

# A review of spatio-temporal modelling of quadrat count data with application to striga occurrence in a pearl millet field

Dale Hess<sup>1</sup>, Marie-Colette van Lieshout<sup>2</sup>, Bill Payne<sup>3</sup> and Alfred Stein<sup>4</sup>

<sup>1</sup> International Crops Research Institute for the Semi-Arid Tropics, BP 320, Bamako, Mali

<sup>2</sup> Centrum voor Wiskunde en Informatica, PO Box 94079, 1090 GB Amsterdam, The Netherlands

<sup>3</sup> Texas A&M University Agricultural Research and Extension Center, 6500 Amarillo Boulevard West, Amarillo, TX 79106, USA

<sup>4</sup> International Institute for Aerospace Survey and Earth Sciences, PO Box 6, 7500 AA Enschede, The Netherlands (e-mail: stein@itc.nl)

**KEYWORDS:** quadrat counts, spatial statistics, striga.

## ABSTRACT

This paper describes how spatial statistical techniques may be used to analyse weed occurrence in tropical fields. Quadrat counts of weed numbers are available over a series of years, as well as data on explanatory variables, and the aim is to smooth the data and assess spatial and temporal trends. We review a range of models for correlated count data. As an illustration, we consider data on striga infestation of a 60 × 24 m<sup>2</sup> millet field in Niger collected from 1985 until 1991, modelled by independent Poisson counts and a prior auto regression term enforcing spatial coherence. The smoothed fields show the presence of a seed bank, the estimated model parameters indicate a decay in the striga numbers over time, as well as a clear correlation with the amount of rainfall in 15 consecutive days following the sowing date. Such results could contribute to precision agriculture as a guide to more cost-effective striga control strategies.

## INTRODUCTION

In semi-arid and sub-humid tropical countries in Africa and parts of Asia, *Striga hermonthica* (Del.) Benth. (commonly known as witch weed) is an economically important parasitic weed that severely attacks cereal crops including pearl millet, sorghum and maize. In fact, it has been called the major biological threat to cereal crops in Africa [Ejeta *et al*, 1993], and has infested large portions of cereal production regions. The striga problem is particularly severe in areas with erratic rainfall and relatively poor soils; the greatest damage occurs in the Sahelian and Savannah zones. Due to the withdrawal of water, minerals and organic compounds from their hosts, and to putative phytotoxic effects, parasitic weeds have a very strong negative effect on both the quality and the quantity of the grain and straw yield. Recent trends away from traditional fallowing practices towards continuous mono cropping to meet the increasing population pressure have intensified the striga problem [Parker & Riches, 1993].

Striga control is very difficult. A contributing factor is early underground development during which time much of the damage to the susceptible host plant is done. The use of herbicides has several drawbacks. Firstly, herbicides are as likely to influence the host plant as the striga, and may damage other crops in the system. Secondly, as the farmers on the striga infested soils are generally poor, herbicides may not be affordable. Thirdly, the farmers do not immediately observe the results of their actions, as most of the harm has been done before control measures could be taken, and many seeds remain dormant in the soil, thus causing problems in the next growing season. Therefore, striga control requires a form of integrated crop management, where measures focused on the prevention of striga seed production and exhaustion of the supply of seeds in the soil [Hausmann *et al*, 2000] are implemented over a range of years.

At the moment, it is not fully understood what factors determine the number and distribution of emerged striga. It has been recorded that additional use of nitrate may increase the crop quality and reduce the striga damage [Boukar *et al*, 1996]. Moreover, a positive correlation was found between crop yield and striga occurrence [Manu *et al*, 1991], suggesting that a healthy, well fed host may be able to tolerate more emerged striga plants than an unhealthy one, so that striga is more likely to be found in the more fertile parts of a field. Striga seed are spread by wind, rainfall, through animal and human activities as well as contaminated crop seed [Hess & Haussmann, 1999]. Preseason rainfall serves to precondition striga seed for germination. Heavy showers with runoff are also likely to spread the striga laterally towards adjacent fields, and beyond. Nevertheless, patterns of weed spread have been little studied and are poorly understood, and stochastic models aimed at explaining covariate effects, as well as spatial and temporal variation, are useful.

The plan of this paper is as follows. We first describe a data set on the occurrence of striga in a millet field col-

lected at the ICRISAT Sahelian Centre. We review statistical models for spatially correlated longitudinal count data, and finally illustrate one such model on the data mentioned before.

**DATA**

In the appendix, we list counts of striga infestation in a 60 × 24 m<sup>2</sup> experimental field at the ICRISAT Sahelian Centre near Niamey (Niger). In 1985, the field was prepared for the cultivation of pearl millet (*Pennisetum glaucum* (L.) R. Br.), a drought resistant cereal crop adapted to light sandy soils [Dave, 1987]. Permanent metal stakes were placed at 1 m intervals around the field perimeter, in accordance with the recommended sowing density of pearl millet for the region. Debris was cleared from the field surface at the outset of the cropping season (May), and a basal fertiliser (15 kg per hectare each of N, P<sub>2</sub>O<sub>5</sub> and K) was applied manually and incorporated by harrow. Millet of the variety *Composite Inter-Variétal de Tarna* (CIVT) was planted in hills at the grid points indicated by the stakes on the 18<sup>th</sup> of June, following the first sizeable rainfall of the year [Sivakumar, 1982] (4.9 mm the day before and 19.4 mm on the sowing date). Many seeds were sown per hill, and thinned two to three weeks after sowing to a maximum of three plants per hill. Mechanical weeding was carried out two and four weeks after sowing. Subsequent weeding, if necessary to facilitate counts, was done by hand, taking care not to disturb emerged or emerging striga plants. Striga counts in quadrats of 1 × 1 m<sup>2</sup> around each hill were recorded at various stages of the growing season. In Table 1 we present summary statistics for the final count carried out after the millet harvest. The full data are listed in the appendix.

Since the striga were left untouched, seeds spread naturally, a phenomenon which may lead to the formation of localised seed banks causing spatial inhomogeneity. The native soil fertility also exhibits spatial variation [Manu *et al*, 1991]. Note that since millet is planted in a regular grid, no heterogeneity due to availability of hosts is caused. The resulting spatial pattern is given in Figure 1.

Finally, rainfall was measured at the central meteorological station of the Centre, as summarised in Table 2.

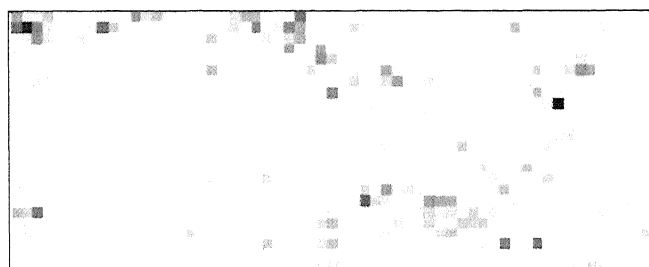
**STATISTICAL MODELS FOR SPATIALLY CORRELATED COUNT DATA**

In this section, we give an overview of stochastic models for spatially correlated count data arising from small area sampling of some underlying point process. Thus, the nature of the data collection process leads us into the realm of a missing data framework, and inference should be aimed at characteristics of the complete data (in our

case the emerged striga pattern) based on the observations (here cumulative striga counts per hill).

**TABLE 1:** Quantiles, minimum, maximum and median of striga counts in 1 × 1 m<sup>2</sup> quadrats in a 60 × 24 m<sup>2</sup> millet field at the ICRISAT Sahelian Centre near Niamey (Niger) as observed 105 days after the sowing date (18 June 1985). Full data are listed in the appendix; for graphical representation, see Figure 1. The acronym DAS stands for days after sowing.

| Year | DAS | Min. | 1 <sup>st</sup> Qu. | Median | Mean | 3 <sup>rd</sup> Qu. | Max. |
|------|-----|------|---------------------|--------|------|---------------------|------|
| 1985 | 105 | 0    | 1                   | 4.5    | 7.96 | 11                  | 103  |



**FIGURE 1:** Striga counts in 1 × 1 m<sup>2</sup> quadrats in a 60 × 24 m<sup>2</sup> millet field at the ICRISAT Sahelian Centre near Niamey (Niger) as observed 105 days after the sowing date (18 June 1985). Low counts correspond to a white color; black is used for high values.

**TABLE 2:** Summary statistics for rainfall (mm) measured at the central meteorological station of the ICRISAT Sahelian Centre near Niamey (Niger) for the year 1985. The acronym DAS stands for days after sowing.

| Year | Sowing date | 10 DAS | 15 DAS | 20 DAS | 73 DAS | 105 DAS |
|------|-------------|--------|--------|--------|--------|---------|
| 1985 | 18 June     | 48.5   | 49.6   | 97.8   | 406.9  | 516.8   |

We shall begin by fixing some notation. Write *I* for the index set of quadrats, and denote by *y<sub>i</sub>* the count in the quadrat *C<sub>i</sub>* indexed by *i* ∈ *I*. Then, the data can be described by the vector **y** = (*y<sub>i</sub>*)<sub>*i* ∈ *I*</sub>, *y<sub>i</sub>* ∈ {0, 1, ...}.

For counts, it is natural to assume a Poisson law. However, the most straightforward route of specifying a Poissonian auto regression with the conditional mean of a particular quadrat depending on its neighbouring quadrats allows for negative correlation only [Besag, 1974], while in the present context the counts at neighbouring quadrats should be similar. Hence, we are led towards a Cox formulation where, conditional on the intensity, the counts are independent, and smoothness is taken into account by the intensity. Hence, the conditional distribution of the observed counts is

$$f(\mathbf{y} | \boldsymbol{\mu}) = \prod_{i \in I} e^{-\mu(C_i)} \frac{\mu(C_i)^{y_i}}{y_i!} \tag{1}$$

where  $\boldsymbol{\mu}$  denotes the complete data intensity. Note that the maximum likelihood estimator of  $\boldsymbol{\mu}(C_i)$  is equal to

the observed  $y_i$ , and no smoothing is achieved. Consequently, we adopt a Bayesian point of view, and proceed to formulate a prior model for  $\mu$ . As an aside, although agricultural applications have played a major role in the statistical literature from its beginning [eg, Fisher, 1928], most studies took a frequentist point of view, despite the obvious spatial correlation present in fields [Besag & Higdon, 1993; 1999]. Only recently, advances from spatial statistics and image analysis, which have both been predominantly Bayesian in nature, have spilled over to agriculture and other branches of statistics.

The prior distribution of the random intensity  $\mu$  should reflect the spatial correlations between counts in different quadrats, as well as take into account covariates and quadrat geometry. For instance, Besag, York & Mollié [1991] propose a log linear Gaussian model of the form

$$\log \mu(C_i) = U_i + V_i + w_i + \log c_i \quad (2)$$

where  $c_i$  is a base rate of expected counts (based on for instance the size of  $C_i$ ),  $U = (U_i)_{i \in I}$  is a smoothing random field capturing spatial structure,  $V = (V_i)_{i \in I}$  is independent of  $U$  and captures any extra-Poisson variation, and  $w = (w_i)_{i \in I}$  incorporates the covariates. Commonly,  $w = A\theta$  for a known matrix  $A$  and parameter vector  $\theta$ , or

in other words  $w_i = \sum_{j=1}^k A_{ij} \theta_j$ . The field  $V$  is assumed to

be Gaussian white noise, and  $U$  is an intrinsic auto regression [Künsch, 1987]

$$p(u) \propto \kappa^{-n/2} \sum_{i \sim j} \exp \left[ -\frac{1}{2\kappa} \sum_{i \sim j} (u_i - u_j)^2 \right]. \quad (3)$$

where  $\sim$  is the neighbourhood relation on  $I$  and  $\kappa > 0$ . Sampling issues are considered in Knorr-Held & Rue [2000] and in Rue [2001]. Knorr-Held & Besag [1998] extend (2) to spatio-temporal observations as follows:

$$\log \mu_t(C_i) = \alpha_t + U_i + V_i + W_{t,i} \quad (4)$$

where  $t$  is the time index, and  $W = (W_{t,i})$  a linear model for the time-dependent effect of the covariates (as an approximation to binomial counts with logit as given above). The temporal trend  $\alpha_t$  is modelled by a random walk with independent Gaussian increments. Similar ideas can be found in Mugglin *et al* [2000]. Diggle, Moyeed & Tawn [1998] consider a generalized linear mixed model as in (2), but ignore the  $V$ -field and employ a (continuous) stationary Gaussian process  $U$  with covariance structure of the form  $\kappa\rho(u-u')$  for some  $\kappa > 0$  and isotropic function  $\rho$ , eg,  $\rho(u) = \exp\{-\beta |u|^\delta\}$  for some  $0 < \delta < 2$  and  $\beta > 0$ . Link functions other than the logarithmic transformation could be considered as well. The same set-up is considered by Christensen, Møller & Waagepetersen [2000], who also discuss simulation.

Omitting the covariates and base rate yields a log Gaussian Cox process [Møller *et al*, 1998].

Note that all log Gaussian models discussed above contain hyper priors that may be modelled by dispersed but proper distributions.

A disadvantage of the logarithmic transform is that it does not readily scale with respect to the quadrat size. For this reason, Wolpert & Ickstadt [1998] define

$$\mu(C_i) = \int_S k(C_{i,s}) d\Gamma(s) \quad (5)$$

to be a kernel mixture of a Gamma process  $\Gamma(s)$  with shape measure  $a(ds)$  and scale function  $b(s) > 0$ . The space  $S$  can be either continuous or discrete. For the mixture kernel, they propose

$$k(x, s) \propto e^{A\theta} \exp \left( \frac{-1}{2\kappa} |x - s|^2 \right) \quad (6)$$

It should be noted that the model is strictly speaking overparameterised (only the ratio  $k(x,s)/b(s)$  matters) but since covariates only affect the kernel  $k(\dots)$  the present formulation may be clearer.

To conclude our review, Green & Richardson [2000] suggest a Gamma mixture model

$$\mu(C_i) = c_i e^{w_i} \lambda_i \quad (7)$$

where  $c_i$  and  $w_i$  are as in (2) and  $\lambda_i = \lambda_{Z_i}$  is a random allocation  $Z_i$  to one of  $k$  components  $\lambda_1, \dots, \lambda_k$ . The allocation variables are smoothed by assuming  $(Z_i)_{i \in I}$  to form a Potts model

$$p(\mathbf{z}) \propto \exp \left[ \beta \sum_{i \sim j} 1 \{z_i = z_j\} \right] \quad (8)$$

with  $\beta \geq 0$ . The  $\lambda_j$  are independently Gamma distributed with shape parameter  $a = I$  and size parameter  $b = \sum_{i \in I} c_i \sum_{i \in I} y_i$ . More generally, the number of components may be random as well, eg, Stephens [2000].

## RESULTS

To illustrate the developed methods, we return to the data on witch weed in a millet field discussed in the data section. For the classical Poisson/lognormal approach (1)-(2), the prior distribution of the smoothing field is improper. Here, we replace the intrinsic auto regression by a proper conditional auto regression [eg, Ripley, 1988], and examine the effect of its parameters. More specifically, recalling that rainfall is measured centrally, no specific fertility information is available, and all quadrats are of equal size, we set

$$\log \mu(C_i) = \log \beta + U_i + V_i \quad (9)$$

Here  $\beta$  is the base rate,  $V = (V_i)_{i \in I}$  a Gaussian white

noise field with precision  $\sigma > 0$ , and  $U = (U_i)_{i \in I}$  a Gaussian conditional auto regression smoothing field. The conditional mean is

$$E[U_i | U_j = u_j, j \neq i] = \frac{\gamma}{n_i} \sum_{j \sim i} u_j, \quad (10)$$

the conditional variance

$$Var[U_i | U_j = u_j, j \neq i] = \frac{1}{n_i \tau}, \quad (11)$$

for each  $i \in I$ . The quadrats  $C_i$  and  $C_j$  are neighbours ( $i \sim j$ ) if and only if they are horizontally or vertically adjacent, and  $n_i$  denotes the number of neighbours of the quadrat represented by index  $i$  (ie,  $n_i = 4$  except at the borders). The parameter  $\gamma \in (0,1)$  encourages near quadrats to have similar fertility and hence striga incidence,  $\tau$  controls the precision.

Time could be taken into account by replacing  $\log \beta$  by  $\alpha_t = \theta w_t$  where  $w_t$  is the rainfall covariate (15DAS) in year

$t$ , and  $\alpha_t$  the year effect.

Since for fixed initial base rate  $\beta = \frac{1}{60 \times 24} \sum_{i \in I} y_i$  and smoothing parameters  $\gamma$ ,  $\sigma$  and  $\tau$  the log posterior density is concave in  $U$  and  $V$ , standard numerical procedures can be used to find the optimum [Besag et al, 1991]. Some results are given in Figure 2. It can be seen that an increase in  $\gamma$  leads to smoother  $U$ ; increasing the precisions  $\tau$  and  $\sigma$  similarly yields smoother reconstructions, with less variation in both  $U$  and  $V$ . Finally, the updated base rates

$$\sum_{i \in I} y_i / \sum_{i \in I} e^{u_i + v_i} \quad (12)$$

are 7.90, 7.60, 7.93 and 7.56 for the parameter values considered in Figure 2 (from top to bottom).

A general pattern is observed in the smoothing field being a dark band from the top left towards a dark coloring in the central right part of the figures. Some details occur as well, where in particular the choice for

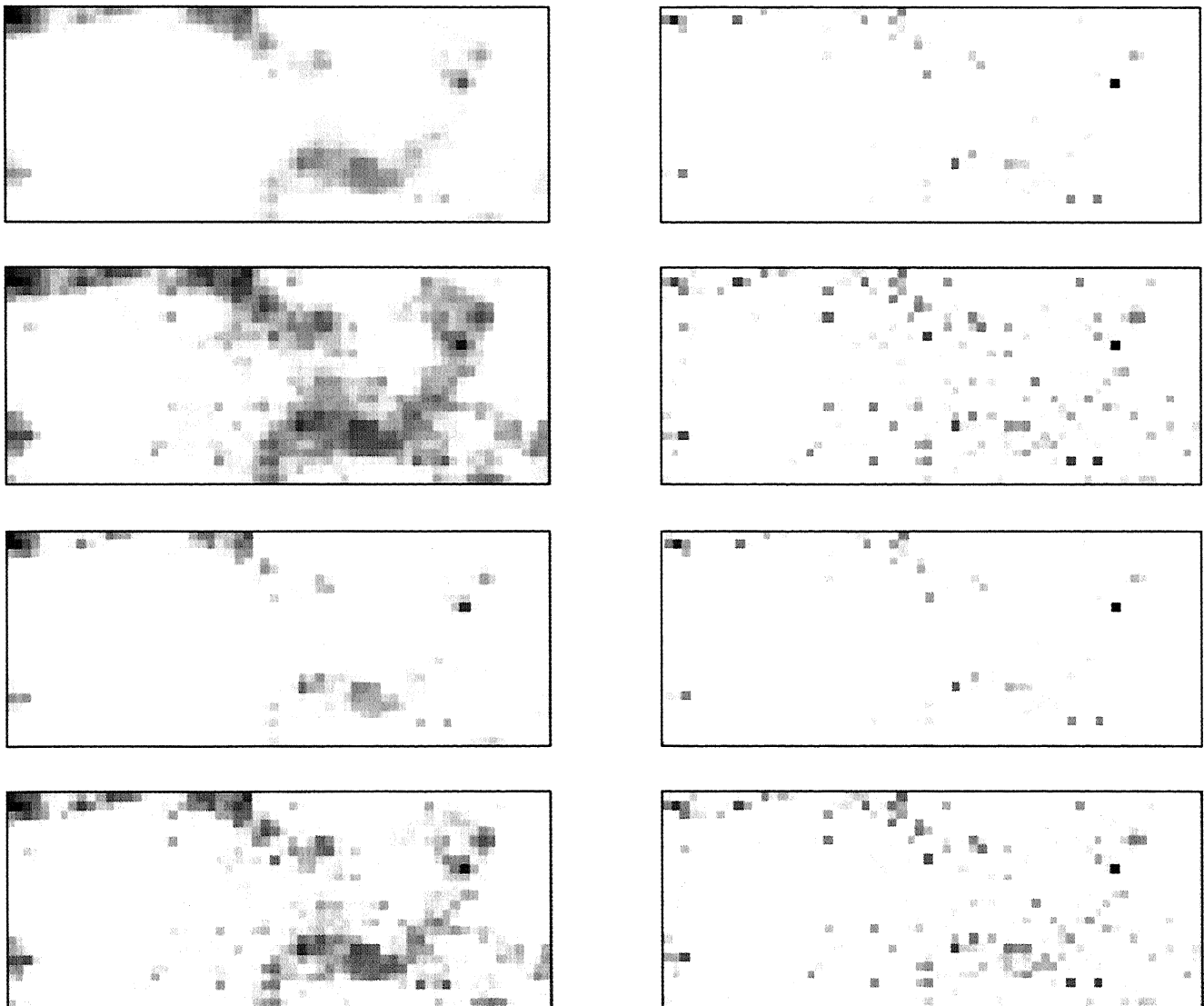


FIGURE 2: Maximum a posteriori estimates of the smoothing (left) and residual variation (right) fields given the data of Figure 1 for  $\gamma = 0.9$  (top two lines) and  $\gamma = 0.7$  (bottom two lines),  $\tau = \sigma/4 = 100.0$  (odd lines) and  $\tau = \sigma/4 = 10.0$  (even lines).

$\sigma=4\tau=100.0$  leads to a less noisy residual pattern than a choice for  $\sigma=4\tau=10.0$ . A choice for  $\gamma=0.9$  also leads to more smoothness than a choice for  $\gamma=0.7$ . The noise, though, is largest where the pattern has the largest values. It can be concluded that the presence of seedbanks is identified by application of a spatial pattern analysis.

## DISCUSSION

So far, we have focused on data obtained as a single image from one field in a single year, leading to a study of spatial variation on count data. However, data from within the same season as well as from different years are available as well. We aim to consider data collected over a range of years. In terms of modeling, we replace  $\log \beta$  by  $\alpha_t + \theta w_t$ , where  $w_t$  is the rainfall covariate (15 DAS) in year  $t$  and  $\alpha_t$  the year effect.

In several other studies, such an analysis may be interesting, for example for the analysis of count data collected by a Geiger counter from the plane. An interesting example concerns the combination of the chemical elements Th, K and U [Schetselaar, 2000]. Related to such data collection and modeling is interpolation to obtain an even distribution of the data over an area of land. Observations along flight lines are relatively simple to collect, but spatial interpolation is not trivial, because the sampling pattern is very dense along the flight lines, but absent between them.

Practical consequences from an analysis as the current one can be large, in particular in relation to reduction in herbicide use. For optimal use of manpower and herbicides, smoothed maps showing the presence of seedbanks are usually of a larger benefit than the original maps, which show large incidental spatial variation that is difficult to handle. A quantitative analysis of these results, though, still has to be done. The smoothed map gives a good impression of how to apply these efforts for weed management in an optimal way, namely where the seed bank is present. This may lead to a better use of scarce resources.

## CONCLUSION

This paper presents a review of spatial models of quadrat count data with possible extensions towards spatio-temporal modelling. It is based on an application to striga occurrence in a pearl millet field. We conclude that the presented models have interesting consequences for agricultural applications.

## ACKNOWLEDGEMENT

This work was initiated during a sabbatical of the fourth author at CWI in May-June 2000. He is grateful for the

support offered in this period. We would also like to thank Aad van Ast for many helpful discussions.

## REFERENCES

- Besag, J.E., 1974. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B* 36: 192-236.
- Besag, J.E. & D.M. Higdon, 1993. Bayesian inference for agricultural field experiments. *Bulletin of the International Statistical Institute* 55: 121-136.
- Besag, J.E. & D.M. Higdon, 1999. Bayesian analysis of agricultural field experiments (with discussion). *Journal of the Royal Statistical Society, Series B* 61: 691-746.
- Besag, J.E., J. York & A. Mollié, 1991. Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Mathematical Statistics* 43: 1-59.
- Boukar, I., D.E. Hess & W.A. Payne, 1996. Dynamics of moisture, nitrogen, and striga infestation on pearl millet transpiration and growth. *Agronomy Journal* 88: 545-549.
- Christensen, O.F., J. Møller & R.P. Waagepetersen, 2000. Analysis of spatial data using generalized linear mixed models and Langevin-type Markov chain Monte Carlo. Technical Report R-00-2009, Department of Mathematical Sciences, Aalborg University.
- Dave, H.R., 1987. Pearl millet hybrids. In: J.R. Witcombe & S.R. Beckerman (Eds), *Proceedings of the International Pearl Millet workshop, ICRISAT, Niamey*, pp. 121-126.
- Diggle, P.J., R.A. Moeed & J.A. Tawn, 1998. Model-based geostatistics (with discussion). *Applied Statistics* 47:299-350.
- Ejeta, G., & L.G. Butler. 1993. Host-parasite interactions throughout the striga life cycle. and their contributions to Striga resistance. *African Crop Science Journal* 1: 75-80.
- Fisher, R.A., 1928. *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh, 269 pp.
- Green, P.J. & S. Richardson, 2000. Spatially correlated allocation models for count data. Department of Mathematics, University of Bristol, <http://www.stats.bris.ac.uk/~peter/papers/smix.ps>
- Hausmann, B.I.G., D.E. Hess, H.G. Welz & H.H. Geiger, 2000. Improved methodologies for breeding striga-resistant sorghums. *Field Crops Research* 66: 195-211.
- Hess, D.E. & B.I.G. Hausmann, 1999. Status quo of striga control: prevention, mechanical and biological control methods and host plant resistance. In J. Kroschel, H. Mercer-Quarshie & J. Sauerborn (Eds), *Advances in Parasitic Weed Control at On-farm Level. Volume 1, Joint Action to Control Striga in Africa*. Margraf Verlag, Weikersheim, pp. 75-87.
- Knorr-Held, L. & J.E. Besag, 1998. Modelling risk from a disease in time and space. *Statistics in Medicine* 17: 2045-2060.
- Knorr-Held, L. & H. Rue, 2000. On block updating in Markov random field models for disease mapping. Discussion Paper 210. Institute of Statistics, University Munich.
- Künsch, H.R., 1987. Intrinsic autoregression and related models on the two-dimensional lattice. *Biometrika* 74: 517-524.
- Manu, A., A. Bationo & S.C. Geiger. 1991. Fertility status of millet producing soils of West Africa with emphasis on phosphorus. *Soil Science* 152: 315-320.
- Mugglin, A.S., N.A. Cressie & I. Gemmell, 2000. Hierarchical statistical modelling of influenza-epidemic dynamics in space and time. Technical Report 662. Department of Statistics, Ohio State University.
- Møller, J., A.R. Syversveen & R.P. Waagepetersen, 1998. Log Gaussian Cox processes. *Scandinavian Journal of Statistics* 25: 451-482.
- Parker, C. & C.R. Riches, 1993. *Parasitic weeds of the world: biology and control*. CAB International, Wallingford, 332 pp.

