between longitude and latitude and interaction between space and time. Therefore the third and most promising solution is to model the remaining autocorrelation. For example this could be done by introducing a random effect for an autoregressive process in space and time; parameters of this autoregressive process could be estimated from a three-dimensional semi-variogram, with dimensions *cdist*, *gdist* and *date*. This could possibly be done via generalized additive mixed models (GAMMs) as proposed by Lin and Zhang (1999). To our knowledge GAMMs have not been applied to spatio-temporal data. Also, it is in principle possible to account for unmodelled spatial correlation by introducing an autocovariate, i.e. an extra covariate depending on egg density in surrounding survey squares (Besag, 1974; Augustin et al., 1996, see also chapter 3). Adopting a more time-series oriented approach, such as dynamic generalised linear models (Lindsey, 1999; Meyer and Millar, 1999), might also improve the fit and give more reliable variance estimates. It should also be investigated whether the model could be improved by incorporating a stage-structured population dynamics model (Wood, 1994). There are five identifiable mackerel egg stages and the model described uses only stage I eggs (the youngest stage) and assumes zero mortality. Counts of eggs at all stages before hatching are available for these data and by utilizing this additional information, the estimation of egg production could be improved. By incorporating a mortality model, the age distribution can be predicted backwards and forwards in time at locations where no samples were taken.

The application of the GAM method illustrated here shows it to be a viable alternative to the traditional method. It is an alternative which has many desirable features not available with the traditional method, not least of which is a potential for a substantial saving in survey costs. With surveys designed specifically with a GAM method analysis in mind, the potential gains can increase further.

2.3 Dangers of overfitting: A simulation study

When locational covariates are combined with flexible models, such as GAMs, it is generally possible to produce a fitted surface whose residuals exhibit no detectable spatial

correlation. The very flexibility of these models can, however, lead to overfitting of the data: fitted models often 'reveal' relationships that do not exist. To illustrate this problem, and to show how model selection uncertainty can be incorporated into statistical inference using the bootstrap, two artificial examples are presented in which GAMs are fitted to spatial data. For both examples count data was simulated on a 20×20 grid using a negative binomial distribution, with probability mass function

$$\Pr(Z = z; \beta, k) = \frac{(z+k-1)!}{z!(k-1)!} \left(\frac{\beta}{\beta+1}\right)^k \left(\frac{1}{\beta+1}\right)^z$$

where $E(Z) = \mu = \frac{k}{\beta}$ and $Var(Z) = \mu + \mu^2/k$. The first example (*process 1*) contains geographical variation. A single realization is simulated from a negative binomial distribution with $\mu = \sqrt{(x+y)}$ and $k = 3$. Here x and y are locational covariates representing eastings and northings, respectively.

The second example (*process 2*) contains no geographical variation and no spatial correlation and is generated from a negative binomial distribution with parameters $\mu = 3$ and $k = 3$.

To each of the realizations of *process 1* and *2* a GAM is fitted assuming a negative binomial distribution for the response with a fixed parameter $k = 3$ and a log link function. Forward stepwise selection based on Akaike's information criterion (AIC) was used to test for inclusion of terms in $s(x), s(y)$ and $s(xy)$, where $s(.)$ represents a smoothing spline with 4 degrees of freedom. For *process 1* the fitted model is $E(z) = exp(\beta + s(x) + s(y))$ (β is the intercept) with a deviance of 424 on 391 df; for *process 2* the fitted model is $E(z) = exp(\beta + s(xy))$, with a deviance of 428 on 395 df.

Bootstrap resamples were generated for both processes using three different methods. Methods (a) and (b) both use the parametric bootstrap, generating bootstrap resamples from the fitted model of the form $z^* \sim NB(\hat{\mu}_i, k = 3)$ where $\hat{\mu}_i$ is the fitted value at location i. Models were fitted to the bootstrap resamples by (a) applying the stepwise model selection procedure and (b) refitting the model which was originally selected. In case (a), model selection uncertainty is incorporated into inference (Buckland, 1982; Hjorth, 1994; Buckland et al., 1997). The bootstrapping methods used in (a) and (b) were compared with a more naive, non-parametric alternative, (c), in which resampling from the observations with replacement was done. In this case, associated covariates were resampled along with the observations, violating the principle that analysis should be conditional on co-

variates. Then forward stepwise model selection is carried out, applying a weighted GAM to each bootstrap resample with weights equal to the number of times each observation was selected for that resample. This method, which preserves the original (x, y) locations associated with the resampled observations, also incorporates model selection uncertainty.

Bootstrap methods (a), (b) and (c) were used to calculate 95% bootstrap percentile confidence intervals (Buckland, 1984; Efron and Tibshirani, 1993) for both the overall mean and the fitted value at a specified location ($x = y = 20$), using 200 bootstrap resamples for each method. The bootstrap percentile confidence intervals are percentiles from the ordered set of bootstrap estimates of the parameter of interest. Thus a $100(1 - 2\alpha)\%$ confidence interval for the mean is obtained by estimating the mean in each resample, and ordering the estimates in increasing order. Then the rth and sth values in the list provide approximate lower and upper $100(1 - 2\alpha)\%$ confidence limits for the actual range, where $r = \alpha(B + 1)$ and $s = (1 - \alpha)(B + 1)$ (Buckland, 1984).

The results for one realization of each process are shown in Table 2.2. In terms of estimating the overall mean of the data, the confidence intervals of the three methods are very similar. In fact methods (a) and (b) are identical for estimating the overall mean, if the same bootstrap samples are used. The confidence intervals for predicted values for $x = y = 20$ show more variation between the methods. For *process 1*, method (a) and (b) produce narrower confidence intervals than does method (c), whereas for *process 2* method (c) produces the narrowest confidence interval.

The number of times different models are selected is shown in Table 2.3. This table shows how bootstrap methods (a) and (c), which take account of model selection, can be used to investigate whether a null model should be rejected. In the case of *process 2*, in which the null model was the true model, it was selected by AIC 103 times in the analysis of 200 bootstrap resamples generated under method (a), whereas the model from which the boostrap samples were generated was selected just 48 times. The corresponding figures for method (c), in which no advantage is given to the model selected by the original data, were 79 times for the true (null) model and 62 times for the model selected for the real data. In both cases, there is a clear message that we cannot rule out the null model.

Coverage accuracy for the three bootstrap methods is assessed by simulating 100 realizations from the model for both processes. Then confidence intervals were estimated for each

Table 2.2: Estimates and bootstrap 95% percentile confidence intervals obtained from models involving locational covariates: estimates are for one realization of *process 1* and *process 2* respectively; expected values are shown in italics; see text for details of processes *1* and *2* and bootstrap methods (a), (b) and (c).

Process	Bootstrap	Overall Mean		Observation at $x = y = 20$	
		Estimate	(C.I.)	Estimate	(C.I.)
1		*7.810*	—	*13.000*	—
	(a)	7.891	(7.352, 8.485)	17.661	(12.714, 22.324)
	(b)	7.891	(7.352, 8.485)	17.661	(12.714, 22.323)
	(c)	7.891	(7.292, 8.599)	17.661	(9.913, 21.434)
2		*3.000*	—	*3.000*	—
	(a)	2.941	(2.703, 3.240)	3.555	(1.593, 5.846)
	(b)	2.941	(2.703, 3.240)	3.555	(1.270, 5.846)
	(c)	2.941	(2.728, 3.193)	3.555	(2.537, 5.087)

Table 2.3: Frequency of model selection in 200 bootstrap resamples using methods (a) and (c): see text for details of processes *1* and *2* and bootstrap methods; † initial model selected for *process 1*; ‡ initial model selected for *process 2*.

Selected Covariates	Process 1		Process 2	
	(a)	(c)	(a)	(c)
Null Model	0	0	103	79
$s(x)$	0	0	9	11
$s(y)$	0	0	19	15
‡ $s(xy)$	0	0	48	62
† $s(x) + s(y)$	198	181	3	1
$s(x) + s(xy)$	0	0	13	12
$s(y) + s(xy)$	0	0	5	11
$s(x) + s(y) + s(xy)$	2	19	0	9

Table 2.4: Coverage accuracy of bootstrap 95% percentile confidence intervals based on 100 Monte Carlo realizations of processes *1* and *2* respectively; see text for details of processes *1* and *2* and bootstrap methods (a), (b) and (c); †$shape = \frac{r-\hat{\mu}}{\hat{\mu}-s}$, ‡$shape = \frac{r-\hat{y}}{\hat{y}-s}$.

Process	Bootstrap	C.I. for Overall Mean			C.I. for Obs. at $x = y = 20$		
		Coverage %	mean length	mean shape†	Coverage %	mean length	mean shape‡
1	(a)	98	1.107	1.136	72	11.568	0.762
	(b)	98	1.107	1.147	68	10.764	0.796
	(c)	98	1.172	1.434	84	12.538	1.189
2	(a)	92	0.494	0.986	100	2.781	0.883
	(b)	93	0.494	0.975	92	0.858	0.949
	(c)	94	0.477	1.166	100	2.600	1.681

of these realizations and the number of times they contained the true value was counted. Coverage (percentage of intervals that cover the true parameter value) along with average length and shape (Efron and Tibshirani, 1993), a measurement for asymmetry of the confidence intervals for μ, where $shape = \frac{r-\hat{\mu}}{\hat{\mu}-s}$, are shown in Table 2.4. For confidence intervals of the overall mean there is not much difference between the methods: coverage agrees well with the nominal value of 95%. In the case of *process 2* method (c) come closest to the nominal coverage. There are more differences in results of confidence intervals for predicted value at $x = y = 20$. For *process 1* all methods have substantially lower coverage than the nominal value although, method (c) agrees best with the nominal value, and method (b), which has the shortest confidence intervals on average, performs worst. Overall, method (c) yields the most asymmetrical confidence intervals. It is noteworthy that the naive method (c) performs best.

A slight improvement in coverage might be anticipated by using BC_a or ABC confidence intervals (Efron and Tibshirani, 1993) in place of the simpler percentile method.

2.4 Concluding remarks

The example of modelling mackerel eggs showed that it is very difficult to eliminate all spatial and spatio-temporal correlation in residuals. Even if there is no intrinsic autocorrelation present, it is rarely the case that all explanatory information is available to completely model the trend in data. Often to measure all the relevant explanatory information is too costly. The result is bad model fit, indicated by autocorrelated residuals. Ignoring this autocorrelation leads to a negatively biased variance, if the autocorrelation is positive. Another source of variability is model uncertainty. The parametric bootstrap we used in the application in section 2.2 assumed that the fitted model was the true model, and we fitted the same model to all bootstrap samples. We have shown in the simulation study in section 2.3 that for estimating the overall mean, which is comparable to abundance estimation as in the mackerel example, the incorporation of model uncertainty into variance estimation does not have a big effect.

Chapter 3

Auto-models for discrete spatio(-temporal) data

3.1 Introduction

When investigating the dynamics of vegetation and ecological processes some of the important issues are:

- local interaction among neighbouring plots for response variables; this can be none, competitive, antagonistic or co-operative;
- nature and the extent of spatial interaction (for example how big is the resource capture area of a plant, and how far do plants have to be apart in order not to compete with each other);
- the types of relationship between spatially referenced response variables and covariates (for example, investigating whether the spatial distribution of plants or animals is related to spatial variation in environmental factors such as temperature).

The last of these points can be dealt with using conventional regression methods such as GLMs and GAMs as described in the previous chapter. If we want to describe ecological processes, modelling local interaction (autocorrelation) and underlying trend simultaneously is preferred as both features are often present.

There are several different types of model which involve an explicit form of intrinsic autocorrelation structure, the interaction between neighbours, in the modelling process. In addition, these models can incorporate the effects of covariates, i.e. model trend, predict the value of a response variable at new locations, and estimate global characteristics of a spatial distribution. For example, they can use data from a few sample points in an agricultural crop to map the severity of a fungal infection throughout the crop. A global characteristic can be the yield of a crop under certain competition conditions induced by plant density. Even if no autocorrelation is present in the data, these types of model can provide a better fit, because they may capture the influence of unmeasured spatial covariates. Many of the existing approaches for this task have been developed for use with continuous data. They include regression models based on simultaneous auto-regressive schemes (Whittle, 1954), moving average models, conditional auto-regressive (auto-normal) schemes (Cressie, 1991) and various forms of kriging (Cressie, 1991).

More recently generalised additive models have been used to account for trend and correlation in normally distributed spatial data (Durbán et al., 1998; Gampe, 1998). With these methods cubic smoothing splines and spatial correlation parameters can be estimated simultaneously. This is due to the work of Wang (1998), who represents cubic smoothing splines as a mixed model. His work allows estimation of the smoothing parameters as a ratio of variance components which can be estimated via restricted maximum likelihood (REML) alongside the other parameters in the model. These methods are currently being further developed for non-normal data. Lin and Zhang (1999) propose generalized additive mixed models (GAMMs) which are an extension of Wang's Gaussian nonparametric mixed models (Wang, 1998). Parameters are estimated using marginal quasi-likelihood. The GAMMs can model trend and in addition take account for overdispersion and correlation in the data. So far they have not been applied to spatial data.

Another approach for modelling spatially correlated counts is to use generalized linear mixed models (GLMMs). The random-effects Poisson model first introduced by Besag et al. (1991) belongs to this class and has mainly been applied in epidemiology for disease mapping (Breslow and Clayton, 1993). This model can be used to account for autocorrelation, overdispersion and covariate effects for trend. In addition to the usual fixed linear covariate effects, the model contains an independent random location effect for heterogeneity and a spatially autocorrelated random effect. For the latter an auto-normal

process (a Gaussian random field, see section 3.2) is assumed. The parameter for autocorrelation is often assumed to be positive, fixed and autocorrelation extends over a defined neighbourhood structure. In theory such a model can be extended to have a parameter estimating positive or negative autocorrelation, which would be important for the application in vegetation dynamics. This has been done by Ickstadt and Wolpert (1996) who constructed a Bayesian hierarchical Gamma-Poisson model (Wolpert and Ickstadt, 1998) which deals with positive and negative autocorrelation as well as dependence on other spatial covariates. For binary response variables there are several applications similar to image restoration where the auto-logistic model is used for the random location effect as a prior and the autocorrelation parameter is estimated (Besag et al., 1991; Högmander and Møller, 1995; Heikkinen and Högmander, 1994; Hoeting et al., 1999). Diggle et al. (1998) extend geostatistical methods to deal with non-normal data by formulating kriging as a generalised linear prediction model. The motivation of their work is in estimating spatial surfaces of non-normal variables, such as intensity of radioactivity and risk of infection. Another approach is to define a general covariance structure that accounts for autocorrelation. Gotway and Stroup (1997) use GLMs with a quasi likelihood, defined with a general variance-covariance matrix for spatially correlated data. They estimate the autocorrelation structure by fitting semi-variograms to residuals, after having fitted a GLM with trend only. As in the application of Diggle et al. (1998) their main concern is to obtain a better fit, rather than explaining the underlying process.

Figure 3.1 shows simulated count data with different degrees of local autocorrelation. The model used for simulating the counts assumes different underlying processes (see also section 4.3.6) including independence, trend along rows, positive autocorrelation and negative autocorrelation. It is not easy to establish different underlying processes by eye. In particular, the simulated data from the model assuming independence between squares does not seem much different to the simulated data from the model assuming positive autocorrelation. Thus there is clearly a need to estimate autocorrelation using modelling techniques. When dealing with survey or field experiment data (Figure 3.2), there are potentially many effects causing trends, e.g. soil gradients, altitude and treatments. These show up as positive autocorrelation, which makes it difficult to distinguish autocorrelation and trend in estimation. In contrast to the random-effects approaches described above (Besag et al., 1991; Breslow and Clayton, 1993; Wolpert and Ickstadt, 1998), we attempt to estimate autocorrelation directly as a fixed effect from the response. Also our emphasis lies in

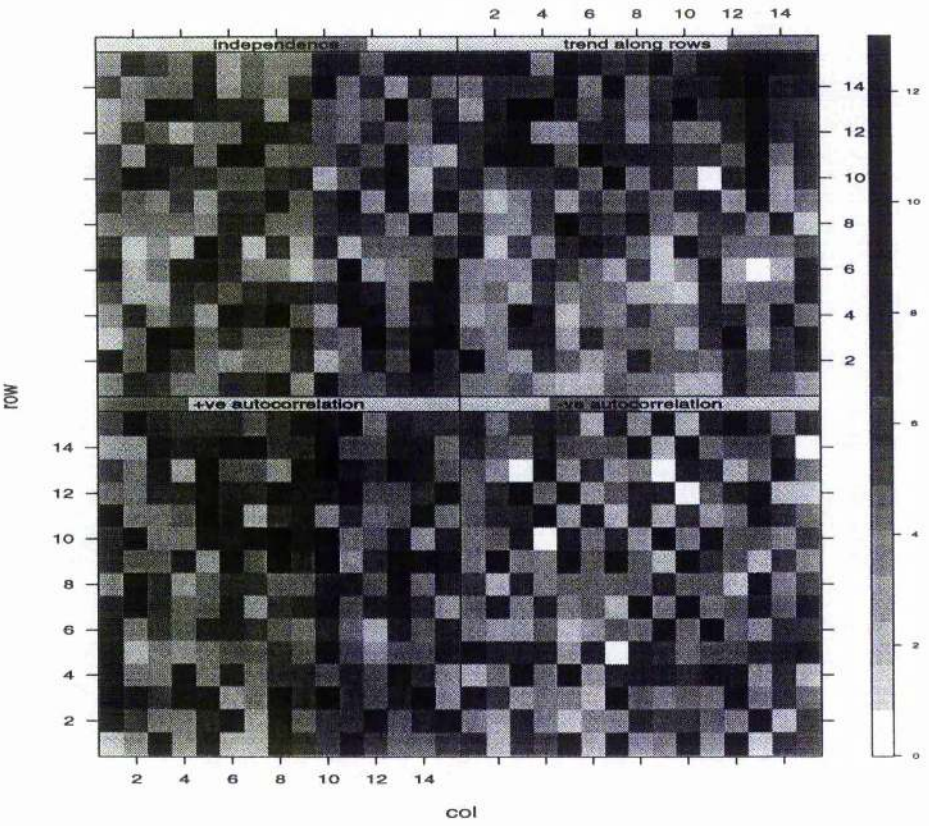


Figure 3.1: Simulated counts from models assuming different degrees of autocorrelation. (Note: These graphs correspond from top left to bottom right to *Example 5, 6, 1 and 2* in section 4.3.6.)

estimation of the autocorrelation parameter and the specific neighbourhood structure it applies to, in contrast to epidemiological applications, where the emphasis is on mapping the spatial distribution. Our interest lies also in finding a simple method for exploratory investigation of autocorrelation, which can routinely be used by biologists. For this we use the so-called auto-models first introduced by Besag (1972), to distinguish between the two components, trend and autocorrelation, with the aim of describing the underlying process. Auto-models include models for counts, the auto-Poisson, auto-binomial and auto-negative binomial model. They also include models for continuous data such as the auto-normal model mentioned above. The models are specified by the conditional probability of each observation given its set of neighbouring sites. The joint probability density, also called a Markov Random Field, is constructed using these conditional probabilities (Besag, 1974). These models incorporate a parameter for neighbour interaction (autocorrelation), caused

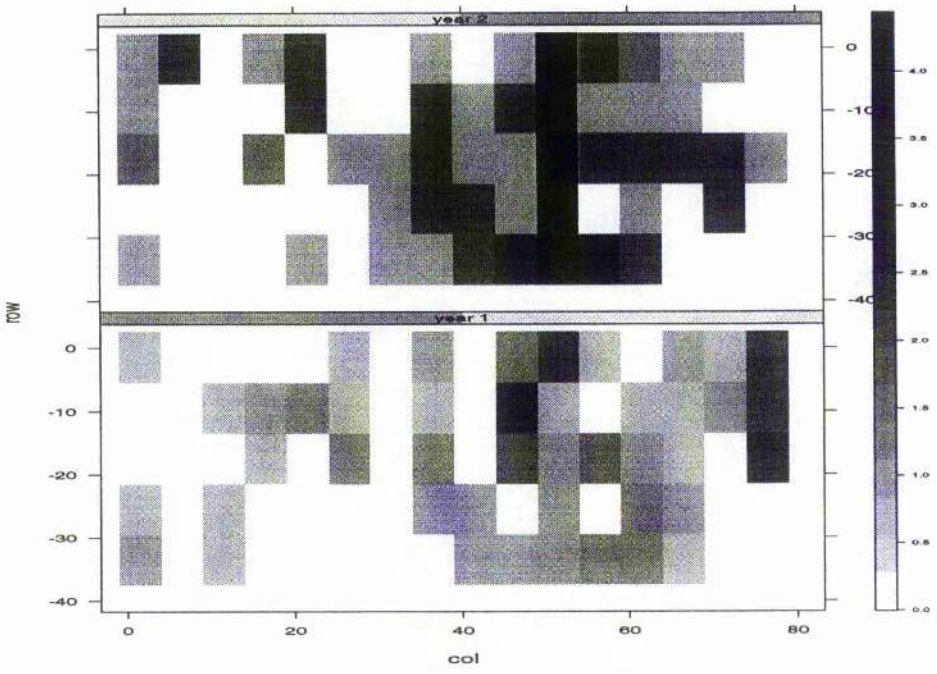


Figure 3.2: Observed seed counts (log(count +1)) of the weed *Poa* sp. in an agricultural experiment observed in two consecutive years. See also section 5.3.

by competitive or co-operative effects, which can influence global characteristics. Auto-models aim to describe the underlying process, which is the main objective for applications in vegetation dynamics and this is the reason why we have chosen these types of model. We concentrate on investigating the properties and practicability of the auto-Poisson model. A major drawback of this model is that it is only able to model negative autocorrelation. We investigate how this restriction can be avoided by right truncating the distribution. Similar work has been done by Kaiser and Cressie (1997) who used the winsorized Poisson distribution. Ferrándiz et al. (1995) used the truncated auto-Poisson distribution to model mortality counts in different counties in Spain. We investigate the properties and interpretation of the model and show how modification of the auto-Poisson model leads to different properties and biological interpretation.

We explain the derivation of auto-models in general in section 3.2. Then we review the auto-logistic model for binary and binomial data (section 3.3). In section 3.4 we derive and investigate properties of the truncated auto-Poisson and auto-negative binomial model.

3.2 Definition of auto-models / Markov Random Fields

Markov Random Fields (MRFs) were originally developed in physics for magnetic fields. The classic version of the first-order auto-logistic model is called the Ising model. Auto-models are constructed by specifying the conditional probability of each observation given its set of neighbouring sites (Besag, 1974). The joint probability density, also called a Markov Random Field, is constructed using the conditional probability of each observation. In the following a summary of Markov random fields (Cressie, 1991; Besag, 1974) is given. Matrices and vectors are given in bold.

Let $\boldsymbol{Y} = (Y(s_i) : s_i \in D)$ be a random process on the lattice $D = (s_i : i = 1, ..., n)$ defined by sites i. Then define a *Markov Random Field (MRF)* by specifying a neighbourhood N_i for each component $Y(s_i)$ of $\boldsymbol{Y}$. The neighbourhood N_i is the set of neighbours over which the interaction of $Y(s_i)$ extends, such that s_j is a neighbour of site s_i if

$$Pr(y(s_i)|y(s_1), ..., y(s_{i-1}), y(s_{i+1}), ..., y(s_n))$$

depends only on neighbouring sites $y(s_j)$ and can be reduced to

$$Pr(y(s_i)|y(s_1), ..., y(s_{i-1}), y(s_{i+1}), ..., y(s_n)) = Pr(y(s_i)|y(s_j); s_j \in N_i).$$

The neighbourhood N_i could be for example the "first-order" type neighbourhood structure such as:

```
0 1 0
1 i 1
0 1 0
```

Sites in the lattice denoted by `1` are neighbours of site `i` and comprise the neighbourhood N_i. Note that N_i excludes site `i`.

Now the ratio of two joint probability mass functions can be obtained via the *Factorisation Theorem*:

$$\frac{Pr(\boldsymbol{y})}{Pr(\boldsymbol{z})} = \prod_{i=1}^{n} \frac{Pr(y(s_i)|y(s_1), ..., y(s_{i-1}), z(s_{i+1}), ..., z(s_n))}{Pr(z(s_i)|y(s_1), ..., y(s_{i-1}), z(s_{i+1}), ..., z(s_n))} \tag{3.1}$$

where $\boldsymbol{y} = (y(s_1), ..., y(s_n))'$ and $\boldsymbol{z} = (z(s_1), ..., z(s_n))'$ are possible realisations of $\boldsymbol{Y}$. All possible configurations of $\boldsymbol{y}$ must theoretically be possible (*positivity condition*): If $\boldsymbol{y} \in \zeta$,

$\zeta \equiv \{\boldsymbol{y} : Pr(\boldsymbol{y}) > 0\}$ and $y_i \in \zeta_i$, $\zeta_i \equiv \{y_i : Pr(y_i) > 0\}$ for $i = 1, ..., n$, then the positivity condition is satisfied if the set ζ is the Cartesian product for all sets ζ_i.

Except for the auto-normal MRF, of which details are given below, the joint density cannot be obtained directly via the factorisation theorem. For most other distributions the negpotential (negative potential energy) or $Q()$-function is needed in order to derive the joint density. The Q()-function is defined as follows

$$Q(\boldsymbol{y}) \equiv \log(Pr(\boldsymbol{y})/Pr(\boldsymbol{0}))$$

where $\boldsymbol{y} \in \zeta$ and $Pr(\boldsymbol{0})$ is the probability that all events $y(s_i) = 0$. $Q(\boldsymbol{y})$ can then be constructed from equation 3.1 using the following result. Note that here the product of the different probability ratios of equation 3.1 simplifies to one ratio because all the other terms cancel out.

$$\frac{Pr(y(s_i) \mid \{y(s_j) : j \neq i\})}{Pr(0(s_i) \mid \{y(s_j) : j \neq i\})} = \frac{Pr(\boldsymbol{y})}{Pr(\boldsymbol{y}_i)} = exp(Q(\boldsymbol{y}) - Q(\boldsymbol{y}_i)),$$

where $0(s_i)$ denotes $y(s_i) = 0$ and $\boldsymbol{y}_i = (y(s_1), ..., y(s_{i-1}), 0, y(s_{i+1}), ..., y(s_n))$. Expanding $Q(.)$ gives

$$Q(\boldsymbol{y}) = \sum_{i=1}^{n} y(s_i)\alpha_i(y(s_i)) + \sum_{i=1}^{n-1}\sum_{j>i}^{n} y(s_i)y(s_j)\gamma_{ij}(y(s_i), y(s_j)). \tag{3.2}$$

Here only two-way interactions are assumed, $\gamma_{ij} = \gamma_{ji}$ and multi-way interactions are assumed to be zero, i.e. $\gamma_{ij...s}(y(s_i), y(s_j), ..., y(s_s)) = 0$. This expansion is obtained by defining the terms referring to trend, $y(s_i)\alpha_i(y(s_i))$, and local neighbour interaction, $y(s_i)y(s_j)\gamma_{ij}(y(s_i), y(s_j))$, in terms of different $Q(.)$-functions (see Cressie, 1991, p. 415). For instance $y(s_i)\alpha_i(y(s_i)) = Q(0, ..., 0, y(s_i), 0, ..., 0)$. In the models of interest here the terms $\alpha_i(y(s_i))$ and $\gamma_{ij}(y(s_i), y(s_j))$ in expansion 3.2 simplify to constants α_i and γ_{ij} (since only two-way interactions are considered and conditional distributions of $y(s_i)$ given N_i are that from the exponential family). Then for a homogeneous process where there is no dependence on spatially referenced explanatory variables, we have $\alpha_i = \alpha$ for all i, whereas in a heterogeneous process $\alpha_i = \boldsymbol{x}_i'\boldsymbol{\beta}$ is the parameter for the underlying trend, the vector $\boldsymbol{x}_i$ is the i'th row of the design matrix and $\boldsymbol{\beta}$ is the vector of coefficients.

To ensure a unique joint probability structure it is necessary to restrict the functional form of $Q(\boldsymbol{y})$. Defined in the theorem by Hammersley and Clifford (1971), the restrictions are that $\gamma_{ij...s} = 0$ if sites $i, j, ..., s$ are not all mutual neighbours. Then the probability mass

function is

$$Pr(\boldsymbol{y}) = exp(Q(\boldsymbol{y}))/\sum_{\boldsymbol{z}\in\zeta} exp(Q(\boldsymbol{z}))$$

provided the summability condition $\sum_{\boldsymbol{z}\in\zeta} exp(Q(\boldsymbol{z})) < \infty$ holds. If $\sum_{\boldsymbol{z}\in\zeta} exp(Q(\boldsymbol{z}))$ is available in closed form, full maximum likelihood estimation is possible.

Markov random fields theorem (Besag, 1974)

If the response $y(s_i)$ given N_i has any distribution of the exponential family, e.g. $\text{Poi}(\mu_i|y(s_j); j \in N_i)$, this theorem gives the Q(.)-function in terms of exponential family parameters. The parameterisation for members of the exponential family is

$$Pr(y_i|y_j; j \in N_i) = exp(\ \theta_i(y_j; j \in N_i)B_i(y_i) + C_i(y_i) + D_i(y_j; j \in N_i)\ ).$$

Note that to avoid clutter we now refer to location using index i only and $y_i = y(s_i)$. The parameter θ_i is called the natural parameter. The first derivative of the likelihood of any exponential distribution with respect to θ_i is equal to $\mu_i = E(y_i)$. The functions $B_i()$ and $C_i()$ have some specific form and θ_i and $D_i()$ are functions of the neighbouring sites. Besag's theorem (Besag, 1974) assumes the above parameterisation and pairwise-only dependence between sites. As a consequence $Q(\boldsymbol{y})$ can be defined in terms of the exponential family parameters up to an additive constant (Cressie, 1991).

$$Q(y) = \sum_{i=1}^{n}(B_i(y_i)\boldsymbol{x}_i'\beta + C_i(y_i)) + \sum_{i=1}^{n-1}\sum_{j>i}^{n} B_i(y_i)B_j(y_j)\gamma_{ij}. \tag{3.3}$$

with i=1,...,n, $\gamma_{ij} = \gamma_{ji}, \gamma_{ii} = 0$ and $\gamma_{ik} = 0$ for $k \notin N_i$. This gives the "recipe" to construct auto-models and makes it possible to formulate the model in terms of the natural parameter θ_i:

$$\theta_i(y_j; j \in N_i) = \boldsymbol{x}_i'\beta + \sum_{j\in N_i}\gamma_{ij}B_j(y_j). \tag{3.4}$$

This formulation is analogous to a GLM with a canonical link, except that here the observations themselves are explanatory variables. We call these *autocovariates*, with

$$autocovariate_i = \sum_{j\in N_i} B_j(y_j)$$

and often we assume that neighbourhood effects are the same in all directions, $\gamma = \gamma_{ij}$, so that

$$\sum_{j\in N_i}\gamma_{ij}B_j(y_j) = \gamma\ autocovariate_i.$$

The normal case: an exception

There is a continuous version of the factorisation theorem and with this the joint density $f(\boldsymbol{y})$ can be inferred for the auto-normal model as a multivariate normal. In the auto-normal model the responses $\boldsymbol{Y}_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$, with $\mu_i = \mu_i(y_j; j \in N_i)$ and

$$\mu_i = \alpha_i + \sum_{j \in N_i} \gamma_{ij}(z_j - \alpha_j),$$

with $i = 1, ..., n$, $\gamma_{ij}\sigma_j^2 = \gamma_{ji}\sigma_i^2$, $\gamma_{ii} = 0$, $\gamma_{ik} = 0$ for $k \notin N_i$ and the parameters α_i referring to trend and γ_{ij} to local neighbour interaction. In the proof establishing the joint density (see Cressie, 1991, p. 413) the factorisation theorem is applied by substituting $\boldsymbol{y} = \boldsymbol{\mu}$, where $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)'$, the mean vector of a multivariate normal distribution. The ratio then has the following form:

$$\frac{f(\boldsymbol{y})}{f(\boldsymbol{\mu})} = \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu})'\boldsymbol{M}^{-1}(\boldsymbol{I} - \boldsymbol{C})(\boldsymbol{y} - \boldsymbol{\mu})\right)$$

where $\boldsymbol{y} = (y_1, ..., y_n)$, $\boldsymbol{C}$ is a n x n matrix with γ_{ij} as the (i,j)th element and $\boldsymbol{M} = diag(\sigma_1^2, ..., \sigma_n^2)$. Using the fact that $\int_{\boldsymbol{y} \in \zeta} f(\boldsymbol{y})d\boldsymbol{y} = 1$ and $f(\boldsymbol{\mu}) = (2\pi)^{\frac{n}{2}}$ implies that

$$f(\boldsymbol{y}) = (2\pi)^{\frac{n}{2}}|\boldsymbol{M}|^{-\frac{1}{2}}\ |\boldsymbol{I} - \boldsymbol{C}|^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu})'\boldsymbol{M}^{-1}(\boldsymbol{I} - \boldsymbol{C})(\boldsymbol{y} - \boldsymbol{\mu})\right)$$

and hence, provided $(\boldsymbol{I} - \boldsymbol{C})$ is invertible and $(\boldsymbol{I} - \boldsymbol{C})^{-1}\boldsymbol{M}$ is symmetric and positive-definite

$$\boldsymbol{Y} \sim \mathrm{MVN}(\boldsymbol{\mu}, (\boldsymbol{I} - \boldsymbol{C})^{-1}\boldsymbol{M}). \tag{3.5}$$

In summary the auto-normal MRF is an exception in several ways. First, the joint density is known and available in closed form. Second, it can be derived easily via the factorisation theorem. Third, the components of the asymptotic covariance matrix referring to the covariance between trend (α_i) and neighbourhood interaction (γ_{ij}) are zero (Haining, 1990).

3.3 The auto-logistic model for binary and binomial data

We give a short review of the auto-logistic model. For further details see Cressie (1991) or Besag (1974). The auto-logistic model was first introduced by Ising (1925) in the field of statistical mechanics. The response y_i at each site of a finite lattice D with locations

$i = 1, ..., n$ is a Bernoulli random variate $Y_i \sim \text{Bernoulli}(p_i)$ with conditional density

$$Pr(Y_i = 1|y_j, j \in N_i) = p_i^{y_i}(1 - p_i)^{1-y_i}$$

where $p_i = Pr(Y_i = 1|y_j, j \in N_i)$. Then the auto-logistic model is defined as

$$\text{logit}(Pr(y_i = 1|y_j; j \in N_i)) = \log\left(\frac{Pr(y_i|y_j; j \in N_i)}{1 - Pr(y_i|y_j; j \in N_i)}\right) = \alpha_i y_i + \sum_{j \in N_i} \gamma_{ij} y_i y_j, \quad (3.6)$$

where $\text{logit}(Pr(y_i|y_j; j \in N_i))$ is the natural parameter of the Bernoulli distribution. This model has the same form as the classical logistic model, except that the response variable is part of the set of explanatory variables. Applying the procedure given in section 3.2, it can be shown that the $Q(.)$-function is defined as

$$Q(\boldsymbol{y}) = \sum_{i=1}^{n}(\alpha_i y_i) + \sum_{i=1}^{n-1}\sum_{j>i}^{n} \gamma_{ij} y_i y_j,$$

with i=1,...,n, $\gamma_{ij} = \gamma_{ji}, \gamma_{ii} = 0$ and $\gamma_{ik} = 0$ for $k \notin N_i$. In section 3.4 the derivation for the auto-Poisson and auto-negative binomial model will be given in detail. For binomial data, where $y_i \sim \text{Bin}(p_i, n_i)$ the form of the auto-binomial model is identical to the autologistic model in equation 3.6, but the $Q(.)$ function has an additional term:

$$Q(\boldsymbol{y}) = \sum_{i=1}^{n}\left(\alpha_i y_i + \log\begin{pmatrix} n_i \\ y_i \end{pmatrix}\right) + \sum_{i=1}^{n-1}\sum_{j>i}^{n} \gamma_{ij} y_i y_j,$$

with γ_{ij} defined as above. In theory the joint probability mass function for both models is available in closed form, but for most practical applications evaluation of the normalising term $\sum_{z \in \zeta} exp(Q(z))$ is not feasible. For instance for data on a 20 x 20 grid this would mean summing over $Q(.)$-functions of 2^{400} possible arrangements of zeros and ones! As will be demonstrated in chapter 4 section 4.3 it is much easier to approximate this normalising constant using samples drawn from a Markov Chain.

Properties and applications

For illustration let us assume the simplest possible auto-logistic model with a first order neighbourhood and no heterogeneity:

$$logit(Pr(y_i|y_j; j \in N_i)) = \alpha y_i + \sum_{j \in N_i} \gamma y_i y_j.$$

The auto-logistic model can deal with positive and negative auto-correlation, but as Figure 3.3 shows there are regions in the parameter space of γ for which different parameter

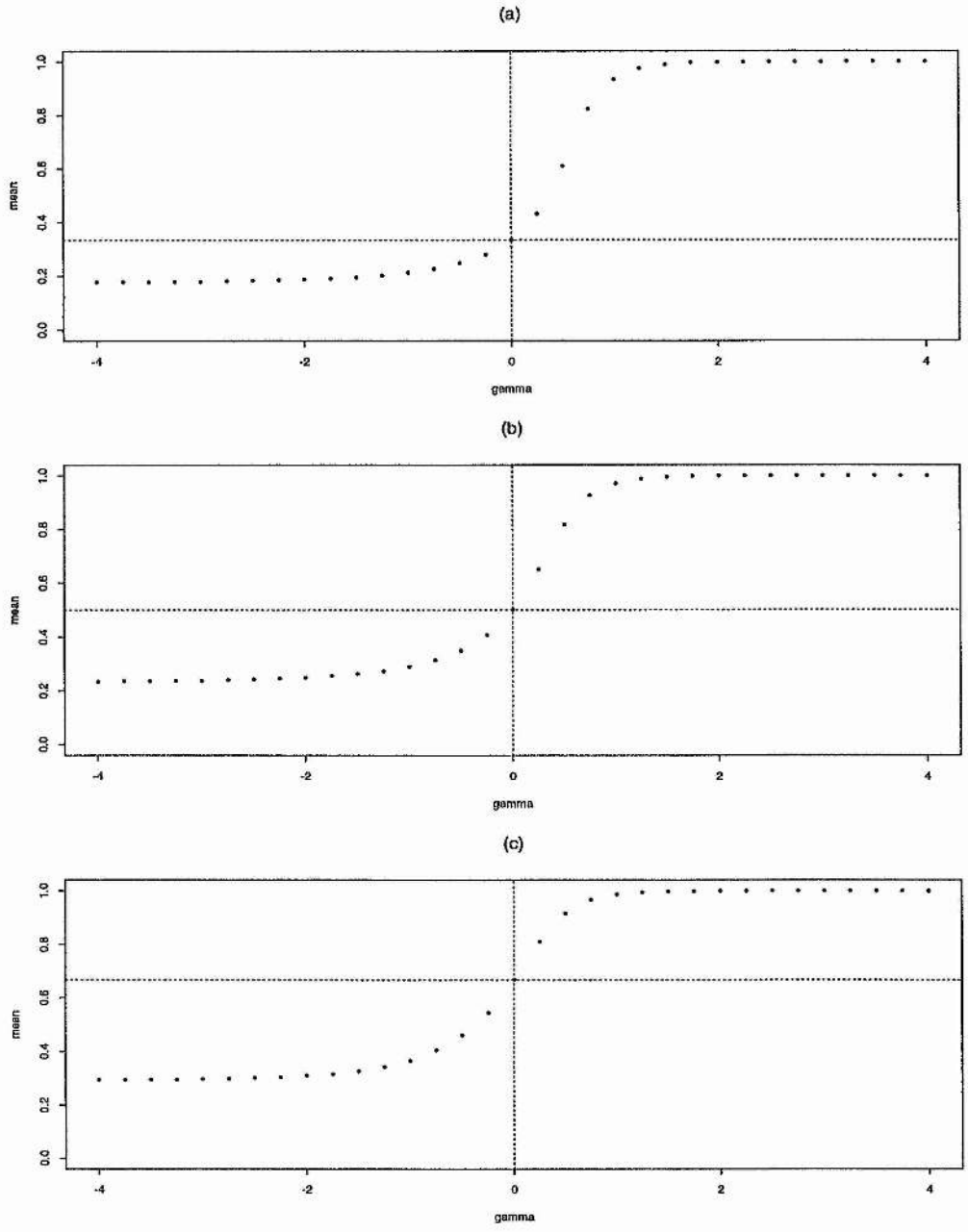


Figure 3.3: Effect of γ on $\bar{y}$ derived from 500 realisations of the auto-logistic model. (a) $\alpha = \log(0.5)$, (b) $\alpha = \log(1)$ and (c) $\alpha = \log(2)$.

values give the same conditional probabilities and hence yield a process with the same mean response $\bar{y}$. This phenomenon is called a phase transition (Pickard, 1987). For $\alpha = \log(1)$, which corresponds to $Pr(y_i) = 0.5$ in the case of independent observations ($\gamma = 0$), Figure 3.3b shows that critical values for γ for such a phase transition are around $\gamma \geq 1$ and $\gamma \leq -1$. The limiting densities yield $\bar{y} = 1$ and $\bar{y} = 0.25$ respectively. In the phase transition for $\gamma \leq -1$ the pattern can be pictured as a chess board with patches of white, where white squares are the zeros. Then the probability of presence at a location i, which is surrounded by zeros, is

$$Pr(y_i|y_j; j \in N_i) = \frac{exp(\alpha)}{1 + exp(\alpha)} = 0.5$$

which can be interpreted as the *persistence* rate given in terms of the inverse link function, here a probability. Different values for α shift the values where phase transition occurs. For $\alpha = \log(0.5)$ (Figure 3.3a) it occurs for negative γ closer to 0 than for $\alpha = \log(1)$, for positive γ it is further from 0. This shift is reversed for $\alpha = \log(2)$ (Figure 3.3c).

The auto-logistic model has been extensively used in image analysis for accurate restoration of binary images (Geman and Geman, 1984). Typically the restoration is based on the posterior probabilities $Pr(\boldsymbol{x}|\boldsymbol{y})$ of the true image $\boldsymbol{x}$ given the observed image $\boldsymbol{y}$. The prior distribution of the true image is assumed to be a MRF, with the conditional auto-logistic model formulation. Heikkinen and Högmander (1994) and Högmander and Møller (1995) apply this approach for atlas mapping of wildlife, the common toad and birds, assuming a homogeneous habitat across the study areas, but incorporating effort of voluntary observers as a covariate. Hoeting et al. (1999) extended their approach by introducing covariate information, hence allowing for habitat heterogeneity.

Additionally to the above applications the auto-logistic model has been fitted to data directly. The wide range of applications covers meteorology (Hughes and Guttorp, 1994) and biology including genetics (Geyer and Thompson, 1992; Abel et al., 1993), plant epidemiology (Preisler, 1993; Preisler and Mitchell, 1993; Gumpertz et al., 1997), plant species mapping (Huffer and Wu, 1998; Wu and Huffer, 1997), and wildlife population mapping Augustin et al. (1996, 1998b). Augustin et al. (1996) fitted an auto-logistic model to presence/absence deer survey data obtained from a random sample of sites by incorporating the Gibbs sampler. Their modelling strategy can be used not only to fit the auto-logistic model to sites included in the survey, but also to estimate the probability that the species is present in the unsurveyed sites. In section 5.4 we revisit this example

in more detail.

3.4 The (truncated) auto-Poisson and auto-negative binomial model for counts

3.4.1 The (truncated) auto-Poisson model

Assume that in an experiment or a survey we observe counts y_i at locations $i = 1, ..., n$. The counts could be for example the number of a particular species in a grid square or the number of fruits on a plant. If the distances between the locations are large, we can assume that the data are independent and that there is no neighbour interaction. We can fit a GLM with a Poisson or negative binomial distribution and a log link to these data:

$$\log(\boldsymbol{\mu}) \quad = \quad X\boldsymbol{\beta}$$

with $y_i \sim \text{Poi}(\mu_i)$ or negative binomial $y_i \sim \text{NegBin}(\mu_i, k)$, for which $Var(y) = \mu_i + \mu_i^2/k$, where k is the aggregation parameter. The negative binomial distribution allows for more variation in the observations than does the Poisson distribution. For now we will concentrate on Poisson data. The matrix $\boldsymbol{X}$ is the design matrix containing explanatory variables and/or factors. With this model we can investigate relationships between spatially referenced response variables and explanatory variables as discussed in chapter 2.

If the counts are observed at adjacent locations, on a lattice structure D with sites i, $(i = 1, ..., n)$, we cannot always assume independence. In this case the auto-Poisson model can be used to describe the actual process underlying $\boldsymbol{Y}$ and to test for the type of interaction present. The auto-Poisson model is a MRF, as described in section 3.2. It is assumed that the data observed are a result of a random process $\boldsymbol{Y} = (Y_i : i \in D)$. The response at each site is a Poisson random variate $Y_i \sim \text{Poi}(\mu_i | y_j; j \in N_i)$ with conditional density function

$$Pr(Y_i = y_i | y_j, j \in N_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}$$

where $\mu_i = \mu_i(y_j; j \in N_i)$ and the auto-Poisson model formulated in terms of the natural parameter is

$$\log(\mu_i | y_j, j \in N_i) = \alpha_i + \sum_{j \in N_i} \gamma_{ij} y_j \tag{3.7}$$

with $\alpha_i = \boldsymbol{x}_i'\boldsymbol{\beta}$. In order to derive the likelihood, we derive the $Q(.)$-function following the steps given in section 3.2. The construction of $Q(\boldsymbol{y})$ is facilitated by the fact that it can be expressed in terms of the conditional probabilities $Pr(y_i|y_j; j \in N_i)$.

$$\begin{aligned} Q(\boldsymbol{y}) &\equiv \log(Pr(\boldsymbol{y})/Pr(\boldsymbol{0})) \\ &= \sum_{i=1}^{n}(\alpha_i y_i - \log(y_i!)) + \sum_{i=1}^{n-1}\sum_{j>i}^{n}\gamma_{ij}y_i y_j \end{aligned}$$

where $\gamma_{ij} = \gamma_{ji}$, $\gamma_{ii} = 0$ and $\gamma_{ij} = 0$ for all $j \notin N_i$.

Unfortunately this model can describe negative correlation only, as the interaction parameter γ_{ij} is constrained to be negative (Besag, 1974): $\sum_{\boldsymbol{z}\in\zeta} exp(Q(\boldsymbol{z})) < \infty$ for all $\gamma_{ij} \leq 0$ only. We now propose a simple way to avoid having a normalising constant diverging to infinity: we truncate the distribution with a nominal truncation value r. We will give the detailed proof for the truncated auto-Poisson model below. The *right truncated Poisson* conditional probability mass function is then

$$Pr(Y_i = y_i|y_j; j \in N_i) = \frac{\mu_i^{y_i}}{y_i!}\left(\sum_{k=0}^{r}\frac{\mu_i^k}{k!}\right)^{-1},$$

with $y_i = 0, 1, ..., r$ and the natural parameter $\log(\mu_i)$. We can then construct $Q(\boldsymbol{y})$ as follows

$$\begin{aligned} Q(\boldsymbol{y}) - Q(\boldsymbol{y}_i) &= \log\left(\frac{Pr(\boldsymbol{y})}{Pr(\boldsymbol{y}_i)}\right) = \log\left(\frac{Pr(y_i|y_j; j \in N_i)}{Pr(0_i|y_j; j \in N_i)}\right) \\ &= \log\left(\frac{\mu_i^{y_i}}{y_i!}\left(\sum_{k=0}^{r}\frac{\mu_i^k}{k!}\right)^{-1}\frac{0!}{\mu_i^0}\left(\sum_{k=0}^{r}\frac{\mu_i^k}{k!}\right)\right) \\ &= y_i\log(\mu_i) - \log(y_i!), \end{aligned}$$

where $\boldsymbol{y}_i = (y_1, ..., y_{i-1}, 0, y_{i+1}, ..., y_n)$ and $\gamma_{ik} = 0$ for $k \notin N_i$. We now insert the model formulation of the natural parameter given in equation 3.7 from Besag's random field theorem (3.4):

$$Q(\boldsymbol{y}) - Q(\boldsymbol{y}_i) = \alpha_i y_i + \sum_{j=1}^{n}\gamma_{ij}y_i y_j - \log(y_i!), \tag{3.8}$$

with i=1,...,n, $\gamma_{ij} = \gamma_{ji}, \gamma_{ii} = 0$ and $\gamma_{ik} = 0$ for $k \notin N_i$. Then from equation 3.2 the form of $Q(.)$-function is a sum of the terms in 3.8 over i. Hence we have

$$Q(\boldsymbol{y}) = \sum_{i=1}^{n}(\alpha_i y_i - \log(y_i!)) + \sum_{i=1}^{n-1}\sum_{j>i}^{n}\gamma_{ij}y_i y_j. \tag{3.9}$$

The $Q(.)$-function of the truncated auto-Poisson model we derived above has the same expression as for the auto-Poisson model, but here $\sum_{z\in\zeta} exp(Q(\boldsymbol{z})) < \infty$ for all finite γ_{ij}. As demonstrated in Kaiser and Cressie (1997) an alternative derivation of the $Q(.)$-function would be to start with the definitions of terms $B_i(y_i)\boldsymbol{x}_i'\beta$ and $B_i(y_i)B_j(y_j)\gamma_{ij}$ given in terms of conditional probabilities (see Cressie, 1991, p. 416). The most straighforward way of deriving the Q()-funtion is to use Besag's theorem directly (equation 3.3).

An alternative to the truncated Poisson distribution is the Winsorized Poisson distribution that is used by Kaiser and Cressie (1997) for modelling Poisson variables with positive auto-correlation. A Winsorized variable is defined as

$$W = Y\ I(Y \leq r) + r\ I(Y > r)$$

with

$$I(Y \leq r) = \begin{cases} 0 & : Y > r \\ 1 & : Y \leq r \end{cases} \quad \text{and} \quad I(Y > r) = \begin{cases} 0 & : Y \leq r \\ 1 & : Y > r \end{cases}$$

for a fixed integer value $r < \infty$. Then the probability mass function is:

$$Pr(W) = \left\{ \frac{\mu^W}{W!} e^{-\mu} \right\} I(W \leq r-1) + \left\{ 1 - \sum_{k=0}^{r-1} \frac{\mu^k}{k!} e^{-\mu} \right\} I(W = r)$$

$$I(W \leq r-1) = \begin{cases} 0 & : W > r-1 \\ 1 & : W \leq r-1 \end{cases} \quad \text{and} \quad I(W = r) = \begin{cases} 1 & : W = r \\ 0 & : W \neq r \end{cases}$$

Using the Lagrange form of the remainder of Taylor's expansion for e^{μ} this can be further simplified to

$$Pr(W) = \left\{ \frac{\mu^W}{W!} e^{-\mu} \right\} I(W \leq r-1) + \left\{ \frac{\mu^r}{r!} e^{\psi-\mu} \right\} I(W = r)$$

where $(\psi - \mu) < 0$. The normalising constant of the likelihood (or joint probability function) $\sum_{\boldsymbol{z}\in\zeta} exp(Q(\boldsymbol{z}))$ is taking the sum of the $Q(.)$-functions over the whole support of the distribution and as a consequence the likelihoods and parameter estimates resulting from maximising these likelihoods could differ between models assuming a truncated or a Winsorized auto-Poisson model. Figure 3.4 shows histograms of data where truncation and Winsorization was applied and illustrates the difference in shape of the two distributions. The data were distributed as $\boldsymbol{y} \sim$ Poisson(5) before the truncation or Winsorization was applied. If the value r is large relative to the mean of the distribution the difference between parameter estimates will be small; the shapes of the distributions in the bottom panel of Figure 3.4 with $r = 30$ do not differ. But if r is small ($r = 9$, top panel), shapes

differ more and it is best to use the most appropriate distribution. For instance if counts greater than r are censored, the Winsorized Poisson distribution is more appropriate.

Properties of the truncated auto-Poisson model

From now on we refer to the truncated auto-Poisson model as the **basic model:**

$$\log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i} y_j. \tag{3.10}$$

for $y_j = 0, ..., r$ assuming $\gamma_{ij} = \gamma$ for all i and j and assuming a homogeneous process with $\alpha_i = \alpha$. Investigating the mean response $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ aids biological interpretation of parameters and determining whether the model is appropriate for explaining processes in vegetation dynamics. The auto-correlation parameter γ has a direct influence on $E(\bar{y}) = \mu_i$:

$$\mu_i \begin{cases} < exp(\alpha) & \text{if } \gamma < 0 \\ = exp(\alpha) & \text{if } \gamma = 0 \\ > exp(\alpha) & \text{if } \gamma > 0 \end{cases}$$

In the case where the observations in all neighbours are zero, $\sum_{j \in N_i} y_j = 0$:

$$\log(\mu_i|y_j; j \in N_i) = \alpha_i.$$

Thus $exp(\alpha_i)$ can be interpreted as the underlying *persistence rate*, given on the response scale. In Figure 3.5(a) the average $\bar{y}$ ($\bar{y}$ is the mean of a single realisation) from 500 simulated data sets, i.e. 500 realisations of a MRF as specified in equation 3.10, is plotted versus different values of γ. In section 4.2 we explain how MRF's are simulated. A first-order neighbourhood is assumed and the *persistence rate* is $exp(\alpha) = 5$. For $\gamma < -1$ the average $\bar{y}$ is roughly equal to half the *persistence rate*. This is because for such strong negative correlation the pattern is similar to a chess board and half of the sites are surrounded by zeros (see also Figure 3.6 top graph). The dip in 3.5(a) correponds to an image similar to Figure 3.6 (top) but with some patches with y-values near zero.

Correlation does not exceed an absolute value of one, so it would make sense to restrict the parameter γ in a similar way. An *ad hoc* calculation can be used to assess the *upper limit* for γ, which defines also the point where phase transition occurs. For example with $\alpha_i = \log(5) = 1.61$ for all i, $\gamma_{ij} = \gamma$ for all i, j and truncation value $r = 100$, μ_i should not exceed a maximum value μ_{max}. Set μ_{max} with the rationale that $\Pr(y_i \geq r) < 0.000001$

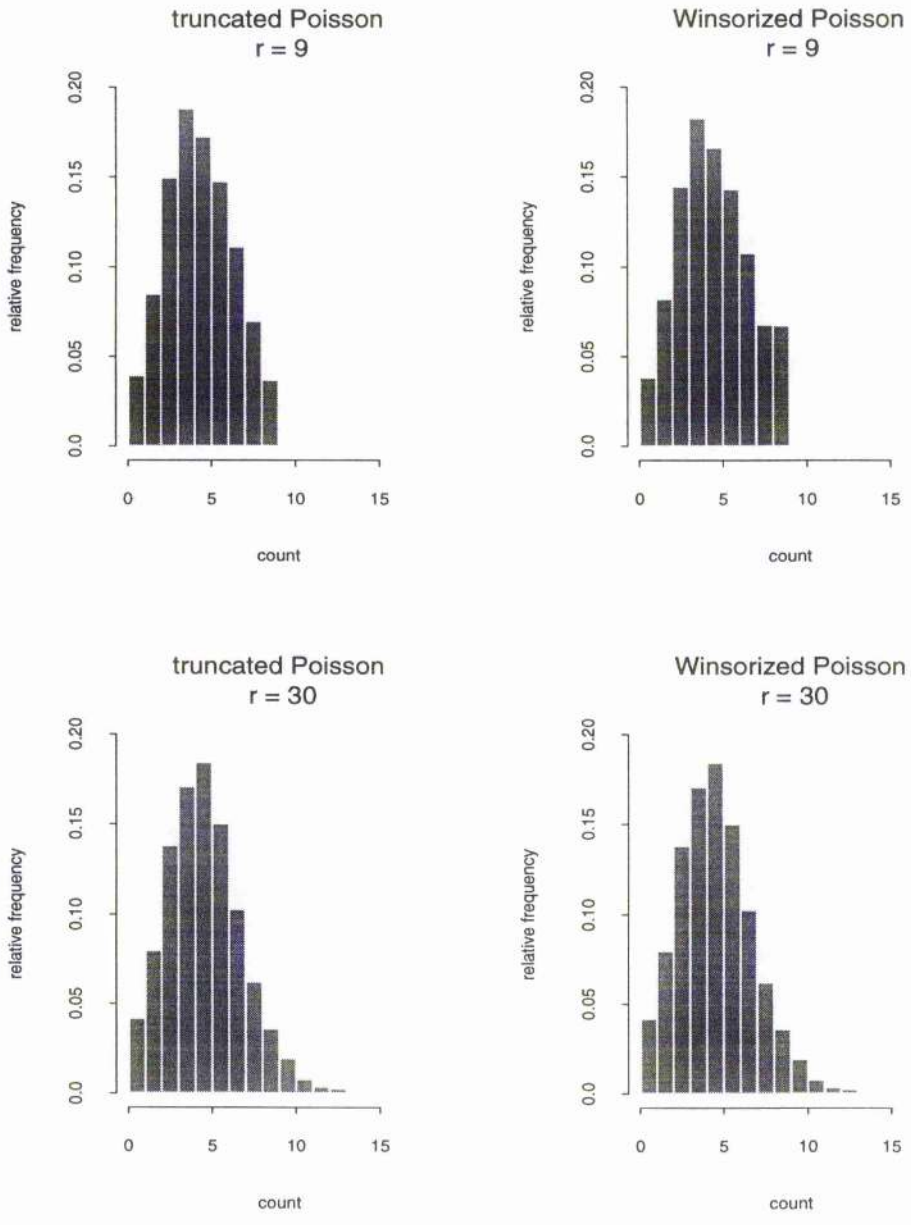


Figure 3.4: Histograms of data $\boldsymbol{y} \sim$ Poisson(5) before the truncation (left) and Winsorization (right) was applied. Truncation values are $r = 10$ (top panel) and $r = 30$ (bottom panel).

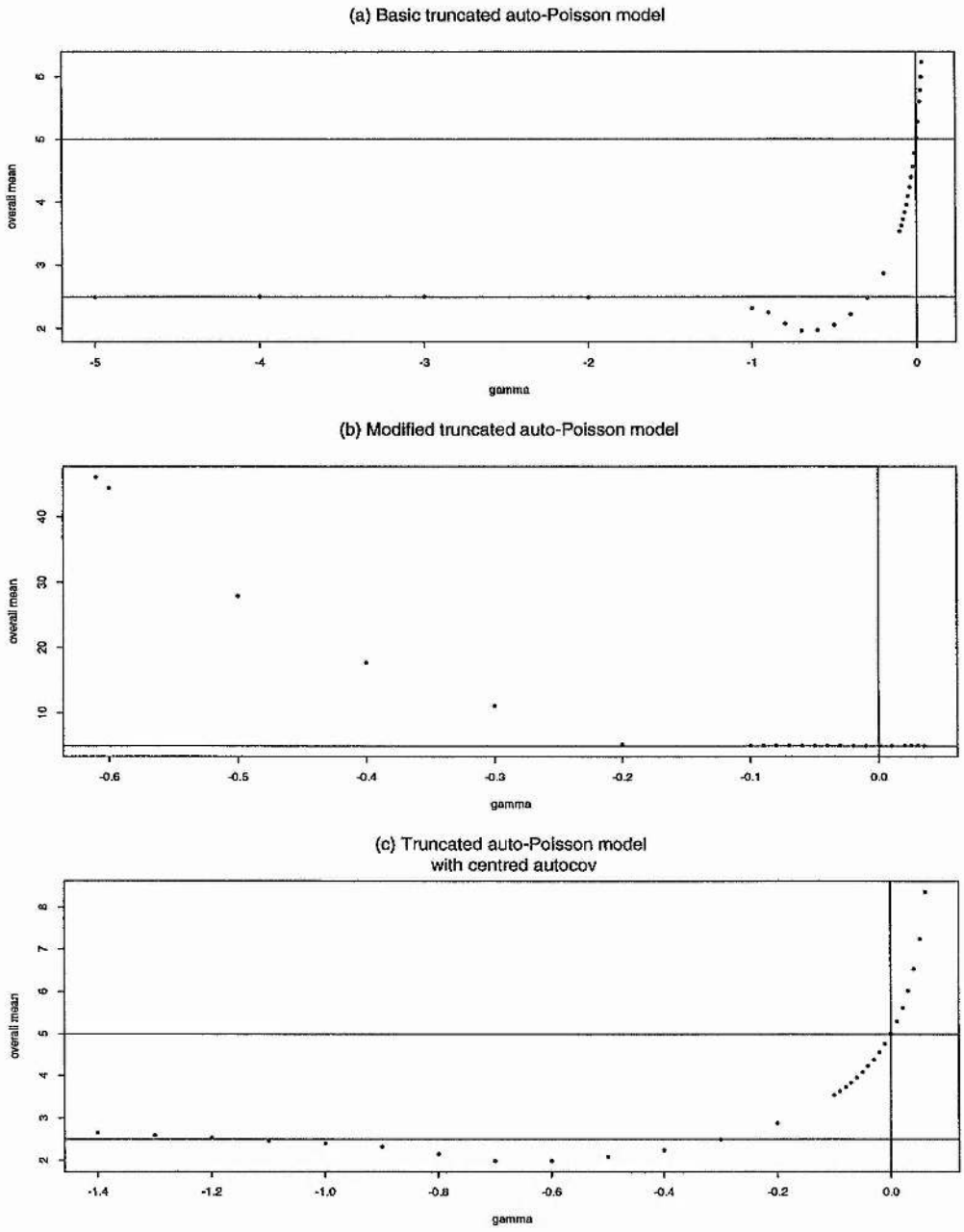


Figure 3.5: Average of $\bar{y}$ from 500 realisations from (a) the *basic truncated auto-Poisson model* with parameters $\alpha = \log(5)$ and different values of γ; (b) the *modified truncated auto-Poisson model* with parameters as in (a), (c) the *truncated auto-Poisson model* with a centred *autocovariate*, parameters as in (a).

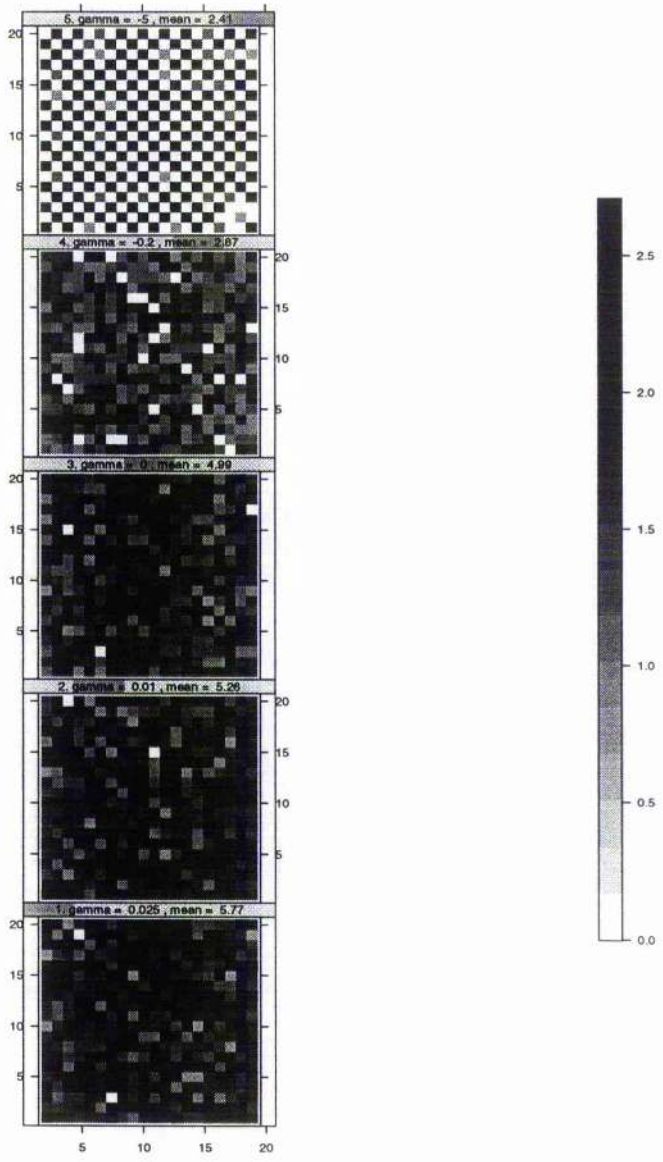


Figure 3.6: log(count+1) from single realisations using the *basic truncated auto-Poisson model* with $\alpha = \log(5)$ and $\gamma = -5, -0.5, 0, 0.01, 0.02$ (from top to bottom) applied to a first-order neighbourhood.

Table 3.1: *Ad hoc* calculation of *upper limits* for γ for different truncation values r for the **basic**, **modified** and **centred** truncated auto-Poisson model.

model	α_i	r	μ_{max}	$\sum_{j \in N_i} y_j$	*upper limit*	*lower limit*	*persistence rate*
basic	$\log(5) = 1.61$	100	60.0	400	0.0062	$-\infty$	$exp(\alpha_i) = 5$
	$\log(5) = 1.61$	50	23.0	200	0.0076	$-\infty$	5
	$\log(5) = 1.61$	30	10.5	120	0.0062	$-\infty$	5
	$\log(10) = 2.3$	100	60.0	400	0.0045	$-\infty$	5
modified	$\log(5) = 1.61$	100	60.0	400	0.0063	-0.124	-
centred	$\log(5) = 1.61$	100	60.0	400	*	*	$exp(\alpha_i - \gamma \sum_{j \in N_i} g)$

* *Upper* and *lower limits* cannot be calculated.

for Poi(μ_{max}). Set $\sum_{j \in N_i} y_j = 4 * r$, assuming a "first-order" neighbourhood for which there are, except at edge squares, four neighbours. Then

$$\begin{aligned} \log(\mu_{max}) &\geq \alpha_i + \gamma \sum_{j \in N_i} y_j \\ \frac{\log(\mu_{max}) - \alpha_i}{\sum_{j \in N_i} y_j} &\geq \gamma \\ \textit{upper limit} &\geq \gamma. \end{aligned}$$

Table 3.1 shows that $r = 100$ with $\mu_{max} = 60$ gives an *upper limit* of 0.0062. The *upper limit* decreases with increasing α_i, e.g. for $\alpha_i = \log(10)$ for all i, $0.0045 \geq \gamma$. Decreasing r does not change the *upper limit* dramatically because μ_{max} is reduced, e.g. taking $r = 50$ increases the *upper limit* to $0.0076 \geq \gamma$, but taking $r = 30$ results in a small $\mu_{max} = 10.5$ and $0.0062 \geq \gamma$. There is no *lower limit* on γ due to the exponential link function, but as illustrated in Figure 3.5(a), any value of $\gamma < -1$ yields the same effect. To summarise, $\gamma < -1$ leads to phase transition and $\gamma >$ *upper limit* leads to explosion. Finally we show that the auto-Poisson model can describe a range of scenarios and patterns within limits of γ.

1. $\gamma < 0 \equiv$ competition, which yields regular patterns, and in extreme cases, a chess board effect (see Figure 3.6 top);

2. $\gamma > 0 \equiv$ co-operative effects, which yields clustering (see Figure 3.6 bottom);

3. different values of γ in different directions yields directional effects, for instance clustering along columns;

4. $\gamma > 0$ in nearest neighbourhood and $\gamma < 0$ in higher order neighbourhoods yields clusters of y_i's with similar values extending over the nearest neighourhood.

Modifying the auto-Poisson model: The fact that γ can influence global characteristics is a desirable feature for data where competition is expected, but might not always be appropriate for other situations, for example positive correlation caused by latent effects, such as unobserved explanatory information. We can modify the (truncated) auto-Poisson model in such a way that the auto-correlation has no direct influence on the mean:

$$\log(\mu_i|y_j; j \in N_i) = \alpha_i + \sum_{j \in N_i} \gamma_{ij}(y_j - exp(\alpha_j))$$

This model is similar to the auto-regressive model for normal data, where the autocorrelation acts on the response in the neighbourhood after subtracting the trend. We now derive the $Q(.)$-function for this model (note that the first line here follows from equation 3.8):

$$\begin{aligned} Q(\boldsymbol{y}) - Q(\boldsymbol{y_i}) &= y_i \log(\mu_i|y_j; j \in N_i) - \log(y_i!) \\ &= y_i \left\{ \alpha_i + \sum_{j=1}^{n} \gamma_{ij}(y_j - exp(\alpha_j)) \right\} - \log(y_i!), \end{aligned}$$

with $\gamma_{ij} = \gamma_{ji}, \gamma_{ii} = 0$ and $\gamma_{ik} = 0$ for $k \notin N_i$. This implies

$$Q(\boldsymbol{y}) = \sum_{i=1}^{n}(y_i\alpha_i - \log(y_i!)) + \sum_{i=1}^{n-1}\sum_{j>i}^{n} \gamma_{ij}y_i(y_j - exp(\alpha_j)).$$

Comparing the above function with the $Q(.)$-function given in Besag's theorem in equation 3.3 gives $B_i(y_i) = y_i$ and $B_j(y_j) = y_j$, assuming that the sum $\sum_{j>i}^{n} \gamma_{ij}exp(\alpha_j)$ is absorbed into the the parameter α_i. Thus the proposed modified auto-Poisson model is a valid MRF. The summability condition $\sum_{z \in \zeta} exp(Q(z)) < \infty$ holds for all γ_{ij}.

Here auto-correlation does not affect $\bar{y}$ for γ within limits. Figure 3.5(b) shows this. In the case where all neighbours are zero ($\sum_{j \in N_i} y_j = 0$) and $\gamma_{ij} = \gamma$ for all i, j:

$$\log(\mu_i|y_j; j \in N_i) = \alpha_i - \gamma \sum_{j \in N_i} exp(\alpha_j).$$

For positive γ, $\mu_i < exp(\alpha_i)$ and for negative γ, $\mu_i > exp(\alpha_i)$. Hence for this model there is *no* unambiguous *persistence rate.*

The *ad hoc* calculation with $\alpha_i = \log(5)$ for all i, $\gamma_{ij} = \gamma$ for all i, j, truncation value $r = 100$ and $\mu_{max} = 60$ is applied for the case where all neighbours are zero and this yields a *lower limit*:

$$\begin{aligned} \log(\mu_{max}) &\geq \alpha_i - \gamma \sum_{j \in N_i} exp(\alpha_j) \\ \frac{\log(\mu_{max}) - \alpha_i}{-\sum_{j \in N_i} exp(\alpha_j)} &\leq \gamma \\ -0.124 &\leq \gamma \end{aligned}$$

The *ad hoc* calculation (Table 3.1) also yields an *upper limit*:

$$\begin{aligned} \frac{\log(\mu_{max}) - \alpha_i}{\sum_{j \in N_i}(y_j - exp(\alpha))} &\geq \gamma \\ 0.0063 &\geq \gamma \end{aligned}$$

which is similar to the result for the basic model. Again by decreasing the truncation value r the *upper limit* does not increase substantially.

Another possibility for modifying the basic auto-Poisson model is

$$\log(\mu_i | y_j : j \neq i) = \alpha_i + \gamma_{ij} \sum_{j \in N_i} (y_j - \bar{y})$$

where $\bar{y}$ is the average of observations y. This amounts to centring the *autocovariate*. Along the same lines as above we can prove that this model is a valid MRF. The mean response y_i is affected by γ as shown in Figure 3.5(c), which shows the behaviour is very similar to the basic (truncated) auto-Poisson model. The only difference is that this model has a lower limit. In the case where all neighbours are zero ($\sum_{j \in N_i} y_j = 0$) and $\gamma_{ij} = \gamma$ for all i, j:

$$\log(\mu_i | y_j; j \in N_i) = \alpha_i - \gamma \sum_{j \in N_i} \bar{y}.$$

the *underlying persistence rate* is $exp(\alpha_i - \gamma \sum_{j \in N_i} \bar{y})$. As for the modified model *lower* and *upper limits* for γ exist, but cannot be evaluated exactly (Table 3.1).
The *lower limit* is

$$\begin{aligned} \log(\mu_{max}) &\geq \alpha_i - \gamma \sum_{j \in N_i} \bar{y} \\ \frac{\log(\mu_{max}) - \alpha_i}{-\sum_{j \in N_i} \bar{y}} &\leq \gamma \end{aligned}$$

and the *upper limit* is

$$\frac{\log(\mu_{max}) - \alpha_i}{\sum_{j \in N_i}(y_j - \bar{y})} \geq \gamma.$$

3.4.2 Temporal truncated auto-Poisson model

If data at different time points are available, the model can be easily extended to a spatio-temporal model by including counts at site i from previous time points as explanatory variables. Let y_{it} be the count at site i at time point t and assume again the simple case $\alpha_i = \log(5)$ for all i and $\gamma_{ij} = \gamma$ for all i, j and

$$\log(\mu_{i,t}|y_{i,t-1}, y_{j,t}; j \in N_i) = \theta \log(y_{i,t-1} + 1) + \alpha_i + \gamma \sum_{j \in N_i} y_j;$$

with $Y_{i,t} \sim \text{Poi}(\mu_{i,t}|y_{i,t-1}, y_{j,t}; j \in N_i)$, θ is the coefficient for $\log(y_{i,t-1} + 1)$.

3.4.3 The (truncated) auto-negative binomial model

For data which exhibit more variation than Poisson variables the negative binomial distribution is more appropriate. If counts have a Poisson distribution with variable rate, the rate itself having a gamma distribution, then the counts can be shown to have a negative binomial distribution (Johnson et al., 1992). Let $y \sim \text{Poi}(\lambda)$ with probability mass function (p.m.f)

$$Pr(y) = \frac{\lambda^y e^{-\lambda}}{y!}$$

with $\lambda \sim \text{Gamma}(\beta, k)$, $\beta > 0, k > 0$ and probability density function (p.d.f.)

$$f(\lambda) = \frac{\beta^k \lambda^{k-1} e^{-\beta\lambda}}{\Gamma(k)}.$$

The parameters k and β could be described as the shape and scale parameter, i.e. varying k changes the shape of the distribution and varying β changes the units of measurement. The marginal distribution of this mixture is a negative binomial with p.m.f.

$$\begin{aligned} Pr(y) &= \frac{\beta^k}{\Gamma(k)} \int_0^\infty \frac{\lambda^y e^{-\lambda}}{y!} \lambda^{k-1} e^{-\beta\lambda} d\lambda \\ &= \frac{\Gamma(y+k)}{y!\Gamma(k)} \left(\frac{\beta}{\beta+1}\right)^k \left(\frac{1}{\beta+1}\right)^y, \; y = 0, 1, ... \end{aligned}$$

Thus $Y \sim \text{NegBin}(\mu, k)$, with $\mu = \frac{k}{\beta}$ and $Var(\mu) = \mu + \frac{\mu^2}{k}$. Again the parameter k can be described as a shape or aggregation parameter. As $k \mapsto \infty$, the negative binomial (NegBin) distribution tends towards a Poisson distribution. The parameter k can be estimated using an extra iterative step in the GLM procedure. If there is need to account for auto-correlation, a MRF can be specified with a NegBin distribution, yielding the

auto-negative binomial model:

$$Pr(y_i|y_j; j \neq i) = \frac{\Gamma(y_i+k)}{y_i!\Gamma(k)}\left(\frac{\beta_i}{\beta_i+1}\right)^k\left(\frac{1}{\beta_i+1}\right)^{y_i},$$

with mean $\mu_i = \frac{k}{\beta_i}$ and variance $= \frac{k}{\beta_i}(1+\frac{1}{\beta_i})$. Note that we define $\beta_i = \beta_i(y_j; j \in N_i)$ to simplify notation. It follows that the natural parameter model is

$$\log(\frac{1}{\beta_i+1}) = \alpha_i + \sum_{j\in N_i}\gamma_{ij}y_j$$

where $\gamma_{ij} = \gamma_{ji}, \gamma_{ii} = 0$ and expressed in terms of the mean

$$\mu_{ij} = \frac{k}{\beta_i} = \frac{k\exp\left(\alpha_i + \sum_{j\in N_i}\gamma_{ij}y_j\right)}{1-\exp\left(\alpha_i+\sum_{j\in N_i}\gamma_{ij}y_j\right)}.$$

Note that only the parameter β_i is a function of the response in the neighbourhood of site i, not k. We now derive the $Q(.)$-function as for the auto-Poisson model:

$$\begin{aligned} Q(\boldsymbol{y}) - Q(\boldsymbol{y}_i) &= \log\left(\frac{Pr(\boldsymbol{y})}{Pr(\boldsymbol{y}_i)}\right) = \log\left(\frac{Pr(y_i|y_j; j\in N_i)}{Pr(0_i|y_j; j\in N_i)}\right) \\ &= y_i\log\left(\frac{1}{\beta_i+1}\right) + \log\left(\frac{\Gamma(y_i+k)}{y_i!\Gamma(k)}\right) \\ &= y_i\alpha_i + \sum_{j=1}^{n}\gamma_{ij}y_iy_j + \log\left(\frac{\Gamma(y_i+k)}{y_i!\Gamma(k)}\right) \end{aligned}$$

where 0_i denotes the event $y_i = 0$ and $\boldsymbol{y}_i \equiv (y_1, ..., y_{i-1}, 0, y_{i+1}, ..., y_n)$ and $\gamma_{ij} = \gamma_{ji}, \gamma_{ii} = 0$ and $\gamma_{ik} = 0$ for $k \notin N_i$. Hence

$$Q(\boldsymbol{y}) = \sum_{i=1}^{n}y_i\alpha_i + \sum_{i=1}^{n-1}\sum_{j>i}^{n}\gamma_{ij}y_iy_j + \sum_{i=1}^{n}\log\left(\frac{\Gamma(y_i+k)}{y_i!\Gamma(k)}\right)$$

As for the auto-Poisson case the model is only valid for negative auto-correlation and by truncating the distribution at the right end this can be circumvented. The truncated p.m.f is

$$Pr(y_i|y_j; j\in N_i) = \frac{\Gamma(y_i+k)}{y_i!}\left(\frac{1}{\beta_i+1}\right)^{y_i}\left(\sum_{l=0}^{r}\frac{\Gamma(l+k)}{l!}\left(\frac{1}{\beta_i+1}\right)^l\right)^{-1}$$

with $y_i = 0, 1, ..., r$ and the natural parameter $\log(\frac{1}{\beta_i+1})$. Then

$$\begin{aligned} Q(\boldsymbol{y}) - Q(\boldsymbol{y}_i) &= \log\left(\frac{Pr(y_i|y_j; j\in N_i)}{Pr(0_i|y_j; j\in N_i)}\right) \\ &= \log\left(\frac{\Gamma(y_i+k)}{y_i!}\left(\frac{1}{\beta_i+1}\right)^{y_i}c\right) - \log\left(\frac{\Gamma(0+k)}{0!}\left(\frac{1}{\beta_i+1}\right)^0 c\right) \\ &= \log\left(\frac{\Gamma(y_i+k)}{y_i!\Gamma(k)}\right) + y_i\log\left(\frac{1}{\beta_i+1}\right) \end{aligned}$$

where

$$c = \left(\sum_{l=0}^{r} \frac{\Gamma(l+k)}{l!} \left(\frac{1}{\beta_i + 1} \right)^l \right)^{-1}$$

and

$$Q(\boldsymbol{y}) = \sum_{i=1}^{n} \alpha_i y_i + \sum_{i=1}^{n-1} \sum_{j>i}^{n} \gamma_{ij} y_i y_j - \sum_{i=1}^{n} \log \left(\frac{\Gamma(y_i + k)}{y_i! \Gamma(k)} \right),$$

with $\gamma_{ij} = \gamma_{ji}, \gamma_{ii} = 0$ and $\gamma_{ik} = 0$ for $k \notin N_i$. As for the auto-negative binomial model the model expressed in terms of the mean is:

$$\mu_{ij} = \frac{k}{\beta_i} = \frac{k \exp\left(\alpha_i + \sum_{j=1}^{n} \gamma_{ij} y_j\right)}{1 - \exp\left(\alpha_i + \sum_{j=1}^{n} \gamma_{ij} y_j\right)}.$$

with $\beta_i = \beta_i(y_j; j \in N_i) > 0$ and $k > 0$ for all i.

Properties of the truncated auto-negative binomial model

Assume the following model

$$\log\left(\frac{1}{\beta_i(y_j; j \in N_i) + 1}\right) = \alpha + \sum_{j \in N_i} \gamma y_j$$

with $\gamma_{ij} = \gamma$ for all i and j. Unfortunately the restriction $\beta > 0$ leads to a very low upper limit for γ. In Figure 3.7 we show the average $\bar{y}$ from 500 simulated count data sets with the truncated auto-negative binomial model for a range of γ values. The other parameters were $k = 5$ and $\alpha = \log(0.5)$, corresponding to $\mu = \frac{k}{\beta} = 5$ for the case of independent observations and an untruncated NegBin distribution. The plot shows the same picture as for the auto-Poisson case, where we have an upper boundary for γ. Again there is phase transition at the lower region of γ and explosion at the upper boundary. Here the upper boundary is dictated by the restriction $\beta > 0$.

This makes the (truncated) auto-negative model for positive correlation extremely cumbersome to work with, and for highly overdispersed count data with auto-correlation, other methods such as gamma-Poisson random fields (Wolpert and Ickstadt, 1998; Brix, 1998) or Gaussian-Poisson random fields (Besag et al., 1991; Breslow and Clayton, 1993) are recommended.

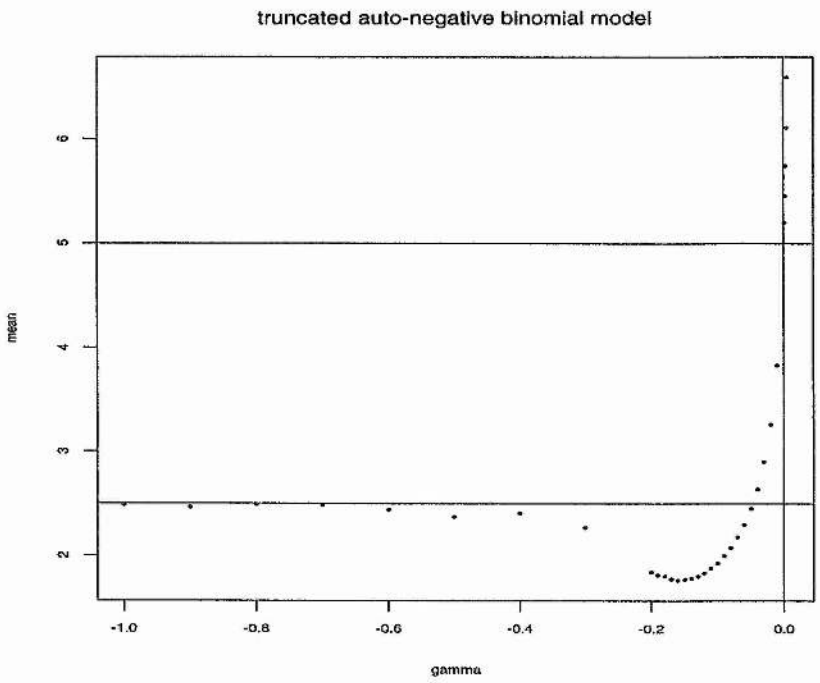


Figure 3.7: Average of $\bar{y}$ from 500 realisations from the truncated auto-negative binomial model with parameters $k = 5$, and $\alpha = \log(0.5)$ corresponding to $\mu = \frac{k}{\beta} = 5$ for the case of independent observations and for a range of γ values.

3.5 Summary

In this chapter auto-models for discrete data, in particular counts, were defined. We have derived the Q(.)-functions of the truncated auto-Poisson and the auto-NegBin (truncated and untruncated) model. We have shown that the problem of the restriction of the autocorrelation parameter γ to be negative can be solved by truncating the Poisson or NegBin distribution. But investigating the properties of the diverse auto-models for discrete data showed that both the truncated auto-Poisson and the truncated auto-NegBin model still have an upper boundary for the autocorrelation parameter γ. There are no restrictions on γ for the auto-logistic model.

Chapter 4

Parameter estimation for auto-models

4.1 Introduction

Since the auto-models incorporate dependence between responses, the likelihood is not always available in closed form and parameter estimation is not straightforward. One method of parameter estimation is maximum pseudo likelihood (MPL). This method ignores the fact that observations are correlated and constructs the likelihood as usual by multiplying the conditional probabilities. An advantage is that it is a very cheap method in computing time and estimates can be obtained from a GLM package. An alternative is Monte Carlo maximum likelihood (MCML) (Geyer and Thompson, 1992), which approximates the term $\sum_{z\in\zeta} exp(Q(z))$ of the likelihood from equation 3.2 using samples drawn from a Markov chain. We compare the two methods to see whether the "noddy" MPL method is good enough for our purpose. We also check whether the MCML method converges.

MCML has been compared with MPL for the auto-logistic model by Wu and Huffer (1997), who conclude that the MCML method yields more accurate results in terms of standard errors and mean square error than the MPL method, but MCML estimates are more biased than MPL estimates. Huffer and Wu (1998) investigate the distributional behaviour of MCML estimates of parameters from the auto-logistic model and find that

the Fisher information of the MCML can be used for variance estimation. In both papers the auto-logistic model is applied to modelling the distribution of plant species in Florida. Wu and Huffer (1997) and Huffer and Wu (1998) do not investigate convergence of the MCML.

In section 4.2 we introduce an important prerequisite of the MCML: the simulation of MRF data, which is also essential for fitted values and goodness-of-fit statistics. In section 4.3 we describe the diverse parameter estimation methods and illustrate them using simulated data examples. We compare the different parameter estimation methods by a simulation study in section 4.4.

4.2 Simulating MRF data using the Gibbs sampler

Data simulation is essential for the methods discussed here. In particular it is required for the MCML method described in section 4.3, and for obtaining fitted values and predictions. For a Gaussian MRF, also called an auto-normal model, data simulation is straightforward; it only involves simulating random white noise (see also Cressie, 1991, p. 202). The joint density distribution is a multivariate normal distribution with $\boldsymbol{Y} \sim \text{MVN}(\boldsymbol{\mu}, (\boldsymbol{I}-\boldsymbol{C})^{-1}\boldsymbol{M})$. For simulation we use the equation $\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{\epsilon}$ where $\boldsymbol{L}$ is the Cholesky decomposition of the variance $Var(\boldsymbol{Y}) = (\boldsymbol{I}-\boldsymbol{C})^{-1}\boldsymbol{M}$ and $\boldsymbol{\epsilon}$ is a vector of $N(0,1)$ independently distributed variables.

For other distributions simulation is not so straightforward. The joint density is generally not available in closed form and even when it is, it can be very computer intensive to simulate directly from it.

Given that our distribution of interest, the MRF defined by the auto-model, is not available in closed form and the fact that the full conditional densities for the response at locations i are available, the Gibbs sampler (Geman and Geman, 1984), a Markov chain Monte Carlo scheme (MCMC), is a natural choice for data simulation. In general the Gibbs sampler applies when the distribution of interest $\pi(\boldsymbol{\theta})$ cannot be sampled from, but full conditional distributions $\pi(\theta_i|\theta_j; j \neq i), i = 1, ..., p$ are available. The components θ_i can be scalars, vectors or matrices. As summarised by Gamerman (1997), the algorithm is based on successive generations from the full conditional distributions:

1. Set the counter of the chain t to 1 and set initial values

$$\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, ..., \theta_p^{(0)})$$

2. Obtain a new value $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, ..., \theta_p^{(t)})$ through successive updating of values

$$\begin{aligned}
\theta_1^{(t)} &\sim \quad \pi(\theta_1|\theta_2^{(t-1)}, ..., \theta_p^{(t-1)}) \\
\theta_2^{(t)} &\sim \quad \pi(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, ..., \theta_p^{(t-1)}) \\
&\cdot \\
&\cdot \\
&\cdot \\
\theta_p^{(t)} &\sim \quad \pi(\theta_p|\theta_1^{(t)}, ..., \theta_{p-1}^{(t)})
\end{aligned}$$

3. Set $t = t + 1$ and return to step 2 until convergence is reached.

Note that convergence in step 3 is not defined strictly. There are different ways to test for convergence and some of these will be discussed below in section 4.2.1.

Now we translate the general algorithm to the case of the auto-Poisson model with $Y_i \sim \text{truncPoi}(\mu_i|y_j; j \in N_i)$, $i = 1, ..., n$:

$$\log(\mu_i|y_j; j \in N_i) = \boldsymbol{x}_i'\boldsymbol{\beta} + \sum_{j \in N_i} \gamma_{ij} y_j. \tag{4.1}$$

Here our distribution of interest is the joint density $Pr(\boldsymbol{y})$ of the MRF with the above conditional specification. Thus the Gibbs sampler is implemented as follows:

1. Set the counter of the chain t to 1 and set initial values $\boldsymbol{y}^{(0)}$ by generating independent $\boldsymbol{Y}$ where $y_i^{(0)} \sim \text{Poi}(\mu_i)$ with $\log(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$. The term $\boldsymbol{x}_i'\boldsymbol{\beta}$ is the part of the model in equation 4.1 describing the trend only.

2. Update $\boldsymbol{y}$ in the lattice successively by systematically scanning through the lattice or by picking locations at random and without replacement:

$$y_i^{(t)} \sim \text{truncPoi}(\mu_i|y_j^{(t)}, y_j^{(t-1)}; j \in N_i)$$

with μ_i as defined in equation 4.1. The neighbourhood N_i is calculated using the current realisation of the lattice, i.e. it can contain previously updated values $y_j^{(t)}$ and $y_j^{(t-1)}$, values which have not yet been updated.

3. Set $t = t + 1$ and return to step 2 until convergence is reached. Once convergence is reached, save realisations $\boldsymbol{y}^{(t)}$ at every lth step yielding the sample of size m $\boldsymbol{y}^{(c)}, \boldsymbol{y}^{(c+l)}, ..., \boldsymbol{y}^{(c+(m-1)l)}$, where c is the point of convergence.

To simulate truncated Poisson variables we use the Poisson distribution and reject values which are greater than the truncation value r. When consecutive realisations $\boldsymbol{y}^{(t)}$ are highly correlated and each realisation is saved, a higher sample size is required than for the case of low correlation between iterations. Instead, only every lth realisation of the Markov Chain can be stored, where l is chosen so that realisations are independent. This practice reduces the amount of storage and is referred to as *thinning* the chain (Raftery and Steven, 1995).

For samples $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, ..., \boldsymbol{y}^{(m)}$ generated by the Gibbs sampler to converge to the stationary distribution $P(\boldsymbol{y}) = \frac{exp(Q(\boldsymbol{y}))}{\sum_{\boldsymbol{z} \in \zeta} exp(Q(\boldsymbol{z}))}$ the Markov chain must be *ergodic*, that is irreducible, positive recurrent and aperiodic (Roberts, 1995). Ergodicity ensures that taking an average of the sampled realisations $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, ..., \boldsymbol{y}^{(m)}$ is an approximation for $E(\boldsymbol{y})$, which is also called an *ergodic average*.

The probability of changing to state $y_i^{(t)}$ at step t from state $y_i^{(t-1)}$ at step $t-1$ is the transition probability for the Gibbs sampler. Here the transition probabilities which we denote in shorthand notation by $Pr(y_i|y_j)$ are:

$$Pr(y_i|y_j) = Pr(Y_i^{(t)} = y_i^{(t)}|y_j^{(t)}, y_j^{(t-1)}; j \in N_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}.$$

Note that $\mu_i = \mu_i(y_j; j \in N_i)$ and that to simplify notation we refer to the Poisson distribution rather than the truncated Poisson distribution. Then

1. A chain is irreducible if any state can be reached from any starting point. This implies that all transition probabilities have to be greater than zero, which is the case here. $Pr(y_i|y_j) > 0$ since $exp(-\mu_i) > 0; \mu^{y_i} > 0; y_i! > 0$.

2. The chain is positive recurrent:

   Define τ_{ii} as the time of the first return to state i. The chain is positive recurrent if $Pr(\tau_{ii} < \infty) = 1$ and $E(\tau_{ii}) < \infty$. This condition is fulfilled because all $Pr(y_i|y_j) > 0$. This also implies that the chain is aperiodic, the greatest common divider of t the time of reoccurrence of state i, $t > 0 : Pr_{ii}(t) > 0$, is one.

Simulation of auto-logistic and auto-negative binomial MRFs follows the same procedure. For the auto-logistic MRF the conditional distribution is Bernoulli$(p_i|y_j; j \in N_i)$ and the model for p_i is as defined in equation 3.6.

For data simulation of the auto-negative binomial MRF we make use of the fact that the NegBin distribution can be derived as a mixture of Poisson distributions with rates distributed as Gamma$(\beta_i(y_j; j \in N_i), k)$ (see section 3.4.3 or Johnson et al. (1992)). For sampling from $Y_i \sim$ trunc. NegBin$(\beta_i(y_j; j \in N_i), k)$, we first sample

$$\lambda_i \sim \text{Gamma}(\beta_i(y_j; j \in N_i), k),$$

for $\beta > 0, k > 0$ and then generate

$$y_i \sim \text{truncPoi}(\lambda_i|y_j; j \in N_i).$$

4.2.1 Convergence of the Gibbs sampler

Convergence of the Gibbs sampler was assessed by monitoring the following summary statistics of the likelihood (see section 4.3.3 for more details):

$$\boldsymbol{t}(\mathbf{y}) = (\boldsymbol{t}_1, t_2)' = (\sum_i^n \boldsymbol{x}_i y_i, \frac{1}{2}\sum_i^n y_i autocovariate_i)',$$

where $\boldsymbol{t}_1$ corresponds to the summary statistic of the parameter $\boldsymbol{\beta}$ and t_2 corresponds to summary statistic of the parameter γ, assuming $\gamma = \gamma_{ij}$ and $autocovariate_i = \sum_{j \in N_i} y_j$ in the Q()-function of the truncated auto-Poisson model:

$$Q(\boldsymbol{y}) = \sum_{i=1}^{n}(\boldsymbol{x}_i'\boldsymbol{\beta} y_i - \log(y_i!)) + \sum_{i=1}^{n-1}\sum_{j>i}^{n}\gamma_{ij} y_i y_j.$$

Computationally it is more practical to use the autocovariate in the auto-Poisson model (equation 4.1) to calculate $t_2 = \sum_{i=1}^{n-1}\sum_{j>i}^{n} y_i y_j = \frac{1}{2}\sum y_i autocovariate_i$.

We also evaluate the autocorrelation in the time series of summary statistics at different lags. This is to determine the length of the l steps for *thinning* the chain. For moderate levels of positive autocorrelation in the auto-Poisson model (equation 4.1), where a homogeneous process is assumed with $x_i'\beta = \alpha = log(5)$ and $\gamma = 0.015$, there is almost zero correlation of summary statistics between samples of the Markov chain (Figure 4.1). For higher levels of positive autocorrelation ($\alpha = log(5), \gamma = 0.05$) the correlation of summary

statistics between iterations increases and is positive (Figure 4.2). For negative autocorrelation ($\gamma = -0.5$), there is low negative correlation between samples (Figure 4.3). In all cases the mixing of the chain is fast. Mixing is the rate the sampler moves around the target distribution. Figures 4.1 to 4.3 show that the summary statistics from each iteration to not remain on similar levels for any length of time, thus the Gibbs sampler moves around the target distribution very fast.

For a more formal assessment of convergence we use a method given in Gelman and Rubin (1992) and Gelman (1995). This method is based on comparing the between and within variance of parallel sequences of monitored summary statistics from a Markov chain. The idea is that with convergence the empirical distributions obtained separately from each sequence is approximately the same as the distribution obtained by mixing all sequences together. Before the sequences have converged the simulation collected within each single sequence will be much less variable than the simulation collected from all the sequences combined. Here we monitor both summary statistics t_k, $k = 1, 2$ in m parallel sequences of length n with $i = 1, ..., m$ and $j = 1, ..., n$, yielding t_{ijk}. The between-sequence variances B and within-sequence variance W are calculated as follows (Gelman, 1995):

$$B = \frac{n}{m-1}\sum_{i=1}^{m}(\bar{t}_{ki.} - \bar{t}_{k..})^2, \text{ where } \bar{t}_{ki.} = \frac{1}{n}\sum_{j=1}^{n} t_{kij} \text{ and } \bar{t}_{k..} = \frac{1}{m}\sum_{i=1}^{m}\bar{t}_{ki.},$$

$$W = \frac{1}{m}\sum_{i=1}^{m} s_i^2, \text{ where } s_i^2 = \frac{1}{(n-1)}\sum_{j=1}^{n}(t_{kij} - \bar{t}_{ki.})^2$$

A conservative variance estimate of the target distribution is

$$\widehat{var}(t_k) = \frac{n-1}{n}W + \frac{1}{n}B.$$

With the above components we can estimate the ratio between the upper and lower bound for the standard deviation of t_k, in our case t_1 and t_2:

$$\sqrt{\hat{R}} = \sqrt{\frac{\widehat{var}(t_k)}{W}}.$$

With convergence this ratio declines to 1. In our case the ratio is close to 1 very rapidly, after only 100 iterations.

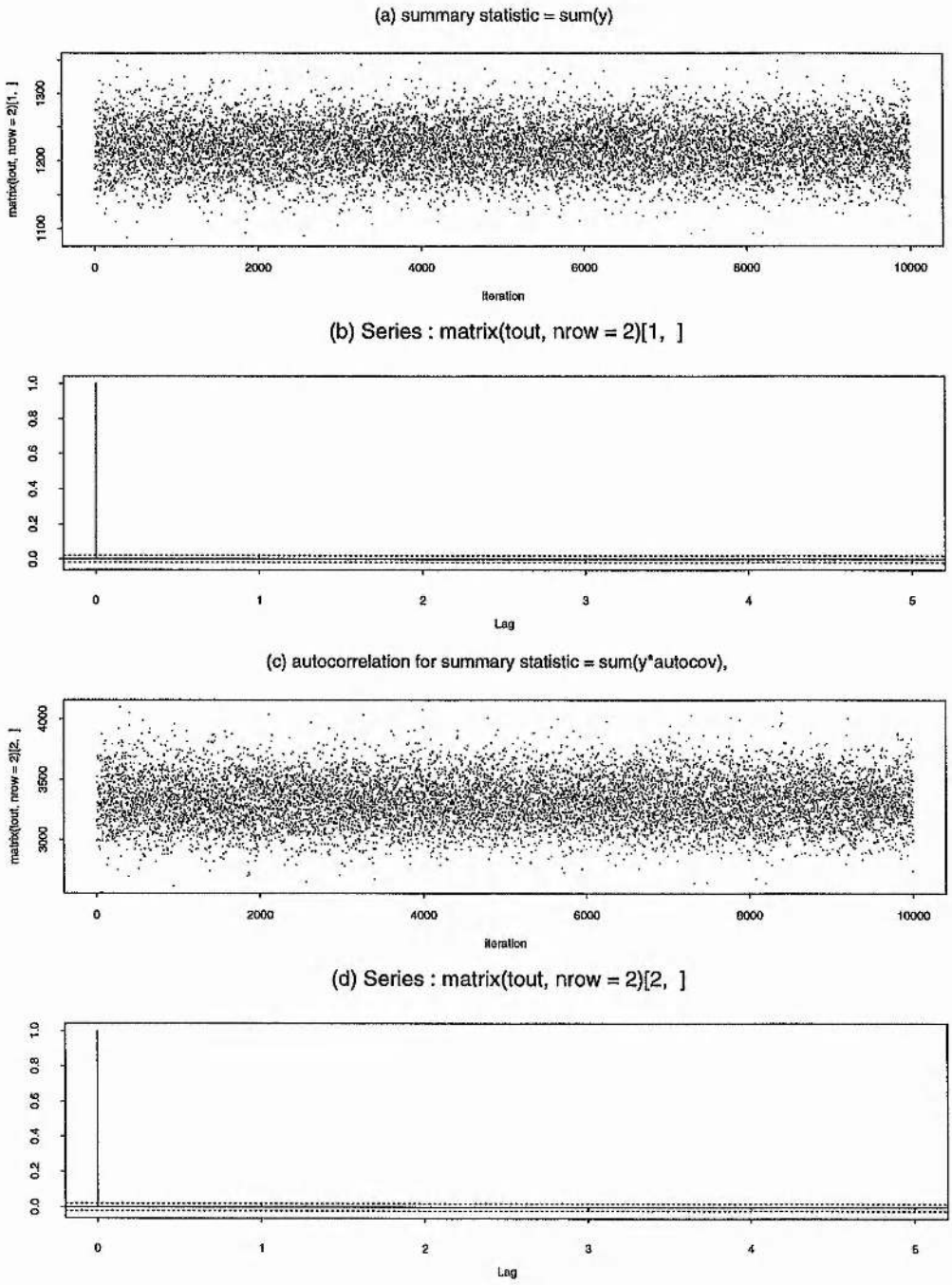


Figure 4.1: Convergence of auto-Poisson model with $\alpha = log(5)$ and $\gamma = 0.015$. Graphs (a) and (c) show the sequence of summary statistic t_1 and t_2 (y-axis) respectively from each iteration (x-axis) of the Gibbs sampler. Graph (b) and (d) show the corresponding autocorrelation function at different iteration lags 0 to 5. The dotted lines correspond to approximate 95% confidence interval bands about the line of zero correlation.

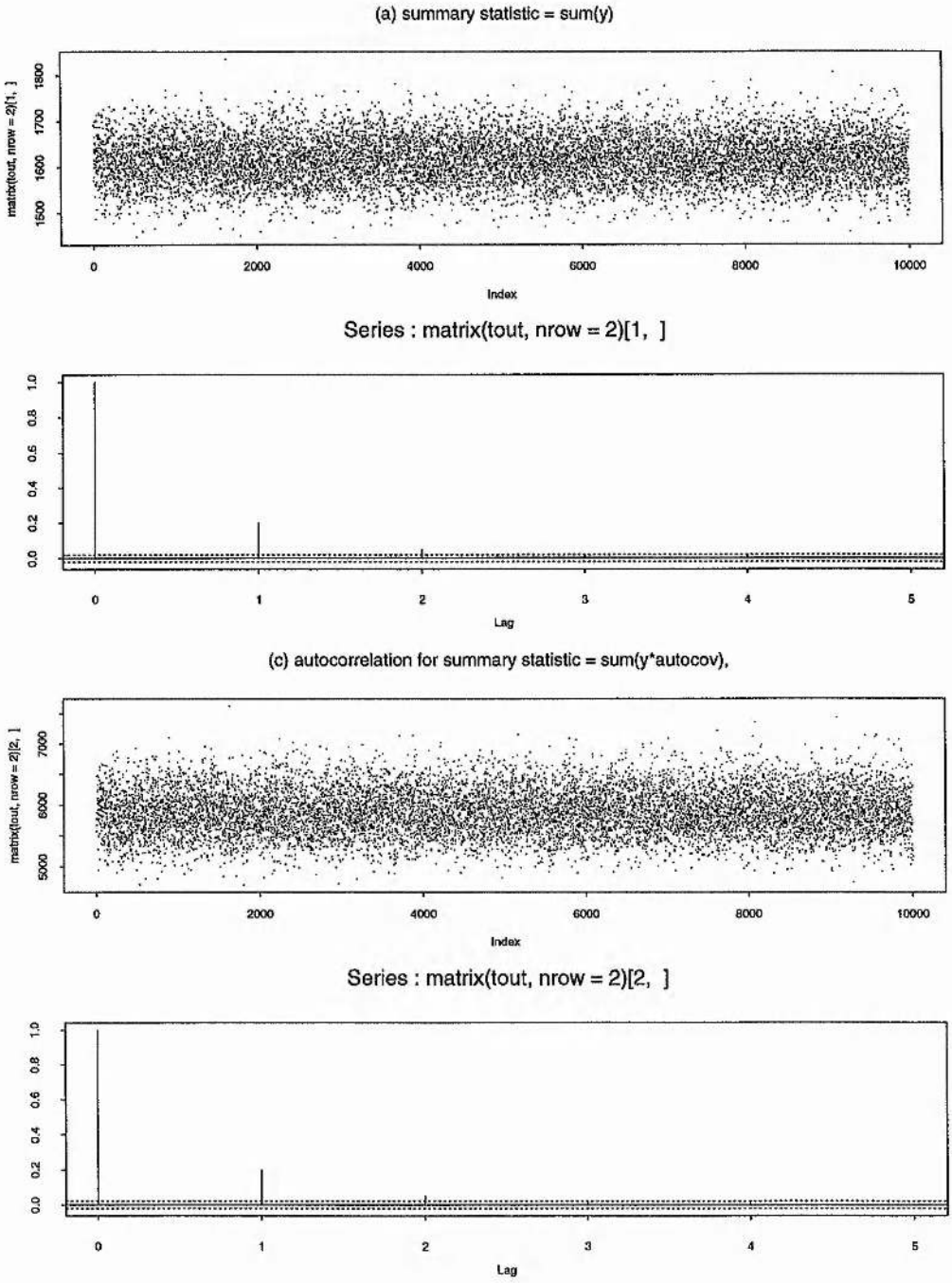


Figure 4.2: Convergence of auto-Poisson model with $\alpha = log(5)$ and $\gamma = 0.05$. Graphs (a) and (c) show the sequence of summary statistic t_1 and t_2 (y-axis) respectively from each iteration (x-axis) of the Gibbs sampler. Graph (b) and (d) show the corresponding autocorrelation function at different iteration lags 0 to 5. The dotted lines correspond to approximate 95% confidence interval bands about the line of zero correlation.

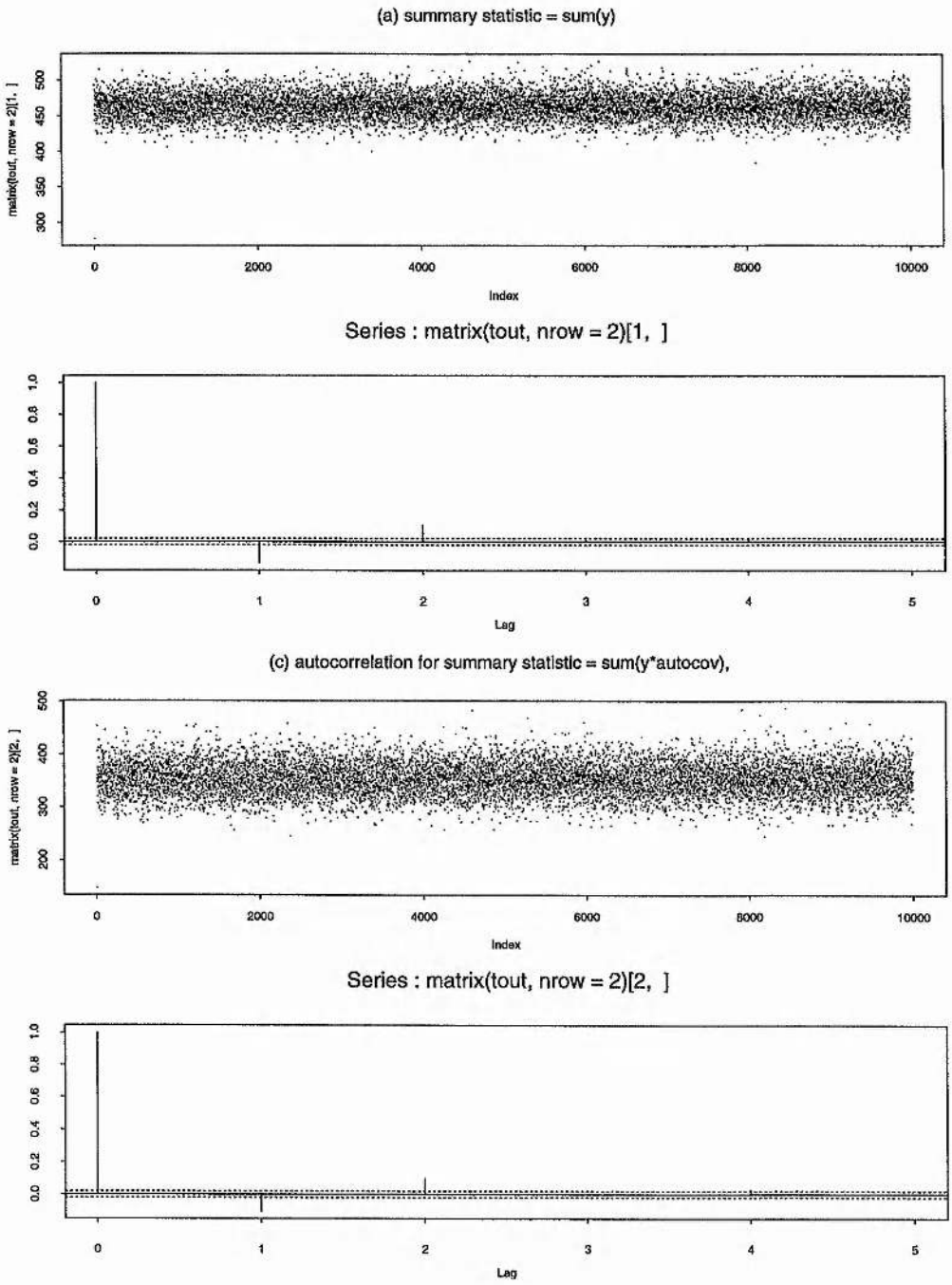


Figure 4.3: Convergence of auto-Poisson model with $\alpha = log(5)$ and $\gamma = -0.5$. Graphs (a) and (c) show the sequence of summary statistic t_1 and t_2 (y-axis) respectively from each iteration (x-axis) of the Gibbs sampler. Graph (b) and (d) show the corresponding autocorrelation function at different iteration lags 0 to 5. The dotted lines correspond to approximate 95% confidence interval bands about the line of zero correlation.

4.2.2 Utilising simulation for fitted values, residuals and goodness-of-fit statistics

When we fit an auto-model to data and use estimated parameters to simulate a MRF as described above, the ergodic average $\bar{y}$ of $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, ..., \boldsymbol{y}^{(m)}$ represents the approximate fitted values $\hat{\boldsymbol{\mu}}$ with components $\hat{\mu}_i$. This is justified by the following result, which can be found in any standard probability textbook:

$$\mu_i = E_i(y_i) = E_j(E_i(y_i|y_j; j \in N_i)) = \sum_{y_i \in \zeta} y_i Pr(y_i)$$

where $\zeta \equiv \{\boldsymbol{y} : Pr(\boldsymbol{y}) > 0\}$. The conditional expected value $E_i(y_i|y_j; j \in N_i)$ is estimated with the fitted value from the auto-model. For instance for the auto-Poisson model we use

$$\hat{E}(y_i|y_j, j \in N_i) = exp(\hat{\alpha}_i + \sum_{j \in N_i} \hat{\gamma}_{ij} y_j).$$

If the likelihood was available in terms of μ_i, deviance residuals and residual deviance could be used to assess goodness-of-fit. Instead we can plug the approximate fitted values $\bar{y}_i$ from ergodic averaging into the deviance function derived from likelihood functions assuming independence of observations. This likelihood corresponds to the pseudo likelihood function, which will be described in section 4.3.1, but fitted values are not conditional on neighbourhood values. For example the deviance of the Poisson log-likelihood function is

$$D(y; \hat{\boldsymbol{\mu}}) = 2 \sum_{i=1}^{n} y_i log\left(\frac{y_i}{\mu_i}\right) - (y_i - \mu_i).$$

We can approximate the deviance by

$$D(y; \bar{\boldsymbol{y}}) = 2 \sum_{i=1}^{n} y_i log\left(\frac{y_i}{\bar{y}_i}\right) - (y_i - \bar{y}_i),$$

and calculate approximate deviance residuals

$$r_{iD} = sign(y_i - \bar{y}_i)\sqrt{2 y_i log\left(\frac{y_i}{\bar{y}_i}\right) - (y_i - \bar{y}_i)}.$$

$D(y; \bar{\boldsymbol{y}})$ can be used as a goodness-of-fit statistic since it can be shown (McCullagh and Nelder, 1989) that $D(y; \hat{\boldsymbol{\mu}}) \sim \chi^2_{df}$, $df = n - p$. We assume that this is also the case for $D(y; \bar{\boldsymbol{y}})$. Also for comparing nested models the change of the approximated deviance $D(y; \bar{\boldsymbol{y}_0}) - D(y; \bar{\boldsymbol{y}_A})$ can be used as usual to test for the significance of a covariate. Here $\bar{\boldsymbol{y}_0}$ corresponds to fitted values of the null model and $\bar{\boldsymbol{y}_A}$ corresponds to fitted values with the additional covariate. For the truncated auto-Poisson model the ergodic average $\bar{y}_i$ would similarly be plugged into the deviance of the truncated Poisson log-likelihood function.

Alternatively we can approximate the Pearson residual

$$r_{iP} = \frac{y_i - \hat{E}(y_i)}{\sqrt{(V(\hat{E}(y_i)))}} = \frac{y_i - \hat{E}(y_i)}{\sqrt{(\hat{E}(y_i^2) - \hat{E}(y_i)^2)}}$$

which yields the Pearson χ^2 goodness-of-fit statistic (McCullagh and Nelder, 1989)

$$\sum_{i=1}^{n} r_{iP}^2 = X^2,$$

$X^2 \sim \chi^2_{df}$, $df = n - p$, where n is the number of observations and p the number of parameters. We replace $\hat{\mu}_i = \hat{E}(y_i)$ by its ergodic average $\bar{y}_i$.

In summary, by replacing fitted values $\hat{\mu}_i$ with their ergodic averages $\bar{y}_i$, we can obtain the ingredients needed for classical model checking and goodness-of-fit statistics. Further work is needed to investigate the distributional properties of the deviance and Pearson X^2-statistic.

4.3 Parameter estimation for auto-models

Parameter estimation for auto-models using full maximum likelihood (ML) is only possible if $\sum_{\boldsymbol{z}\in\zeta} exp(Q(\boldsymbol{z}))$ is available in closed form. Then the log-likelihood has the form:

$$l(\boldsymbol{\eta}) = Q(\boldsymbol{y}) - log\left(\sum_{\boldsymbol{z}\in\zeta} exp(Q(\boldsymbol{z}))\right). \tag{4.2}$$

where $\boldsymbol{\eta}$ is the parameter vector and $\boldsymbol{y} \in \zeta$, $\zeta \equiv \{\boldsymbol{y} : Pr(\boldsymbol{y}) > 0\}$. For the auto-logistic model the likelihood is available in closed form (see section 3.3), but for most data sets evaluating the normalising constant is computationally too intensive. For the truncated and Winsorized auto-Poisson model the normalising constant can also be evaluated: Kaiser and Cressie (1997) derived full maximum likelihood estimates for a Winsorized auto-Poisson model. As for the auto-logistic model the procedure is very computer intensive. Kaiser and Cressie (1997) use a simulation example comprising six locations to test the influence of different truncation points on parameter estimates. Their example shows that if r, the value where the distribution is Winsorized (equivalent to the truncation value; see section 3.4), is substantially larger than the largest observation, changes in r have little effect on estimates. Deriving the normalising constant for a data set with more than six observations and a higher value for r, is so expensive that full maximum likelihood estimation becomes infeasible.

Other possibilities for parameter estimation are:

- Pseudo maximum likelihood (Besag, 1974);
- coding (Besag, 1974);
- Monte Carlo maximum likelihood (Geyer and Thompson, 1992);
- Monte Carlo Newton-Raphson (Penttinen, 1984);
- stochastic approximation (Younes, 1988).

A comparison of these different methods (except coding and pseudo likelihood) is given in Geyer (1999).

4.3.1 Pseudo likelihood (PLE)

The pseudo likelihood function (Besag, 1975) is the product of the conditional probabilities, ignoring non-independence of observations. For example for the auto-Poisson model the pseudo log-likelihood has the form:

$$\begin{aligned} l_p(\eta) &= \sum_{i=1}^{n} log(Pr(y_i|y_j; j \in N_i)) \\ &= \sum_{i=1}^{n} log((\mu_i|y_j; j \in N_i) + y_i log(\mu_i|y_j; j \in N_i) - log(y_i!)) \end{aligned}$$

By formulating the auto-model as a GLM with the canonical link, maximum pseudo likelihood (MPL) estimation can be performed in any GLM package using iteratively re-weighted least squares. Because maximum pseudo likelihood assumes observations are independent, standard errors from iteratively re-weighted least squares are likely to be biased when strong autocorrelation is present. The parametric bootstrap can be applied to obtain standard errors, in which case the data are simulated using the Gibbs sampler as described in section 4.2.

4.3.2 Coding

Besag (1974) introduced coding, where the lattice is subdivided into two or more disjoint sub-lattices, so that the sites of the sub-lattices are mutually independent. Coding estimators are obtained by maximising the conditional likelihood of one sub-lattice given

the other. This yields two or more possible coding estimates that can be combined in some way. For example if correlation is assumed only among the four nearest neighbours, there are two sub-lattices D_0 and D_1, whose elements are mutually independent. Then the conditional log-likelihoods are:

$$l_{c,D_0}(\eta) = \sum_{i \in D_o} log(Pr(y_i|y_j : j \in N_i)),$$

$$l_{c,D_1}(\eta) = \sum_{i \in D_1} log(Pr(y_i|y_j : j \in N_i))$$

and estimates from the two conditional likelihoods are combined.

4.3.3 Monte Carlo maximum likelihood (MCML)

Monte Carlo maximum likelihood (MCML) was introduced by Geyer and Thompson (1992), but it is very similar to Monte Carlo Newton-Raphson introduced by Penttinen (1984). It can be applied to any model for dependent data derived from the exponential family such as MRFs. Other examples include spatial models for point processes, models for Markov graphs and conditional maximum likelihood. We illustrate the method with the right truncated auto-Poisson model from equation 4.1, with the random variable $Y_i \sim \text{truncPoi}(\mu_i|y_j; j \in N_i)$ at locations $i = 1, ..., n$ and truncation value r:

$$Pr(Y_i = y_i|y_j; j \in N_i) = \frac{\mu_i^{y_i}}{y_i!} \left(\sum_{k=0}^{r} \frac{\mu_i^k}{k!} \right)^{-1}$$

with $\mu_i = \mu_i(y_j; j \in N_i)$. For explaining the MCML method it is best to rewrite $Q(.)$ and $l(\eta)$ from equation 4.2 as

$$Q(\boldsymbol{y}) = \boldsymbol{t}(\boldsymbol{y})'\boldsymbol{\eta} \quad = \quad \sum_{i=1}^{n} (\alpha_i y_i - log(y_i!)) + \sum_{i=1}^{n-1} \sum_{j>i}^{n} \gamma_{ij} y_i y_j.$$

Considering the log-likelihood as a function of the parameters $l(\boldsymbol{\alpha}, \boldsymbol{\gamma})$ the term $\sum_{i=1}^{n} -log(y_i!)$ is a constant and can be omitted, so that (by redefinition of $\boldsymbol{t}(\boldsymbol{y})$).

$$\boldsymbol{t}(\boldsymbol{y})'\boldsymbol{\eta} \quad = \quad \sum_{i=1}^{n-1} \alpha_i y_i + \sum_{i=1}^{n-1} \sum_{j>i}^{n} \gamma_{ij} y_i y_j,$$

with vectors $\boldsymbol{t}(\boldsymbol{y}) = (y_1, ..., y_n, y_1 y_2, ..., y_{n-1} y_n)'$ and $\boldsymbol{\eta} = (\alpha_1, ..., \alpha_n, \gamma_{12}, ..., \gamma_{n-1,n})'$. Then define the normalising constant

$$c(\boldsymbol{\eta}) = \sum_{\boldsymbol{z} \in \zeta} exp(\boldsymbol{t}(\boldsymbol{z})'\boldsymbol{\eta}).$$

The log-likelihood takes the form

$$l_y(\eta) = log\left(\frac{exp(t(y)'\eta)}{c(\eta)}\right)$$

Now a constant term $log(c(\psi))$ is added to the likelihood. The term $c(\psi)$ has the same form as the normalising constant $c(\eta)$, vectors ψ and η are of the same length and the values of ψ are constant and close to η. The log-likelihood is then

$$\begin{aligned} l_y(\eta) &= log\left(\frac{exp(t(y)'\eta)}{c(\eta)}\right) + log(c(\psi)) \\ &= t(y)'\eta - log\left\{\frac{c(\eta)}{c(\psi)}\right\} \end{aligned}$$

It is now worth noting that the ratio $\frac{c(\eta)}{c(\psi)}$ is the moment-generating function $M_\psi(\eta - \psi)$ of $t(y)$

$$\frac{c(\eta)}{c(\psi)} = M_\psi(\eta - \psi) = E_\psi(exp(t(y)'(\eta - \psi))).$$

N.B. Define $\tau = \eta - \psi$, then

$$\begin{aligned} M_\psi(\tau) &= E_\psi(exp(t(y)'\tau)) \\ &= \sum_y exp(t(y)'\tau)P_\psi(y) \\ &= \sum_y exp(t(y)'\tau)\frac{exp(t(y)'\psi)}{\sum_{z\in\zeta} exp(t(z)'\psi))} \\ &= \sum_y \frac{exp(t(y)'(\tau+\psi))}{c(\psi)} \\ &= \frac{c(\tau+\psi)}{c(\psi)} = \frac{c(\eta)}{c(\psi)} \end{aligned}$$

Thus the ratio $\frac{c(\eta)}{c(\psi)}$ can be expressed as the expected value $E_\psi(exp(t(y)'(\eta - \psi))$, which can be approximated by its ergodic average

$$E_\psi(exp(t(y)'(\eta - \psi)) \approx \frac{\bar{c}(\eta)}{\bar{c}(\psi)} = \frac{1}{m}\sum_{i=1}^{m} exp(t(y^{(i)})'(\eta - \psi))$$

with $t(y^{(i)})$ simulated from an ergodic Markov chain $y^{(1)}, y^{(2)}, ..., y^{(m)}$ with equilibrium $P_\psi(y)$, the joint density of the MRF with the auto-Poisson model defined by parameters ψ. The approximate log-likelihood is then

$$l_{y,m}(\eta) = t(y)'\eta - log\left(\frac{1}{m}\sum_{i=1}^{m} exp(t(y^{(i)})'(\eta - \psi))\right)$$

Note that $t(y), \eta$ and ψ are all vectors of length p, the number of parameters, and the term $exp\left(t(y^{(i)})'(\eta - \psi)\right)$ is a scalar. Maximising this approximate likelihood yields the Monte

Carlo maximum likelihood estimator (MCMLE) $\hat{\eta}_{y,m}$. The approximate likelihood $l_{y,m}(\eta)$ is only a good approximation to $l_y(\eta)$ near ψ, so that iterative constrained maximisation is needed. The derivatives are

$$\begin{aligned}\frac{\delta l}{\delta \eta} &= t(y) - \frac{\frac{1}{m}\sum_{i=1}^{m} t(y^{(i)}) exp[\eta - \psi)' t(y^{(i)})]}{\frac{1}{m}\sum_{i=1}^{m} exp[(\eta - \psi)' t(y^{(i)})]} \\ &\approx t(y) - E_{\eta,\psi}(t(y))\end{aligned}$$

and the Monte Carlo approximation of the Hessian matrix $H(\eta)$ is

$$\begin{aligned}H(\eta) = \frac{\delta^2 l}{\delta \eta^2} &= \frac{\frac{1}{m}\sum_{i=1}^{m} t(y^{(i)}) exp[(\eta - \psi)' t(y^{(i)})] \; \frac{1}{m}\sum_{i=1}^{m} t(y^{(i)})' exp[t(y^{(i)})'(\eta - \psi)]}{\left\{\frac{1}{m}\sum_{i=1}^{m} exp[(\eta - \psi)' t(y^{(i)})]\right\}^2} \\ &\quad - \frac{\frac{1}{m}\sum_{i=1}^{m} t(y^{(i)}) t(y^{(i)})' exp[(\eta - \psi)' t(y^{(i)})]}{\frac{1}{m}\sum_{i=1}^{m} exp[(\eta - \psi)' t(y^{(i)})]} \\ &\approx E_{\eta,\psi}(t(y))[E_{\eta,\psi}(t(y))]' - E_{\eta,\psi}(t(y)t(y)') = -Var(t(y)).\end{aligned}$$

The (j,k)th element of $H(\eta)$ is

$$\begin{aligned}\frac{\delta^2 l}{\delta \eta_j \eta_k} &= \frac{\frac{1}{m}\sum_{i=1}^{m} t_j(y^{(i)}) exp[(\eta - \psi)' t_j(y^{(i)})] \; \frac{1}{m}\sum_{i=1}^{m} t_k(y^{(i)})' exp[t(y^{(i)})'(\eta - \psi)]}{\left\{\frac{1}{m}\sum_{i=1}^{m} exp[(\eta - \psi)' t(y^{(i)})]\right\}^2} \\ &\quad - \frac{\frac{1}{m}\sum_{i=1}^{m} t_j(y^{(i)}) t_k(y^{(i)}) exp[(\eta - \psi)' t(y^{(i)})]}{\frac{1}{m}\sum_{i=1}^{m} exp[(\eta - \psi)' t(y^{(i)})]} \\ &\approx E_{\eta,\psi}(t_j(y)) E_{\eta,\psi}(t_k(y)) - E_{\eta,\psi}(t_j(y) t_k(y))\end{aligned}$$

The estimated standard error of the MLE can be derived as the square root of the diagonal elements of the inverse observed Fisher information, which is the negative Hessian matrix $-H(\eta)$. The standard errors are valid under the usual regularity conditions for maximum likelihood estimation. They refer to the difference between the exact MLE and the true parameter value $(\hat{\eta} - \eta)$. The standard error of the Monte Carlo calculation, referring to the difference between the MCMLE and MLE, $(\hat{\eta}_{y,n} - \hat{\eta})$, can be estimated from the following variance matrix (for details see Geyer (1994), section 3):

$$m^{-1}(-H(\eta))^{-1} A(\eta) (-H(\eta))^{-1}$$

where $A(\eta)$ is estimated using the empirical variances and covariances of the samples from the Markov Chain $y^{(1)}, y^{(2)}, ..., y^{(n)}$.

We *do not know* the variance relating to the difference between the MCMLE and the true value

$$Var(\hat{\eta}_{y,m} - \eta) = Var(\hat{\eta}_{y,m} - \hat{\eta}) + Var(\hat{\eta} - \eta) + 2Cov((\hat{\eta}_{y,m} - \hat{\eta}), (\hat{\eta} - \eta)) \qquad (4.3)$$

because $Cov((\hat{\eta}_{y,m} - \hat{\eta}), (\hat{\eta} - \eta))$ is unknown. In cases where the asymptotics do not hold, i.e. the standardised parameters do not follow a normal distribution, we cannot use the observed Fisher information for variance estimation. Instead the bootstrap can be applied (Geyer, 1995). Using the bootstrap has the benefit of estimating the variance referring to the difference between the MCMLE and the true value, $Var(\hat{\eta}_{y,n} - \eta)$. Bootstrap samples can be taken from one run of the Markov chain with estimated parameters $\hat{\eta}_r$ and their likelihood can be estimated using the same Markov chain sample in the normalising constant. Thus bootstrapping can be done from the same sample $\boldsymbol{y}^{(1)}, ..., \boldsymbol{y}^{(m)}$ used for the approximate likelihood at no extra computational expense.

Figure 4.4 gives a schematic overview of how MCML is performed in practice. For the initial approximation of the likelihood, ψ_r is set to parameter estimates from another method such as MPL, coding or ML of a model assuming independence (i.e. without the parameter γ). Then auto-Poisson data with parameters ψ_r are simulated using the Gibbs sampler and using these simulated data in the normalising constant the approximated likelihood can be derived and maximised. The maximisation is constrained to be within an upper and lower limit around ψ_r. If the obtained estimates $\hat{\eta}_r$ are close to ψ the approximation should be good enough and the $\hat{\eta}_r$ are the MCML estimates. If they are not, we start again simulating data with $\psi_{r+1} = \hat{\eta}_r$ and so forth. We start from the MPL estimates in most examples. These are usually quite close to the MCML estimates, and convergence can be rapidly achieved.

4.3.4 Monte Carlo Newton-Raphson

The Monte Carlo Newton-Raphson method proposed by Penttinen (1984) is similar to Geyer's (Geyer and Thompson, 1992) MCML method. It has recently been applied to a Winsorised auto-Poisson model by Lee and Kaiser (1997). A good description of the algorithm is given in Heikkinen and Penttinen (1999) or Geyer (1999). Rather than approximating the likelihood as in MCML, here only the gradient and Hessian matrix are

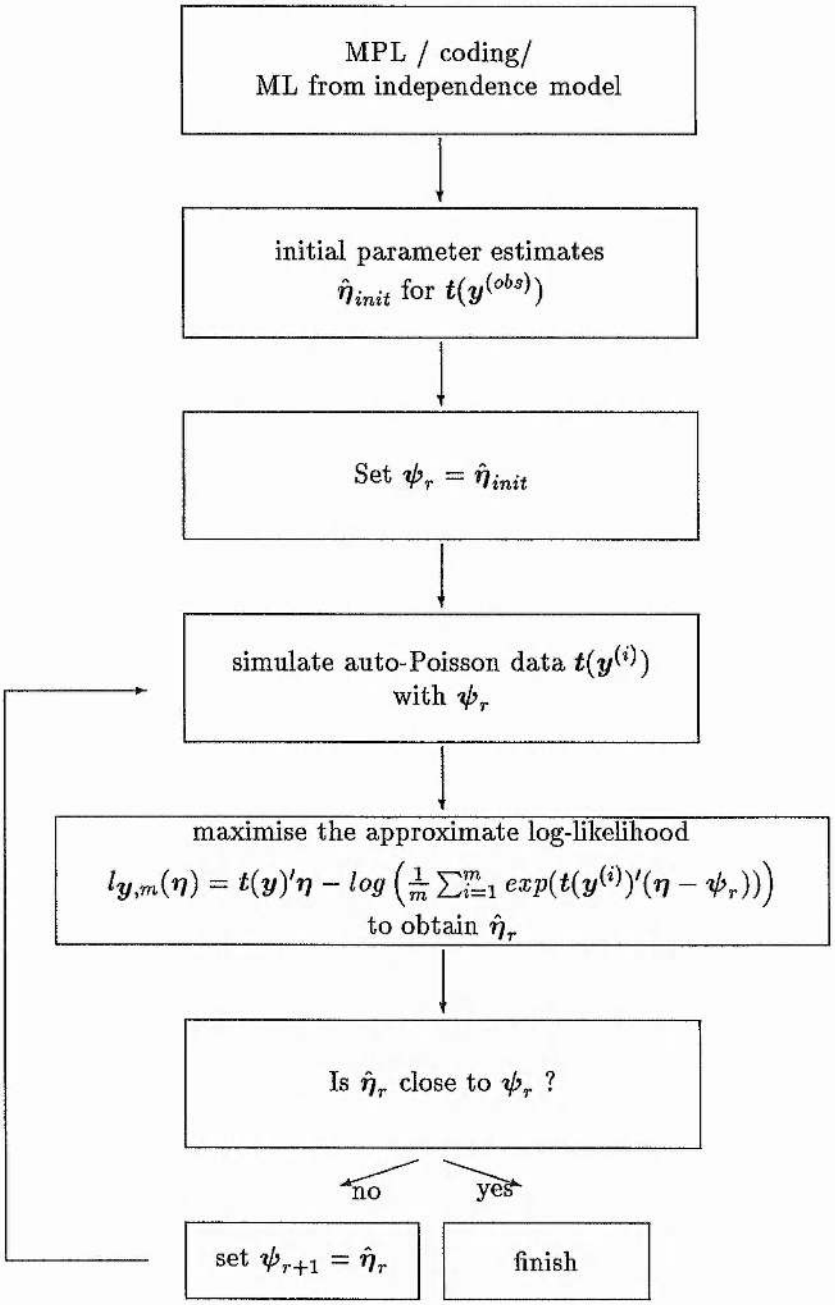


Figure 4.4: Flowchart for Monte Carlo maximum likelihood (MCML). The vector $t(y^{(obs)})$ refers to summary statistics from the observed data and $t(y^{(i)})$ to summary statistics for simulated data.

approximated. For the update,

$$\eta_{r+1} = \eta - (H(\eta_r))^{-1}\frac{\delta l}{\delta \eta_r},$$

where

$$\frac{\delta l}{\delta \eta} = t(y) - E_\eta(t(y))$$

and

$$H(\eta) = -Var_\eta(t(y)).$$

In each update of the Newton-Raphson algorithm a sample $y^{(1)}, y^{(2)}, ..., y^{(n)}$ from a Markov Chain with equilibrium $P_{\eta_r}(y)$ is drawn. The gradient $\frac{\delta l}{\delta \eta}$ is replaced by $t(y) - t(\bar{y})$, where $t(\bar{y})$ is the ergodic average of $t(y)$ and $H(\eta)$ is replaced by the negative sample covariance matrix.

4.3.5 Stochastic approximation

Younes (1988) first introduced a stochastic gradient algorithm based on the Gibbs sampler to obtain maximum likelihood estimates. Geyer (1999) proposes using this method as a starting point for MCML. The idea is to run a Markov chain with non-stationary transition probabilities, adjusting the parameter η in each iteration to move towards the MLE. A simplified version of the algorithm is as follows:

From the sequence of simulated data $t(y^{(i)})$ from a Markov Chain using an auto-Poisson model with parameter η_i, a crude estimate of the score is:

$$t(y^{(obs)}) - t(y^{(i)})$$

Then the parameter η_r is moved in the direction of the score, which should on average move it toward the MLE. η is updated by

$$\eta_{r+1} = \eta_r + \epsilon_r(t(y^{(obs)}) - t(y^{(r)})),$$

with

$$\epsilon_r = \frac{\epsilon_0}{1 + \rho r}, r = 1, 2, ...$$

where $\rho > 0$ and $\epsilon_0 > 0$ are constants. ϵ_0 and ρ are chosen arbitrarily, which makes this method rather ad hoc. In addition the method does not supply estimates for standard errors, which is why we do not pursue this method further here.

4.3.6 Simulated data examples for illustration

We use simulated data examples mainly to illustrate the MCML method applied to the truncated auto-Poisson model, but also to check whether this method converges and to illustrate one possible way in which data simulation can be used to assess goodness of fit. Six data sets were simulated on a 15 x 15 lattice using the Gibbs sampler with a truncated auto-Poisson model with six different sets of parameters. We simulated $Y_i \sim$ truncPoi$(\mu_i|y_j; j \in N_i)$, $i = 1, ..., n$ and $r = 100$ defined by the model:

$$\log(\mu_i|y_j; j \in N_i) = \alpha + \beta x_i + \gamma \frac{1}{n_i} \sum_{j \in N_i} y_j. \tag{4.4}$$

where α is the intercept, x_i is a covariate for row effect with $x_i = row_i - \frac{1}{n}\sum_{i=1}^{n} row_i$, with slope parameter β. The parameter γ is the interaction parameter defining spatial autocorrelation, N_i are the four nearest neighbours and $n_i = 4$, except for edge sites. Then summary statistics are

$$t(\mathbf{y}) = (t_1, t_2, t_3)' = (\sum y_i, \sum y_i x_i, \frac{1}{2} \sum y_i autocovariate_i)',$$

where $autocovariate_i = \frac{1}{n_i} \sum_{j \in N_i} y_j$ and the parameter vector is $\eta = (\alpha, \beta, \gamma)'$. For the approximate likelihood in the MCML method $m = 1000$ samples were used, by extracting every 2nd sample of the Markov chain, after discarding the first 100 iterations. The first 4 examples are all processes without trend, that is $\beta = 0$. Parameter estimates from the MPL and MCML (Table 4.1) agree fairly well. Standard errors are smaller from MPL for Examples 1 and 3 where γ, the parameter for auto-correlation, is positive. In Examples 2 and 4 the MPL method gives larger standard errors than the MCML. For the MCML method Monte Carlo standard errors are also given. These measure additional variation introduced by MCML as defined in Section 4.3.3 and are very small compared with the asymptotic standard errors of MCML. Thus, the MCML estimate does not differ much from the true maximum likelihood estimate. The standard errors for the coding method are highest and in Example 4 coding did not work very well. The two coding estimates for α are very far apart and the second one has a very high standard error. This is because the model used to simulate Example 4 had an autocorrelation parameter $\gamma = -0.5$ which created a very regular pattern of zeros and high numbers. The data for the second coding estimate included most all of the zeros and a few ones. Therefore α could not be estimated very well from the second coding data set of Example 4.

Table 4.1: Auto-Poisson parameter estimates for *Examples 1* to *4* with different parameter values using pseudo likelihood (MPL), coding (CODE1, CODE2) and Monte Carlo maximum likelihood (MCML). The parameter β is zero for these examples. Asymptotic standard errors of MCML are denoted as [se] and Monte Carlo standard errors of MCML are denoted as (mcse). The Pearson goodness-of-fit statistic X^2 is given; for MPL and MCML the corresponding *df* are 223, for CODE1 *df* = 111, for CODE2 *df* = 110.

Parameter	α	[se]	(mcse)	γ	[se]	(mcse)	X^2
Ex 1: true values	*1.609*			*0.015*			
MPL	1.520	[0.142]		0.026	[0.026]	()	200.323
CODE1	1.512	[0.205]	-	0.026	[0.038]	()	97.614
CODE2	1.526	[0.197]	-	0.026	[0.037]	-	88.959
MCML	1.530	[0.170]	(0.0057)	0.023	[0.032]	(0.00097)	200.878
Ex 2: true values	*2.303*			*-0.2*			
MPL	2.311	[0.102]	()	-0.209	[0.025]	()	174.410
CODE1	2.325	[0.167]	-	-0.214	[0.037]	-	88.959
CODE2	2.283	[0.140]	-	-0.199	[0.037]	-	87.288
MCML	2.260	[0.086]	(0.0021)	-0.195	[0.021]	(0.00047)	186.881
Ex 3: true values	*2.303*			*0.0001*			
MPL	2.407	[0.136]	-	-0.008	[0.013]	-	208.656
CODE1	2.384	[0.199]	-	-0.003	[0.020]		98.289
CODE2	2.319	[0.197]	-	-0.003	[0.018]		107.142
MCML	2.440	[0.170]	(0.0042)	-0.012	[0.017]	(0.00042)	208.702
Ex 4: true values	*2.303*			*-0.5*			
MPL	2.270	[0.031]	-	-0.545	[0.038]	-	192.335
CODE1	2.270	[0.036]	-	-0.538	[0.228]	-	111.015
CODE2	1.428	[1.700]	-	-0.443	[0.206]	-	91.015
MCML	2.270	[0.030]	(0.00110)	-0.546	[0.038]	(0.00084)	194.404

The Pearson's χ^2 goodness-of-fit statistic, as defined in section 4.2.2, does indicate good fits for all models. We have not checked whether the χ^2 distribution is appropriate for X^2, this might be a problem for example 4 which contains many zeros (102 out of 225 of the observations are zero). Figures 4.5 to 4.8 (top panel) show the sequence of $\hat{\eta}_r$ for the MCML procedure on the same plot with the estimate from MPL and the two Coding estimates for Examples 1 to 4. Assuming a first order neighbourhood N_i means two coding estimates (CODE1, CODE2) are obtained, corresponding to two sublattices. In all cases the MCML estimate appears to converge.

The summary statistics $\boldsymbol{t}(\mathbf{y}^{(i)}) = (t_{1i}, t_{2i})$ in Figures 4.5 to 4.8 correspond to Examples 1 to 4 and were calculated from realisations of a simulated ergodic Markov chain with equilibrium $P_{\boldsymbol{\psi}}(\mathbf{y})$, using $\boldsymbol{\psi}$ equal to the estimates of MCML, MPL and coding respectively. These $\boldsymbol{t}(\mathbf{y}^{(i)})$ can be used as a goodness of fit diagnostic, by plotting t_{1i} versus t_{2i} and checking whether the observed t_1 and t_2 are in the centre of the scatterplot (Geyer and Thompson, 1992). For all examples the observed t_1 and t_2 are well centred in the convex hull of the $\boldsymbol{t}(\mathbf{y}^{(i)})$ using MCML, MPL and Coding. In Example 2 the observed summary statistics appear more centred for the MCML method than for the other two methods. In the case of Example 4 the simulated $\boldsymbol{t}(\mathbf{y}^{(i)})$ from coding estimates do not coincide at all with the observed summary statistic; this is because the second coding estimate is very far from the true value. In Example 5 (Table 4.2) $\boldsymbol{Y}_i \sim \text{Poi}(\mu)$, with $\mu = 5$, and no trend or autocorrelation is present ($\beta = 0$ and $\gamma = 0$). Estimates and standard errors from both methods are almost identical. Example 6 represents a process with trend but without auto-correlation ($\gamma = 0$). In this case estimates are similar but standard errors from MPL are smaller than from MCML, because γ was estimated. Likelihood functions were plotted for the MPL and the MCML method. The contours of the pseudo likelihood and Monte Carlo likelihood are shown in Figures 4.9 to 4.11 for Examples 1 to 5. Note that likelihood windows of the same width are shown, but depending on the actual estimates the windows are shifted. The shapes of the likelihoods are fairly similar when comparing MPL and MCML, except for Examples 2 and 4 with $\gamma < 0$ (Figures 4.9 and 4.10) where the shapes differ. In particular the likelihood from MCML for Example 4 has a plateau in the negative direction of γ. This is not the case for the MPL surface. In Section 3.4.1 (Figure 3.5a) we showed that any value for $\gamma < -1$ produces the same chess board effect, hence the plateau in the likelihood surface in the negative γ direction.

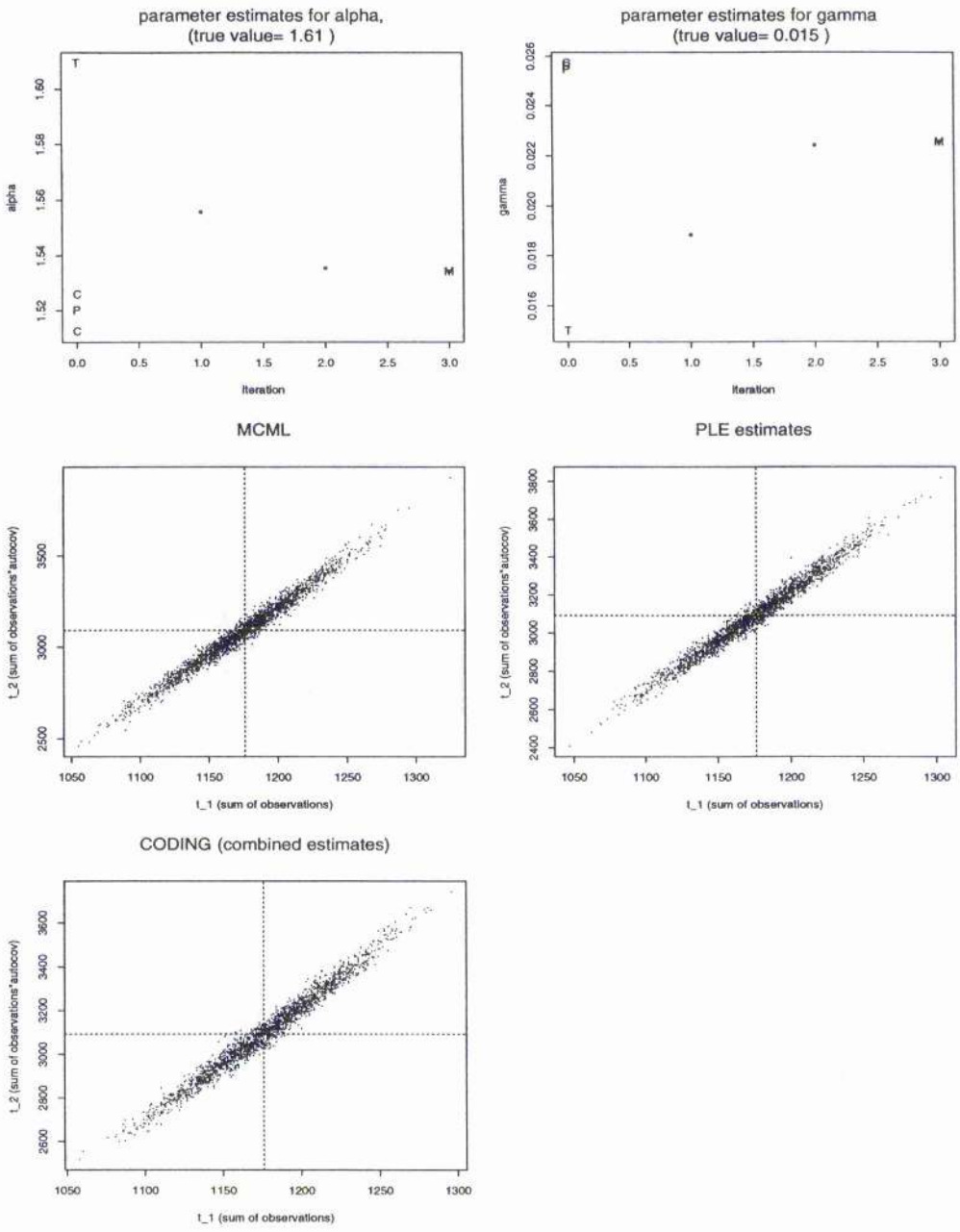


Figure 4.5: Example 1. From top left to bottom right: Parameter estimates for α and β for the iterative process for MCML (*, final estimate marked as M), parameter estimates for MPL (marked as P), Coding (CODE1 and CODE2 marked both as C) and true values (marked as T). Plots of simulated summary statistics t_{1i} versus t_{2i} using parameters from MCML, MPL and Coding. Where the two dotted lines cross corresponds to observed t_1 and t_2.

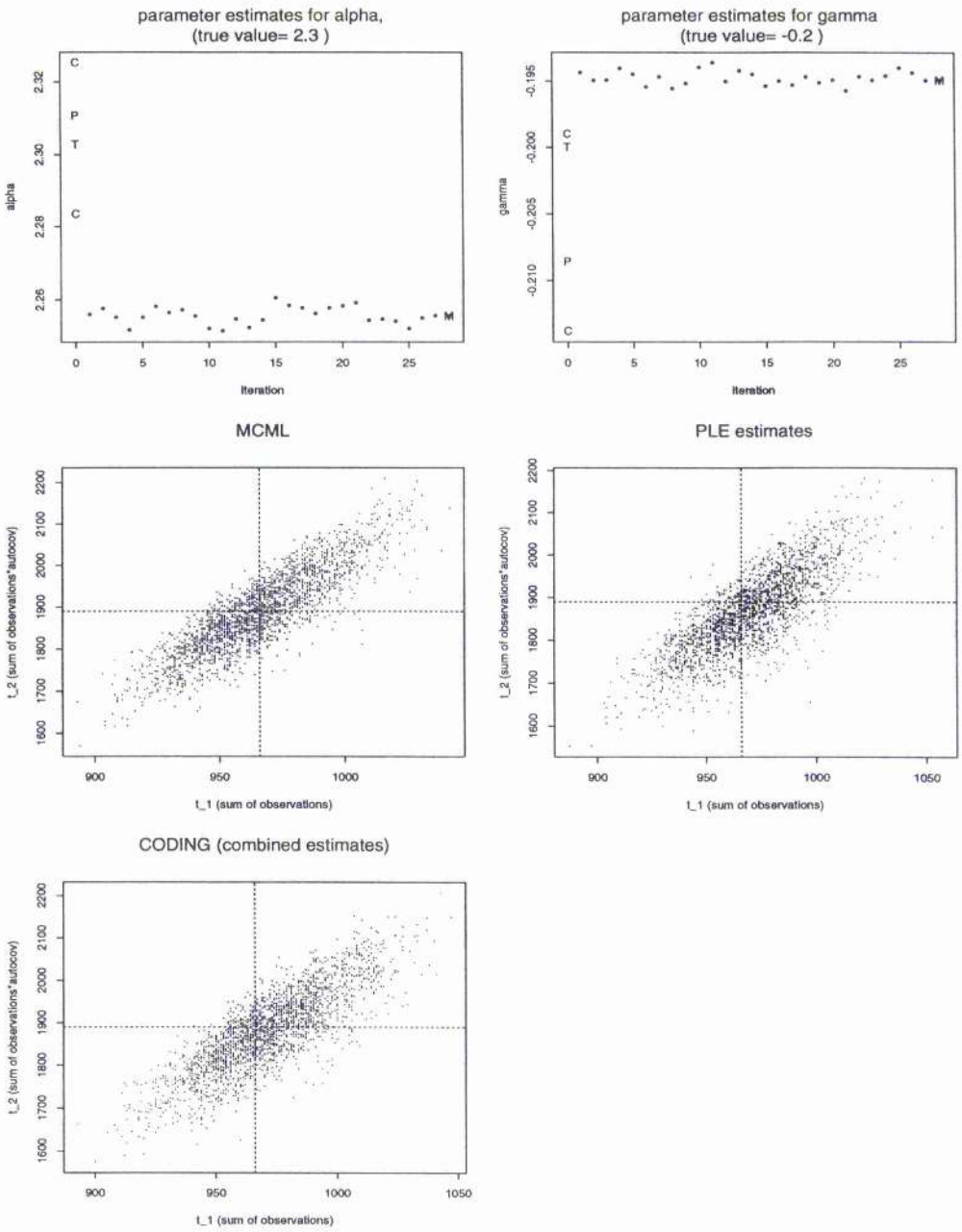


Figure 4.6: Example 2. From top left to bottom right: Parameter estimates for α and β for the iterative process for MCML (*, final estimate marked as M), parameter estimates for MPL (P), Coding (CODE1 and CODE2 marked as C) and true values (T). Plots of simulated summary statistics t_{1i} versus t_{2i} using parameters from MCML, MPL and Coding. Where the two dotted lines cross corresponds to observed t_1 and t_2.

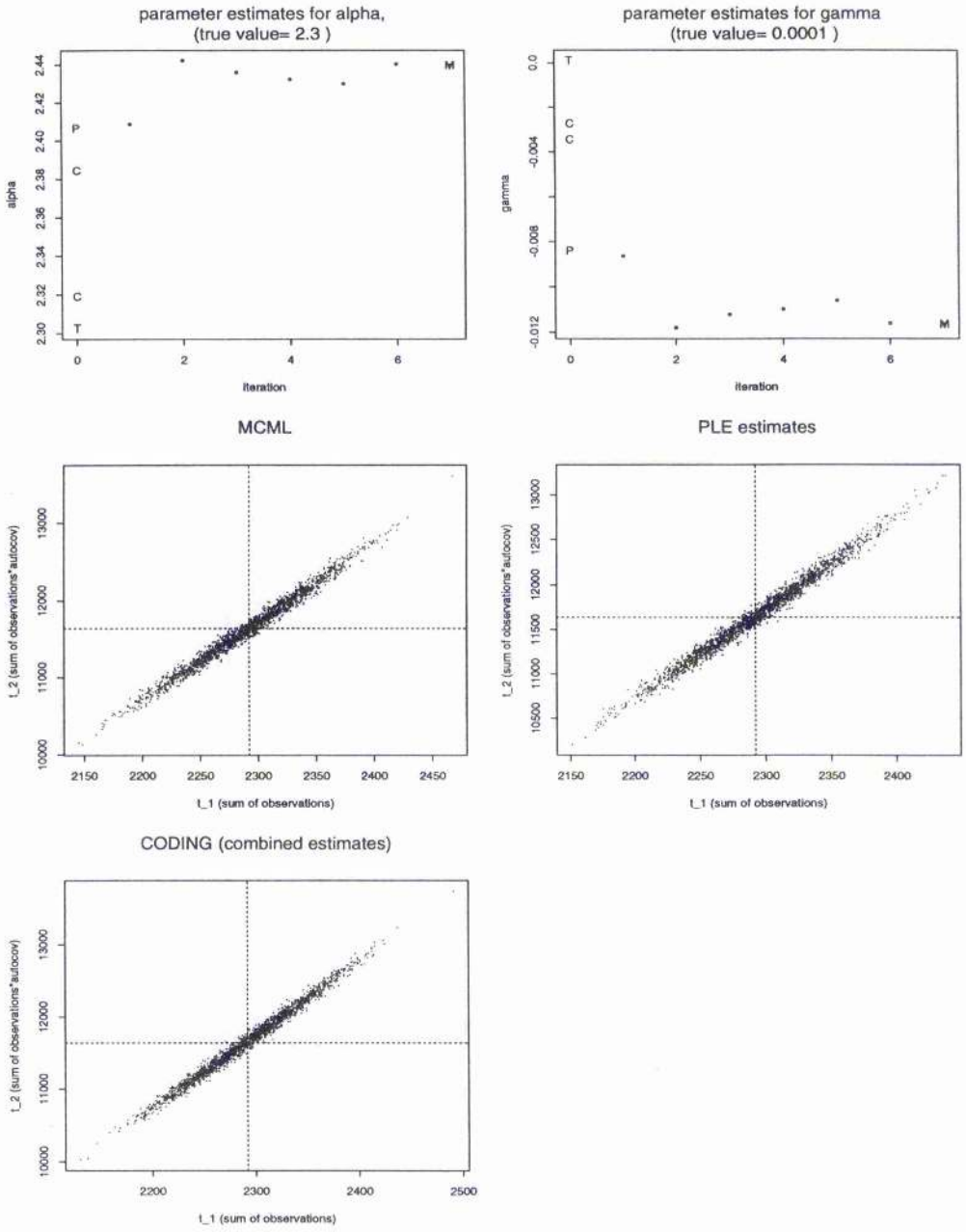


Figure 4.7: Example 3. From top left to bottom right: Parameter estimates for α and β for the iterative process for MCML (marked as *, final estimate as M), parameter estimates for MPL (P), Coding (CODE1 and CODE2 marked as C) and true values (T). Plots of simulated summary statistics t_{1i} versus t_{2i} using parameters from MCML, MPL and Coding. Where the two dotted lines cross corresponds to observed t_1 and t_2.

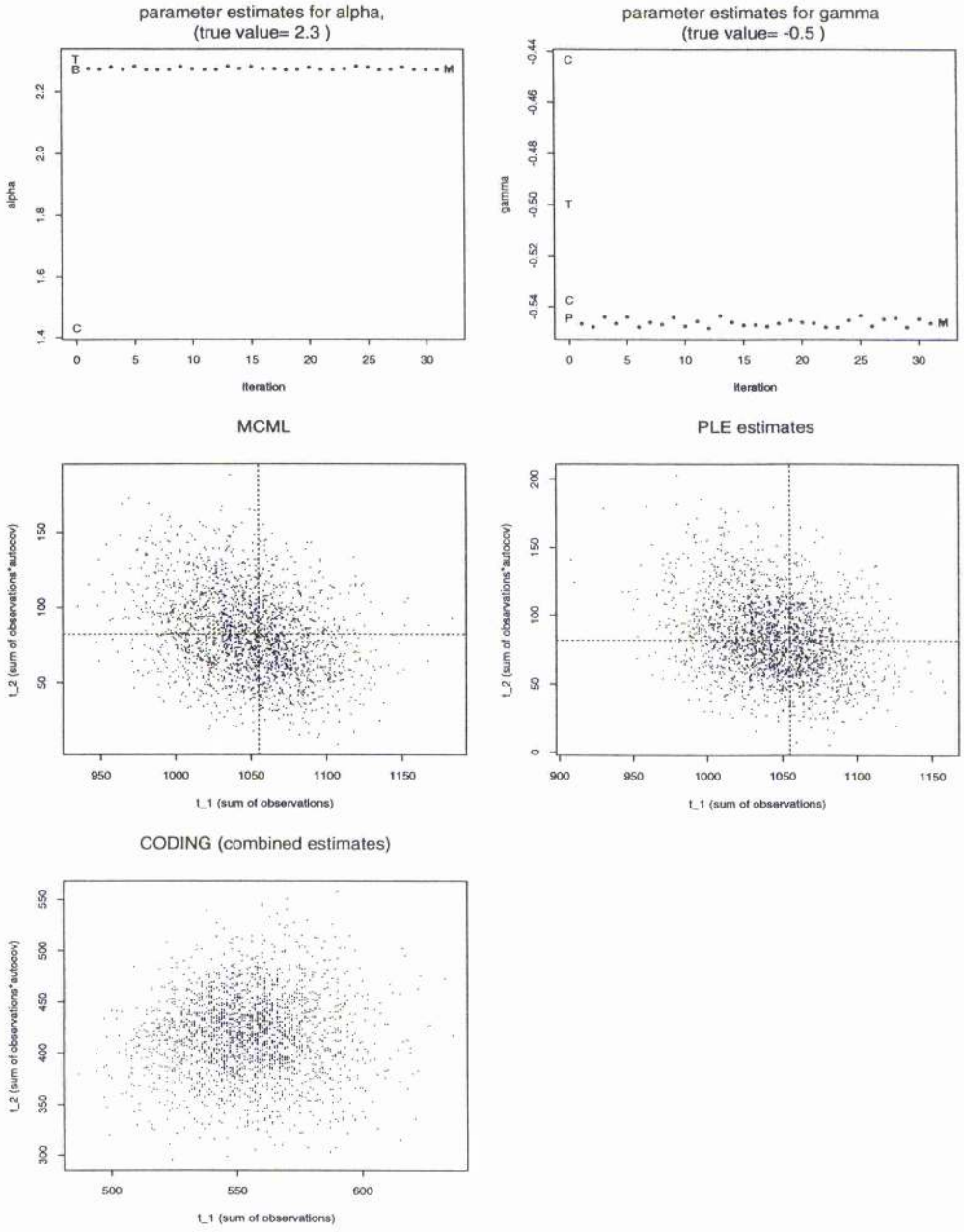
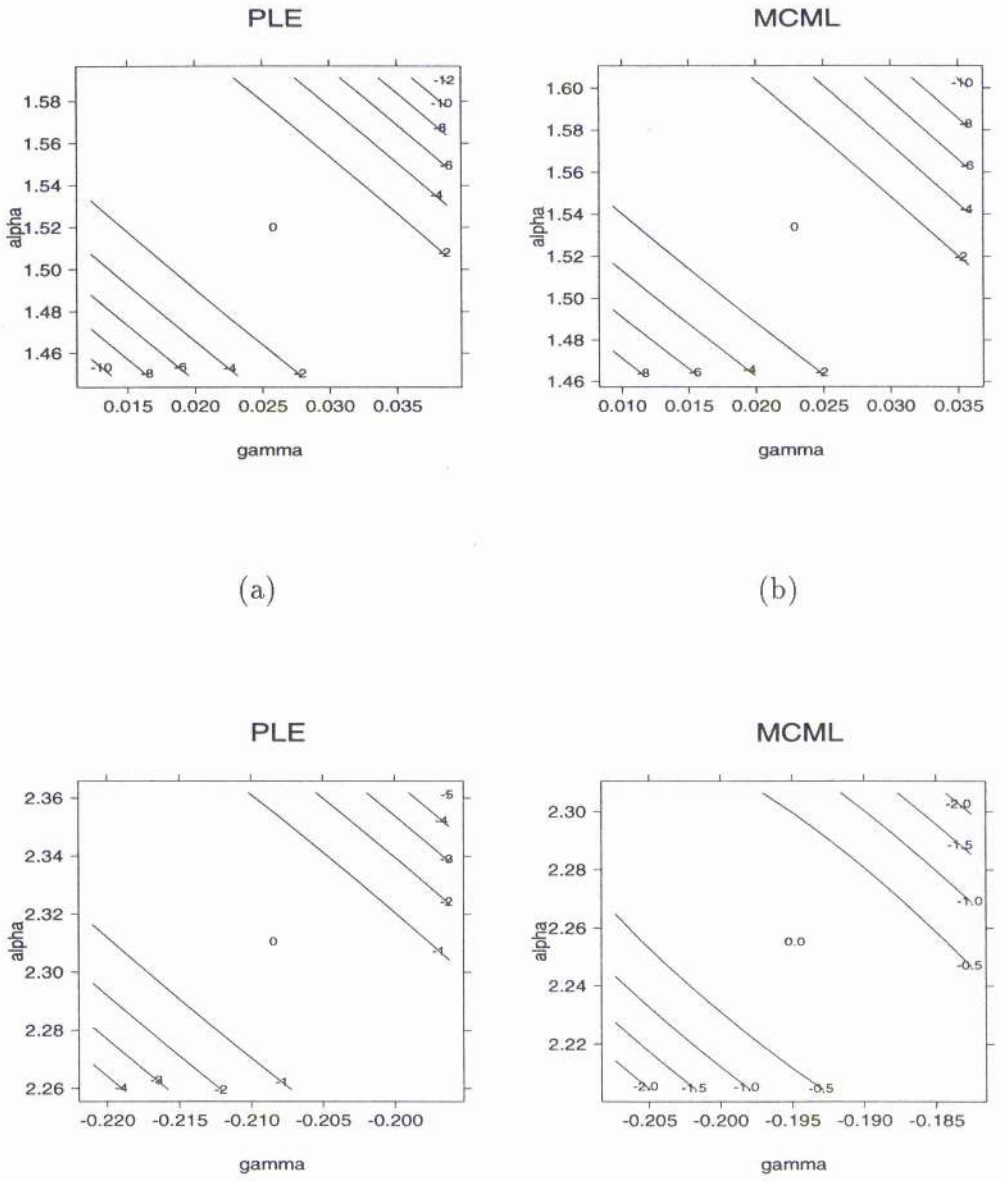


Figure 4.8: Example 4. From top left to bottom right: Parameter estimates for α and β for the iterative process for MCML (marked as *, final estimate as M), parameter estimates for MPL (P), Coding (C) and true values (T). Plots of simulated summary statistics t_{1i} versus t_{2i} using parameters from MCML, MPL and coding. Where the two dotted lines cross corresponds to observed t_1 and t_2. Note that for coding the scatterplot does not contain the observed t_1 and t_2.



(a) (b)

(c) (d)

Figure 4.9: Contours of MPL (left) and MCML (right), Example 1 ((a) and (b)), Example 2 ((c) and (d)). The y-axis represents the intercept parameter α and the x-axis represents the auto-correlation parameter γ. Note that axes of plots for each example are of the same width, but do not show the same window of the log-likelihood.

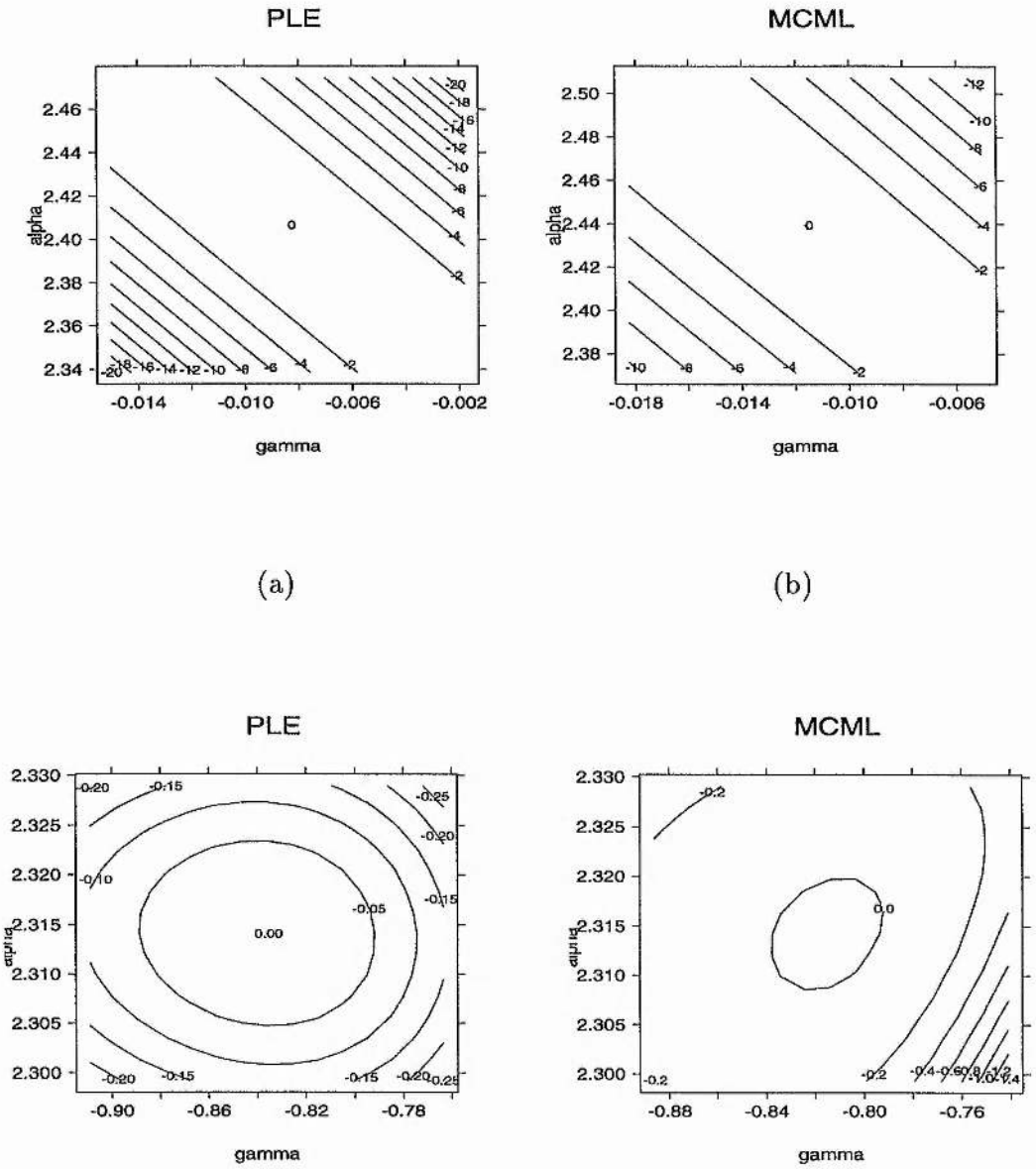


Figure 4.10: Contours for (a) MPL and (b) MCML for Example 3 and (c) MPL and (d) MCML for Example 4. The y-axis represents the intercept parameter α and the x-axis represents the auto-correlation parameter γ. Note that axes of plots for each example are of the same width, but do not show the same window of the log-likelihood.

Table 4.2: Auto-Poisson parameter estimates for different parameter values using pseudo likelihood (MPL), coding (CODE1, CODE2) and Monte Carlo maximum likelihood (MCML). Asymptotic standard errors are denoted as [se]. The Pearson goodness-of-fit statistic X^2 is given, for MPL and MCML the corresponding df are 222, for Coding 1 $df = 110$, for coding 2 $df = 109$.

Parameter	α	[s.e.]	β	[s.e.]	γ	[s.e.]	X^2
Ex 5: true values	*1.609*		*0*		*0*		
MPL	1.544	[0.031]					
MCML	1.540	[0.031]					
Ex 6: true values	*1.609*		*0.05*		*0*		
MPL	1.320	[0.151]	0.034	[0.009]	0.045	[0.031]	178.462
CODE1	1.357	[0.197]	0.034	[0.013]	0.038	[0.040]	71.312
CODE2	1.268	[0.235]	0.033	[0.014]	0.055	[0.049]	108.642
MCML	1.420	[0.180]	0.038	[0.011]	0.025	[0.038]	180.484

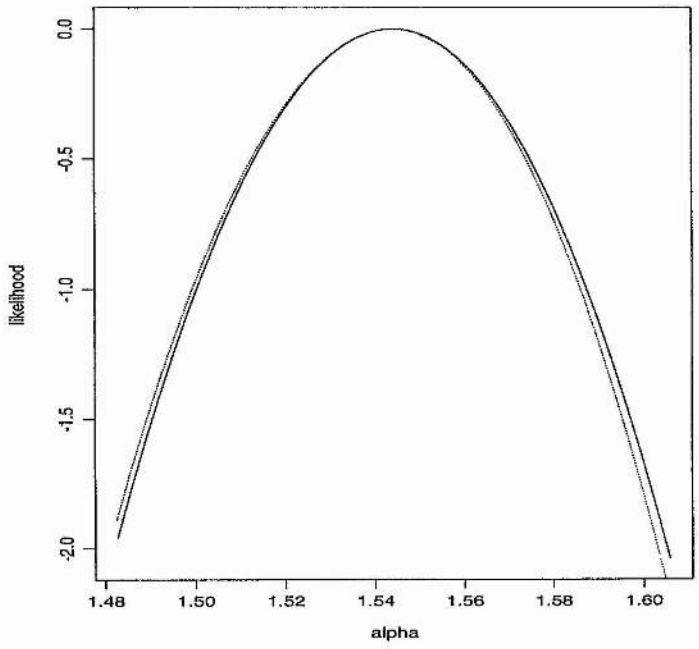


Figure 4.11: Example 5, likelihood for MPL (continous line) and MCML (dotted line).

4.4 Comparison of different parameter estimation methods: A simulation study

4.4.1 Comparison of the three parameter estimation methods

The three parameter estimation methods, pseudo likelihood (MPL), coding (CODE1, CODE2) and Monte Carlo maximum likelihood (MCML) were compared in a simulation study. The same model as in equation 4.4 is used for data simulation. Respective parameter values for *studies 1* to *3* are given in Table 4.3, and those for *study 4* in 4.4. In *studies 1* to *3* the parameter β was zero, so that no trend was present. A truncation value of $r = 100$ for the MCML method was assumed, except for the second simulation study with $r = 20$. For the MCML method the initial value for η_{init} was taken from the model fitted assuming independence for those studies where the model was without auto-correlation (study 1 and 4). For studies 2 and 3, where the model included auto-correlation, the starting value was taken from the MPL estimates. This choice of η_{init} was based on the speed of convergence.

The simulation study consolidates the findings from the illustrative examples in section 4.3.6. When comparing the mean asymptotic standard errors with empirical standard errors in Table 4.3 and 4.4, it appears that MPL underestimates standard errors in the simulation studies which incorporated zero or positive auto-correlation. In studies 1, 2 and 4 the mean asymptotic standard errors are substantially lower than empirical standard errors for the MPL method. This is also the case for the MCML method, probably caused by the extra component of Monte Carlo variation, measured by the Monte Carlo standard error (mcse), but differences there are smaller. Unfortunately the mcse cannot be used to inflate the variance correctly, because the covariance $Cov((\hat{\eta}_{y,m} - \hat{\eta}), (\hat{\eta} - \eta))$ from equation 4.3 is not known. In simulation *study 3*, where negative auto-correlation is incorporated, asymptotic standard errors from MPL are higher than their empirical counterparts.

The bias in parameter estimates in all methods appears to be negligible, given the associated standard errors. It is surprising to see that the coding method gives the lowest bias only in *study 2* for which $\gamma = 0.05$. This is probably because the coding estimates are derived from only half of the sample size. (See also Example 4 of the previous section.)

On average MPL gives the lowest bias in this limited investigation. In study 3, where $\gamma = -0.5$, MCML gives the largest bias, underestimating α and overestimating γ. For investigating asymptotic normality parameter estimates were centred and standardised in the following way:

$$\frac{\hat{\eta}_i - \frac{1}{100}\sum \hat{\eta}_i}{\sqrt{\widehat{var}(\hat{\eta}_i)}}$$

Checking histograms and qq-normal plots of these estimates in Figures A.1 to A.8, show that their distribution appears to be normal. Estimates from MPL plotted versus estimates MCML are shown in Figure 4.12 and 4.13. These appear to be fairly close, except for *study 3* where strong negative auto-correlation was present and MCML had the largest bias.

4.4.2 Investigation of correlated parameter estimates

So far the parameters of the models have been investigated only in a univariate sense. Consider the model used in simulation *studies 1* to *3*:

$$\log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \frac{1}{n_i} \sum_{j \in N_i} y_j.$$

When the parameters are considered in a multivariate sense, for auto-correlation greater than or equal to zero, the parameter estimates of α and γ from *studies 1* to *3* are strongly correlated. This is in contrast to the auto-normal model, which has the property that parameters α and γ are orthogonal. The component of the asymptotic covariance matrix from the auto-normal model referring to the covariance between α and γ is zero (Haining, 1990).

Plotting pseudo likelihood estimates from simulation *study 1* to *3* of α versus γ (Figures 4.14(a) to 4.16(a)) shows strong negative correlation. For the modified model as discussed in section 3.4.1

$$log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i} (y_j - exp(\alpha))$$

fitted to the data from studies 1 and 2, the pseudo likelihood estimates also show a strong relationship (Figure 4.14(b) to 4.16(b)). Here the correlation is negative for $\hat{\gamma} > 0$ and positive for $\hat{\gamma} < 0$.

If we "centre" the *autocovariate* as in section 3.4.1,

$$log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i} (y_j - \bar{y})$$

Table 4.3: Means of parameter estimates, using pseudo likelihood (MPL), coding (CODE1, CODE2) and Monte Carlo maximum likelihood (MCML)), from 100 simulated data sets for each of simulation *studies 1-3*. The mean of asymptotic standard errors are given in brackets, empirical standard errors are given in square brackets.

Parameter	α	(s. e.)	[s.e.]	bias	γ	(s. e.)	[s.e.]	bias
Study 1								
($r = 100$)								
True values	*1.609*				*0*			
MPL	1.594	(0.131)	[0.187]	-0.016	0.003	(0.026)	[0.036]	0.003
CODE1	1.578	(0.189)	[0.201]	-0.031	0.006	(0.037)	[0.038]	0.006
CODE2	1.568	(0.188)	[0.189]	-0.042	0.008	(0.037)	[0.037]	0.008
MCML	1.592	(0.175)	[0.187]	-0.017	0.003	(0.034)	[0.036]	0.003
Study 2								
($r = 20$)								
True values	*1.609*				*0.05*			
MPL	1.640	(0.128)	[0.173]	0.030	0.045	(0.017)	[0.024]	-0.005
CODE1	1.618	(0.183)	[0.178]	0.009	0.048	(0.025)	[0.024]	-0.002
CODE2	1.611	(0.185)	[0.194]	0.002	0.049	(0.025)	[0.027]	-0.001
MCML	1.644	(0.157)	[0.172]	0.035	0.045	(0.021)	[0.024]	-0.005
Study 3								
($r = 100$)								
True values	*1.609*				*-0.5*			
MPL	1.600	(0.091)	[0.082]	-0.010	-0.499	(0.051)	[0.050]	0.001
CODE1	1.590	(0.135)	[0.136]	-0.019	-0.497	(0.074)	[0.077]	0.003
CODE2	1.606	(0.131)	[0.128]	-0.004	-0.503	(0.074)	[0.075]	-0.003
MCML	1.585	(0.074)	[0.076]	-0.024	-0.491	(0.042)	[0.042]	0.009

Table 4.4: Means of parameter estimates from *study 4*, using pseudo likelihood (MPL), coding (CODE1, CODE2) and Monte Carlo maximum likelihood (MCML)), from 100 simulated data sets. The mean of asymptotic standard errors are given in brackets, empirical standard errors are given in square brackets.

Parameter	α	(s. e.)	[s.e.]	bias	β	(s. e.)	[s.e.]	bias	γ	(s. e.)	[s.e.]	bias
Study 4 ($r = 100$)												
True values	*1.609*				*0.05*				*0*			
MPL	1.603	(0.130)	[0.178]	-0.006	0.050	(0.009)	[0.011]	0.000	0.002	(0.025)	[0.033]	0.002
CODE1	1.567	(0.188)	[0.190]	-0.042	0.048	(0.013)	[0.013]	-0.002	0.008	(0.036)	[0.034]	0.008
CODE2	1.567	(0.190)	[0.185]	-0.043	0.048	(0.014)	[0.014]	-0.002	0.009	(0.036)	[0.036]	0.009
MCML	1.601	(0.167)	[0.173]	-0.008	0.050	(0.011)	[0.011]	0.000	0.002	(0.032)	[0.032]	0.002

where $\bar{y}$ is the average of observations y, parameter estimates are no longer correlated for the cases with $\gamma \geq 0$, as can be seen from pseudo likelihood estimates shown in Figure 4.14(c) and 4.15(c). This is not the case for $\gamma < 0$ (Figure 4.16(c)). In all three studies the range of parameter estimates for α is narrower in comparison to the range from the basic and modified model.

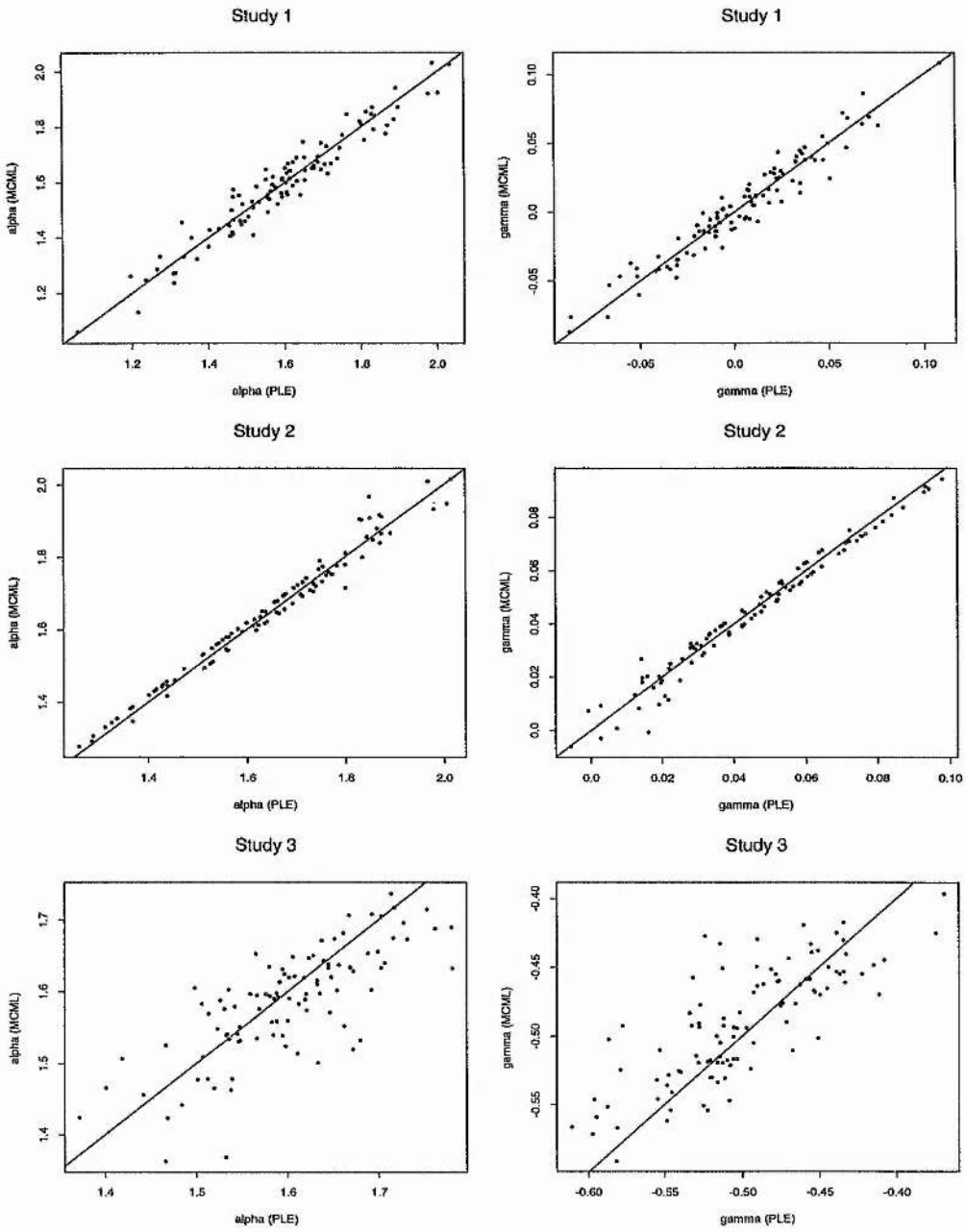


Figure 4.12: Comparison of parameter estimates from simulation *studies 1* to *3*. Estimates α (left) and γ (right) of MCML plotted versus estimates of PLE.

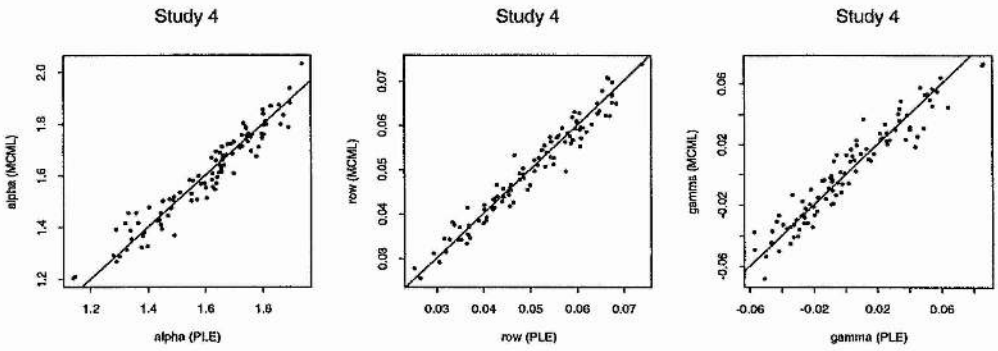


Figure 4.13: Comparison of parameter estimates from *study 4*. Estimates α (left), β (centre) and γ (right) of MCML plotted versus estimates of PLE.

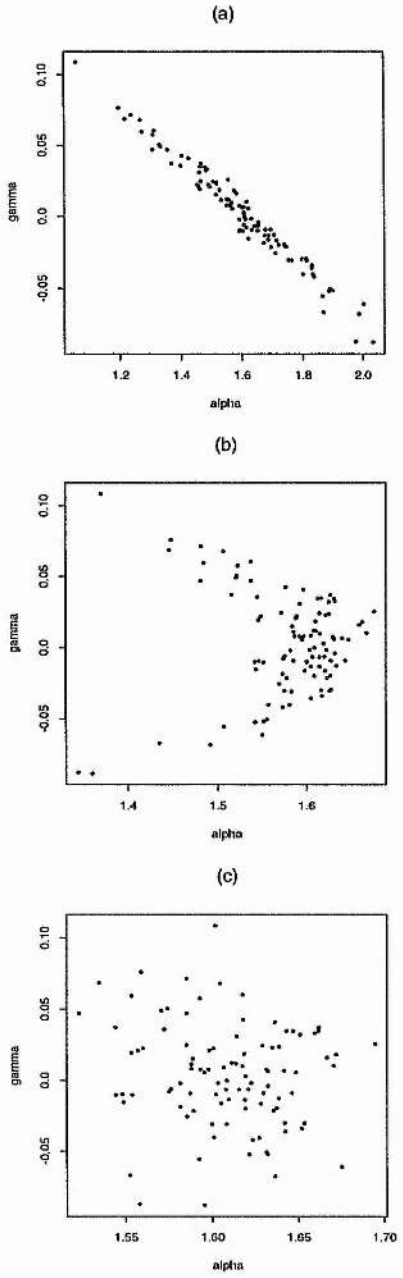


Figure 4.14: Parameter estimates using pseudo likelihood from simulation *study 1.* (a) $log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i} y_j$, (b) $log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i}(y_j - exp(\alpha))$ and (c) $log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i}(y_j - \bar{y})$.

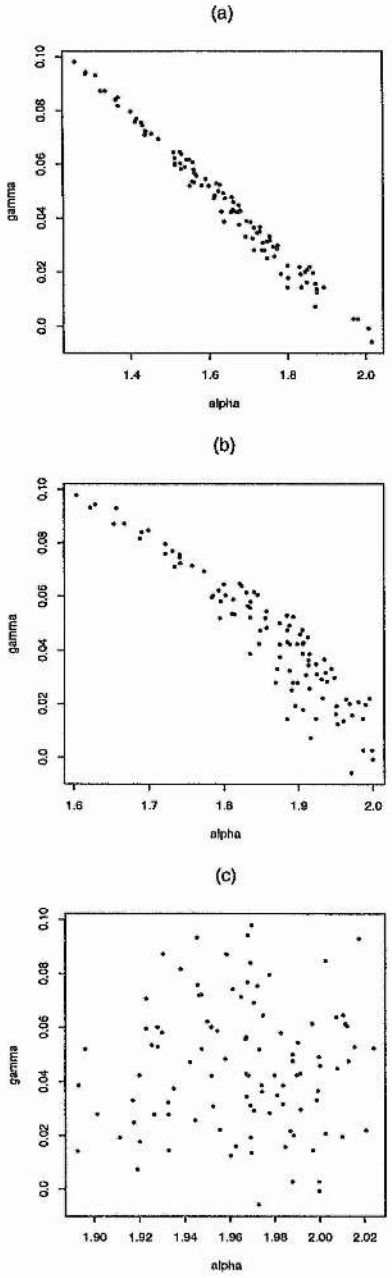


Figure 4.15: Parameter estimates using pseudo likelihood from simulation *study 2.* (a) $log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i} y_j$, (b) $log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i}(y_j - exp(\alpha))$ and (c) $log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i}(y_j - \bar{y})$.

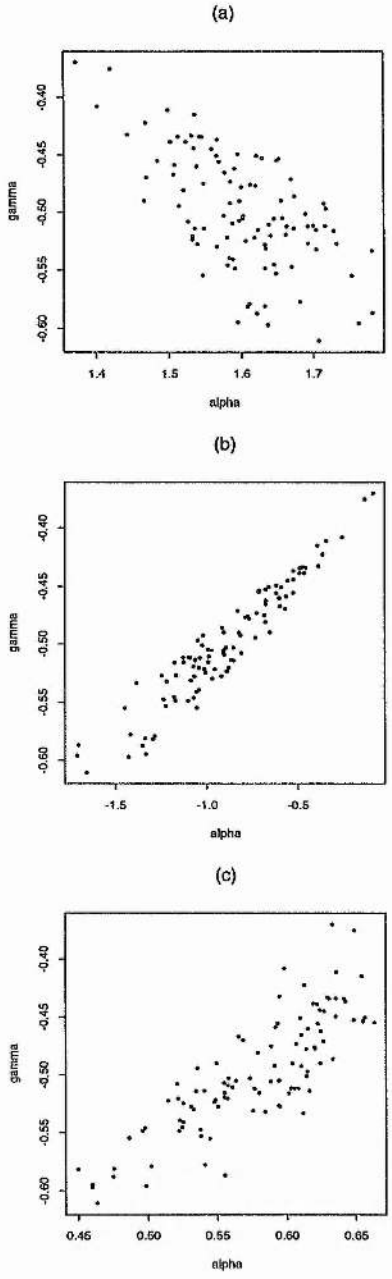


Figure 4.16: Parameter estimates using pseudo likelihood from simulation *study 3.* (a) $log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i} y_j$, (b) $log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i} (y_j - exp(\alpha))$ and (c) $log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i} (y_j - \bar{y})$.

Chapter 5

Applications of auto-models

5.1 Introduction

In the previous two chapters we have covered the theory, properties and parameter estimation methods for auto-models. Here we present several different applications of auto-models. In the literature not much advice on model selection and validation for auto-models is given. We illustrate a range of possible techniques.

1. Investigation of deviance or Pearson residuals as defined in section 4.2.2.

    - Distribution validation. Plot residuals versus fitted values to assess whether the distribution is adequate.
    - Spatial analysis of residuals. We use the correlogram (alternatively the variogram could be used) to check whether there is remaining autocorrelation in the residuals. Another option is to calculate a correlation coefficient, for instance Moran's coefficient of correlation (Moran, 1950), which assumes a certain neighbourhood structure. Then a permutation test is performed, calculating the correlation coefficients for permutations of the data, to test whether the observed correlation is significant.
    - Overall goodness of fit. The sum of squared deviance or Pearson's residuals provide the deviance and the X^2 statistic respectively, which can be compared with the χ^2 distribution to assess overall goodness-of-fit. The change in deviance can be used as usual in model selection.

2. Investigation of simulated summary statistics, such as $\sum y_i^{(t)}$ and $\sum y_i^{(t)} autocovariate_i^{(t)}$, where $y_i^{(t)}$ are response values of the simulated data set. If the cloud of simulated summary statistics in a scatter plot contains the observed summary statistics then the model is plausible (see section 4.3.6 or Geyer and Thompson, 1992).

3. Misclassification rates of spatial predictions (not suitable for counts). We investigate the number of misclassified squares from simulated data.

In section 5.2 we compare estimates from the auto-Poisson model for the spatial distribution of mites using the MCML parameter estimation method with estimates from the Monte Carlo Newton-Raphson and MPL method. Section 5.3 shows how the auto-Poisson models can be used to improve the understanding of the spatial distribution of weeds in a field. Finally section 5.4 shows how the auto-logistic model can be used to improve spatial prediction of deer presence in the Grampian region of Scotland for the case where only presence/absence in a sample of squares is available.

5.2 Mite count data example

The mite data are used by Lee and Kaiser (1997) to illustrate their proposed Monte Carlo Newton-Raphson (MCNR) estimation method for the Winsorized auto-Poisson model. The data are taken from Hairston et al. (1971) who examined distribution of soil arthropods. This is a good example for two reasons. First, the MCML method seems to work well because the positive auto-correlation is not very strong and therefore not close to the upper boundary of the autocorrelation parameter. Second we can use this example to compare the Lee and Kaiser (1997) parameter estimates with estimates obtained using the very similar MCML method (Geyer and Thompson, 1992).

Herbivorous mite counts were observed in 1965 on an 8 x 8 inch square of an abandoned field with continuous grass cover . A square was subdivided into 64 one inch cubes. These were labeled and taken to the lab, where the mites in each cube were counted. The data are shown below:

2 1 2 1 0 0 1 2

1 1 1 1 3 4 1 4

0 1 0 2 2 1 3 1

0 0 0 3 3 0 1 2

2 1 0 1 1 1 0 0

1 1 0 1 2 1 0 1

0 3 1 0 1 3 3 3

0 0 0 0 1 5 0 1.

Similarly to Lee and Kaiser (1997) who fit a Winsorized auto-Poisson model, we fit a truncated auto-Poisson model to these data, using a truncation value $r = 7$ and assuming a first-order neighourhood interaction:

$$\log(\mu_i|y_j; j \in N_i) = \alpha + \gamma \sum_{j \in N_i} y_j \tag{5.1}$$

Figures 5.1 (a) and (b) show the sequence of $\hat{\alpha}$ and $\hat{\gamma}$ for the MCML procedure, together with the MPL estimate and that of a model assuming independence, that is equation 5.1 without the term $\gamma \sum_{j \in N_i} y_j$. The MCML method clearly converges. The MCML result is also very similar to Lee and Kaiser's result (Table 5.2). The standard errors from MCML

Table 5.1: Mite data results. Results for MCNR are taken from Lee and Kaiser (1997). X^2 is the Pearson statistic of goodness-of-fit.

parameter	MCML	MCNR	MPL	ML
X^2_{df}	72.18_{62}	-	71.96_{62}	74.62_{61}
$\hat{\alpha}$	-0.199	-0.206	-0.215	0.198
s.e.	0.270	0.270	0.239	0.113
$\hat{\gamma}$	0.087	0.088	0.090	–
s.e.	0.051	0.052	0.043	–

and MCNR are almost identical and if a 95% confidence interval for $\hat{\gamma}$ was constructed using the standard error (assuming standard maximum likelihood theory) then it would only just include zero. This leads us to the conclusion that both the truncated auto-Poisson model and the model assuming independence might be adequate. Standard errors from MPL are smaller, but MPL possibly underestimates the variance when positive auto-

correlation is present. The summary statistics $t_1^{(t)} = \sum y_i^{(t)}$ and $t_1^{(t)} = \sum y_i^{(t)} \sum_{j \in N_i} y_j^{(t)}$, calculated from realizations of a simulated ergodic Markov chain using MCML, MPL for the model in equation 5.1 and ML (independence) estimates (Figure 5.1(c) - (d)) demonstrate that all of the models are plausible, but the independence model is marginal. In contrast to that, investigation of correlograms of Pearson residuals in Figure 5.2 showed that the positive correlation of 0.15 present in the first lag does not differ between the auto-Poisson and independence model. Using different ways of model validation we get different answers as to which model is appropriate. This example was mainly included for illustrative purposes and further investigation of the appropriate neighbourhood structure over which autocorrelation applies is needed. Also, if the example was to be taken further one should consult a biologist on whether there are mechanisms causing positive autocorrelation in these data.

5.3 Long Ashton seed count data example

This example illustrates how the auto-Poisson model can potentially be used to improve model fit. It also illustrates parameter estimation problems when positive autocorrelation is present, typically caused by an unmodelled trend rather than true autocorrelation. The data are from a field experiment with a randomised block design (Table 5.2) which was conducted at Long Ashton Research Station. There were four blocks (16 columns and 5 rows), two crops (WW = winter wheat, SW = spring wheat), two cultivation types (P = plough, T = tine) and five weed groups. A split-plot design was used, i.e. cultivation type and crop were only randomised among columns. The plot size was 3 x 3 m, and plots were 5 m apart in the north-south direction and 2 m apart in the east-west direction. Here the counts of a perennial weed species *Poa* spp. (Figure 5.3) are analysed. Seed counts refer to the total number of seeds found in six soil cores of volume 0.2 l, taken at 0 to 10 cm depth, which were collected in four consecutive years. Here only the first two years are analysed. The aim of the experiment was to observe persistence of seeds in the soil of the five groups of weed species sown at the start of the experiment, and to examine the persistence of *Poa* seeds which were already present before the experiment took place. Figures 5.3 show the observed seed counts. Given the fact that counts are aggregated from six soil cores per plot, there is not much chance that "true" spatial autocorrelation, such as competition, can be observed. Therefore the analysis strategy is initially to fit a model

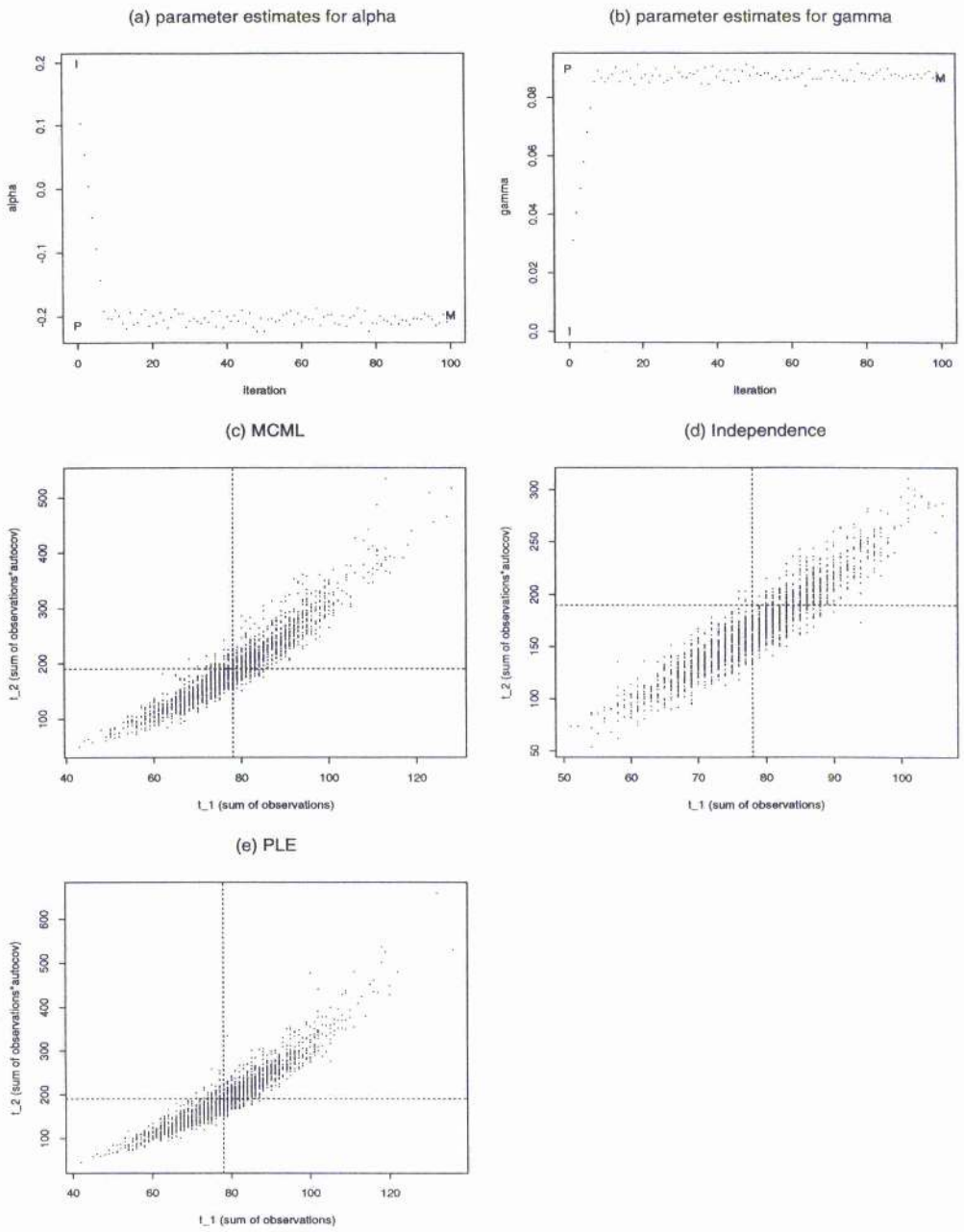


Figure 5.1: Mite example: parameter estimates for (a) α and (b) γ for the iterative process of MCML (marked as dots, final estimate as M) starting from ML estimates of the independence model (marked as I) and parameter estimates for MPL (marked as P). (c) - (e) are plots of simulated summary statistics $t_1^{(t)}$ versus $t_2^{(t)}$ using parameters from MCML, the independence model and MPL. Where the two dotted lines cross corresponds to observed t_1 and t_2.

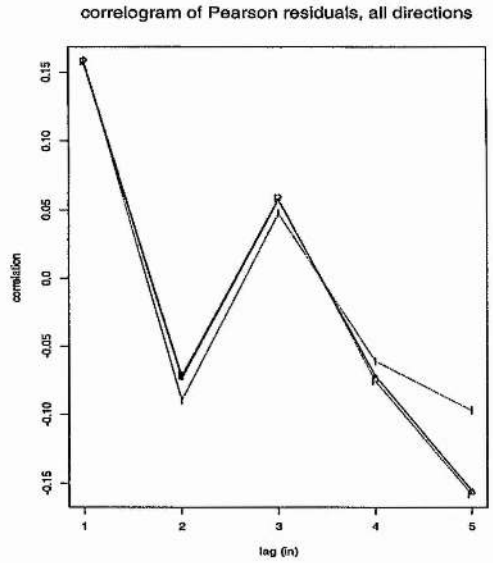


Figure 5.2: Omni-directional correlograms of Pearson's residuals from the auto-Poisson (MCML - solid line, MPL - dotted line) and independence (ML - dashed line) model fitted to the mite counts.

with block and treatment effects, then residuals are checked for remaining autocorrelation. The experiment design is such that effects of cultivation type and crop are confounded with columns. Since we are interested in the effects of the treatments rather than prediction, we fit block and treatment effects rather than column effects.

5.3.1 Spatial models for *Poa* year 1

Poa seed counts were lowest in year 1, with a maximum of 9. We consider a GLM assuming either a Poisson or a negative binomial distribution. Let y_{it} be the seed count at site i in year t. (In the following index t is omitted.) The maximal model considered is:

$$\begin{aligned} \log(\mu_i) &= \mathbf{x}_i'\boldsymbol{\beta} \\ &= block_i + weed_i * crop_i * cult_i \end{aligned}$$

with $y_i \sim \text{Poi}(\mu_i)$ or $y_i \sim \text{NegBin}(\mu_i, k)$, for which $Var(y) = \mu_i + \mu_i^2/k$, where k is the aggregation parameter. As $k \mapsto \infty$, the negative binomial (NegBin) distribution tends towards a Poisson distribution. A low value for k indicates high aggregation and a high value for k indicates little aggregation. With both distributions a log-link is used. The

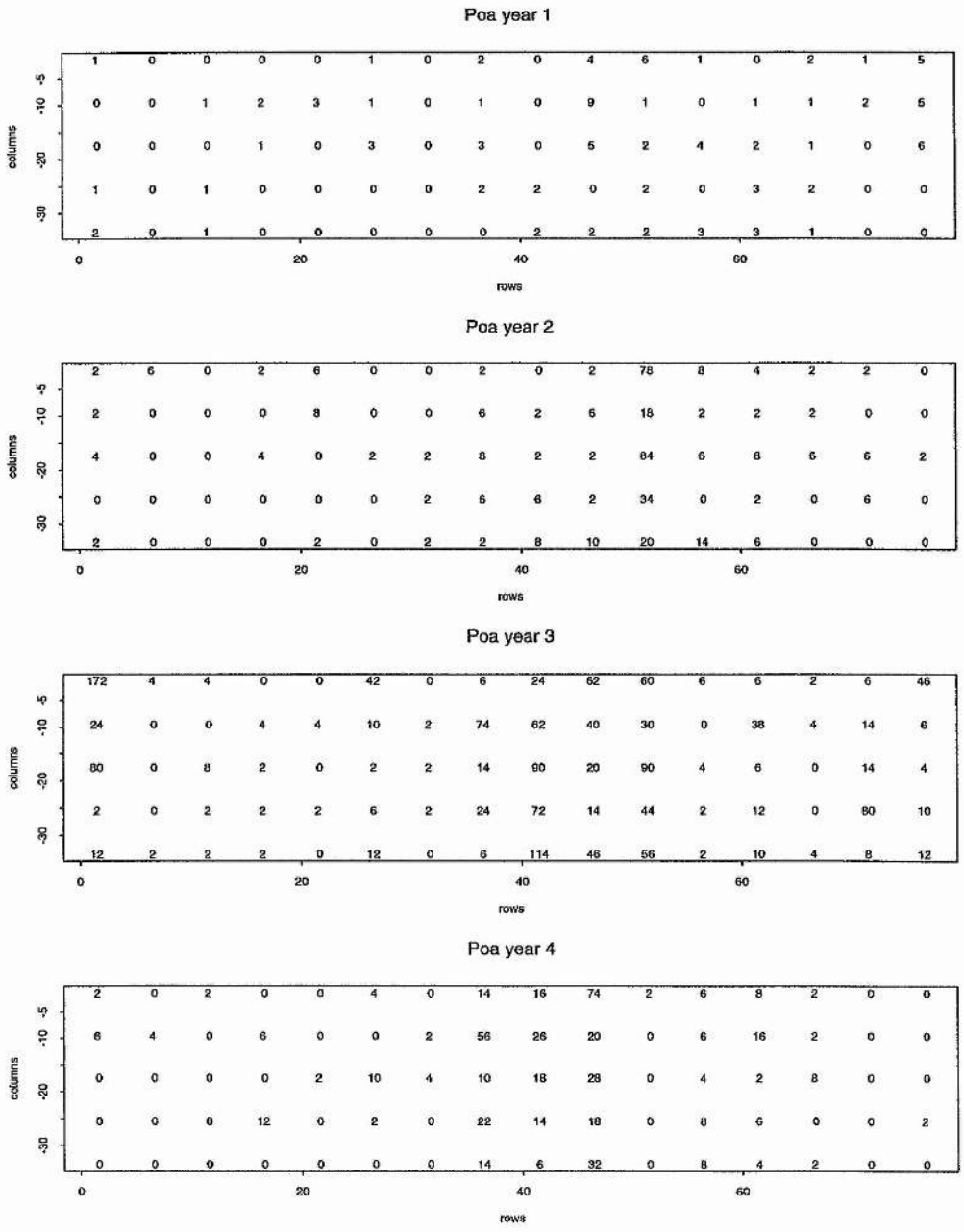


Figure 5.3: *Poa* seed counts in soil cores of volume 1.2 l sampled in four consecutive years.

Table 5.2: Long Ashton experiment. There are four blocks, two crops (WW, SW), two cultivation types (T, P) and five weed types (1-5) within each column. The numbers 1 to 5 symbolise the five weed types. Weed type numbers are shown at their relative position in the blocks.

block	1				2				3				4			
crop	WW	SW	WW	SW	SW	WW	SW	WW	WW	WW	SW	SW	WW	SW	SW	WW
cult	T	P	P	T	T	T	P	P	T	P	T	P	P	P	T	T
weed type	3	5	5	4	4	1	3	4	3	5	2	4	2	3	3	5
	1	3	1	3	2	4	1	5	2	1	3	2	1	2	4	4
	4	1	2	5	5	3	4	1	1	2	1	1	5	4	2	2
	2	4	3	1	1	5	5	2	5	4	4	3	4	5	1	1
	5	2	4	2	3	2	2	3	4	3	5	5	3	1	5	3

best fit for *Poa* seed was obtained for the model

$$\log(\mu_i) = block_i + weed_i + crop_i + cult_i + weed_i.cult_i, \tag{5.2}$$

for both the NegBin and the Poisson distribution. The NegBin model has an estimated aggregation parameter of $\hat{k} = 3.32$ (s.e.= 2.07) and the residual deviance is lower than that from the Poisson model (Table 5.3). The column effect is confounded with crop and cultivation type and not fitted. Deviance residual plots in Figure 5.4 show that there is little difference between the Poisson model and the NegBin model. After adjusting for overdispersion in the Poisson case, estimates and standard errors for the fixed effects are very similar. Spatial correlation in the Pearson residuals using correlograms in the north/south and east/west directions shows that in both models there is autocorrelation remaining in the north/south directions (Figure 5.5), which is slightly weaker for the NegBin model. The autocorrelation within a column could be caused by the tractor trailing along the columns. We can account for the autocorrelation by incorporating a specific term for autocorrelation in the north/south direction using the truncated auto-

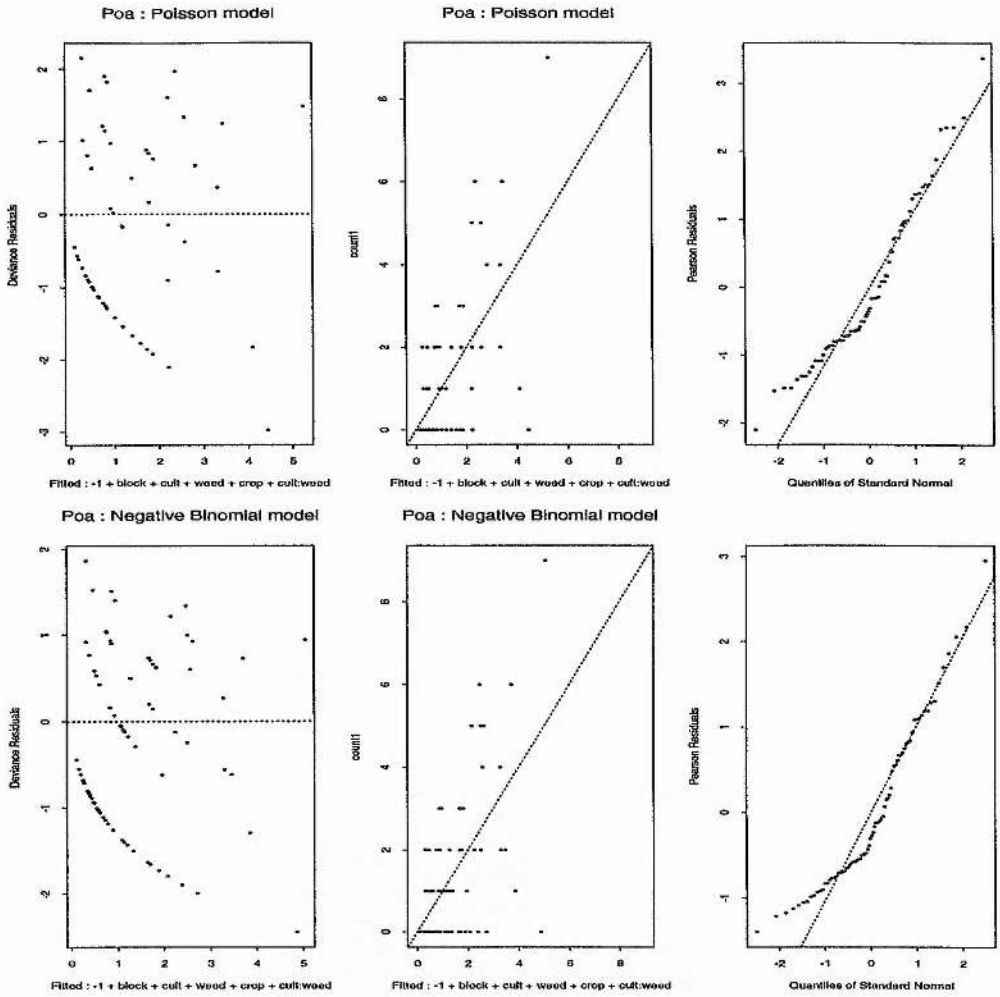


Figure 5.4: Residual plots for the Poisson (top) and negative binomial (bottom) models fitted to *Poa* counts. From left to right: Deviance residuals plotted versus fitted values, count plotted versus fitted values and deviance residuals plotted versus quantiles of the standard normal distribution.

Poisson model or the truncated auto-negative binomial model:

$$\log(\mu_i|y_j : j \neq i) = \mathbf{x}_i'\beta + \frac{1}{n_i}\gamma_{ij}\sum_{j\in N_i} y_j;$$

with $y_i \sim \text{Poi}(\mu_i|y_j; j \in N_i)$ or $y_i \sim \text{NegBin}(\mu_i|y_j; j \in N_i; k)$, n_i is the number of neighbours, N_i are the nearest four neighbours, except for edge plots, and $autocov_i = \frac{1}{n_i}\sum_{j\in N_i} y_j$. We define $\gamma_{ij} = ns$ for y_j which are north/south neighbours and $\gamma_{ij} = ew$ for y_j which are east/west neighbours. Parameter estimation for the auto-models can be carried out using pseudo likelihood, coding or the MCML method described in Section 4.3. When fitting this model using the MPL, only the north/south neighbourhood interaction

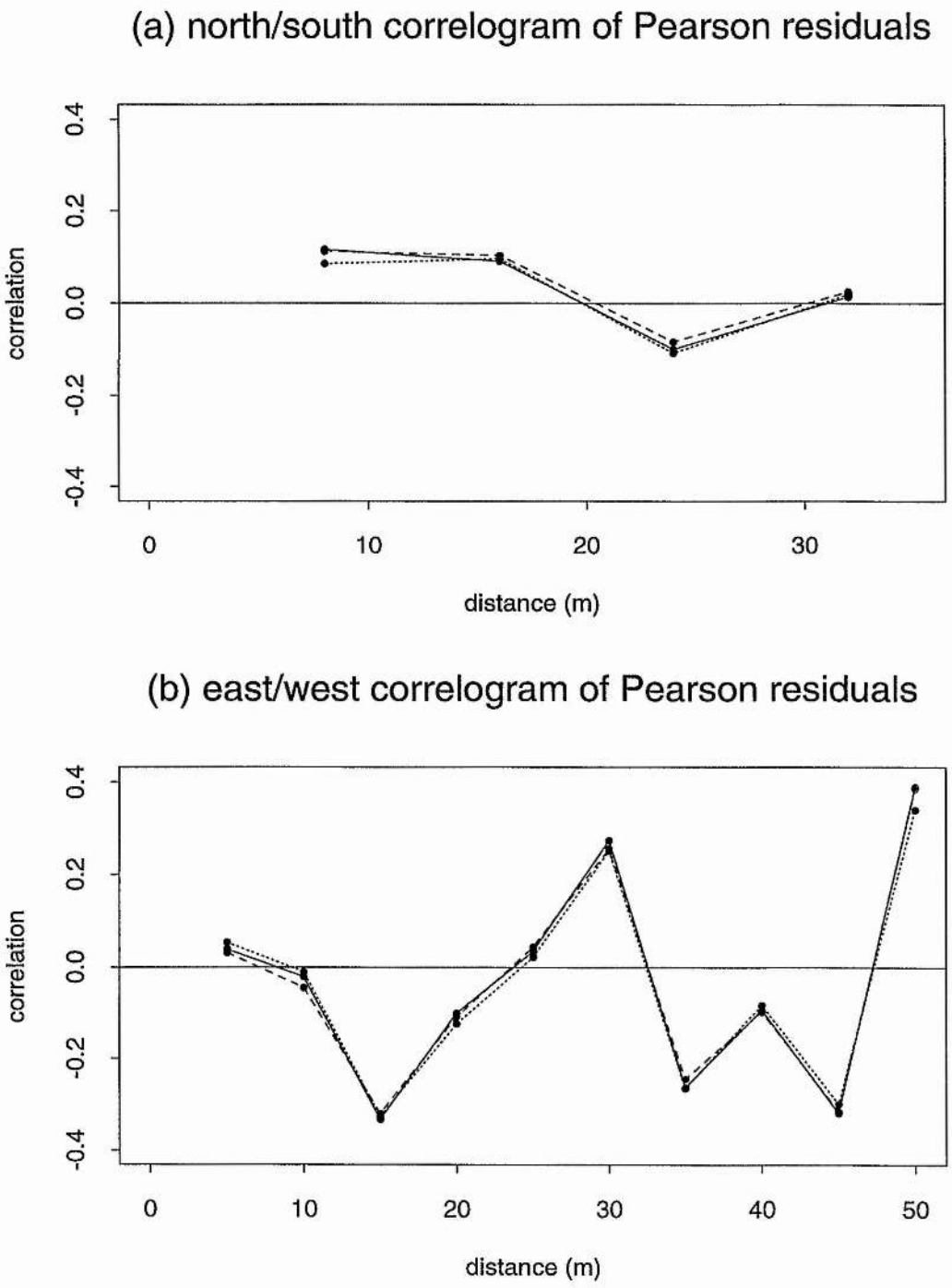


Figure 5.5: Correlograms in (a) north/south direction and (b) east/west direction of Pearson residuals from Poisson (solid line), auto-Poisson (dotted line) and NegBin model (dashed line) fitted to *Poa* year 1 seed counts. The auto-Poisson model is fitted by MCML.

is significant. Although standard GLM software provides the deviance of the pseudo likelihood when fitting the auto-Poisson model using the MPL method, it is not clear whether the deviance is approximately χ^2-distributed. In addition to MPL, the MCML method

Table 5.3: Results for the four different models (Poisson, auto-Poisson, NegBin and auto-NegBin) fitted to seed counts of *Poa* spp (year 1): Residual deviance, *df*, estimates of the coefficient of autocorrelation $\hat{\gamma_{ij}}$ and aggregation parameter $\hat{k}$ (fixed for the NegBin model). All models have as a minimum terms for block, weed, crop, cult and the weed.cult interaction. For MCML deviance residuals and Pearson's residuals for X^2 and Moran's correlation coefficient are defined as in section 4.2.2.

	Poisson	auto Poisson		Negative Binomial	auto NegBin
		MPL	MCML		MPL
df	66	65	65	66	65
X^2	100.2	-	95.2	69.1	-
residual deviance	111.7	102.9	112.4	83.8	79.5
$\hat{ns}$ (north/south)	n.a.	0.169	0.034	n.a.	0.165
{asymptotic s.e.}		{0.072}	{0.056}		{0.075}
$\hat{ew}$ (east/west)	n.a.	-	-	n.a.	-
$\hat{k}$ (aggregation)	n.a.	n.a.	n.a.	3.3	3.3
{asymptotic s.e.}	-	-	-	2.07	n.a.
Can residuals be assumed random in space?	yes	-	yes	yes	-
Moran corr. (first order)	0.047	-	0.046	0.047	-
(p-value perm. test.)	(0.247)	-	(0.209)	(0.238)	-
Can residuals be assumed random in space?	yes	-	yes	yes	-
Moran corr. (north/south)	0.151	-	0.147	0.121	-
(p-value perm. test.)	(0.085)	-	(0.100)	(0.138)	-

was applied to obtain estimates for the auto-Poisson model. For MCML, the vector of summary statistics is

$$t(y) = (\sum y_i x_{i,1}, ..., \sum y_i x_{i,14}, \frac{1}{2} \sum y_i autocov_i)',$$

where

$$autocov_i = \frac{1}{n_i} \sum_{j \in N_i} y_j.$$

We have 14 columns in the design matrix of the independence model from equation 5.2. The parameter vector is

$$\eta = (\beta_1, ..., \beta_{14}, ns)'.$$

One problem in this example is that the estimate for ns is at the upper boundary of the parameter space, which is around 0.1 using the ad-hoc calculation described earlier. This means that the MCML method cannot work very well, because simulation of data needed for the normalizing constant is restricted to values below or equal to the upper boundary. The MCML method breaks down, because estimates for γ reach the boundary and the counts simulated by the Gibbs sampler are at the truncation value. The other problem is that treatment, block and autocorrelation effects cannot be distinguished. When keeping the treatment and block parameter estimates from the Poisson model fixed and then estimating γ with MCML, the estimate converges. Using the converged $\hat{\gamma}$ we do one final run of MCML where all parameters are estimated simultaneously, and the resulting $\hat{\gamma} = 0.034$ ($s.e. = 0.056$) is our final estimate. It is only a quarter of the MPL estimate $\hat{\gamma} = 0.169$ ($s.e. = 0.072$) and a 95% confidence interval ($\gamma + / - 1.96 * s.e.(\gamma)$) for the MCML estimate would include 0, indicating that autocorrelation is not significant. The Gibbs sampler broke down when estimating the auto-negative binomial model, due to the restriction on the scale parameter of β of the NegBin distribution, which has to be greater than zero. Investigating the correlograms in the north/south (Figure 5.5a)) and east/west (Figure 5.5b) directions of Pearson residuals from the three models (Figure 5.5), using the MCML estimates for the auto-Poisson model, shows no big difference between models. This is confirmed by a formal test of spatial correlation on the Pearson residuals using Moran's coefficient of correlation (Moran, 1950), which does not show significant positive correlation for any of the models when based on a first order neighbourhood (Table 5.3). Calculating Moran's coefficient of correlation based on a north/south neighbourhood, residuals from none of the models are significantly correlated, based on a 5% significance level. So in addition to the confidence intervals for the autocorrelation parameter ns from MCML estimates, correlograms and Moran's coefficient indicate that autocorrelation is not significant. In contrast to this is the estimate of ns we obtain from MPL, for which confidence intervals indicate a significant effect. Calculation of Pearson residuals for the auto-models requires simulating data using the Gibbs sampler,

as described in Section 4.2.2. With MPL estimates this could not be achieved, counts simulated by the Gibbs sampler are at the truncation value. It has been quoted in the literature that when estimates are near the critical value where phase transition occurs, MPL estimates are more variable than MCML estimates (Geyer, 1991). All these encountered problems add to the evidence that the model with autocorrelation as it stands is not appropriate. Figure 5.5b) shows that there is negative correlation in east/west direction at a lag of 15 m. Including an additional autocovariate for the counts of the third plot in east and west direction might improve the model fit. Further work in this respect is required.

5.3.2 Spatio-temporal model for *Poa* year 2

The counts for *poa* seeds in year 2 are much higher than in year 1 with a mean of 5.3 and maximum count of 84 (Figure 5.3). When fitting a Poisson model, overdispersion was very high, pointing to the negative binomial as the appropriate distribution, and the fitted model is:

$$\begin{aligned} \log(\mu_{it}) &= \theta log(y_{it-1}+1) + \mathbf{x}_i'\boldsymbol{\beta} \\ &= \theta log(y_{i1}+1) + block_i + crop_i + cult_i + crop_i.cult_i \end{aligned}$$

where the response y_{it} is the seed count at site i in year t, with $t = 2$ and $y_{it} \sim \text{NegBin}(\mu_{it}, k)$. The model fit could not be improved by adding terms for the autocorrelation. Parameter estimates are given in Table 5.4.

5.4 Modelling the presence/absence of red deer

In this section we summarise the work on modelling red deer data of which a substantial part was submitted for the degree of Master of Science in Biometry at Reading University (Augustin, 1993). This example was published in Augustin et al. (1996, 1998b) and is included for several reasons. Firstly, it illustrates the auto-logistic model for the case in which the response represents presence/absence of a species in squares of a sampling grid, where only a simple random sample is available. Secondly, since the original work was carried out, more insight has been gained on several aspects of the work. One aspect is the difference between theoretical approaches taken to derive abundance estimators in

Table 5.4: Results for the NegBin model fitted to seed counts of *Poa* spp. (year 2) with and without a temporal component ($log(y_{i1}+1)$): Residual deviance, *df*, estimates of the coefficient for the temporal effect (θ) and aggregation parameter $\hat{k}$ (fixed for the temporal NegBin model). Both models have as a minimum the terms block, crop, cult and the crop.cult interaction fitted. For the temporal NegBin model k is kept fixed using estimate from the NegBin model.

	NegBin	temp. NegBin
df	73	72
residual deviance	88.4	82.7
$\hat{\theta}$	-	0.67
{s.e.}		{0.26}
$\hat{k}$ (aggregation)	0.82	0.82
{s.e.}	{0.21}	-

the comparison. Thirdly, an improvement to the bootstrap algorithm used for variance estimation is proposed.

The appropriate GLM for the presence/absence of deer in a sample of squares is a logistic model, which on adding the autocovariate becomes an *auto*-logistic model. If data are available only from a sample of grid squares, we need a mechanism for generating 'observations' for unsurveyed squares so that we can evaluate the autocovariate term. We use simulation in the form of the Gibbs sampler to achieve this. We also show that a modification to the Gibbs sampler, in which we use expected values from neighbouring squares rather than simulated observations in the calculation of the autocovariate, gives lower misclassification rates when used to model Scottish red deer data (Augustin et al., 1996).

5.4.1 The model

Suppose that the study region is divided into a grid of n squares with the response in square i denoted by y_i $(i = 1, \ldots, n)$. If the species being studied is present in square i then $y_i = 1$, otherwise $y_i = 0$. The conditional probability of presence in square i is given

by

$$p_i = \Pr(Y_i = 1 | Y_j = y_j, j \in N_i).$$

When presence/absence is recorded in every square, the auto-logistic model may be fitted and the estimated conditional probability of occupation, $\hat{p}_i$, may be obtained for all the squares using the equation

$$p_i = \frac{\exp(x_i'\beta + \sum_{j \in N_i} \gamma_{ij} y_j)}{1 + \exp(x_i'\beta + \sum_{j \in N_i} \gamma_{ij} y_j)} \tag{5.3}$$

(Besag, 1972; Preisler, 1993). The term $x_i'\beta$ represents a collection of effects associated with a set of spatial covariates, where x_i is the i'th row of the design matrix and β is the vector of coefficients. Omitting the term involving γ_{ij} from the right hand side of the equation reduces the model to a classical logistic model.

It is also of interest to predict presence/absence in each square for comparison with the observed spatial distribution. Buckland and Elston (1993) describe two ways of doing this. Their first method is a deterministic prediction in which those squares which have the highest estimated probability of occupation are marked as occupied, subject to an overall number of $\sum_i \hat{p}_i$ occupied squares. Their alternative method is a stochastic prediction, or realisation, in which the outcome in square i is obtained by simulating from Bernoulli($\hat{p}_i$).

It is straightforward to fit an ordinary logistic model to data from a sample of, say, 20% of the squares, because the only terms in this model are the spatial covariates which are known for every square. Estimated probabilities of occupation can, therefore, be calculated for every square and predictions of presence/absence can be obtained by retaining the observed 'real' values for the 20% of squares included in the sample and using the deterministic/stochastic methods to obtain predictions for the remaining squares. Fitting an auto-logistic model to a random sample of squares is complicated by the fact that the neighbourhoods associated with sampled squares contain missing values. One solution is to use the Gibbs sampler to estimate presence/absence in unsampled squares as described below.

5.4.2 Implementing the Gibbs sampler

The specification of the auto-logistic model is inherently conditional, since y_i depends on presence/absence in neighbouring squares. Our real objective, however, is to model the

joint distribution of all the y_i's. The Gibbs sampler allows sampling from the conditional distribution to generate the joint distribution by successively updating each response according to the conditional probability given the current values of all the other variables in the model. We can use the Gibbs sampler to overcome the problem of having surveyed just some of the squares by including the imputed presence/absence responses for the unsurveyed squares as additional variables in the model. The main benefit of this is that the autocovariate for each surveyed square can be computed, and so the auto-logistic model can be fitted to these squares. As a by-product, we obtain estimated probabilities of occupation for *all* the squares so that we can map the estimated spatial distribution of the species and estimate the total number of occupied squares.

The algorithm for implementing the Gibbs sampler is as follows. We assume that the autocorrelation parameter $\gamma_{ij} = \gamma$.

1. Fit an ordinary logistic regression model to the surveyed sample of squares. Calculate the fitted probability,
$$\hat{p}_i = \frac{\exp(x_i'\hat{\boldsymbol{\beta}})}{1 + \exp(x_i'\hat{\boldsymbol{\beta}})}$$
for all of the squares.

2. Set $t = 1$ and initialise $\boldsymbol{Y}$ using the stochastic prediction method, where
$$y_i^{(0)} \sim \text{Bernoulli}(\hat{p}_i).$$

3. Calculate the autocovariate for each square using the map obtained in the previous step.

4. Fit the auto-logistic model to the 20% sample
$$\log \frac{p_i}{1 - p_i} = x_i'\boldsymbol{\beta} + \gamma \sum_{j \in N_i} y_j$$
to obtain $\hat{\boldsymbol{\beta}}$.

5. Update $\boldsymbol{y}$ successively:
$$y_i^{(t)} \sim \text{Bernoulli}(\hat{p}_i | y_j; j \in N_i)$$
with $N_i = (y_j^{(t)}, y_k^{(t-1)})$, where $y_j^{(t)}$ are already updated squares in the set and $y_k^{(t-1)}$ are squares not yet updated, and
$$(\hat{p}_i | y_j; j \in N_i) = \frac{\exp(x_i'\hat{\boldsymbol{\beta}} + \hat{\gamma} \sum_{j \in N_i} y_j)}{1 + \exp(x_i'\hat{\boldsymbol{\beta}} + \hat{\gamma} \sum_{j \in N_i} y_j)}$$

6. Set $t = t + 1$ and return to step 4. Once convergence is reached realisations $\boldsymbol{y}^{(t)}$, $\boldsymbol{y}^{(t+1)}, ..., \boldsymbol{y}^{(t+m)}$ and the corresponding $\hat{\mathbf{p}}^{(t)}$, $\hat{\mathbf{p}}^{(t+1)}$,..., $\hat{\mathbf{p}}^{(t+m)}$ are saved.

Note that steps 1 to 5 describe the process of fitting an auto-logistic model on its own and then generating a stochastic prediction of the distribution (with one iteration of step 5), whereas step 5 describes the Gibbs sampling in addition. The repeated application of steps 5 and 6 allows us to assess the Monte Carlo variability introduced by the Gibbs sampler. The final map gives the predicted distribution of the species allowing for spatial autocorrelation. The selection of T, the number of iterations of the Gibbs sampler, depends on how much computing time is feasible and how quickly the sampler converges (see Section 5.4.4). In step 4 there are different possibilities to estimate parameters: MCML, MPL or coding as discussed in section 4.3. For computational ease we choose MPL in this application. We also implemented a modification to the auto-logistic model used in the EM/Gibbs sampler in step 4 of the algorithm, replacing y_j by $\hat{p}_j$ in equation (5.5).

Steps 1 to 6 of our algorithm combined with the modified version of the Gibbs sampler in step 4 are similar to the steps involved in the EM algorithm. The computation of the autocovariate using $\hat{p}_j$ is analogous to the 'E-step' (estimation) and the maximisation of the pseudo likelihood is analogous to the 'M-step' (maximisation). Using the unmodified version of the Gibbs sampler in step 4 is analogous to the 'S-step' (simulation) in the stochastic EM algorithm (see Diebolt and Ip, 1995, for details). The main difference between our algorithm and the various EM algorithms is that the Gibbs sampler is needed to generate observations for all the unsampled squares in turn, taking into account their interdependencies, whereas the EM algorithm approach could be relevant if it were possible to generate values of the auto-covariate to replace missing values simultaneously. Due to the hybrid form of our algorithm, where the Gibbs sampler is used for the simulation or estimation step of the EM algorithm, we call our described algorithm EM/Gibbs sampler.

5.4.3 Variance estimation - bootstrap

For variance estimation the following bootstrap procedure is used:

1. Perform the EM/Gibbs sampler algorithm as described above on the observed 20% random sample of sites, yielding

$y^{(t)}$, $y^{(t+1)}$, ..., $y^{(t+m)}$ and the corresponding $\hat{\mathbf{p}}^{(t)}$, $\hat{\mathbf{p}}^{(t+1)}$,..., $\hat{\mathbf{p}}^{(t+m)}$.

2. Generate new presence/absence data from $\hat{\mathbf{p}}^{(t+m)}$ in step 1.

3. Select a 20% simple random sample from the presence/absence deer data simulated in step 2.

4. Perform the EM/Gibbs sampler algorithm as described in the previous section on the bootstrapped 20% sample from step 3, yielding $y^{(t)}$, $y^{(t+1)}$, ..., $y^{(t+m)}$ and the corresponding $\hat{\mathbf{p}}^{(t)}$, $\hat{\mathbf{p}}^{(t+1)}$,..., $\hat{\mathbf{p}}^{(t+m)}$.

5. Calculate the estimated number of expected occupied squares $\hat{o}_b$ from $\hat{o}_b = \sum_{i=1}^{n} \hat{p}_i^{(t+m)}$, where $n = 1277$ and m is the number of realisations created by the EM/Gibbs sampler algorithm.

6. Repeat steps 3 to 5 120 times, yielding 120 $\hat{o}_b$'s. The variance $Var(\hat{o}_b)$ is the bootstrap estimate of variance for the number of occupied squares.

The bootstrap for the estimated number of occupied squares using the logistic model is very similar: all steps involving the EM/Gibbs sampler are replaced by a single step of fitting the logistic model. The estimated number of expected occupied squares $\hat{o}$ is the sum of fitted probabilities from the logistic model. An improvement to the above bootstrap procedure is to use the ergodic average $\bar{\mathbf{y}}$ from $y^{(t)}$, $y^{(t+1)}$, ..., $y^{(t+m)}$ to estimate $\hat{\mathbf{p}}$ instead of using $\hat{\mathbf{p}}^{(t+m)}$ in step 2 and in step 5 for estimating $\hat{o}_b = \sum_{i=1}^{n} \bar{\mathbf{y}}$.

5.4.4 The model fitted to the deer data

Buckland and Elston (1993) model the spatial distribution of red deer in the Grampian Region of Scotland using Red Deer Commission census data. Their response variable is the number of deer counted per 1 km grid square and their spatial covariates represent physical and habitat-related attributes. The distribution of counts is heavily skewed and so they use a two-stage modelling strategy (Aitchison, 1955): first presence/absence of deer is modelled using logistic regression and an estimate of the total number of occupied squares is calculated; this estimate is then multiplied by the average number of deer per occupied square to calculate overall abundance. Buckland & Elston select a simple random sample of 20% of the 1277 1 km squares in the study area and use this to assess the

effectiveness of logistic regression coupled with Aitchison's method when only a sample of squares is surveyed. They compare estimates of overall abundance and maps of the predicted distribution of deer with the 'true' values from the complete census.

Buckland & Elston's logistic regression model is based on

$$x_i'\beta = \alpha + \beta_1 altitude_i^2 + \beta_2 northing_i + \beta_3 mires_i + \beta_4 easting_i + \beta_5 pine_i, \qquad (5.4)$$

where *altitude*, *northing*, *easting*, *mires* and *pine* are all variates (*mires* and *pine* are, respectively, the areas covered by mire and native pine woodland in each square). This model was selected using a forward stepwise procedure applied to a specific 20% sample of squares. Our comparisons are based on the same 20% sample and we use the logistic model implied by equation (5.4) as a starting point for our auto-logistic model.

A suitable autocovariate for use with the deer data is

$$autocov_i = \frac{1}{h_i} \sum_{j \in N_i} \frac{1}{h_{ij}} y_j, \qquad (5.5)$$

where h_{ij} is the Euclidean distance between squares i and j measured from the square centre, $h_i = \sum_{j \in N_i} (1/h_{ij})$ and N_i is the set of neighbours of square i. Thus $autocov_i$ is a weighted average of the number of occupied squares in the neighbourhood associated with square i. This autocovariate is equivalent to using $\gamma_{ij} = \beta_6 \frac{1}{h_i} \frac{1}{h_{ij}}$ in equation (5.3) and the full auto-logistic model is given by

$$\log \frac{p_i}{1 - p_i} = \alpha + \beta_1 altitude_i^2 + \beta_2 northing_i + \beta_3 mires_i + \beta_4 easting_i + \beta_5 pine_i + \beta_6 autocov_i.$$

Autocovariates corresponding to various neighbourhood sizes, ranging from Besag's (1974) isotropic first-order scheme involving the four nearest neighbours of square i to a square of side 9 km, were computed according to equation (5.5). The most suitable autocovariate for the deer data was determined by examining the reduction in deviance of the pseudo likelihood obtained by adding a specified autocovariate to Buckland & Elston's logistic model. In all cases fitting the autocovariate gave a significant reduction in deviance (see Table 5.5). We also tried unweighted versions of the autocovariates (equivalent to using $\gamma_{ij} = \beta_6 \frac{1}{h_i}$, where h_i is now the total number of squares in the ith clique), but these were always outperformed by their weighted counterparts. The weighted autocovariate corresponding to a clique of side 7 km produced the greatest reduction in deviance (7.264 compared to χ_1^2) relative to the amount of extra computation required and was used in all subsequent analyses. Once this autocovariate had been fitted, *easting*, *northing* and *mires*

Table 5.5: Results of fitting different auto-logistic models: all models contain the explanatory variables *altitude*^{2}, *northing*, *mires*, *easting* and *pine* in addition to the specified autocovariate.

Autocovariate	Degrees of Freedom	Residual Deviance
—	250	210.858
Besag First-Order	249	205.782
Besag Second-Order	249	205.504
5×5 Square, Unweighted	249	204.401
5×5 Square, Weighted	249	203.594
7×7 Square, Unweighted	249	201.129
7×7 Square, Weighted	249	200.882
9×9 Square, Unweighted	249	201.471
9×9 Square, Weighted	249	200.741

were no longer significant. We examined the effect of dropping these terms from the auto-logistic model before implementing the EM/Gibbs sampler when we compared the ability of different models to predict the total number of occupied squares (see Section 5.4.5).

Our model assumes that the probability of a square having deer present only depends on the presence of deer in neighbouring squares. The justification of this is that deer tend to occur in groups of similar sizes. Nevertheless it might be more adequate to let the probability of deer present in a square depend on the actual deer counts observed in neighbouring squares. Such an alternative model deserves investigation.

When running the algorithm we conditioned on the observed presence/absence from the 20% sample at each step, as we wished to predict presence/absence only for the 80% of 'unobserved' squares. Thus in steps 2 and 5 of the algorithm, the stochastic method for generating presence/absence data is applied only to squares excluded from the 20% sample. At the time the work was carried out we found that $T = 20$ iterations of the EM/Gibbs sampler was practically feasible, although the parameter estimates (Figure 5.6) and the predicted probabilities (Figure 5.7a) continued to show considerable variation after 20 iterations. This is because the autocovariate is calculated from a stochastic realisation of presence/absence data at each iteration. We therefore implemented a modification to the

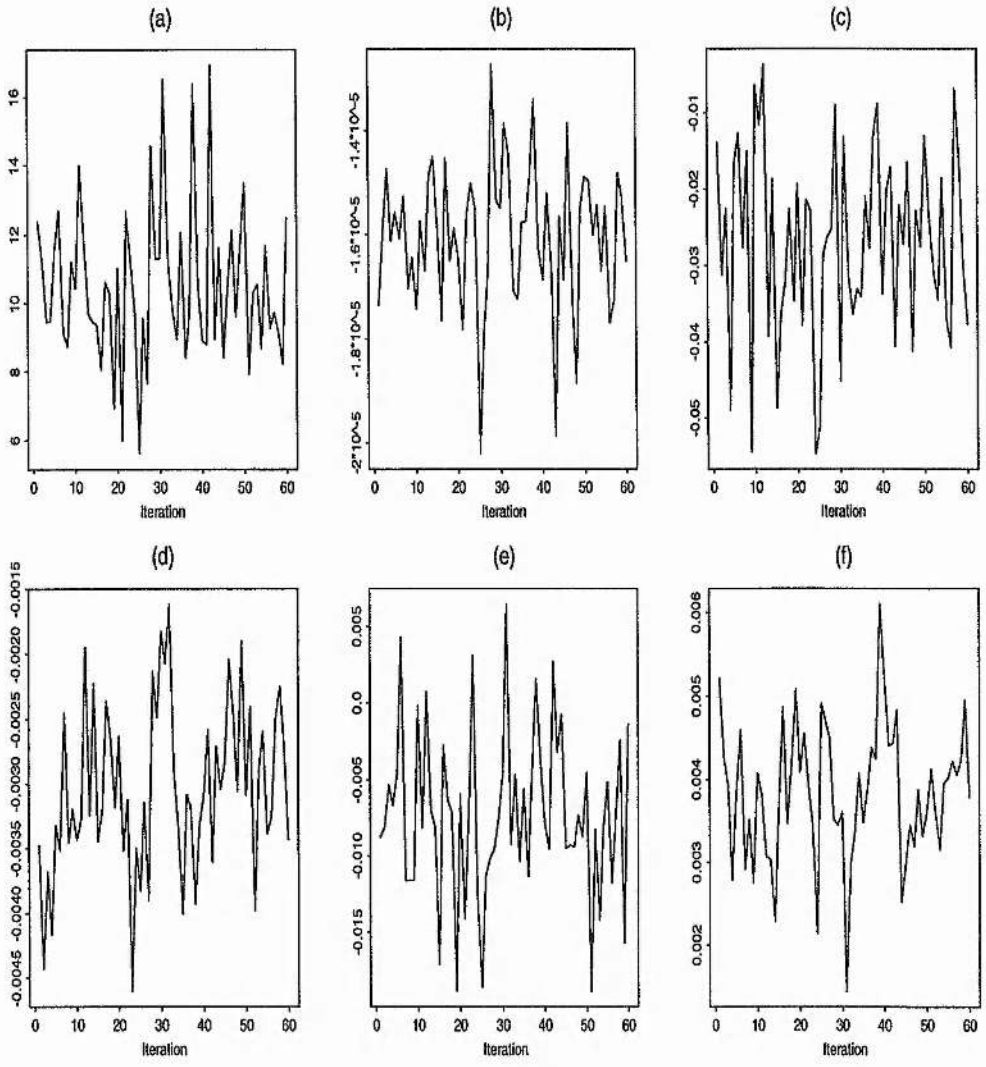
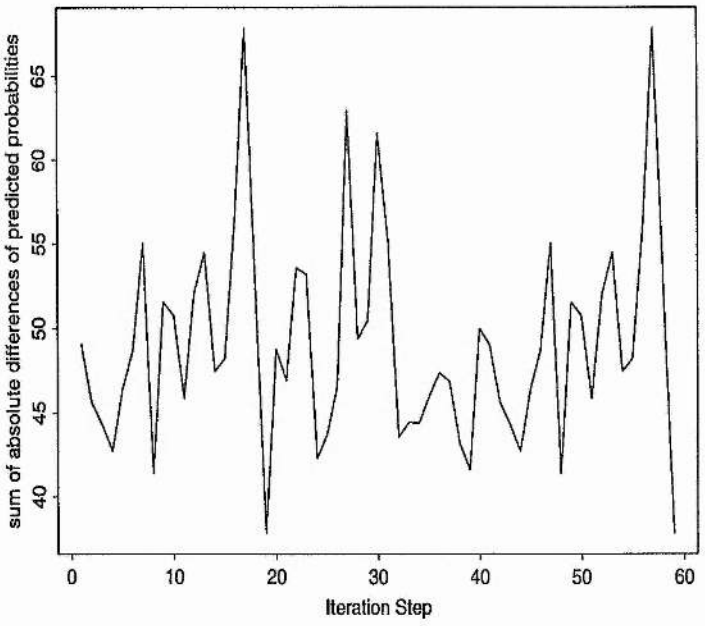


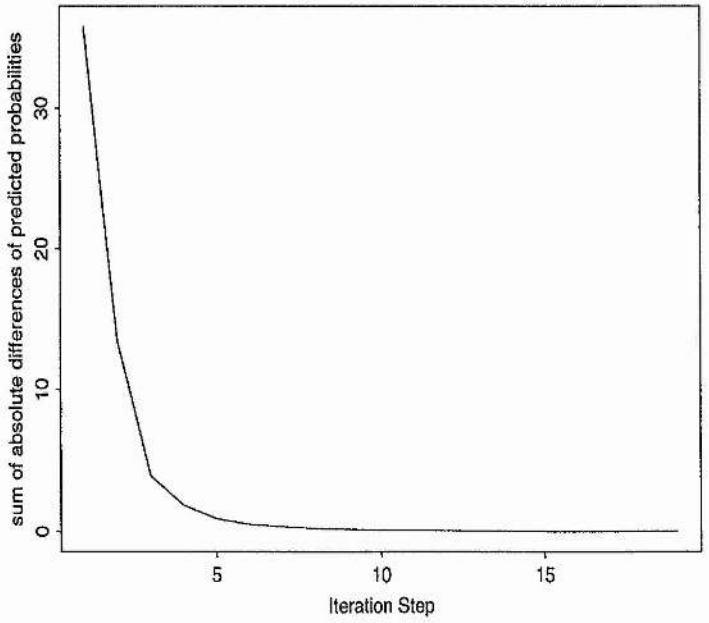
Figure 5.6: Convergence of parameter estimates for the EM/Gibbs sampler: (a) to (f) show estimates for *autocov*, *altitude*^{2}, *northing*, *mires*, *easting* and *pine*, respectively.

auto-logistic model used in the EM/Gibbs sampler in step 4 of the algorithm, replacing y_j by $\hat{p}_j$ in equation (5.5). This modified EM/Gibbs sampler converged within 10 iterations (see Figures 5.8 and 5.7b). We did not prove whether the auto-logistic model with the autocovariate defined in such a way is a valid Markov random field.

Figures 5.8 and 5.7b show that convergence of the modified EM/Gibbs sampler is towards fixed values of final parameter estimates. The unmodified EM/Gibbs sampler (Figures 5.6 and 5.7a) has possibly converged too, not to a fixed quantity, but to the equilibrium distribution of the parameters. Although we have not done any formal convergence tests



(a)



(b)

Figure 5.7: Convergence of fitted probabilities (a) for the EM/Gibbs sampler and (b) for the modified EM/Gibbs sampler.

for this particular version of the algorithm, we have shown in section 4.2.1 that these types of algorithms converge very fast. The estimated coefficient for the autocovariate is around 12 for both methods, but variability about this value is considerable for the EM/Gibbs sampler (Figure 5.6). This variability in the parameter estimates of each realisation of the Markov chain allows for the uncertainty in the missing values. In the modified EM/Gibbs sampler this information on uncertainty of the missing values is lost. By dividing the coefficient $\hat{\beta}_6 = 12$ by $h_i = 48$ (there are 48 squares in N_i), we obtain $\hat{\gamma}_{ij} = 0.25$. Relating $\hat{\gamma}_{ij}$ to Figure 3.3 in section 3.3 indicates that $\hat{\gamma}_{ij} = 0.25$ is not close to a value where "phase transition" would be a problem. Since this model here is a lot more complicated than the model assumed for Figure 3.3, this is only a very crude check.

The modified EM/Gibbs sampler requires much less computation than the original version since there is no need to generate presence/absence data at every iteration. Another option, which requires even less computation, is to perform just one iteration of the algorithm. This gives us a total of four methods for modelling the spatial distribution of the deer, namely Buckland & Elston's logistic model, the basic auto-logistic model after one iteration of the algorithm, the EM/Gibbs sampler and the modified EM/Gibbs sampler. The performance of the methods is compared in the next section.

5.4.5 Comparison of methods

The effectiveness of the methods for predicting the spatial distribution of deer can be assessed by comparing predicted probabilities and stochastic realisations of presence/absence data with the 'true' distribution of deer. Formal comparisons based on misclassification rates are summarised in Table 5.6. Results are based on 120 stochastic realisations from the final map of fitted probabilities for each method. The modified EM/Gibbs sampler is the best of the four methods, giving the highest and least variable matching coefficient (the proportion of squares classified correctly). All the new methods are better than the ordinary logistic model, although the standard, unmodified EM/Gibbs sampler performed marginally worse than the basic auto-logistic model. Visual comparison of maps of estimated probabilities, based on ergodic averages of the 120 stochastic realisations ($\boldsymbol{y}^{(t)}$, $\boldsymbol{y}^{(t+1)}, ..., \boldsymbol{y}^{(t+m)}$) in Figures 5.9 (a) and (b) and stochastic realisations of presence/absence data in Figure 5.9 (c) and (d) with the observed distribution (Figure 5.10) reveal that the new methods produce less uniform estimated occupation probabilities and more

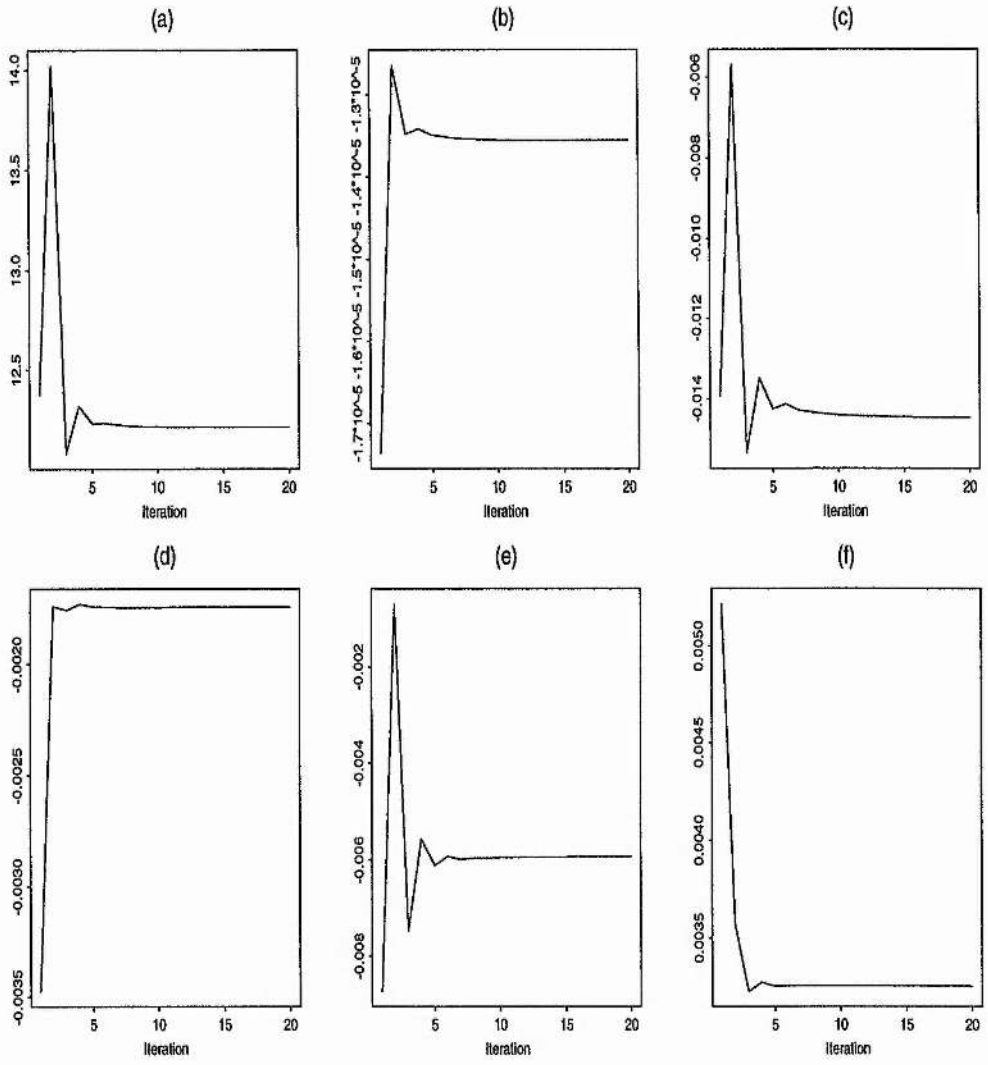


Figure 5.8: Convergence of parameter estimates for the modified EM/Gibbs sampler: (a) to (f) as in Figure 5.6.

realistic clustering than the ordinary logistic model. This is because the estimates from the logistic model reflect the average response to the habitat-related covariates across the entire region, whereas the auto-logistic approach incorporates local variations by adjusting for the response in neighbouring squares.

Comparison of the modified EM/Gibbs sampler and the logistic model in terms of their ability to estimate the total number of occupied squares is shown in Table 5.7. We generated 120 parametric bootstrap samples using the predicted probabilites from the logistic model and from the algorithm with the modified EM/Gibbs sampler in step 4. We used the

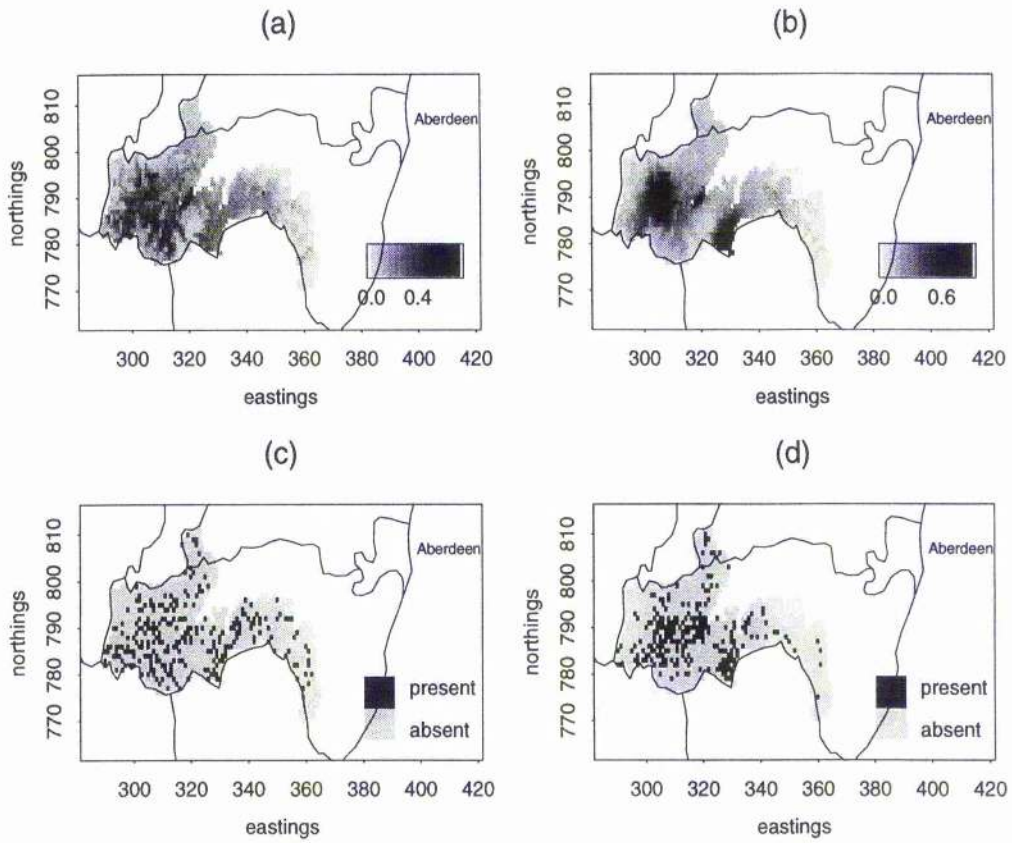


Figure 5.9: Estimated probabilities of deer being present obtained using: (a) the logistic model; (b) the auto-logistic model combined with the modified EM/Gibbs sampler. Simulated presence/absence of deer obtained using: (c) the logistic model; (d) the auto-logistic model combined with the EM/Gibbs sampler.

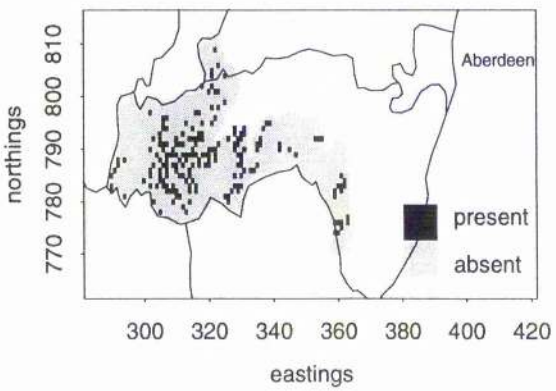


Figure 5.10: Observed spatial distribution of red deer in the intersection of the West Grampian Management Group areas with Grampian Region (excluding Moray District).

Table 5.6: Comparison of mis-classification rates based on 120 stochastic realisations from the final map of fitted probabilities for each method. The matching coefficient is the proportion of squares classified correctly.

Comparison Based on	Logistic Model		Auto-Logistic Model		EM/ Gibbs Sampler		Modified EM/ Gibbs Sampler	
Frequencies of	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
True=1, Predicted=0	107.6	5.4	102.5	5.0	101.0	6.2	99.6	5.1
True=0, Predicted=1	157.8	11.2	148.6	9.9	150.6	15.6	144.7	9.3
Matching Coefficient	0.792		0.803		0.808		0.809	

bootstrap procedure without the improvement presented in section 5.4.3. Computational constraints, to do with programming of loops in S-PLUS, did not allow the application of the improved bootstrap version. Dropping the non-significant spatial covariates from the initial logistic model before implementing the modified EM/Gibbs sampler makes little difference to the standard error of the number of occupied squares. The ordinary logistic model clearly gives more precise estimates of the number of occupied squares than the modified EM/Gibbs sampler (and the same applies to overall abundance since the factor used in scaling up is the same whichever method is used to predict the number of occupied squares). These differences in precision arise because the estimator from the auto-logistic model (with or without the EM/Gibbs sampler) is model-based. It allows for an additional source of variation: it assumes that the observed distribution is a realisation from an underlying super-population in which the locations of clusters are generated by a random process. By contrast, the logistic model is appropriate for quantifying the number of animals actually present at the time of the survey. The logistic model estimator (and the SRS estimator) can be seen as a design-based approach where observed presence/absence values are regarded as a sample from a fixed population and inference for estimation is based on the distribution of estimates generated by the sampling design. Section 7.2 in the general discussion gives more details on these two approaches.

Our comparisons suggest that the auto-logistic modelling approach combined with the modified EM/Gibbs sampler should be used when the main objective of a particular investigation is to map the spatial distribution of a given species but that the ordinary

Table 5.7: Comparison of estimates of the total number of occupied squares: the comparison is based on 120 bootstrap samples for each method; the reduced model just has *altitude*^{2} and *pine* in addition to the autocovariate; corresponding estimates obtained using a simple random sample which ignores covariates are also shown, together with the true value.

Method Used	Estimated Number of Occupied Squares	
	Mean	Standard Error
Logistic Model	241	25.0
Modified Gibbs Sampler with Full Model	237	59.4
Modified Gibbs Sampler with Reduced Model	240	59.5
Simple Random Sample	240	28.0
True Value	190	—

logistic model is to be preferred when priority is to be given to estimating global characteristics of wildlife distributions such as total abundance.

5.5 Conclusions and discussion on auto-models and their applications

Biological understanding

When choosing a spatial model there has to be some understanding of the underlying biological process. For instance as for the auto-logistic model for binary data, the basic auto-Poisson model for counts features an underlying persistence rate for locations with zero counts in their neighourhood. It depends on the biological context whether this is a reasonable assumption. If the counts are from vegetation, then it is reasonable. But when

the counts are from animals it is possibly inappropriate. An example for such a case is a population of mammals on an island where migration can be excluded.

One could claim that auto-correlation only makes sense for stationary processes. For a heterogeneous process things become more complicated; in particular positive auto-correlation cannot be distinguished from trend. As illustrated by the weed seed bank example, what appears as positive auto-correlation is often just caused by an underlying trend. Often the information to model this trend adequately is not available, because data collection on soil, habitat and physical characteristics would be too expensive. Generalized additive models (GAMs) provide a flexible tool to fit two-dimensional surfaces, which should be fairly good for estimating the trend. The combination of GAMs and models with an explicit auto-correlation structure, such as auto-models, deserves further investigation.

Is the auto-Poisson model useful for other non-biological applications?
Although the auto-Poisson model has been applied in other fields such as epidemiology, its features and interpretation have not been investigated. In the application of the basic truncated auto-Poisson model to mortality counts in Spain by Ferrándiz et al. (1995), it is questionable whether the choice of model was appropriate, especially since the data did not have the depth to make inference about "true" auto-correlation. Also, the interpretation of the auto-correlation parameter in such an application is not obvious. In this case, the modified version of the auto-Poisson model or mixed Poisson models (Breslow and Clayton, 1993) could be more adequate. Often, in empirical disease mapping and other epidemiological applications, counts are aggregated and there is not much reason for "true" auto-correlation to be observed, except for infectious diseases.

Correlated parameter estimates
If estimation of the auto-correlation parameters rather than prediction is the main aim, correlated parameter estimates in the auto-Poisson model is a problem. There are a few examples in the literature which exemplify that correlated parameter estimates are a common problem in spatial statistics. Kaiser and Cressie (1997) analyse simulated data using the Winsorized auto-Poisson model using maximum likelihood and obtain a correlation of -0.939, as estimated from the Fisher information matrix, between the estimates of intercept and autocorrelation parameter. There is an example in Geyer (1999) for a spatial model fitted to point process data where the correlation between estimates of trend and autocorrelation is -0.983. Because of such high correlation between estimates of the trend

parameter α_i and the auto-correlation parameter γ, as we demonstrated in section 4.4.2, a different method for constructing confidence intervals should be considered. If one of the parameters can be considered as a nuisance parameter, profile likelihood can be used. For instance consider α_i as a nuisance parameter. Obtain $\hat{\alpha}_i$ while keeping γ at a constant, then use the profile likelihood $l_{z,m}(\hat{\alpha}_i, \gamma)$ for constructing likelihood intervals for γ. If the data shows trend, β from the term $x_i'\beta$ could be treated as nuisance parameters or for a homogeneous process the model could be fitted without an intercept. This and other possibilities for solving the problem deserves further investigation.

Parameter estimation

The simulation study showed that MCML works well for estimating the parameters and standard errors. Also standard errors are asymptotically normal. But MCML is quite computer intensive and a lot more complicated than the MPL method. Problems we encountered with MCML turned out to be caused by a constraint on the upper boundary for the auto-correlation parameter γ in the auto-Poisson model. We have also shown that MPL gives unbiased estimates, but that its asymptotic standard errors are possibly biased. For the auto-logistic model of both MPL and MCML, estimates are consistent, but near the critical value of the autocorrelation parameter γ where phase transition starts, the variability of PLE estimates can be much higher than MCML estimates (Geyer, 1991). This means that, if we are interested in estimating parameters and we expect strong autocorrelation, MPL is not advisable. Also MPL tends to overestimate standard errors when there is negative correlation and underestimate standard errors when correlation is positive in the presence of strong correlation. Standard errors of the coding method always exceed those of both MPL and MCML. Although we only assume models with a first-order neighbourhood structure that lead to two sub-lattices in the coding method, this results in a substantial reduction of sample size ($\frac{n}{2}$) leading to the increased standard errors. With higher order neighbourhoods the standard errors will increase even more. Given that generally the neighbourhood structure is unknown and different neighourhood sizes need to be tested, coding is not a pragmatic option.

The truncated auto-Poisson model has the problem that the usual 95% confidence intervals could exceed the upper boundary of the auto-correlation parameter γ. In study 2 the simulated data had only mild positive auto-correlation ($\gamma = 0.05$), so that data simulation for the MCML method was still possible. It is quite likely that for higher values of γ, close

to the upper boundary, asymptotic normality does not hold and confidence intervals will not be symmetric. In such cases alternative ways for quantification of precision should be found. One possibility is likelihood intervals.

In section 3.4.2 we formulated the temporal truncated auto-Poisson model of which we have not investigated parameter estimation techniques. This deserves further investigation.

What the different application examples showed

The mite example showed that the MCML method works well for moderate levels of positive auto-correlation. Also the results are very similar to results from the Monte Carlo Newton-Raphson method proposed by Lee and Kaiser (1997) and are in comparison substantially less computer intensive (and less complicated).

We have seen in the seed bank example that the auto-Poisson models can be useful to improve model fit, even if the auto-correlation observed is just unmodelled trend.

In cases where only a sample of squares in a lattice are available, auto-models can be combined with the Gibbs sampler as a mechanism for spatial prediction. The deer example showed that this works quite well for the auto-logistic model, which is not as problematic as the auto-Poisson model. There are no constraints on the parameter boundary as in the truncated auto-Poisson case. This makes estimation of parameters much easier.

Is the auto-Poisson model useful for vegetation dynamics?

To summarize, the auto-Poisson has proven to be useful for modelling processes in vegetation dynamics. It is very flexible and with slight modifications can accommodate a variety of processes. It can be applied even in cases with weak positive auto-correlation caused by unmodelled trend and can improve precision. When strong positive autocorrelation is present, parameter estimation becomes difficult, because estimates will be close to the upper boundary of the parameter space. Often strong autocorrelation indicates unmodelled trend, so in theory by adequately modelling the trend, boundary problems could be avoided. If data exhibit more variation than Poisson, the auto-negative binomial model can be used, but its properties do not allow positive auto-correlation to be modelled. Then it is advisable to resort to other methods such as Poisson-Gamma random fields (Wolpert and Ickstadt, 1998).

Chapter 6

Modelling change of semi-natural vegetation in Scotland

6.1 Introduction

Spatial categorical data can arise in ecology when information such as vegetation species or vegetation type are recorded per grid cell on a lattice. Often these data are collected at different time points to study the changes and dynamics in the vegetation system. A first step in the analysis is to look at the proportions of cells that have changed to a certain category, given the category in the previous period. These proportions estimate transition probabilities p_{lk} and define the probability of change from category l to k and the matrix listing all possible p_{lk} is called a transition matrix.

The transition matrix and the categories are the main components of Markov chain models which have been extensively used to model dynamic systems in ecology. Their main feature is that under certain assumptions the long-term expected proportions, e.g. the species composition, can be derived. Usually the assumptions are that a stable state exists, that transition probabilities are stationary over time and that they depend only on a defined number of the immediately preceding states. If the number of preceding states is greater than one, the models are referred to as higher order Markov models and obtaining asymptotic results, such as long-term expected proportions, becomes more complicated.

Unfortunately in ecology, transition probabilities are rarely stationary and past history

influences future changes, e.g. vegetation types recorded in year 10 depend on vegetation types present not only in year 9 but also in preceding years. Also, the transition probabilities differ with location and depend on the states in neighbouring locations. Despite these problems, Markov models have been used in simulation studies to mimic different scenarios, test for equilibrium of the system and to predict future vegetation changes under different management strategies. This is done by applying a range of transition matrices representing different scenarios, for example representing different climatic periods in forest (Lerzman, 1995) and grassland dynamics (Scanlan, 1994). Childress et al. (1998) compared observed vegetation dynamics from a successional plant community (Mount St. Helens, Washington) with patterns predicted from Markov models incorporating different assumptions regarding stationarity. Similar applications include Gibson (1997) and van Groenendael et al. (1996). In most applications transition probabilities are estimated from two-way cross classified tables of empirical data and do not take spatial and other continuous explanatory information into account. Often stationarity of transition matrices is assumed. As a consequence most of the models are not very successful in matching the observed population patterns and have limited predictive value.

We propose to improve estimates of transition probabilities p_{lk}, by estimating them as a function of location, neighbourhood information and other factors at time $t-1$. Such a transition model provides a flexible tool to test different assumptions on the dynamics of a system. The probability of transition from category l at time $t-1$ to category k at time t can be modelled using a multinomial logit model (Agresti, 1990). Although transition (Markov) models for categorical data have been applied to longitudinal data (see Diggle et al., 1994, for a review), not many applications on spatial data exist. We illustrate this approach on Scottish vegetation land classification data. The Institute of Terrestrial Ecology surveyed change of semi-natural vegetation in Scotland using aerial photographs during 1946-1986. Figure 6.1 shows one of the 22 squares photographed. The grid cells, each 5 x 5m, classify the semi-natural vegetation type from aerial photographs taken in 1947, 1967 and 1977. Initially we investigate different effects of explanatory variables on change of vegetation using a logistic regression model. We then use a transition model to test significance of the different explanatory variables and compare the predictive performance of different models. Since little information on management types is available and it is hard to retrieve, an additional aim of the analysis is to detect features characterising management styles.

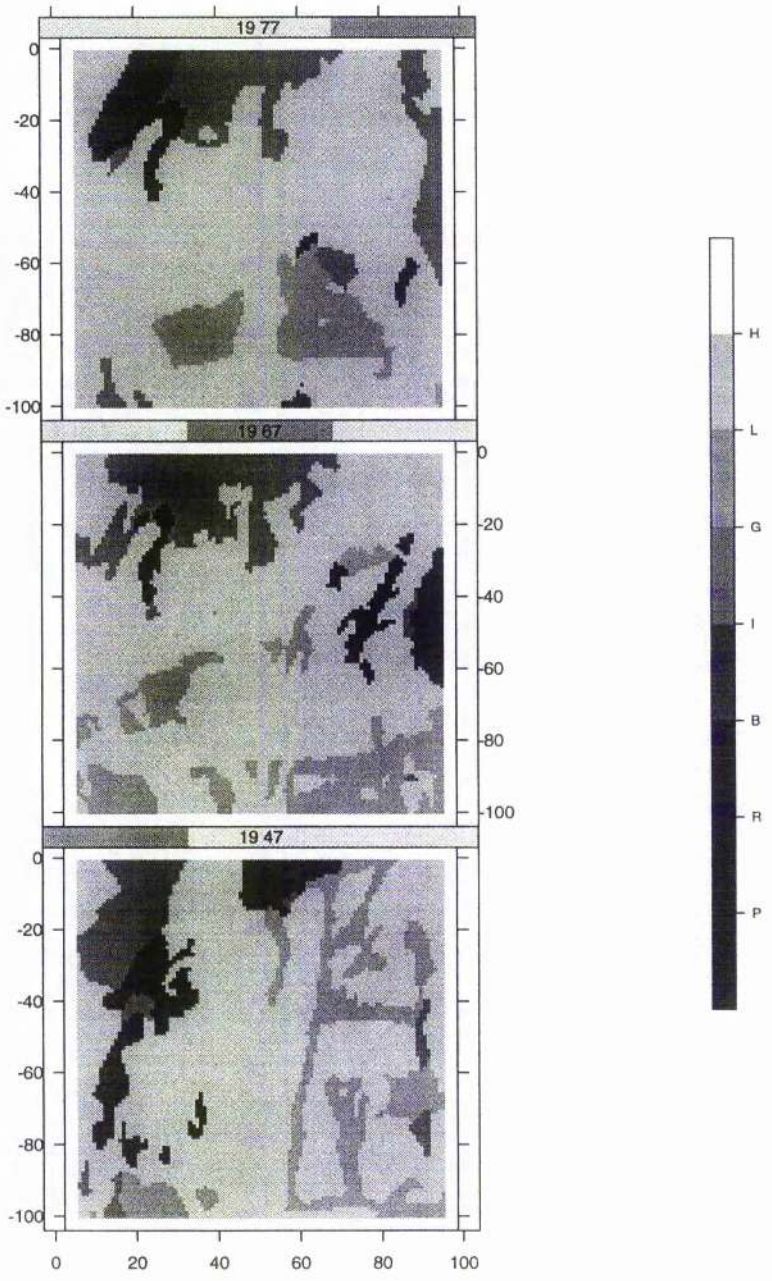


Figure 6.1: Vegetation category of a square of size 0.25 km^2 with semi-natural vegetation near Banchory (Aberdeenshire, Scotland) in the years 1947, 1967 and 1977. Pixels are of size 5 x 5m. Vegetation types are bracken (P), rush (R), bogs (B), graminoids/heath co-dominant (I), graminoids neither enclosed nor improved (G), burned heath (L) and dwarf shrub heath (H).

6.2 Models for change

6.2.1 Possible explanatory variables

Two different models are presented, first a model for the transition probability and second a model for the probability of change. For both models a similar set of explanatory variables can be used. These variables were agreed on with the collaborating ecologists as possible influential factors on transition.

1. *neighbour*: The number of neighbouring pixels with the same vegetation category. Here the neighbourhood consists of the 8 nearest neighbours, but the neighbourhood size could potentially differ between categories. For example neighbour effects for trees are likely to be more long range than for grass, where only adjacent pixels will have a strong influence.

2. *centrality score*: This score is a measure of the position of a pixel within a polygon, which is high if the pixel is in the centre of a polygon. Score 1 corresponds to a *single* pixel, score 2 corresponds to an *edge* pixel, score 3 corresponds to a pixel where the eight nearest neighbours are the same category (*c1*), score 4 is the equivalent for the 24 nearest neighbours (*c2*) and score 5 is a pixel with at least 48 nearest neighbours from the same category (*c3*).

3. *location*: The location within the square, such as *northing* and *easting*, might be a good proxy for other unmeasured variables such as soil type and slope.

4. *edge-ratio*: The ratio of the number of edge pixels over the number of pixels in the polygon, thus characterising the shape and size of the polygon. A small polygon has an *edge-ratio* close to one, a large polygon has an *edge-ratio* close to zero, unless it is a large polygon of very irregular shape.

5. $change_{it-1}$: Change of vegetation category in the previous period time point $t-2$ to $t-1$ on pixel i, where time t refers to the period end point.

6. *management level:* There can be ownership boundaries within a square and a factor can be used to indicate different management types depending on the ownership of the pixel.

6.2.2 Transition/Markov model

In our data example several changes in vegetation could have taken place between each observation. Therefore, for this model, it is assumed that succession of vegetation category does not tend to happen in any particular order and that the categories are strictly nominal. Define the response $\boldsymbol{y}_t$ at time t with $y_{it} = m$ for pixel i and t is the period end point, where m is one of k categories and $i = 1, ..., n$. Then the response $\boldsymbol{y}_t$ is modelled as a realisation at time t of a first-order discrete time Markov chain. For first-order Markov chains, the joint probability mass function has the form (Agresti, 1990)

$$Pr(\boldsymbol{y}_0, ..., \boldsymbol{y}_T) = Pr(\boldsymbol{y}_0)Pr(\boldsymbol{y}_1|\boldsymbol{y}_0)Pr(\boldsymbol{y}_2|\boldsymbol{y}_1)....Pr(\boldsymbol{y}_t|\boldsymbol{y}_{t-1})....Pr(\boldsymbol{y}_T|\boldsymbol{y}_{T-1}),$$

where T is the last observed time point. Define the transition probability as the probability of change from category m at time $t-1$ to category l at time t:

$$p_{ilm}(t) = Pr(y_{it} = l|y_{it-1} = m)$$

with $\sum_{l=1}^{k} p_{ilm}(t) = 1$. We assume a multinomial distribution for the response y_{it} with class probabilities $p_{ilm}(t)$ and a sample size of one. Then assuming that the y_{it} are independent with respect to i, the above likelihood takes the form

$$\prod_{i=1}^{n}\left(\prod_{l=1}^{k} p_{il}\right)\left(\prod_{t=1}^{T}\prod_{m=1}^{k}\prod_{l=1}^{k} p_{ilm}(t)\right) \tag{6.1}$$

where p_{il} is the probability for pixel i to be in category l at time 0. When explanatory information is not taken into account, the maximum likelihood estimates of $p_{ilm}(t)$ is the proportion of pixels which changed to category l, given they were category m at time $t-1$. In the case where explanatory information is categorical or can be summarised into categories, the transition probabilities can also be estimated using a log-linear model with standard GLM software. This has the advantage of providing variance estimates for parameters. If we want to estimate $p_{ilm}(t)$ as a function of continuous explanatory variables, such as *easting* and *northing*, the transition probability $p_{ilm}(t)$ can be estimated using a multinomial logit model (Agresti, 1990). The multinomial logit model is a generalised linear model (GLM) and parameters can be estimated using GLM software catering for the multinomial distribution. The model has the form:

$$p_{ilm} = Pr(y_{it} = l|y_{it-1} = m) = \frac{\exp(x'_{it-1}\beta_l)}{\sum_{s=1}^{k}\exp(x'_{it-1}\beta_s)}, \tag{6.2}$$

with the linear predictor $\eta_{lm} = x'_{it-1}\beta_l$ and the parameter vectors β_l and β_s relate to categories l and s respectively. The coefficient β_1 is constrained to be zero for identifiability.

The vector x_{it-1} is a row of the design matrix containing a factor for y_{it-1} and explanatory variables at time $t-1$ and interactions. Note that x_{it-1} containing the factor y_{it-1} is the minimal model fitted to estimate the transition probabilities. Estimating the above parameters amounts to maximising the right hand side term of the likelihood 6.1, which is a conditional likelihood. This conditional multinomial likelihood is maximised subject to the constraint of equation 6.2. For large sample size parameter estimates are normally distributed and asymptotic standard errors are obtained from the inverse Fisher information matrix (Agresti, 1990). The Newton-Raphson algorithm can be used for maximisation, but we have found that it can be slow to converge when sparse data are combined with fitting models with continuous explanatory variables. Much better results were achieved by maximising the likelihood via neural networks using the S-PLUS function *multinom*() provided by Venables and Ripley (1994) in their function library.

Having fitted the transition model, the future vegetation cover can be predicted, assuming that grazing pressure and other imposed management features remains constant. Summing over the predicted probabilities of transition into categories $1, ..., k$ gives the predicted species compositions in $t+1$.

6.2.3 Modelling change using logistic regression

Instead of transition probabilities, the probability of change can be modelled. The response is then defined as $y_{it} = 1$ for cells changed between time $t-1$ and t or $y_{it} = 0$ otherwise. The response y_{it} has a Bernoulli(p_{it}) distribution and the logistic model has the form

$$\text{logit}(Pr(y_{it} = 1)) = \log\left(\frac{Pr(y_{it} = 1)}{1 - Pr(y_{it} = 1)}\right) = x'_{it-1}\beta, \tag{6.3}$$

where the index i is the pixel i, $i = 1, ..., n$ and t is the sampling time point t. The parameter vector β contains the coefficients corresponding to x_{it-1}, a row of the design matrix containing explanatory variables recorded at time $t-1$ and their interactions. If we include the factor y_{it-1} in the model, we are dealing with a transition or Markov chain model as defined in equation 6.1, and the parameter estimation amounts to maximising the conditional likelihood (see Diggle et al., 1994, for details).

The above model assumes that the responses y_{it} are mutually independent given the information at time $t-1$. Standard GLM software for logistic regression can be used to fit the model.

6.2.4 Model selection and validation

Forward stepwise selection, based on the AIC statistic, was used for model selection of the logistic regression model in section 6.2.3.

For the multinomial logit model prediction is the objective and K-fold cross-validation (Davison and Hinkley, 1997) can be used to assess the aggregate prediction error of the different models. The usual leave-one-out cross-validation procedure removes a single observation at a time and the model is fitted to the remaining data set, called the training set. In K-fold cross-validation, whole groups of observations, here pixels, are omitted. The omitted observations constitute the assessment set and the model fitted to the training set is used to predict observations in the assessment set.

Omitting pixels can alter the values of explanatory variables *neighbour* and *edge-ratio* of the remaining pixels in the training set. Thus the way the assessment set is chosen matters. A reasonable strategy is to omit a group of adjacent pixels rather than a sample of non-adjacent pixels, because there will be fewer changes in explanatory variables of the training set. A large number of alterations in explanatory variables could possibly have an adverse effect on estimation of the prediction error.

The procedure is as follows: The n pixels are divided into a minimum of $\sqrt{n}$ disjoint groups of pixels. These groups define $\sqrt{n}$ splits of the data into assessment and training sets. In our example the lattice consists of $n = 100$ x 100 pixels and each assessment set contains a block of 10 x 10 adjacent pixels, i.e. the 100 x 100 grid is divided into a 10 x 10 grid, with each grid square of size 10 x 10 yielding 100 different assessment sets. In each split the group of $\sqrt{n}$ adjacent pixels is predicted once, leading to a prediction value for each pixel. Then the *K-fold cross-validation estimate of mis-classification* is used to assess the prediction error for the multinomial response:

$$\hat{C}_{mis} = \frac{1}{n}\sum_{i}^{n}(1 - \hat{p}_{ilk}(t)),$$

where $y_{it} = l$ and $y_{it-1} = k$. The mis-classification measure is the categorical version of the commonly used prediction error for other types of data.

6.3 Example: Institute of Terrestrial Ecology survey on change in habitat and vegetation during 1946-1986

6.3.1 Data description and exploration

Each of 22 squares of size ranging between $0.25 - 1\text{km}^2$ in the Grampian region in the north-east of Scotland were photographed three times within a period of 40 years. There are 10 to 20 years between photographs which ranged in scale from 1:7500 to 1:25000. The same person (Roger Cummins) mapped and classified the vegetation seen on the photographs for all three time points. The time periods were chosen so that a change of vegetation could be noticed from aerial photographs. The maps were then converted to a grid format, with each pixel (cell) equivalent to 5 x 5m on the ground, using procedures given in Green et al. (1993). The vegetation category for each pixel was ascribed to its components using six variables:

1. Overall vegetation category;
2. Percentage coverage of predominant vegetation category: > 50%,
3. 26 - 50% coverage,
4. 10 - 26% coverage,
5. < 10% coverage;
6. Mosaic: used where the predominant vegetation category within a polygon occurs as discrete patches rather than as an intimate mixture.

The overall vegetation category analysed here contains the following codes:

- *graminoid (grass-like) dominant vegetation* such as: agricultural (A), enclosed rough grazings (E), graminoids neither enclosed nor improved (G), blanket bog-type vegetation (Y), bogs (B);
- *heath dominant vegetation* such as: dwarf shrub heath (H), graminoids/heath co-dominant (I), heath burned within 4 years of the photograph being taken (L) and burned blanket bog (Z);

- *tall woody vegetation* such as: shrub (S), juniper (J), plantation (T), birch (1), rowan (2), pine (3), mixed coniferous (4), undefined coniferous (5), undefined deciduous (6) and willow (7);
- *bracken dominated* (P);
- *tall rush or reeds* (R);
- other cover types such as: bare (roads, metalled tracks, rocks, etc.) (K), bare but readily colonised (e.g. minor soils) (M) and open water (X).

We use one of the 22 photographed squares as an example for demonstrating the possibilities of exploratory data analysis and modelling change and transition probabilities in this study. This square, situated on a hill near Banchory, is of size 0.25 km^2. It has no pixels classified as heavily managed vegetation (agricultural and plantation) and also has fewer vegetation categories than other squares. Some of the questions related to these data are:

- Do the transition probabilities (the probability that a pixel changes from one category to another) depend on polygon size and location within a polygon?
- Do stable areas remain stable? Are some vegetation categories stable and others not?
- Does the proximity of certain vegetation categories have an influence on the transition at a pixel?
- Are there obvious features characterising certain management types?

The exploratory data analysis should answer some of these questions informally, and lead to suitable models to test different assumptions.

Seven different vegetation categories were observed at the three sampling time points, 1947, 1967 and 1977 (see also Figure 6.1). The pixels were mainly graminoid dominant (B, G) and heath dominant vegetation (H, I, L), although there were a few pixels with bracken (P) and rush (R) (Figure 6.2). Most of the area was covered by dwarf shrub heath (H) in all three years. In 1947 and 1967, the second highest coverage was burned heath (L), but in 1977 it was graminoids (G). The transition probabilities in Table 6.1 are estimated as the proportion of pixels which changed to category l, given they were

category m at time $t-1$. These transition probabilities and other summary statistics help us to describe certain features:

Stability: The square is mostly covered with dwarf shrub heath and more than 70% of pixels remain in this category during both periods. More than half of the pixels with vegetation category P, R, B, L and I change in both periods. The grass/heather co-dominant vegetation category (I) is notably unstable, with only one pixel being so recorded in consecutive photographs, and is clearly a successional stage in transitions to and from H and G.

Grazing pressure: In the first period, 62% of the pixels with graminoid/ heath co-dominant vegetation (I) change into dwarf shrub heath (H) and 29% change into graminoid (G). This is reversed in the second period, where only 25% change into heath (H) and 74% change into graminoid. This may imply that grazing pressure had increased. Due to the small number of I pixels, we need to be cautious not to over-interpret these results. In 1967 there were only 265 (3%) pixels classified as I, divided into two distinct polygons and in 1947 there were 803 pixels (8%), split between 4 polygons.

Rotation cycle for burning heather: Vegetation category 'L' is defined as heathland that was burned within 4 years of the photograph being taken. 136 (9 %) of pixels that were thus classified in 1967 were again classified as 'L' in 1977. This shows that some areas are burned again after 6-14 years. However, this is a relatively short burning cycle for a deliberate management policy and it may be that the areas were burnt accidentally.

Liability to repeated change: Table 6.2 shows that a pixel that changed in the first period is 1.5 times more likely to change in the second period than a pixel which did not change in the first time period.

An important feature for characterising the square is *polygon size*. As Figure 6.1 shows, it increases over time. Figure 6.3 shows centrality score by year, indicating that the number of pixels within a large polygon ($c3$ = surrounded by pixels of the same category within a minimum of 45 x 45m) increases over the years. The transition probabilities of vegetation by centrality score (Table 6.3) for both periods show that the percentage of pixels which are graminoid (G) in both periods is highest for pixels within large polygons ($c3$). This

Table 6.1: Transitions in 1947 - 1967 and 1967 - 1977; $\hat{p}_{lk}$ corresponds to the probability of transition from category k in year $t-1$ to category l in year t.

vegetation	vegetation category in 67 (l)						
category in 1947 (k)	R	B	I	G	L	H	total
P	0	0	0	0	0	45	45
$\hat{p}_{lk}$	0	0	0	0	0	1	
R	0	0	0	6	0	10	16
$\hat{p}_{lk}$	0	0	0	0.37	0	0.63	
B	0	0	0	0	0	29	29
$\hat{p}_{lk}$	0	0	0	0	0	1	
I	64	0	0	233	10	496	803
$\hat{p}_{lk}$	0.08	0	0	0.29	0.01	0.62	
G	38	0	50	373	56	290	807
$\hat{p}_{lk}$	0.05	0	0.06	0.46	0.07	0.36	
L	9	28	34	26	377	1140	1614
$\hat{p}_{lk}$	0.01	0.02	0.02	0.01	0.23	0.71	
H	34	152	181	577	1039	4703	6686
$\hat{p}_{lk}$	0.01	0.02	0.03	0.09	0.16	0.70	
total	145	180	265	1215	1482	6713	10000

category in 1967	vegetation category in 1977						
R	0	75	6	2	0	62	145
$\hat{p}_{lk}$	0	0.52	0.04	0.01	0	0.43	
B	0	0	0	3	0	177	180
$\hat{p}_{lk}$	0	0	0	0.02	0	0.98	
I	0	0	1	197	0	67	265
$\hat{p}_{lk}$	0	0	0.01	0.74	0	0.25	
G	0	72	251	593	0	299	1215
$\hat{p}_{lk}$	0	0.06	0.21	0.49	0	0.24	
L	0	10	23	67	136	1246	1482
$\hat{p}_{lk}$	0	0.01	0.02	0.05	0.09	0.84	
H	0	60	130	411	826	5286	6713
$\hat{p}_{lk}$	0	0.02	0.02	0.06	0.12	0.78	
total	0	217	411	1273	962	7137	10000

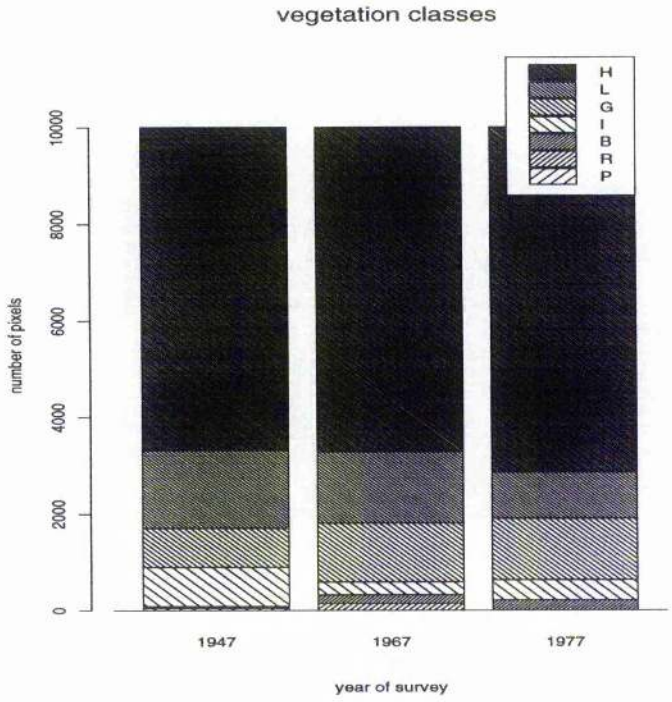


Figure 6.2: Frequency (number of pixels) of vegetation category by year. The category 'rush' (R) was collapsed with category 'bogs' (B).

Table 6.2: Change 1947-67 by change 1967-77.

period	period 1967 - 77	
1947 - 67	change	no change
change	3864	1589
	0.71	*0.29*
no change	2152	2395
	0.47	*0.53*

effect is not discernible for the other categories. In particular, there are too few RR and BB pixels to make any inferences about the effect of polygon size.

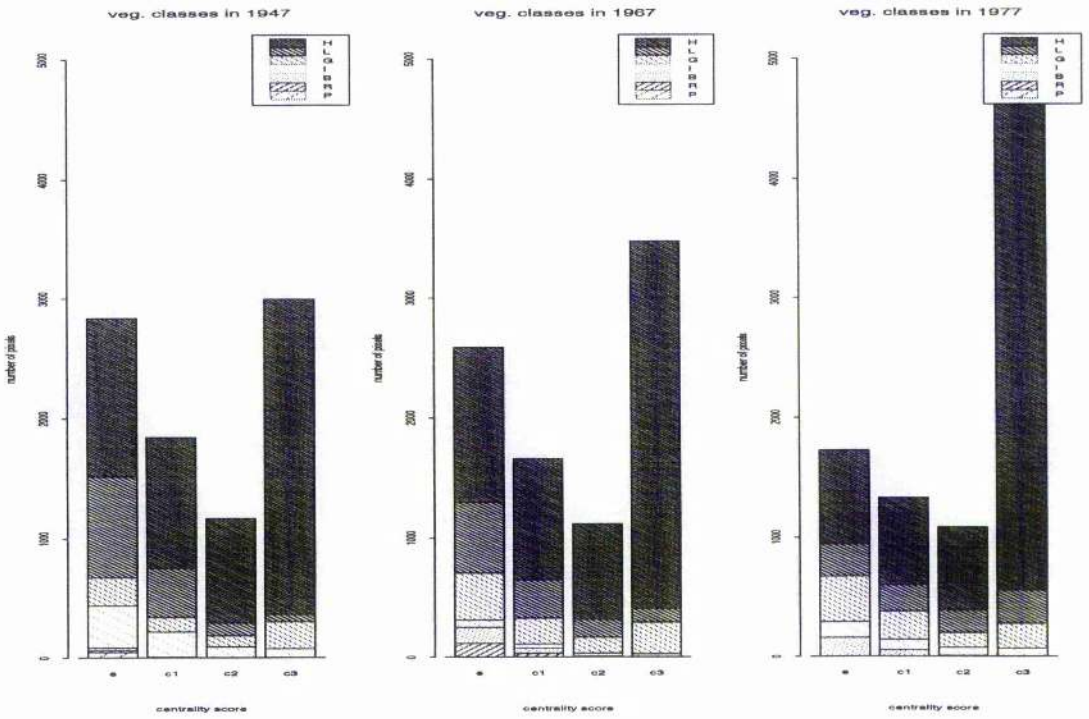


Figure 6.3: Frequency (number of pixels) of vegetation category by year and centrality score. The category 'rush' (R) was collapsed with category 'bogs' (B).

Table 6.3: Percentage of pixels which remained in the same vegetation category.

centrality score	transition 1947-67 (%)						
	PP	RR	BB	II	GG	LL	HH
edge	-	0	0	0	12	19	73
c1	-	-	0	0	30	21	73
c2	-	-	-	0	60	13	70
c3	-	-	-	0	75	31	73
centrality	transition 1967-77 (%)						
	PP	RR	BB	II	GG	LL	HH
edge	-	-	0	2	29	13	75
c1	-	-	0	0	44	8	81
c2	-	-	0	0	57	9	84
c3	-	-	-	0	71	4	76

6.3.2 Model for change in the two periods 1947 - 1967 and 1967 - 1977

A possible maximal model for the probability of change, fitted to each period separately is,

$$\text{logit}(Pr(y_{it} = 1)) = \log\left(\frac{Pr(y_{it} = 1)}{1 - Pr(y_{it} = 1)}\right) = x'_{it-1}\beta \tag{6.4}$$

where the index i represents the pixel i, $i = 1, ..., 10000$ and t is the period end time point, with $t = 1967, 1977$ and periods 1947-1967 and 1967-1977. The design matrix x_{it-1} of the maximum model contains the following explanatory variables recorded at time $t-1$ and their interactions: $neighbour_{it-1}$, $centrality\ score_{it-1}$, $edge\text{-}ratio_{it-1}$, $change_{it-1}$ (only applicable for period 1967 - 1977) and veg_{it-1}, the category at time $t-1$. In this model the effects of the variables *neighbour* and *edge-ratio* were of specific interest and therefore locational variables (*easting* and *northing*) were not fitted.

There were only 16 pixels with category 'rush' (R) in 1947 and 145 pixels in 1967. These small numbers of pixels were not sufficient to make inferences about the interactions between explanatory variables and category at time $t-1$. To avoid this problem we collapsed the category R with category 'bogs' (B). At time t, i.e. at the end of period 2, burned heather (L) is combined with heather (H), because burning heath is imposed management and cannot be predicted with a model of this type. The small sample size did not enable us to distinguish between the effects of $change_{it-1}$ and veg_{it-1}. All pixels with category B and I changed in the second period, thus the effect of $change_{it-1}$ was aliased with veg_{it-1}. Therefore we abandoned the inclusion of $change_{it-1}$ into the set of explanatory variables and could not infer how previous change of a pixel affects its future change.

Initially we fitted a model to change in the second period and forward stepwise selection yielded the model

$$\text{logit}(Pr(y_{it} = 1)) = veg_{it-1} + \beta\ neighbour_{it-1} + veg_{it-1}.\beta\ neighbour_{it-1}, \tag{6.5}$$

where $t = 1977$. The change of deviance when adding the variable *neighbour* is larger than the change in deviance when adding variables *edge-ratio* and *centrality score*. The change in deviance when adding the explanatory variables is 3587 on 9 d.f.. Although forward stepwise selection yielded a model with more terms for the first period, for comparability we also fitted the model of equation 6.5 to change in the first period, yielding a change in deviance of 2904 on 11 d.f..

Figure 6.4 shows the fitted probability of change versus *neighbour* (the number of neighbours with the same category at time $t-1$ as pixel i) for the two periods. For both periods vegetation category H has the lowest probability of change, and it is not very much affected by *neighbour*. In period 2, categories B and G have a much higher probability of change which decreases with the number of neighbours of the same category. For both periods *neighbour* has a similar effect on the two main categories H and G. Figure 6.4 also shows 95% pointwise confidence bars of estimated probabilities of change. In period 1 not much information is available to model change of pixels that have category P or B in 1947. The confidence intervals for these categories are therefore very large (from zero to one), and are not shown on Figure 6.4a. Comparing observed probabilities with fitted probabilities of period 2 in Figure 6.5 shows a fairly good fit, although there are two large polygons of change in the lower half of the square (Figure 6.5a), which the model does not pick up. These changes are due to the occurrence of burned heath (L) within a bigger dwarf shrub heath (H) polygon in 1977. Burning heath is imposed management and therefore unpredictable with a model of our type. There is also an edge effect visible on the top edge of Figure 6.5 due to the *neighbour* variable. This happens because the the edge of the lattice is defined as a polygon edge and hence the number of neighbours of the same category at the lattice edge is the same as at a polygon edge.

6.3.3 Transition/Markov models fitted to the transitions from 1967 to 1977

Due to low frequencies of some transitions for different values of explanatory variables, we reduced the number of vegetation categories by combining R (rush) with B (bogs). For the period end time point t, category L (burned heath) was combined with H (dwarf shrub heath) as for the change model. Explanatory variables *neighbour, edge-ratio* and *centrality score* were derived from the uncollapsed data set.

Fitting a model to transitions between 1967 and 1977 indicates that all of the explanatory variables have an effect on transition probability. The effect of *neighbour* was stronger than the effect of *centrality score*, and centrality score was therefore not included in the cross-validation model selection exercise. In Table 6.4 we give deviance, AIC statistic and K-fold estimated prediction mis-classification rate of different fitted models. Model 7, including the explanatory variables *northing, easting* and *neighbour*, has the lowest K-fold

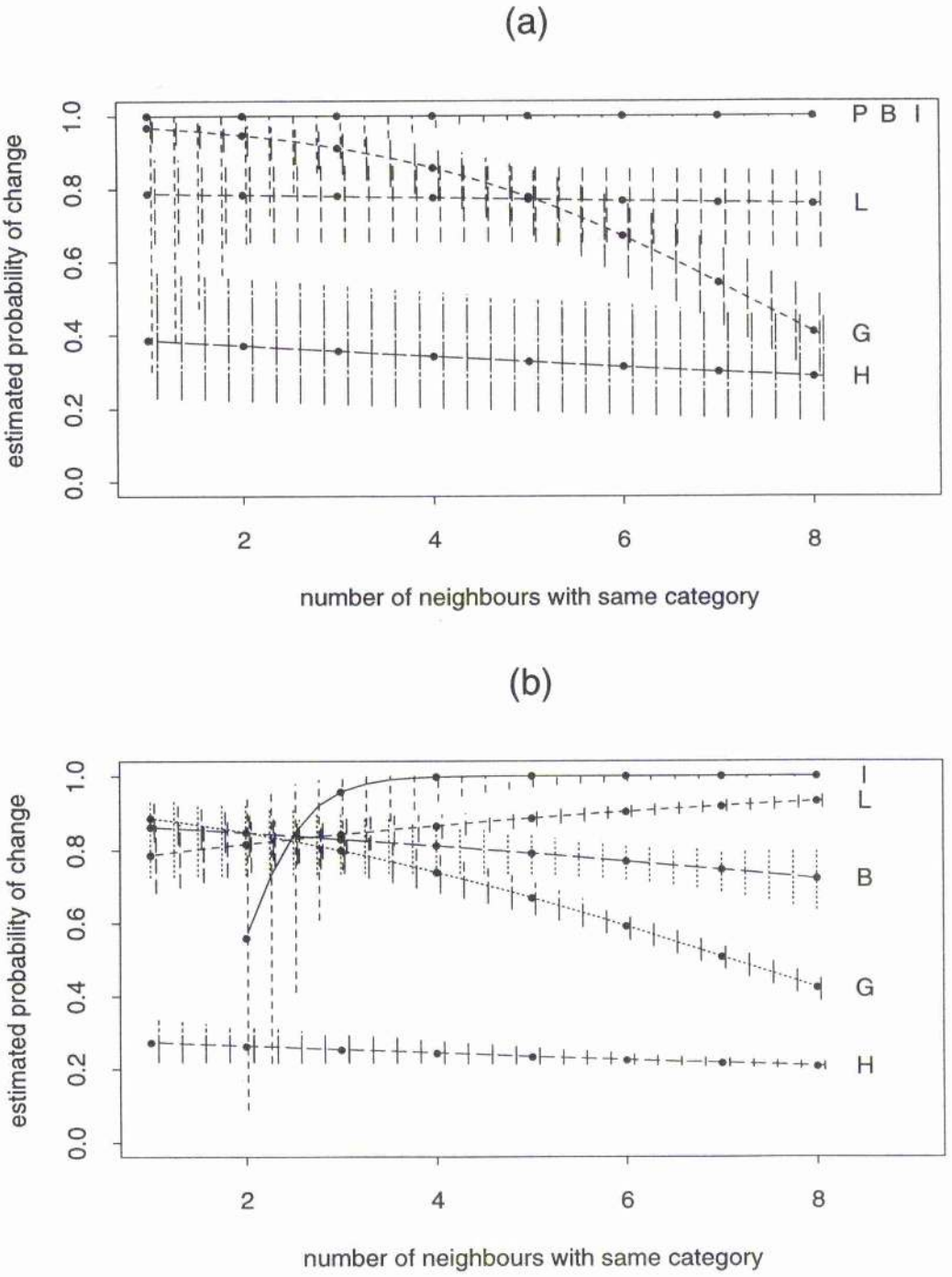
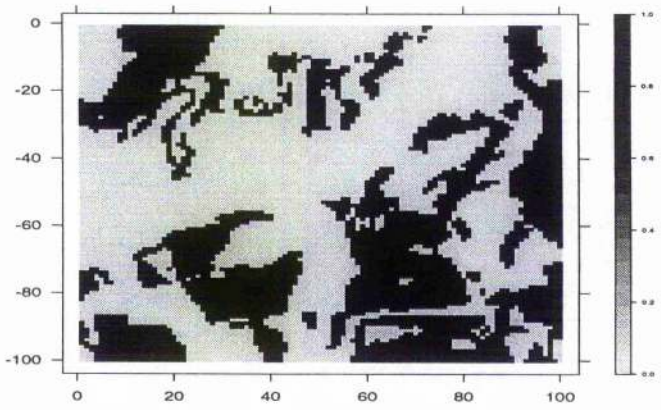
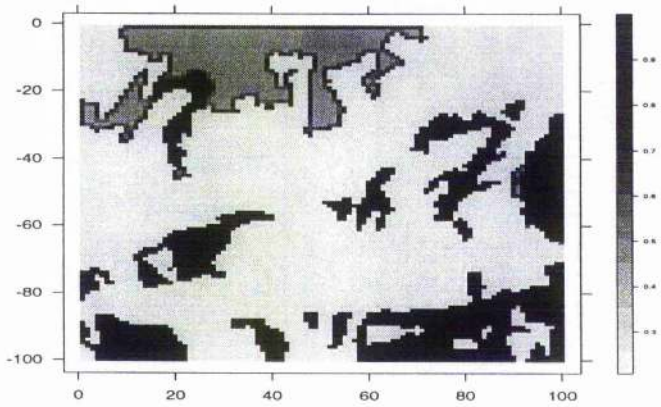


Figure 6.4: Estimated probability of change by *neighbour*. The vertical bars join jittered upper and lower 95% confidence intervals. Estimated probabilities are given for two periods (a) 1947 - 1967 for pixels with categories P, B, I, L, G and H in 1947 (confidence intervals for categories P and B were omitted) and (b) 1967 - 1977 for pixels with categories B, I, L, G and H in 1967. For both periods category 'rush' R is combined with category 'bogs' (B) and at period end time point t burned heather (L) is combined with heather (H).



(a)



(b)

Figure 6.5: (a) Observed change and (b) estimated probability of change during the period 1967 and 1977.

Table 6.4: Transition models fitted to data of the period 1967 ($t-1$) to 1977 (t). The *K-fold cross-validation estimate of mis-classified* pixels out of 10000 is given in the last column ($\hat{C}_{mis}$). The symbol '*' stands for the inclusion of the explanatory variables separately and their interactions.

explanatory variables in x_{it-1} of η_{lm} of equation 6.2	d.f.	residual deviance	AIC	$\hat{C}_{mis}$
1. y_{it-1}	9988	9598.3	9622.3	2389
2. $y_{it-1} * neighbour_{it-1}$	9976	9316.5	9364.5	2338
3. $y_{it-1} * edge - ratio_{it-1}$	9976	9284.0	9332.0	2395
4. $y_{it-1} * change_{it-1}$	9976	8483.8	8519.8	3198
5. $y_{it-1} * row_i * col_i$	9976	7346.7	7442.7	1987
6. $y_{it-1} * neighbour_{it-1} * edge - ratio_{it-1}$	9952	8989.5	9085.5	2336
7. $y_{it-1} * neighbour_{it-1} * row_i * col_i$	9952	6926.2	7118.2	1908

estimated prediction mis-classification rate. 1908 pixels out of 10000 are mis-classified compared to 2389 for model 1, which is just based on the 2-way contingency table. This reduction in mis-classification (4%) shows that there are strong spatial effects present. It is surprising to see that model 4 with the explanatory variable $change_{it-1}$ performs worst in terms of mis-classification, probably due to the fact that $change_{it-1}$ is aliased with y_{it-1}. All of the pixels in category B or I changed in the previous period. Having fitted the transition model to the period between 1967 and 1977, we can use it to predict change in vegetation cover for the period 1977 to 1987, assuming that grazing pressure and other imposed management features remain constant. Table 6.5 gives the predicted species composition for the best (model 7) and second worst model (model 1). It would be ideal to compare predictions with observed data, but these are not available for 1987. In comparison to the observed species composition in 1977, model 7 indicates 3.38% increase in H (heath), whereas for model 1 the percentage of H is predicted to decrease by 1.15%.

The predicted transition probabilities for period 1977 to 1987 from model 2 are shown as a function of *neighbour* in Figure 6.6. Depending on the category at the period start time point $t-1$ (1977), the effects of the *neighbour* can be very different. For example the effect of *neighbour* on transition G to H is strong, but it is not very strong on the transition H to H and B to H.

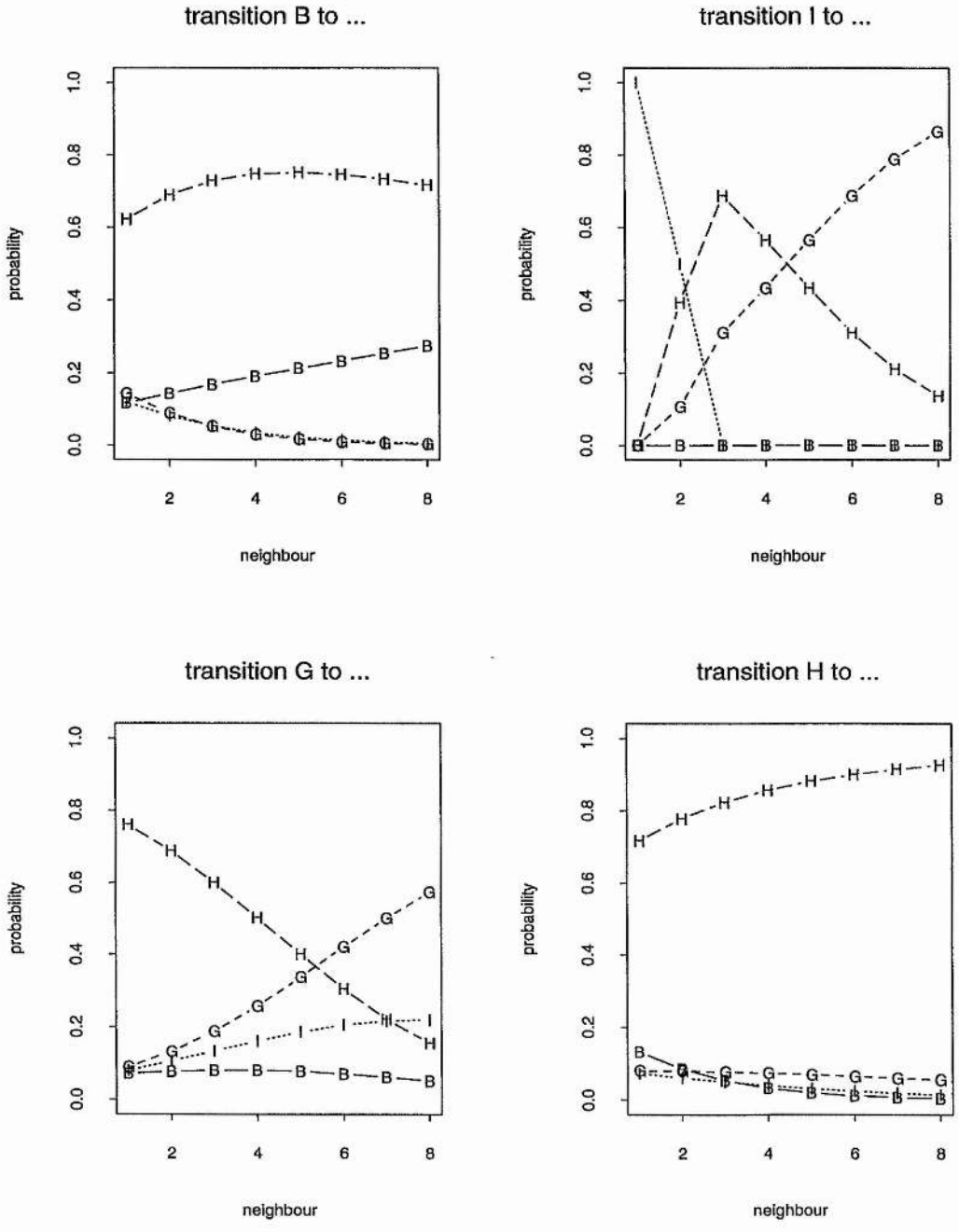


Figure 6.6: Predicted transition probabilities for the period 1977 to 1987 as a function of *neighbour* using model 2.

Table 6.5: Predicted species composition of square 45 for 1987 using different transition models fitted to data for the period 1967 to 1977. The first row gives the observed species composition for the year 1977 (category L collapsed with H).

observed species composition in 1977 (%)				
	B	I	G	H
	2.17	4.11	12.73	80.99
predicted species composition in 1987 (%)				
model	B	I	G	H
1. y_{it-1}	1.95	4.20	14.03	79.84
7. $y_{it-1} * neighbour_{it-1} * row_i * col_i$	2.58	2.20	10.84	84.37

Figure 6.7 shows predicted transition probabilities for the period 1977 to 1987 using model 7. A stochastic realisation of predicted vegetation category in 1987 was simulated in Figure 6.8 by sampling at each pixel from a multinomial distribution with the predicted transition probabilities from Figure 6.7. The many single pixels the simulated image shows are atypical for this kind of vegetation. Also, the observed data in Figure 6.1 has very few single pixels. Thus there is clearly an element of spatial autocorrelation missing in model 7 and improvement in that respect is needed.

6.4 Discussion and further work

We have modelled the land-classification data on the pixel level using regression techniques. Another approach is to model the data on the polygon level, but here the polygons do not change as a unit and it is therefore difficult to see how a polygon level model could be implemented. This requires further investigation.

Our limited investigation has shown that in the case of the Scottish land classification data there are strong spatial effects on the transition of vegetation. Fitting a multinomial logit model to transition probability showed that location (*eastings, northings*), the number of neighbouring pixels from the same category (*neighbour*) and the size and shape of polygon (*edge-ratio*) at time $t-1$ have a strong influence on transition probability. For prediction best results are achieved with a model including *eastings, northings* and *neighbour*. Our model attempts to describe the dynamic process, rather than imposed management. If

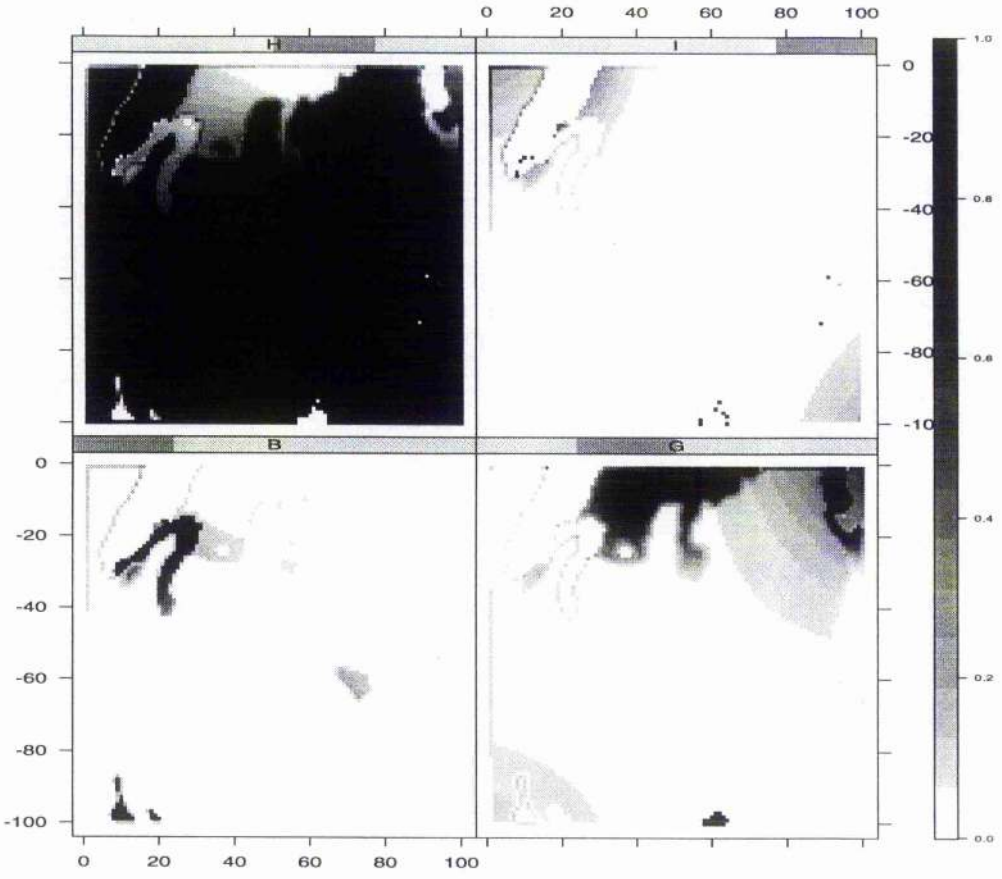


Figure 6.7: Predicted probabilities of transition to the four categories H (top left), I (top right), B (bottom left) and G (bottom right) for the period 1977 to 1987 using model 7.

management is imposed, such as burning heather, this results in bad model fit, as we have seen in section 6.3.2.

If in a long-term study we can measure grazing pressure in addition to recording vegetation cover at several points in time, then transition probabilities can be estimated under a range of grazing pressures. Preferably the square size should be bigger than 0.25 km^2 if the polygon sizes are similar to our example, where on the polygon level the sample size was quite small. This allows prediction of future changes under a range of grazing scenarios, which could prove an invaluable aid for conservation management. In theory we can model several time periods simultaneously rather than merely fit separate models for each time period. In such a model, the responses of the two time points 1967 and 1977 are combined into one variable $\boldsymbol{y}_t$ and the responses of 1947 and 1967 are combined into

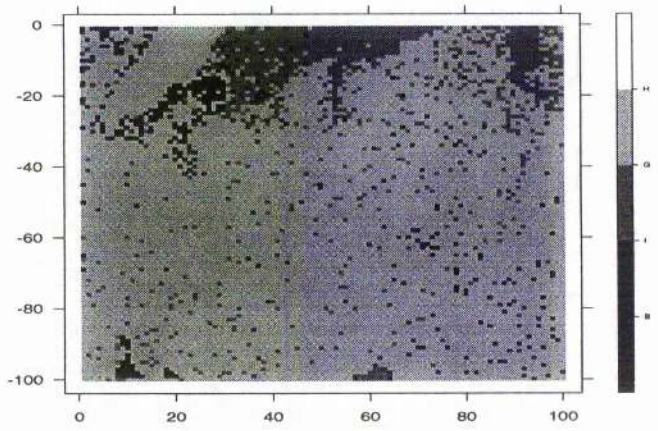


Figure 6.8: A stochastic realisation of predicted vegetation category in 1987. At each pixel one of the four categories B, I, G and H was sampled using a multinomial distribution with the probabilities given in Figure 6.7. Vegetation categories are bogs (B), graminoids/heath co-dominant (I), graminoids neither enclosed nor improved (G) and dwarf shrub heath (H).

another variable $\boldsymbol{y}_{t-1}$. A factor for grazing pressure or period is included in the model. By fitting such a model, forcing $\boldsymbol{y}_{t-1}$ to interact with all other variables, we would get similar results to the two separate models fitted to the two periods.

So far the issue of spatial autocorrelation of transitions has not been addressed and incorporating this into the model could potentially improve estimates of transition probabilities. Ignoring strong positive autocorrelation can lead to underestimating variance. Both, the model for change and the transition model assume that the responses y_{it} are mutually independent given the information at time $t-1$. Ecologically this is an unlikely scenario and estimates of transition probabilities could potentially be improved by incorporating spatial autocorrelation into models. One possible way is to include so-called 'autocovariates' as described in chapter 3. For the change model, the 'autocovariate' would be defined as the number of neighbours which change, $\sum_{j \in N_i} y_{jt}$. The resulting model is called an *auto-logistic* model and parameter estimation methods for such methods were discussed

in chapter 4.3.

Another possible way of accounting for autocorrelation would be to introduce random effects into the multinomial model. For instance a random effect for polygon in year $t-1$ could be introduced. Alternatively the random effect could follow a Gaussian random field, which could have different parameters and neighbourhood structures depending on the response category y_{it}. The model could be fitted within a Bayesian framework and would have the following form

$$\eta_{ilk} = x'_{it-1}\beta_l + r_i$$

where η_{ilk} is the linear predictor of the multinomial model as defined in equation 6.2, and the random effect r_i has distribution

$$r_i \sim N((\mu_i|\mu_j; j \in N_i), \sigma^2)$$

where

$$\mu_i = \gamma \sum_{j \in N_i} r_j.$$

Similar to the above is the random-effects Poisson model first introduced by Besag et al. (1991), which has mainly found applications in epidemiology for disease mapping (Breslow and Clayton, 1993).

The improved estimates of transition probability could feed into Markov models used in simulation studies testing different management strategies, and improve their predictive value.

Chapter 7

General discussion

In this chapter we discuss briefly two general points which are relevant to all the work described in this thesis: Firstly allowing for trend and autocorrelation in the same model; and secondly the inference implications of using a design-based approach rather than a model-based approach.

7.1 Distinguishing trend and autocorrelation

One of the biggest problems encountered is that trend and positive autocorrelation cannot easily be distinguished. In a lot of environmental applications, apparent unmodelled autocorrelation is due to bad model fit, rather than true intrinsic autocorrelation from competitive or enhancing interactions between spatial neighbours. In order to validate the model, spatial autocorrelation in residuals needs to be investigated. We have used descriptive tools, such as the investigation of correlograms of residuals or the visual investigation of residual maps. Also, we have used a permutation test on Moran's correlation coefficient (Moran, 1950), which assumes a certain neighbourhood structure, to determine whether residuals are correlated. Further work in establishing spatial model validation and selection procedures is needed.

Although the spatial analysis of residuals described above aids model validation and selection, there still remains the problem of non-separability of trend and autocorrelation. This problem is not confined to the type of spatial model considered here. It also occurs

elsewhere, for example in the analysis of agricultural field trials. In this area classical competition models have recently been extended (Durbán, 1998; Durbán et al., 1998) to allow the estimation of spatial trends using multi-dimensional smoothers within the framework of semi-parametric additive models (SAMs). Although SAMs incorporate estimation of the smoothing parameters, the underlying theory of additive models assumes that residuals are not correlated. If this is not the case, the use of certain criteria for the selection of smoothing parameters can lead to an under-smoothed surface, (i.e. a too wiggly surface). Thus it is not always possible to distinguish between spatial trend, treatment interference and autocorrelation in the estimation procedure.

Often non-separability of trend and autocorrelation can be avoided by making assumptions about the underlying process causing autocorrelation. For instance it is important to decide whether the underlying process is likely to produce autocorrelation as well as spatial trend and whether the scale of sampling actually allows the observation of autocorrelation. In the seed count example in section 5.3 this was demonstrated. We assumed that observed autocorrelation could not be distinguished from treatment and block effects. Because seed counts were aggregated in the data collection process, autocorrelation was not likely to be observed. Therefore the assumption was made, that most of the autocorrelation was due to treatment and block effects. The estimated autocorrelation parameter was not significant when these effects were kept fixed. We had similar experiences with other spatial count data sets. This approach of keeping treatment and trend effects fixed, while estimating autocorrelation, is common practice in spatial modelling. For example Gotway and Stroup (1997) present an example on weed count modelling using GLMs with a quasi likelihood, defined with a general variance-covariance matrix accounting for autocorrelation. They initially fit a GLM with trend only, then estimate the autocorrelation structure from the semi-variogram of the residuals. The estimated correlation matrix is then inserted into the variance-covariance matrix when fitting the final GLM using quasi-likelihood.

7.2 Design-based versus model-based approach

In the deer modelling example in section 5.4 we compare estimates of total number of occupied squares using estimators based on different theoretical approaches. The estimator from simple random sampling (SRS) is a design-based estimator, whereas the estimators

based on the auto-logistic model are model-based. The estimator from logistic regression can be interpreted as either a design-based or a model-based estimator.

To clarify the difference between the two approaches a summary is given below. Material is drawn from Barnett (1991); Thompson (1992); Brus and DeGruijter (1994) and Gregoire (1998).

Let y_i be the sampled value in area A at location i. In a design-based approach the y_i values are regarded as a sample from a fixed population, that is the observed y-values are regarded as fixed and the sample locations are regarded as the random quantity. Inference is based on the distribution of estimates generated by the sampling design and is free of any assumptions about the distribution of y_i. In contrast to that, the model-based approach regards the y-values as random variables. The population is a realisation of a random process, called a "superpopulation" with an underlying model, the "superpopulation model". Thus $y_1, ..., y_n$ are regarded as realisations of the random variables $Y_1, ..., Y_n$ of the superpopulation. As a consequence only under the model-based approach do we need to model spatial autocorrelation, if present. In the model-based approach inference is based on the model and the model should mimic the superpopulation model as well as possible. In the design-based approach inference is based on the distribution of estimates generated by the sampling design and it is free of any assumptions about the distribution of y_i. The fact that the data of the sample might be correlated does *not* have *any* implication on estimators from the design-based approach.

For both approaches unbiased estimators exist (at least asymptotically). Models can be used in the design-based approach, but they then describe the population itself, whereas in the model-based approach they describe the superpopulation model (data generating process). For illustration we compare expected values of the design-based and model-based approach for the case of estimation of the mean in simple regression.

Let y_i be the sampled value in area A at location i, where either sampling location i is the random variable (design-based approach) or y_i is the random variable for the model-based approach. Let x_i be an explanatory variable at location i which is correlated with y_i and has expected values μ_x. In the design-based approach there is an estimator which accounts for any sampling design: the Horvitz-Thompson estimator (Thompson, 1992). We first illustrate this estimator on estimating the mean of $y_i, ..., y_n$ under the design-based

approach:

$$\hat{\mu}_A = \frac{\sum_{i=1}^{n} \frac{y_i}{\Pi_i}}{N}.$$

The Π_i is the inclusion probability, the probability that location i is included in the sample, and for SRS these are all equal, $\Pi_i = \frac{n}{N}$ yielding $\hat{\mu}_A = \sum_{i=1}^{n} \frac{y_i}{n}$. The expected value $E_s(\hat{\mu}_A) = \sum_{s=1}^{S} p(s)\hat{\mu}_{As}$, where s is one of many possible sampling designs and S is the total number of possible designs.

The design-based regression estimator (with an unknown slope β and intercept α) under SRS is

$$\hat{\mu}_A = \hat{\alpha} + \hat{\beta}\mu_x$$

where

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^n (x_i - \bar{x})^2}$$

and

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

Then substituting for $\hat{\alpha}$

$$\hat{\mu}_A = \bar{y} + \hat{\beta}(\mu_x - \bar{x}).$$

If the model is not known, i.e. β needs to be estimated, then expectation and variance can be shown to be asymptotically (Barnett, 1991, p. 96):

$$E(\hat{\mu}_A) = \mu + O(n^{-1})$$

and

$$Var(\hat{\mu}_A) = \frac{1 - (n/N)}{n} S_{yy}(1 - \rho_{xy}^2) + O(n^{-3/2}),$$

where ρ_{xy} is the correlation of x and y, S_{yy} is the usual sum of squares for y. The quantity $O(f(n)) \to 0$ proportionally to $f(n)$ and with increasing n the bias order $O(n^{-1})$ in $E(\hat{\mu}_A)$ goes to zero more slowly than for $Var(\mu_A)$ with bias of order $O(n^{-3/2})$. This means that only for large samples are regression estimators approximately unbiased and the greater the correlation between x and y the better the improvement in precision in comparison to just calculating the mean. Bias can be a problem for small samples in particular if the relationship is not linear, or x is highly skewed. The design-based variance, omitting the bias term, can be re-expressed as (Thompson, 1992)

$$Var(\hat{\mu}_A) = \frac{1 - (n/N)}{n(n-2)} \sum_i^n (y_i - \hat{\alpha} + \hat{\beta}x_i)^2. \tag{7.1}$$

The model-based variance is the same, but without the finite population correction,

$$Var(\hat{\mu}_A) = \frac{1}{n(n-2)} \sum_{i}^{n} (y_i - \hat{\alpha} + \hat{\beta} x_i)^2, \tag{7.2}$$

which is unbiased under the usual regression model assumptions. Although the formulas 7.1 and 7.2 are the same (ignoring the finite population correction), the estimated parameters have a different meaning. For the design-based approach they relate to the observed population and for the model-based approach they relate to the superpopulation.

For sampling with unequal probability designs, e.g. probabilities proportional to their size, design-based regression estimators are derived by using the inclusion probabilities as weights in a similar fashion to the simple case of estimating the mean (see above).

For multiple regression, design-based estimators are constructed in a similar fashion as described above (see also Thompson, 1992). Further work is needed to determine the properties of design-based estimators based on generalized linear models and generalized additive models. To my knowledge there are no publications on this issue, but there is some work on comparing design and model-based kriging estimators (Brus and DeGruijter, 1994; Brus and deGruijter, 1997).

7.2.1 The deer example in chapter 5 - remark

- From the model-based approach: In the deer example some of the difference of variance between estimates from the logistic and auto-logistic models is due to the fact that the logistic model does not include autocorrelation. There might also be some extra Monte Carlo error introduced by the Gibbs sampling in the process of the bootstrap for the auto-logistic model, which adds to the variance of the autologistic model-based estimate.

- From the design-based approach: The logistic model estimator can be seen as a design-based estimator, and it is expected that variance could be reduced with such an estimator in comparison to simple random sampling. In the bootstrap procedure the finite population correction was included by keeping the observed 20% sample fixed, i.e. only the remaining 80% of the squares were predicted using the model fitted to the bootstrap sample.

7.2.2 The mackerel example in chapter 2 - remark

- From the model-based approach: Examination of residuals indicated spatial and temporal auto-correlation. Thus the model does not entirely describe the "data generation process" or the superpopulation model. Estimators from such a model could be biased, in particular underestimate the variance due to the unmodelled positive autocorrelation.

- From the design-based approach: The sampling design is not random; it is designed so that most of the spawning distribution is sampled, and at the edge of the survey area the sampling design is adaptive. The area sampled changes during the spawning season, for instance in the first period of sampling, only squares in the south of the survey area are sampled, in later periods the area of sampling is expanded. Thus at some time points some of the squares have zero probability of being included in the sample. A design-based approach would require the calculation of inclusion probabilities for a Horvitz-Thompson type estimator. The inclusion probabilities would be proportional to the volume of water sampled. It is not obvious whether inclusion probabilities can be calculated with the survey design as it stands, and as a consequence the design-based approach cannot be applied.

The choice of approach taken depends on the objective of the survey. In most fisheries surveys, as in the mackerel example, the surveys are done to estimate stock size at the time of the survey. If the main aim of the surveys is to quantify what was there at the time of the survey, the design-based approach should be preferred.

Appendix A

Empirical distributions of parameter estimates generated by the simulation study in section 4.4 - Figures

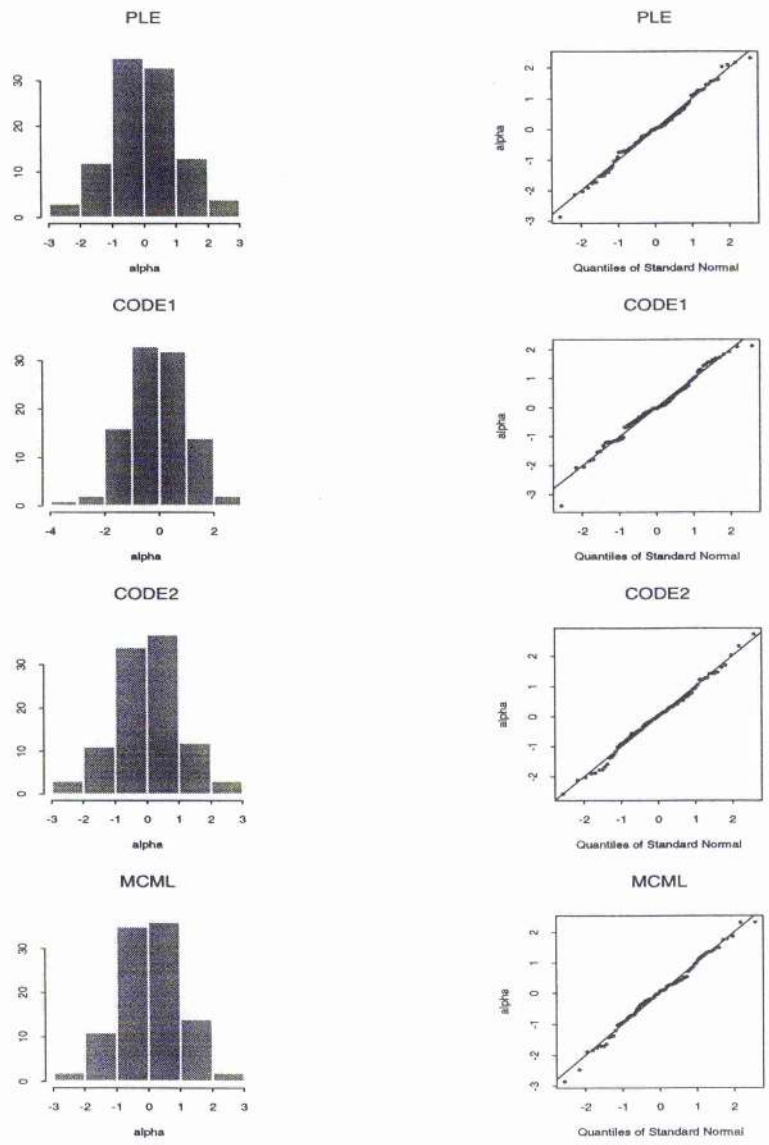


Figure A.1: The empirical distribution of standardised $\hat{\alpha}$ from simulation study 1. On the left histograms and on the right qq-plots are shown. Estimates from PLE, Coding 1, Coding 2 and MCML are shown (top to bottom).

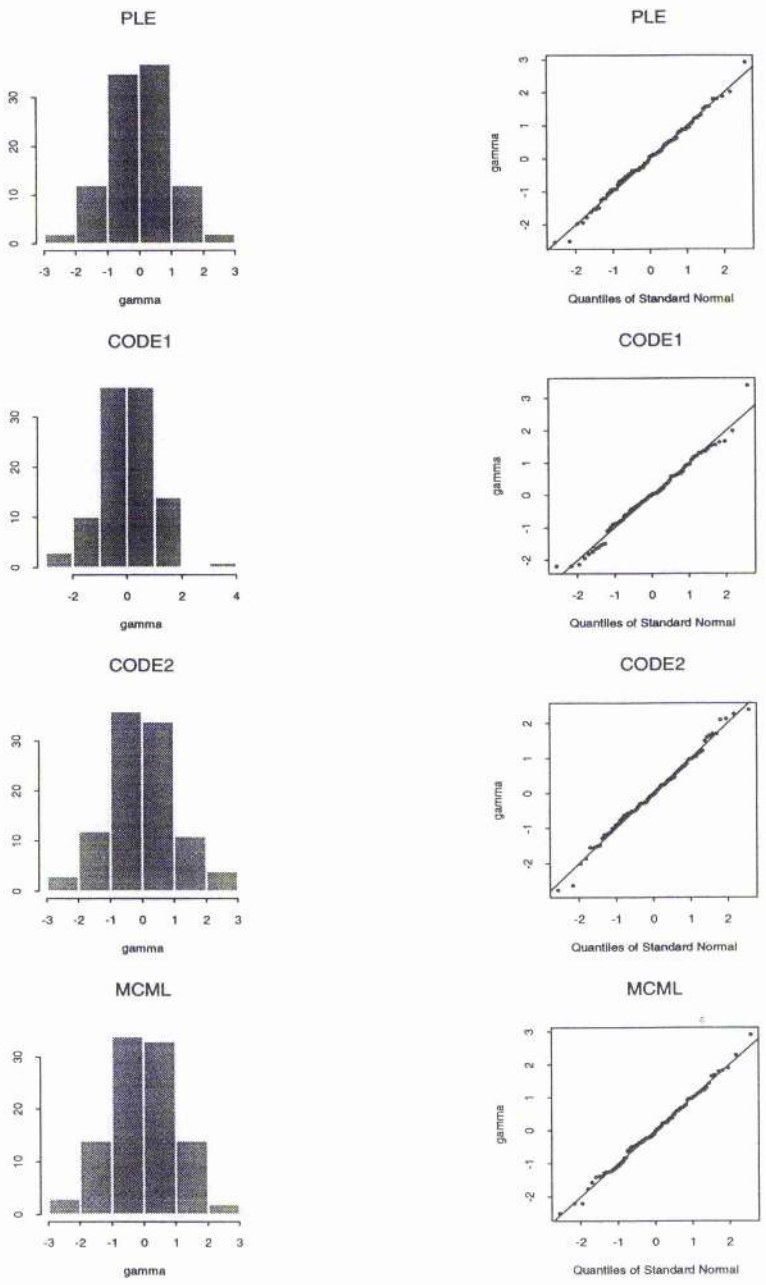


Figure A.2: The empirical distribution of standardised $\hat{\gamma}$ from simulation study 1. On the left histograms and on the right qq-plots are shown. Estimates from PLE, Coding 1, Coding 2 and MCML are shown (top to bottom).

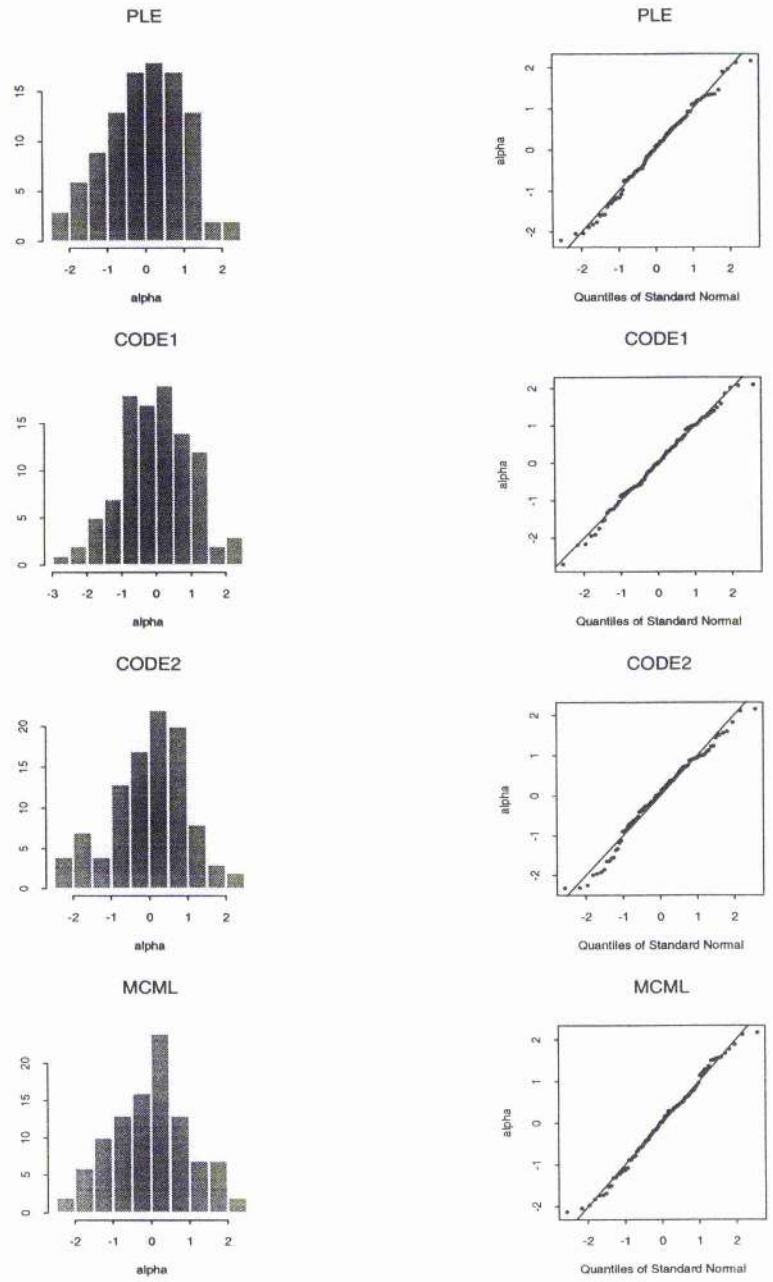


Figure A.3: The empirical distribution of standardised $\hat{\alpha}$ from simulation study 2. On the left histograms and on the right qq-plots are shown. Estimates from PLE, Coding 1, Coding 2 and MCML are shown (top to bottom).

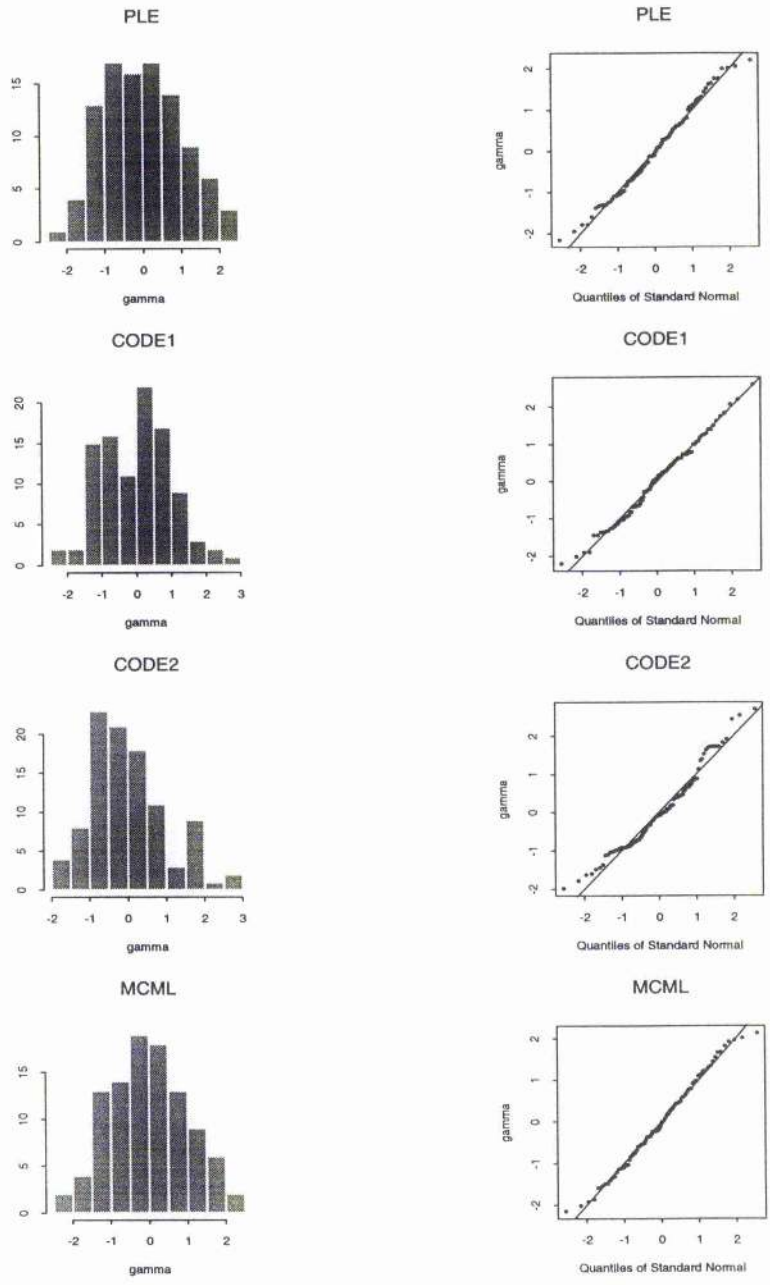


Figure A.4: The empirical distribution of standardised $\hat{\gamma}$ from simulation study 2. On the left histograms and on the right qq-plots are shown. Estimates from PLE, Coding 1, Coding 2 and MCML are shown (top to bottom).

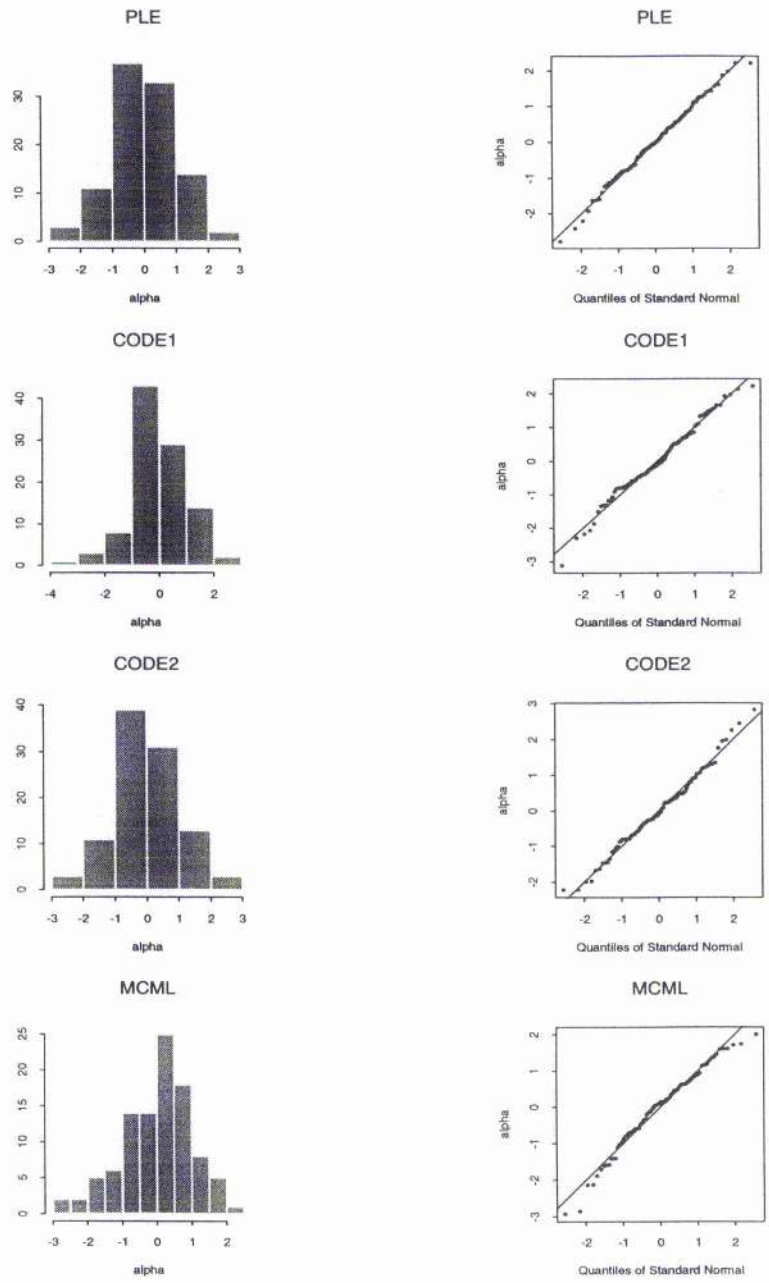


Figure A.5: The empirical distribution of standardised $\hat{\alpha}$ from simulation study 3. On the left histograms and on the right qq-plots are shown. Estimates from PLE, Coding 1, Coding 2 and MCML are shown (top to bottom).

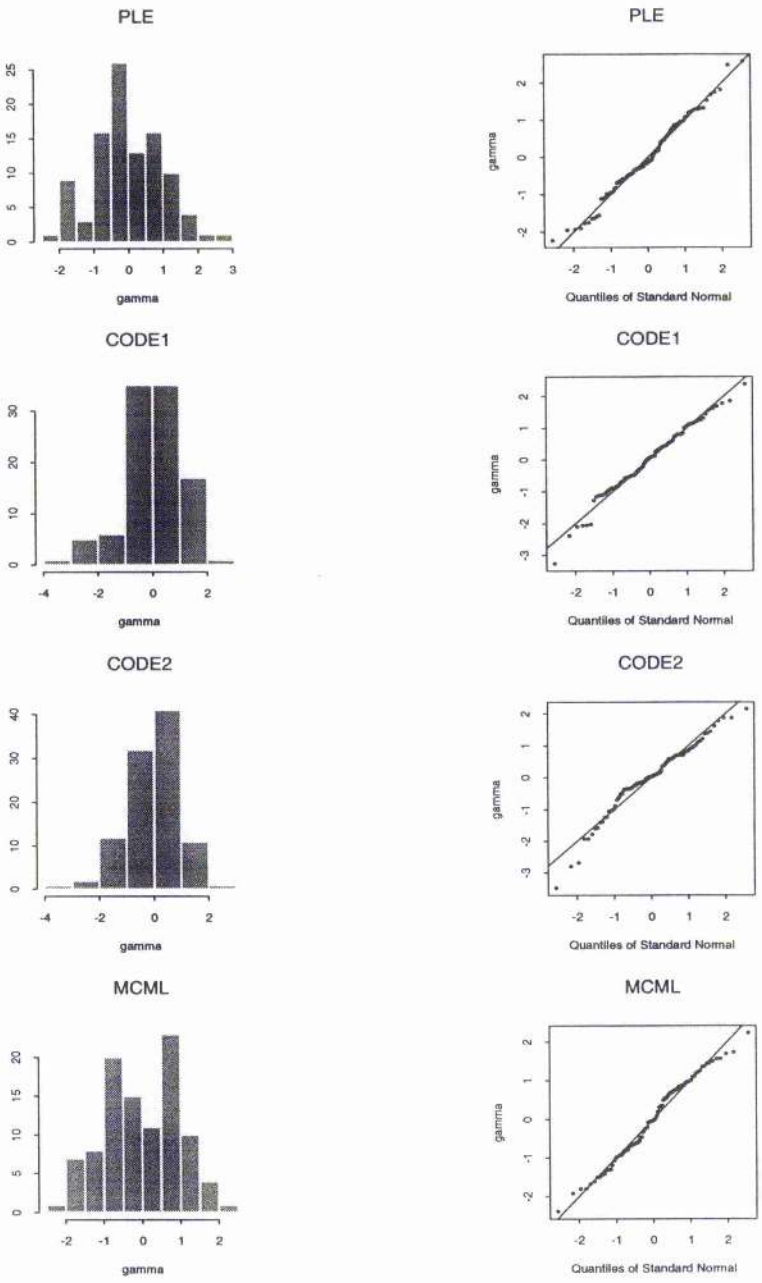


Figure A.6: The empirical distribution of standardised $\hat{\gamma}$ from simulation study 3. On the left histograms and on the right qq-plots are shown. Estimates from PLE, Coding 1, Coding 2 and MCML are shown (top to bottom).

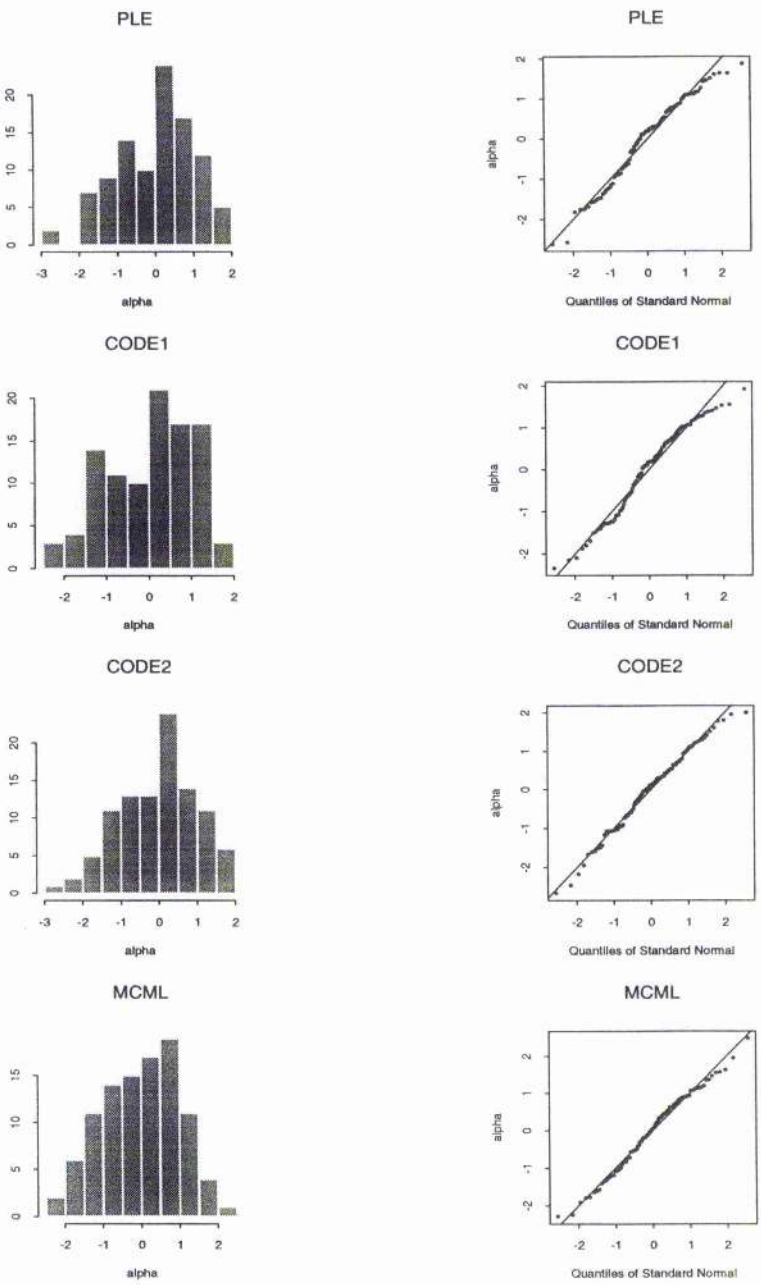


Figure A.7: The empirical distribution of standardised $\hat{\alpha}$ from simulation study 4. On the left histograms and on the right qq-plots are shown. Estimates from PLE, Coding 1, Coding 2 and MCML are shown (top to bottom).

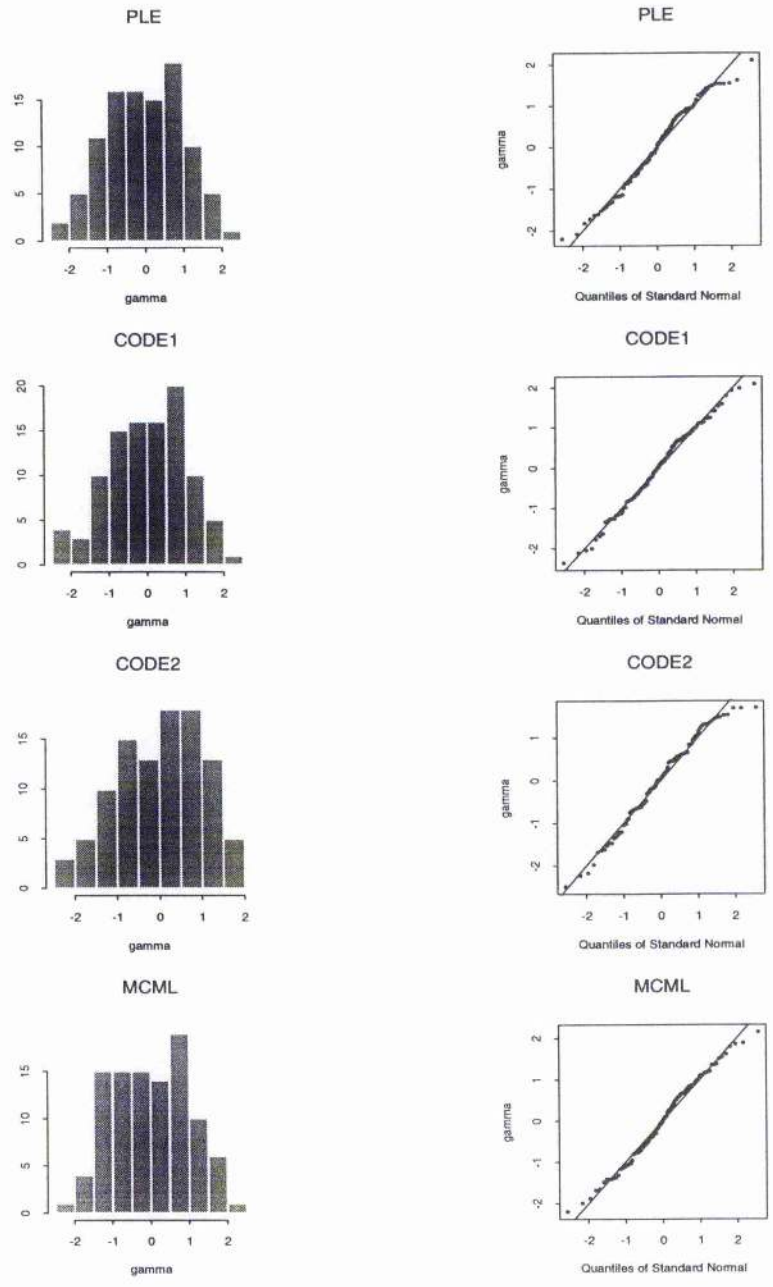


Figure A.8: The empirical distribution of standardised $\hat{\gamma}$ from simulation study 4. On the left histograms and on the right qq-plots are shown. Estimates from PLE, Coding 1, Coding 2 and MCML are shown (top to bottom).

Bibliography

Abel, L., Golmard, J. L., and Mallet, A. (1993). An autologistic model for the genetic-analysis of familial binary data. *American Journal of Human Genetics*, 53(4):894–907.

Agresti, A. (1990). *Categorical Data Analysis*. John Wiley & Sons, New York.

Aitchison, J. (1955). On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association*, 50:901–908.

Anon. (1994). Spawning biology, distribution and abundance of mackerel, scomber scombrus, and horse mackerel, Trachurus trachurus, in the North East Atlantic. A final report to the Directorate General for Fisheries (DG XIV) of the Commission of the European Communities. Technical Report Project Number MA2436, Universty of St Andrews, Research Unit for Wildlife Population Assessment, UK.

Anon. (March, 1996). Report of the working group on mackerel and horse mackerel egg surveys. Technical Report ICES CM 1996/H:2, The International Council of the Exploration of the Seas.

Augustin, N. H. (1993). Modelling the spatial distribution of red deer. Master's thesis, Department of Applied Statistics, University of Reading.

Augustin, N. H., Borchers, D. L., Clarke, E. D., Buckland, S. T., and Walsh, M. (1998a). Spatiotemporal modelling for the annual egg production method of stock assessment using generalized additive models. *Canadian Journal of Fisheries and Aquatic Sciences*, 55:2608–2621.

Augustin, N. H., Mugglestone, M. A., and Buckland, S. T. (1996). An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology*, 33:339–347.

Augustin, N. H., Mugglestone, M. A., and Buckland, S. T. (1998b). The role of simulation in modelling spatially correlated data. *Environmetrics*, 9:175–196.

Barnett, V. (1991). *Sample Survey Principles & Methods.* Edward Arnold, London.

Besag, J. (1972). Nearest-neighbour systems and the auto-logistic model for binary data. *Journal of the Royal Statistical Society, B*, 34:75–83.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, B*, 36:192–236.

Besag, J. (1975). Statistical analysis of non-lattice data. *Statistician*, 24:179–195.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image-restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–59.

Borchers, D. L., Augustin, N. H., and Buckland, S. T. (1997a). Spatiotemporal model development to improve annual egg production assessments of mackerel/horse mackerel. A final report to the Directorate General for Fisheries (DG XIV) of the Commission of the European Community. Technical Report Project Number 94/107, Universty of St Andrews, Research Unit for Wildlife Population Assessment.

Borchers, D. L., Buckland, S. T., Priede, I. G., and Ahmadi, S. (1997b). Improving the precision of the daily egg production method using generalized additive models. *Canadian Journal of Fisheries and Aquatic Sciences*, 54:2727–2742.

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88:9–25.

Brix, A. (1998). *Spatial and Spatio-temporal Models for Weed Abundance.* PhD thesis, Royal Veterinary and Agricultural University Copenhagen, Denmark.

Brus, D. J. and DeGruijter, J. J. (1994). Estimation of nonergodic variograms and their sampling variance by design-based sampling strategies. *Mathematical Geology*, 26:437–454.

Brus, D. J. and deGruijter, J. J. (1997). Random sampling or geostatistical modelling? choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma*, 80:1–59.

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., 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.

Childress, W. M., Crisafulli, C. M., and Rykiel Jr., E. J. (1998). Comparison of Markovian matrix models of a primary successional plant community. *Ecological Modelling*, 107:93–102.

Cressie, N. A. (1991). *Statistics for Spatial Data.* Wiley, New York.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press, Cambridge.

Deriso, R. B., Barnes, J. T., Jacobson, L. D., and Arenas, P. R. (1996). Catch-at-age analysis for Pacific sardine (Sardinops sagax), 1983–1995. Technical report, California Cooperative Oceanic Fisheries Investigations Reports.

Diebolt, J. and Ip, E. H. S. (1995). Stochastic EM: method and application. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in Practice*, pages 259–273. Chapman & Hall, London.

Diggle, P. J., Liang, K., and Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Clarendon Press, Oxford.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics. *Journal of the Royal Statistical Society, C - Applied Statistics*, 47:299–326.

Durbán, M. L. (1998). *Modelling Spatial Trends and Local Competition Effects Using Semiparametric Additive Models.* PhD thesis, Herriot-Watt University.

Durbán, M. L., Hackett, C. A., and Currie, I. (1998). Blocks, trend and interference in field trials. In Marx, B. and Friedl, H., editors, *Statistical Modeling. Proceedings of the 14th International Workshop on Statistical Modeling.*, Graz, Austria.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman & Hall, London.

Ferrándiz, J., López, A., Llopis, A., Morales, M., and Tejerizo, M. L. (1995). Spatial interaction between neighouring counties: cancer mortality data in Valencia (Spain). *Biometrics*, 51:665–678.

Gamerman, D. (1997). *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference.* Chapman & Hall, London.

Gampe, J. (1998). Trend or correlation? - attributing smoothness in nonparametric regression. In Marx, B. and Friedl, H., editors, *Statistical Modeling. Proceedings of the 14th International Workshop on Statistical Modeling*, pages 216–221, New Orleans, Louisiana.

Gelman, A. (1995). Inference and monitoring convergence. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in Practice*, pages 131–161. Chapman & Hall, London.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science*, 7:457–511.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *I. E. E. E. Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Geyer, C. J. (1991). Markov chain Monte Carlo maximum likelihood. In *Computer Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pages 156–163.

Geyer, C. J. (1994). On the convergence of Monte Carlo maximum likelihood calculations. *Journal of the Royal Statistical Society, B*, 56:261–274.

Geyer, C. J. (1995). Estimation and optimization of functions. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in Practice*, pages 241–258. Chapman & Hall, London.

Geyer, C. J. (1999). Likelihood inference for spatial point processess. In Barndorff-Nielsen, O. E. and Kendall, W. S., editors, *Current Trends in Stochastic Geometry and Applications (to appear).* Chapman and Hall, London.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society, B*, 54:657–699.

Gibson, G. J. (1997). Markov chain Monte Carlo methods for fitting spatiotemporal stochastic models in plant epidemiology. *Journal of the Royal Statistical Society, C - Applied Statistics*, 46:215–233.

Gotway, C. A. and Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 2:157–178.

Green, D. G. (1989). Simulated effects of fire, dispersal and spatial pattern on competition within forest mosaics. *Vegetatio*, 82:139–153.

Green, D. R., Cummins, R., Wright, R., and Miles, J. (1993). A methodology for acquiring information on vegetation succession from remotely sensed imagery. In Haines-Young, R., Green, D. R., and Cousens, S. H., editors, *Landscape Ecology and Geographical Information Systems*. Taylor & Francis, London.

Gregoire, T. G. (1998). Design-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, 28:1429–1447.

Gumpertz, M. L., Graham, J. M., and Ristaino, J. B. (1997). Autologistic model of spatial pattern of phytophthora epidemic in bell pepper. *Journal of Agricultural, Biological, and Environmental Statistics*, 2:131–156.

Gunderson, D. R. (1993). *Surveys of Fisheries Resources*. John Wiley & Sons, Inc., London.

Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, London.

Hairston, N. G., Hill, R. W., and Ritte, U. (1971). The interpretation of aggregation patterns. In *Statistical Ecology Volume 1: Spatial Patterns and Statistical Distributions*. The Pennsylvania State University Press, University Park, Pennsylvania.

Hammersley, J. M. and Clifford, P. (1971). Markov fields on finite graphs and lattices. *Unpublished Manuscript, Oxford University*.

Hastie, T. and Tibshirani, R. J. (1990). *Generalized Additive Models.* Chapman and Hall, London.

Heikkinen, J. and Högmander, H. (1994). Fully Bayesian approach to image restoration with an application in biogeography. *Applied Statistics*, 43:569–582.

Heikkinen, J. and Penttinen, A. (1999). Bayesian smoothing in the estimation of the pair potential function of Gibbs point processes. *Bernoulli*, in press.

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

Hoeting, J. A., Van Caster, M., and Bowden, D. (1999). An improved model for spatially correlated binary responses. Technical report, Colorado State University.

Högmander, H. and Møller, J. (1995). Estimating distribution maps from atlas data using methods of statistical image analysis. *Biometrics*, 51:393–404.

Huffer, F. W. and Wu, H. L. (1998). Markov chain Monte Carlo for autologistic regression models with application to the distribution of plant species. *Biometrics*, 54:509–524.

Hughes, J. P. and Guttorp, P. (1994). Incorporating spatial dependence and atmospheric data in a model of precipitation. *Journal of Applied Meteorology*, 33:1503–1515.

Ickstadt, K. and Wolpert, R. L. (1996). Spatial correlation or spatial variation? A comparison of Gamma/Poisson hierarchical models. Technical report, Institute of Statistics and Decision Sciences, Duke University, Durham NC 27708-0251.

Ising, E. (1925). Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31:253–258.

Jeltsch, F., Milton, S. J., Dean, W. R. J., and van Rooyen, N. (1996). Tree spacing and coexistence in semiarid savannas. *Journal of Applied Ecology*, 84:583–595.

Johnson, N. L., Kotz, S., and Kemp, A. W. (1992). *Univariate Discrete Distributions.* John Wiley & Sons, New York.

Kaiser, M. S. and Cressie, N. (1997). Modeling Poisson variables with positive spatial dependence. *Statistics and Probability Letters*, 35:423–432.

Lee, J. and Kaiser, M. S. (1997). Maximum likelihood estimation in conditionally specified statistical models. Technical report, Department of Statistics, Iowa State University.

Lerzman, K. P. (1995). Forest dynamics, differential mortality and variable recruitment probabilities. *Journal of Vegetation Science*, 6:191–204.

Lin, X. and Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society, B*, 61:381–400.

Lindsey, J. K. (1999). *Models for Repeated Measurements.* Oxford University Press, Oxford.

Maravelias, C. D. and Reid, D. G. (1997). Identifying the effects of oceanographic features and zooplankton on pre-spawning herring abundance using generalized additive models. *Marine Ecology-Progress Series*, 147:1–9.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* Chapman and Hall. London.

Meyer, R. and Millar, R. B. (1999). Bayesian stock assessment using a state-space implementation of the delay difference model. *Canadian Journal of Fisheries and Aquatic Sciences*, 56:37–52.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrica*, 37:17–23.

Osborne, P. E. and Tigar, B. J. (1992). Interpreting bird atlas data using logistic models: an example from Lesotho, southern Africa. *Journal of Applied Ecology*, 29:55–62.

Pacala, S. W. (1986). Neighborhood models of plant population dynamics, 2. multi-species models of annuals. *Theoretical Population Biology*, 29:262–292.

Penttinen, A. (1984). Modelling interaction in spatial point patterns: Parameter estimation by the maximum likelihood method. *Jyvaskyla Studies in Computer Science, Economics, and Statistics*, pages 1–107.

Pickard, D. K. (1987). Inference for discrete Markov fields: The simplest nontrivial case. *Journal of the American Statistical Association*, 82:90–96.

Preisler, H. K. (1993). Modelling spatial patterns of trees attacked by bark-beetles. *Applied Statistics*, 42:501–514.

Preisler, H. K. and Mitchell, R. G. (1993). Colonization patterns of the mountain pine-beetle in thinned and unthinned lodgepole pine stands. *Forest Science*, 39:528–545.

Raftery, A. E. and Steven, M. L. (1995). Implementing MCMC. In *Markov Chain Monte Carlo in Practice*, pages 115–143. Chapman & Hall, London.

Roberts, G. O. (1995). Markov chain concepts related to sampling algorithms. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in Practice*, pages 45–57. Chapman & Hall, London.

Scanlan, J. C. (1994). State and transition models for rangelands. 5. the use of state and transition models for predicting vegetation change in rangelands. *Tropical Grasslands*, 28:229–240.

Swartzman, G. (1997). Analysis of the summer distribution of fish schools in the Pacific Eastern Boundary Current. *ICES Journal of Marine Science*, 54:105–116.

Swartzman, G., Huang, C. H., and Kaluzny, S. (1992). Spatial analysis of Bering sea groundfish survey data using generalized additive models. *Canadian Journal of Fisheries and Aquatic Science*, 49:1366–1378.

Swartzman, G., Silverman, E., and Williamson, N. (1995). Relating trends in walleye pollock (Theragra chalcogramma) abundance in the Bering Sea to environmental factors. *Canadian Journal of Fisheries and Aquatic Sciences*, 52:369–380.

Swartzman, G., Stuetzle, W., Kulman, K., and Powojowski, M. (1994). Relating the distribution of pollock schools in the Bering Sea to environmental factors. *ICES Journal of Marine Science*, 51:481–492.

Thompson, S. K. (1992). *Sampling*. Wiley & Sons, New York.

Tongeren, O. v. and Prentice, I. C. (1986). A spatial simulation model for vegetation dynamics. *Vegetation*, 65:163–173.

van Groenendael, J. M., Roepers, R. G., Woltjer, I., and Zweers, H. R. (1996). Vegetation succession in lakes of west Connemara, Ireland: Comparing predicted and actual changes. *Journal of Vegetation Science*, 7:211–218.

Venables, W. N. and Ripley, B. D. (1994). *Modern Applied Statistics with S-Plus*. Springer-Verlag, New York.

Walker, P. A. (1990). Modelling wildlife distributions using a geographic information system: kangaroos in relation to climate. *Journal of Biogeography*, 17:279–289.

Wang, Y. (1998). Smoothing spline models with correlated errors. *Journal of the American Statistical Association*, 93:341–348.

Welch, D. W., Chigirinsky, A. I., and Ishida, Y. (1995). Upper thermal limits on the oceanic distribution of pacific salmon (Oncorhynchus spp.) in the spring. *Canadian Journal of Fisheries and Aquatic Sciences*, 52:489–503.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41:434–439.

Wolpert, R. L. and Ickstadt, K. (1998). Gamma/Poisson random field models for spatial statistics. *Biometrika*, 85:251–267.

Wood, S. N. (1994). Obtaining birth and mortality patterns from structured population trajectories. *Ecological Monographs*, 64:23–44.

Wood, S. N. and Horwood, J. W. (1995). Spatial distribution functions and abundances inferred from sparse noisy plankton data: an application of constrained thinplate splines. *Journal of Plankton Research*, 17:1189–1208.

Wu, H. L. and Huffer, F. W. (1997). Modeling the distribution of plant species using the autologistic regression model. *Environmental and Ecological Statistics*, 4:49–64.

Wu, J. and Levin, S. A. (1994). A spatial patch dynamic modeling approach to pattern and process in an annual grassland. *Ecological Monographs*, 64:447–464.

Younes, L. (1988). Estimation and annealing for Gibbsian fields. *Annales de l'Institut Henri Poincaré*, 24:269–294.