



Spatial prediction of weed intensities from exact count data and image-based estimates

Gilles Guillot,

Institut National de la Recherche Agronomique—AgroParisTech, Paris, France, Chalmers University of Technology, Gothenburg, Sweden, and University of Oslo, Norway

Niklas Lorén

Swedish Institute for Food and Biotechnology, Gothenburg, Sweden

and Mats Rudemo

Chalmers University of Technology, Gothenburg, Sweden

[Received November 2006. Final revision December 2008]

Summary. Collecting weed exact counts in an agricultural field is easy but extremely time consuming. Image analysis algorithms for object extraction applied to pictures of agricultural fields may be used to estimate the weed content with a high resolution (about 1 m²), and pictures that are acquired at a large number of sites can be used to obtain maps of weed content over a whole field at a reasonably low cost. However, these image-based estimates are not perfect and acquiring exact weed counts also is highly useful both for assessing the accuracy of the image-based algorithms and for improving the estimates by use of the combined data. We propose and compare various models for image index and exact weed count and we use them to assess how such data should be combined to obtain reliable maps. The method is applied to a real data set from a 30-ha field. We show that using image estimates in addition to exact counts allows us to improve the accuracy of maps significantly. We also show that the relative performances of the methods depend on the size of the data set and on the specific methodology (full Bayes versus plug-in) that is implemented.

Keywords: Approximate Cox process; Gaussian random field; Image analysis; Model-based geostatistics; Multivariate data; Poisson regression; Precision farming; Spatial prediction

1. Statistical approaches for weed mapping

The increased need to protect the environment, in particular to avoid spreading chemical products where it is not necessary, has prompted farmers to develop practices that have been adapted to the characteristics of soil and vegetation at small scale in the field. Regarding the spread of herbicides, varying the amount of product supplied locally requires a map of weed content in the field at a fairly high resolution.

The statistical approaches that have been developed to obtain such maps may be based on exact weed counts only. The earlier approaches to obtain maps from weed counts have only been based on traditional geostatistical methods. Implementation of such methods can be found in the works of Heisel *et al.* (1996), Rew and Cousens (2001), Rew *et al.* (2001) and Dille *et al.*

Address for correspondence: Gilles Guillot, Centre for Ecological and Evolutionary Synthesis, Department of Biology, University of Oslo, PO Box 1066, Blindern 0316, Oslo, Norway.
E-mail: gilles.guillot@bio.uio.no

(2002), where it is reported that simple interpolation procedures often behave poorly for weed mapping when the number of samples that are available is restricted by economical constraints (fewer than 10 data points per hectare).

Other methods that are based on the use of exact counts and auxiliary variables correlated with the weed content such as organic matter in soil or remote sensing estimates have been considered (see for example Heisel *et al.* (1999) and Lamb *et al.* (1999)), but the improvement compared with the use of exact counts depends only on the strength of the relationship between exact counts and covariates, which is usually quite low.

Strategies that are based on analysing pictures of the ground at small scale (about 1 m²) to obtain estimates of weed (or crop) content have been widely studied; see for example Marchant and Brivot (1995), Brivot and Marchant (1996, 1998), Andreasen *et al.* (1997), Perez *et al.* (2000), Soille (2000), Tillett *et al.* (2001), Åstrand and Baerveldt (2002), Onyango and Marchant (2003) and Gerhards and Christensen (2003). In the previous works, high resolution images are supposed to be available at an arbitrary site. Hence, the question of spatial interpolation has not been tackled in these works; the accuracy of the image estimates at a particular site is assessed by comparing its output with the weed content at this particular site. However, in practice collecting and processing a very large number of pictures can be prohibitive in terms of time, and visiting the whole field at a late stage of growth can damage the crop. Consequently, the question of how a large but limited number (say up to a few tens per hectare) of such image-derived estimates should be used to obtain interpolated maps should be addressed. Additionally, the quality of the estimation of weed content by images only can vary considerably. Hence it can be informative to collect also exact weed count data to assess the accuracy of the image analysis algorithm, and possibly to combine exact count and image-derived estimates in the interpolation scheme to increase its accuracy.

The aim of the present paper is to propose and implement a statistical model to address these two tasks. We base our analysis on a data set of weed counts and image-derived estimates that is described in the next section. In Section 3 we describe two variants of a Bayesian model aimed at capturing joint spatial variations of weed count and image estimates and we also discuss inference and prediction issues. Implementation of the approach on real data can be found in Section 4. We conclude with a discussion on the interest of the various approaches that are considered and give some practical recommendations.

The data that are analysed in the paper can be obtained from

<http://www.blackwellpublishing.com/rss>

2. A weed data set

Our analysis is based on data that were collected at the Bjertorp farm located 58.26° N–13.13° E in the south-west of Sweden. The data have been collected in a cultivated wheatfield of 30 ha and consist of exact counts in frames of 0.5 m × 0.75 m taken at 100 sites and estimates derived from pictures that were taken exactly over these frames. The most abundant species were common chickweed (*Stellaria Media*), field pennycress (*Thlaspi Arvense*), field pansy (*Viola Arvensis*), knotgrass (*Polygonum Aviculare*) and black bindweed (*Fallopia Convolvulus*), but we did not make use of the individual species counts, and numbers that are reported correspond to the total number of non-crop plants that were observed in the frame. Fig. 1 shows the measurement sites (an almost regular grid over the field plus an area with irregularly spaced sample points with higher density) and the spatial variations of weed counts.

Images were recorded with a digital camera mounted on a tripod at breast height. The images were recorded at a resolution of 3008 pixels × 2000 pixels. The image analysis algorithm that

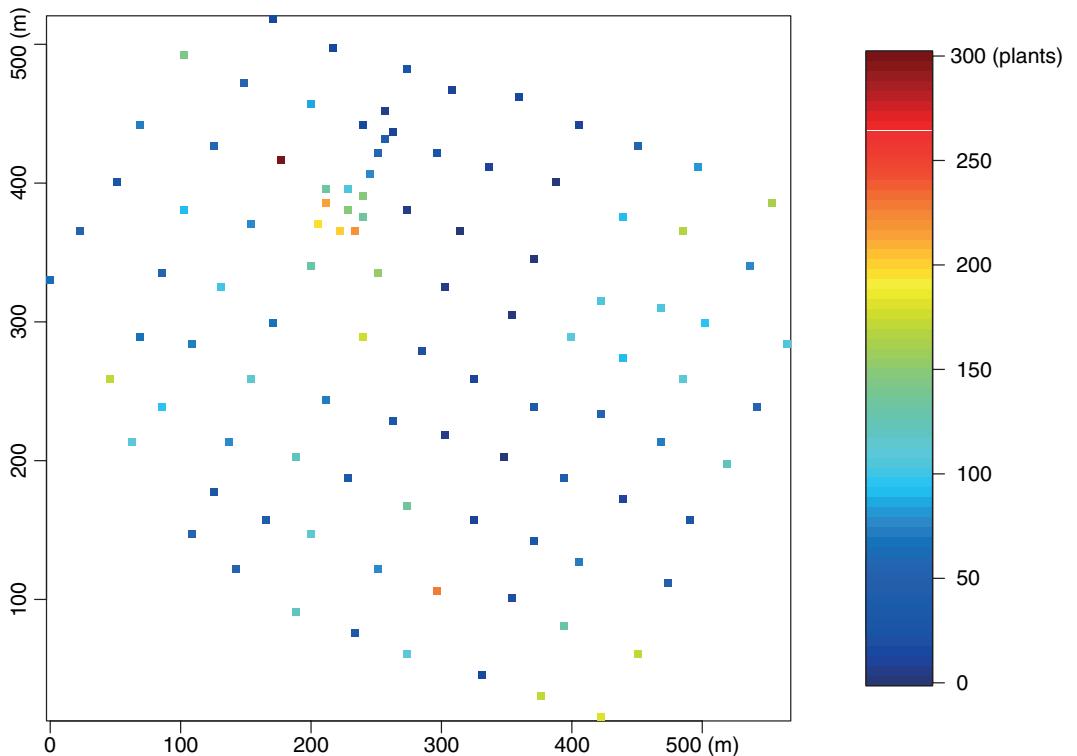


Fig. 1. Spatial distribution of the exact weed count; the colour code denotes the number of observed plants

was implemented to detect and count weeds involves the following steps: soil and plants were segmented by using the excess green transform (Woebbecke *et al.*, 1995) and subsequent thresholding (Otsu, 1979). Small objects were considered as noise and removed from the binary image. Crop rows were detected by using the Hough transform (Marchant, 1996) and weed plants located between the crop rows and uncovered by crop were found by using masking (Russ, 1994). Finally, large weeds covered by crop straws were extracted by using combinations of morphological operations (Soille, 2000). The details about the image analysis algorithm will be presented in a forthcoming paper. Fig. 2 shows a graphical illustration of the main steps of the algorithm.

The exact counts are integers ranging from 4 to 300, and the image estimates are also discrete and range between 7 and 222. The spatially structured nature of the weed counts appears more clearly on the empirical variogram of weed counts that is shown in Fig. 3 for various binnings. It displays a decrease beyond 150 m, a phenomenon which is known as the hole effect in geostatistics (see for example Chilès and Delfiner (1999), page 55) that occurs when high values tend to be systematically surrounded by low values, but a good fit of this empirical variogram at short-to-medium distances is obtained with a theoretical exponential model with a spatial scale parameter equal to 50 m.

Fig. 4 gives the histograms for the exact counts and the image estimates and also pair plots of these two variables. Although these histograms of spatially correlated data can be misleading, they suggest that a crude Gaussian assumption would not be suitable. A non-Gaussian model with asymmetric marginal distribution and cross-correlation will be proposed in the next section.

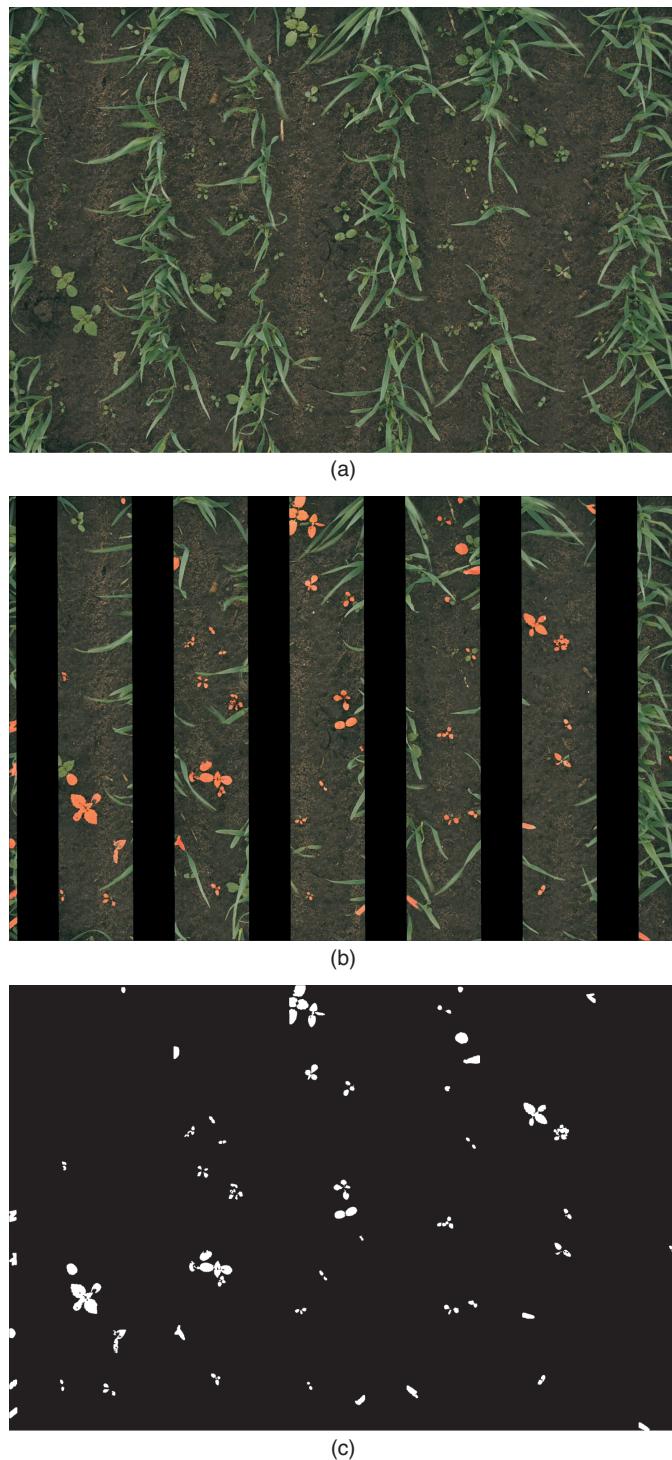


Fig. 2. Illustration of the image processing algorithm: (a) raw colour image; (b) detection of rows by the Hough transform; (c) weeds detected

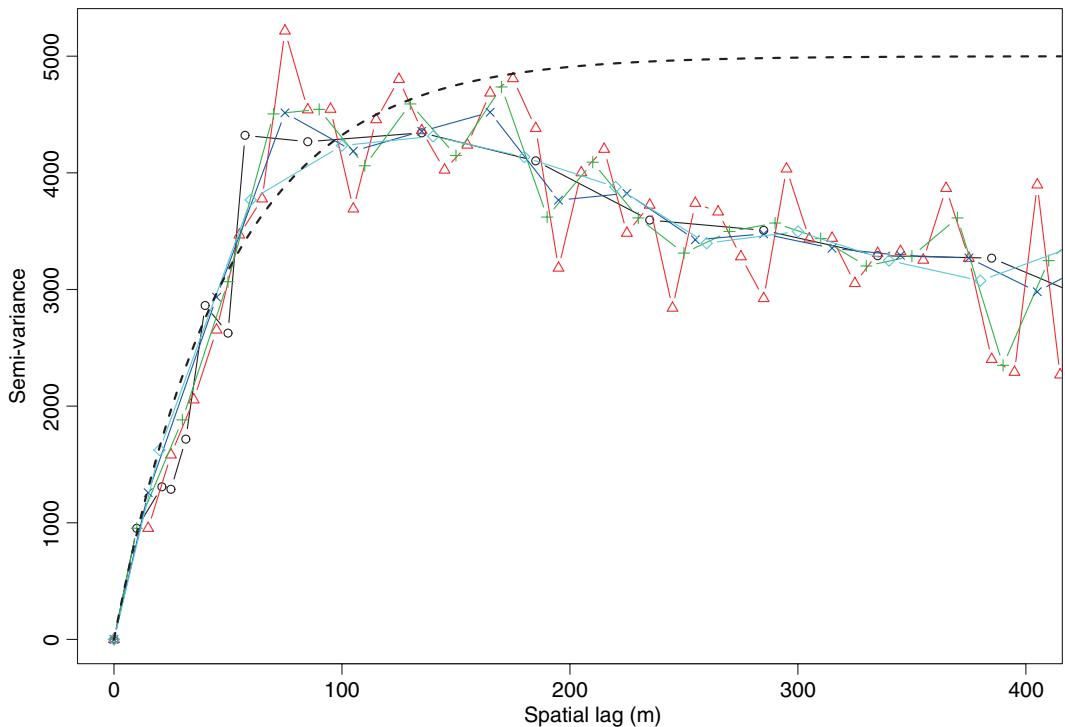


Fig. 3. Empirical variograms of the exact count data $c(s)$ for various binnings of the spatial lag, and exponential model (-----) with scale parameter equal to 50 m

3. Modelling

Exact weed counts are observed at a set of sites $\mathcal{S} = (s_1, \dots, s_{n_s})$ and denoted by $\mathbf{c}_{\mathcal{S}} = (c(s_1), \dots, c(s_{n_s}))$; image estimates are given at a set of sites $\mathcal{T} = (t_1, \dots, t_{n_t})$ and denoted by $\mathbf{i}_{\mathcal{T}} = (i(t_1), \dots, i(t_{n_t}))$. We also denote by $\mathcal{U} = (u_1, \dots, u_{n_u})$ the set of sites at which spatial prediction of the weed field c is sought. In our data set, the sets \mathcal{S} and \mathcal{T} are identical, but we consider the case (which is highly relevant in practice) where the two sets do not necessarily coincide.

3.1. Model for exact count

Here we consider two models where the spatial variations are essentially captured by a Gaussian random field.

3.1.1. Transformed Gaussian random-field model

A first simple model accounting for spatial auto-correlation and non-Gaussian marginal distribution is the transformed Gaussian random-field (TGRF) model. This model assumes that there is a zero-mean, unit variance, Gaussian random field $y(s)$, $s \in \mathbb{R}^2$, and a function $\psi: \mathbb{R} \rightarrow \mathbb{R}$ such that $c(s) = \psi\{y(s)\}$. This assumption is common in geostatistics (De Oliveira *et al.*, 1997; Chilès and Delfiner, 1999; Diggle *et al.*, 2003) and sometimes for data on grids (Allcroft and Glasbey, 2003). It makes it possible to account for non-Gaussian marginal distributions while keeping the parsimony and flexibility of Gaussian random fields for the modelling of spatial variations.

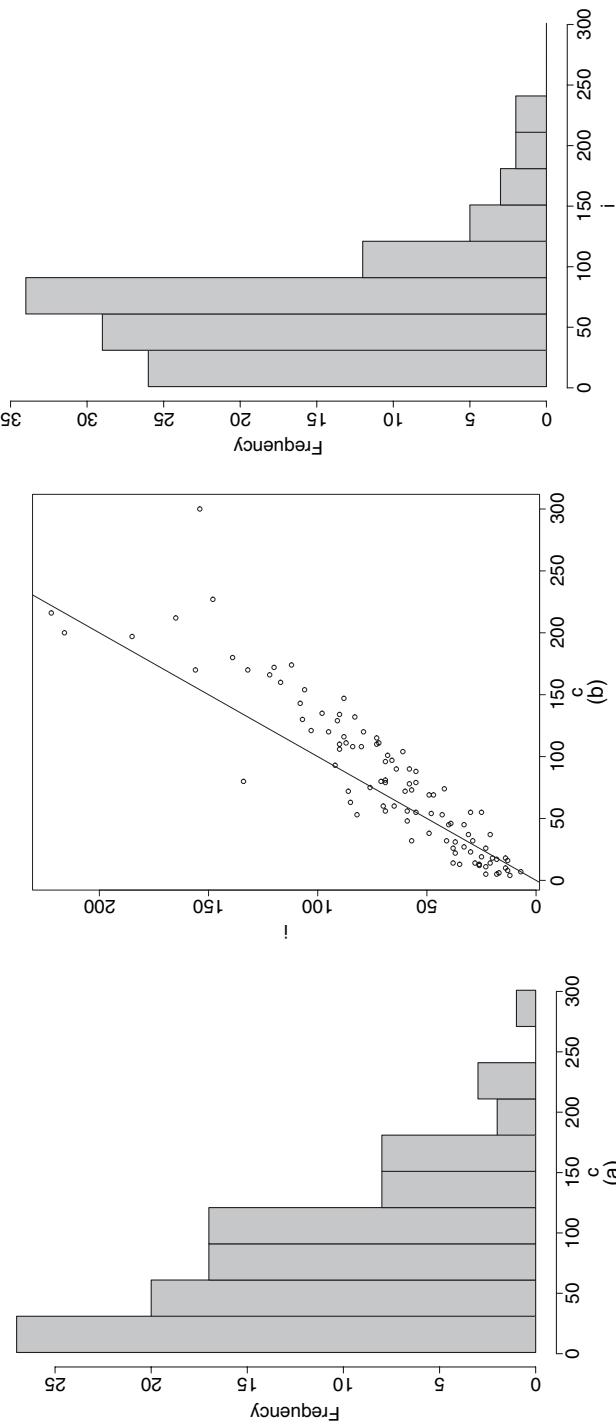


Fig. 4. (a) Histogram of exact weed counts, (b) pair plot of image estimates versus exact counts and (c) histogram of image estimates

We assume that y has a stationary and isotropic correlation function $\rho(h)$, with exponential decay, namely

$$\text{corr}\{y(s), y(s+h)\} = \exp(-||h||/\kappa) \quad (1)$$

for some unknown spatial correlation parameter κ .

This model is common in geostatistics. Although the spatial correlation structure is formulated in terms of y whereas the empirical variogram that is available is that of c (Fig. 3), it is known (see for example De Oliveira (2003)) that, for a large class of continuous transformations ψ , the variograms of the latent Gaussian field and of the observed field are quite similar. Hence, it is expected that the exponential model will give a good fit, at least for short to medium distances (see Section 4.1).

It is sometimes recommended to use an extra parameter to quantify the smoothness properties of the field (e.g. within a model of the Whittle–Matérn family). In addition, it would be natural to consider an anisotropic model (the dispersal process of weed probably being much influenced by the ploughing process). The inference of three extra parameters (a smoothing parameter, the angle and the ratio for geometrical anisotropy) adds no difficulty in principle. But it can be expected that it would not improve the results in terms of inference (in particular with the use of a data set with few exact counts), whereas it would increase the computational burden. Hence we choose to keep a reasonably simple one-parameter model for the correlation function.

3.1.2. Approximate transformed Gaussian Cox process

The previous model places a continuous distribution on c although it is a discrete variable by nature. A simple way to avoid this inconsistency is to assume that the previous transformed Gaussian field defines an intensity field and that, conditionally on $w = \psi(y)$, $c(s)$ has an independent Poisson $\{w(s)\}$ distribution. This turns out to be an approximate Cox process with transformed Gaussian intensity and a special case of a spatial generalized linear model. We refer to this model as a transformed Gaussian Cox process (TGCP) in what follows (disregarding the approximation in the terminology). This second approach is an example of the general modelling framework that was proposed by Diggle *et al.* (1998) and is also similar to that taken by Christensen and Waagepetersen (2002) where the weed counts are assumed to arise from a spatial generalized linear mixed model with a Gaussian random field as random effect, and covariates are used to model a deterministic trend of the intensity.

We note here that, under this model with Poisson regression relating c and w , we have

$$\text{cov}\{c(s), c(s+h)\} = \text{cov}\{w(s), w(s+h)\} \quad \text{for } h > 0 \quad (2)$$

and

$$\text{var}\{c(s)\} = E[w(s)] + \text{var}\{w(s)\} \quad (3)$$

implying a discontinuity of the covariance function at 0 (a property which is known as the nugget effect in geostatistics). The jump at 0 is given by the mean of the intensity field $E[w]$, so this parameter controls jointly two very different aspects of the model.

Regarding ψ , we considered functions such that $\psi(y)$ has gamma or log-normal marginal distribution. We denote by μ and σ^2 the stationary mean and variance of $\psi(y)$. The choice of the marginal distribution will be discussed later, together with a comparison between the TGCP and TGRF models.

3.2. Modelling the relationship between exact counts and image estimates

To capture the relationship between the two variables and the observed heteroscedasticity, we assume a Poisson regression model for the image estimate:

$$i(s)|c \sim \text{Poisson}\{\gamma c(s)^\delta\} \quad (4)$$

where γ and δ are parameters to account for a slight departure from linearity as can be seen in Fig. 4.

3.3. Prior distributions

The vector of unknown parameters is $\theta = (\kappa, \mu, \sigma, \gamma, \delta)$. All parameters take their value on $[0, \infty)$. We assume proper gamma priors for all of them. Each such distribution has shape and rate parameters a_\cdot and b_\cdot (where ‘ \cdot ’ is $\kappa, \mu, \sigma, \gamma$ or δ). We assume *a priori* independence of the components of the vector of parameters, namely

$$\begin{aligned} \pi(\theta) &= \pi(\mu) \pi(\sigma) \pi(\kappa) \pi(\gamma) \pi(\delta) \\ &\propto \mu^{a_\mu-1} \exp(-\mu b_\mu) \sigma^{a_\sigma-1} \exp(-\sigma b_\sigma) \kappa^{a_\kappa-1} \exp(-\kappa b_\kappa) \gamma^{a_\gamma-1} \exp(-\gamma b_\gamma) \delta^{a_\delta-1} \exp(-\delta b_\delta). \end{aligned} \quad (5)$$

Having no previous information on what sensible values for the parameters θ should be, we chose the parameters a_\cdot and b_\cdot in such a way that a_\cdot/b_\cdot^2 is large compared with a_\cdot/b_\cdot (flat priors). We used moment estimates as an expedient to set the values of a_\cdot/b_\cdot ; an approach that is common in practice; for example Eidsvik *et al.* (2006). The parameters a_κ and b_κ were chosen to have a mean of 80 m corresponding to a rough empirical estimate from Fig. 3. We chose a_μ, b_μ, a_σ and b_σ so that $E[\mu]$ and $E[\sigma]$ comply with the moment estimate of μ and σ , ignoring the spatial correlation in the weed counts. Further, $a_\gamma, b_\gamma, a_\delta$ and b_δ were chosen to have $E[\gamma] = E[\delta] = 1$. Each such constraint gives an equation for each pair of parameters (a_\cdot, b_\cdot) , a second equation being obtained by setting the variance of the parameter to an arbitrary large value. Note also that the prior on κ is actually discretized on a fine regular grid and truncated at an upper value κ_{\max} . We checked that the choice of the ratio a_\cdot/b_\cdot does not affect the result of the inference for a broad range of values. Further discussion on the choice of prior for such spatial models can be found in Berger *et al.* (2001) and Diggle *et al.* (2003).

3.4. Inference and prediction

We perform jointly inference on the parameter vector θ and prediction of the vector of weed values \mathbf{c}_U . This is achieved through simulation of the joint distribution $\pi(\theta, \mathbf{c}_U | \mathbf{c}_S, \mathbf{i}_T)$. However, in the two models that were considered, inference and prediction are greatly simplified by the use of model structure and the introduction of auxiliary variables. Instead of making prediction of c at sites \mathcal{U} only, we also make prediction of c at sites $\mathcal{T} \setminus \mathcal{S}$. For the TGCP model, we also introduce the latent variable w which we sample at all the sites in $\mathcal{S} \cup \mathcal{T} \cup \mathcal{U}$. Then the algorithm that we use involves different updates of either Metropolis–Hastings or Gibbs type, performed in a deterministic order as described in Appendix A.

It is known that in such hierarchical models (see for example Diggle *et al.* (2003) for an explicit example) even under proper prior–likelihood models with finite moments, the posterior or the predictive distributions may not have finite first-order moments. Thus we use the median of the simulated sample for inference and prediction.

4. Implementation on real data

4.1. Bjertorp data set: inference, convergence and goodness-of-fit assessment

4.1.1. Gamma versus log-normal transform

We checked that the gamma and the log-normal distribution gave similar results in terms of goodness of fit for the TGCP and the TGRF model. Hence, we assume hereafter that $\psi(y)$ has a gamma marginal distribution both for the TGRF and the TGCP model. The estimated values for the parameter α of the gamma distribution that is involved in the weed counts models are both small ($\hat{\alpha}_{\text{TGRF}} = 1.03$ and $\hat{\alpha}_{\text{TGCP}} = 1.41$), corresponding to a fairly high amount of transformation of the underlying Gaussian distribution.

4.1.2. Convergence assessment

We implemented the previous algorithm on the Bjertorp data set for the TGRF and the TGCP models. We launched 20 independent runs with different initial conditions for the components of θ sampled from the prior. To detect potential flaws in convergence, we computed the Gelman–Rubin statistics testing the significance of the difference between the within-chain variance and the between-chain variance. No flaw in convergence was detected (values under the 97.5% quantile). Histograms of Markov chain Monte Carlo samples of the parameters for the two models are displayed in Fig. 5.

4.1.3. Transformed Gaussian random-field model versus transformed Gaussian Cox process model

We checked the fit of the TGRF model and the TGCP model to data (within the global model including image estimate data) as follows: we plotted the average posterior variogram of weed counts and compared it with the empirical variograms (obtained for various binnings). We compared the distribution of a sample of $\mathbf{h} = L_{\kappa}^{-1}\mathbf{y}$ from the predictive distribution with the assumed normal distribution (where $L_{\kappa}U_{\kappa}$ is the Choleski decomposition of the correlation matrix of \mathbf{y}). We checked the fit between the image index values and a sample from their predictive distribution by a pair plot and a quantile–quantile plot.

This check is not rooted in a formal statistical testing (see for example Waagepetersen (2006) for a more formal approach) but it suggests that the TGCP and TGRF models give a similar fit regarding image index data. In contrast, regarding the weed counts, the TGRF model gives a better fit than the TGCP model (see the fitted variogram in Figs 6(d) and 7). This relatively poor fit of the variogram under the TGCP model might be explained by a bad specification of the behaviour of the variogram near 0. Indeed, the TGCP model assumes a nugget effect whereas such a nugget effect seems to be absent from the empirical variogram whatever the binning that is considered. The overestimation of the scale parameter might be a consequence of the need to reconcile the non-smooth variations that are induced by the nugget effect with the quite smooth variations that are observed on weed counts.

4.2. Assessment of prediction accuracy

4.2.1. Setting for the transformed Gaussian Cox process and transformed Gaussian random-field models

We compared the TGCP model and the TGRF model in terms of accuracy of prediction of weed counts on various subsamples of the Bjertorp data set. We considered random subsamples with various numbers of exact count and image estimate sites, namely 99–0, 30–0, 10–0, 30–99 and 10–99. We considered 250 subsamples for each case, except for the combination 99–0

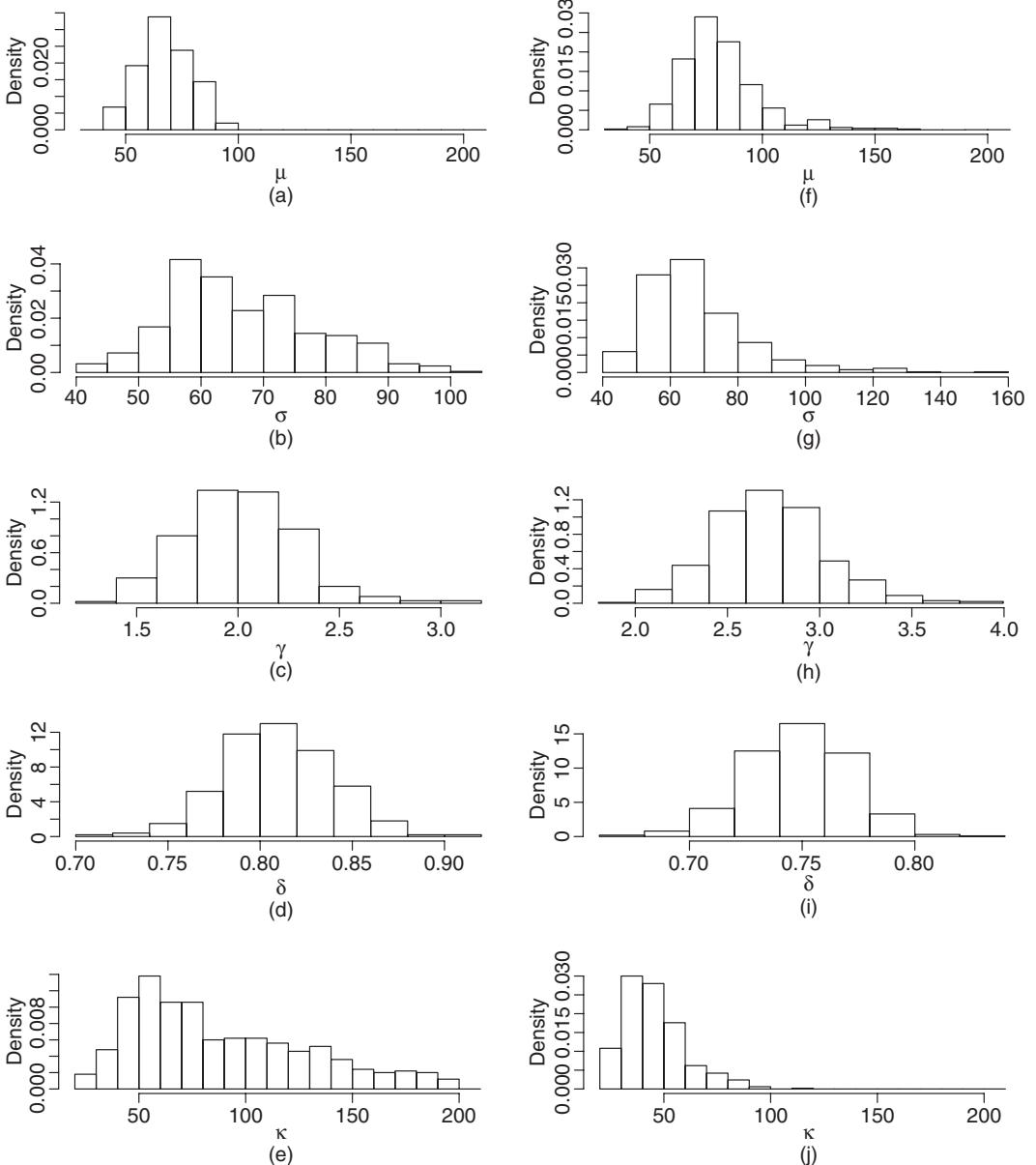


Fig. 5. Posterior distribution of the parameters for a random subsample of the Bjertorp data set with 30 exact counts and 100 images: (a) TGCP, average μ 67.7; (b) TGCP, average σ 66.6; (c) TGCP, average γ 2.01; (d) TGCP, average δ 0.81; (e) TGCP, average κ 89.3; (f) TGRF, average μ 80.5; (g) TGRF, average σ 67.7; (h) TGRF, average γ 2.74; (i) TGRF, average δ 0.75; (j) TGRF, average κ 46.5

where there are only 100 subsamples. Prediction was carried out for the left-out sites (one, 30 or 70 sites) with the TGCP model and the TGRF model. Prediction was carried out in full Bayes mode (parameters considered as unknown) but, to assess the effect of uncertainty about parameters, we also performed prediction with plugged-in parameters from an off-line estimation.

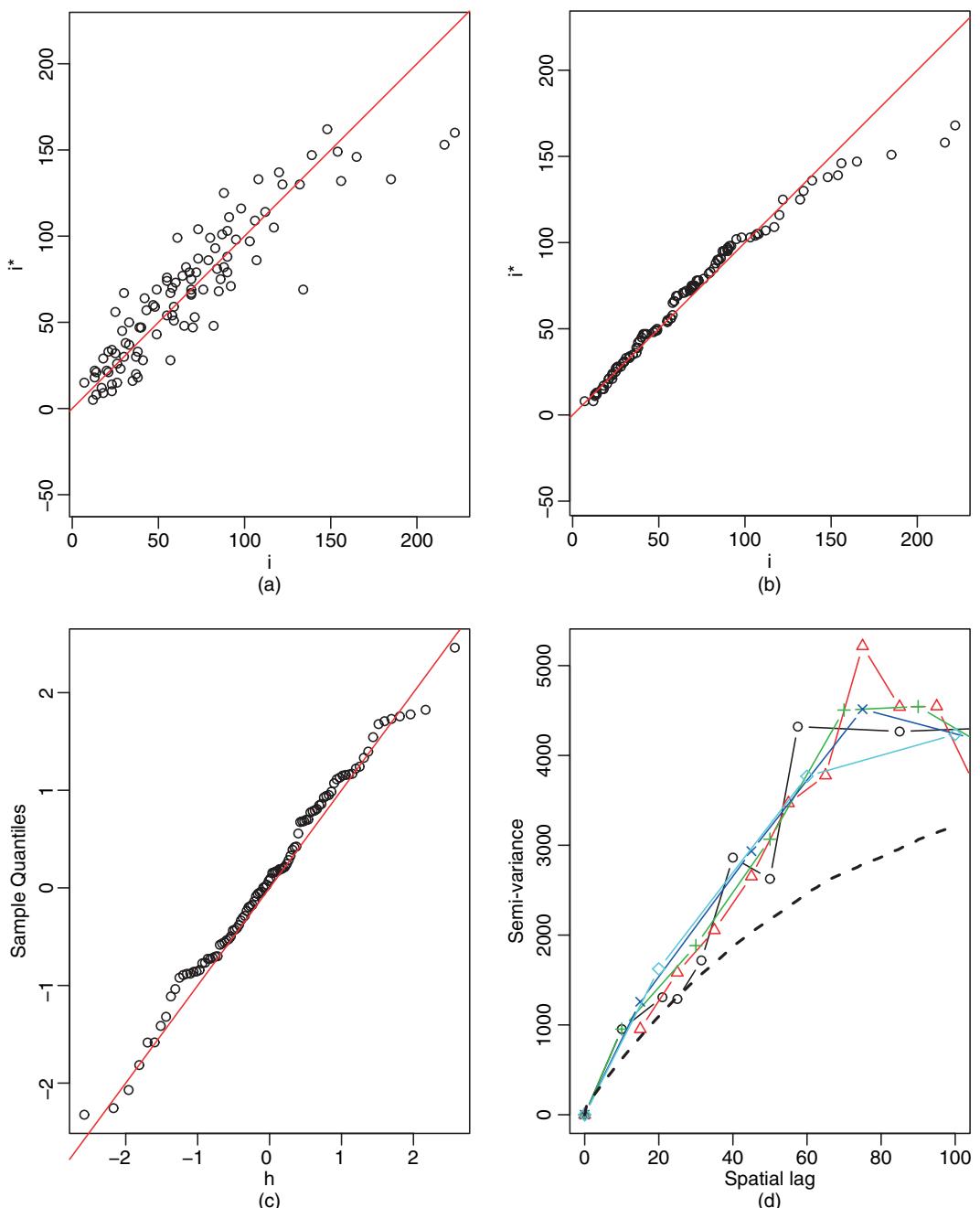


Fig. 6. Goodness-of-fit plots for the TGCP model: (a) pair plot of a sample of the predictive distribution of i against data; (b) quantile-quantile plot of a sample of the predictive distribution of i against data; (c) quantile-quantile norm plot of a sample $\mathbf{h} = L_{\kappa}^{-1}\mathbf{y}$ from the predictive distribution; (d) average posterior variogram of weed counts (— —) and empirical variograms for various binnings

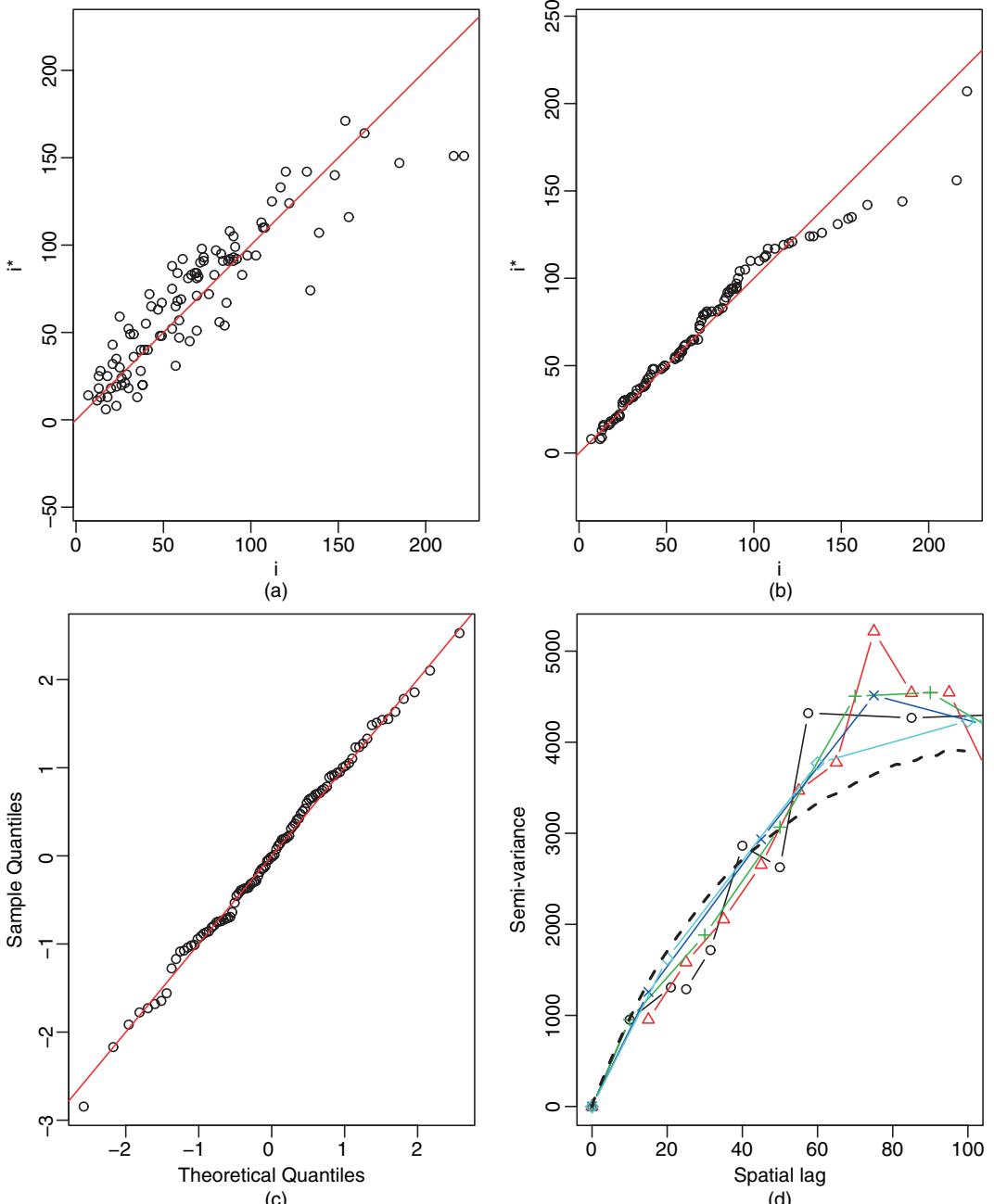


Fig. 7. Goodness-of-fit plots for the TGRF model: (a) pair plot of a sample of the predictive distribution of i against data; (b) quantile-quantile plot of a sample of the predictive distribution of i against data; (c) quantile-quantile norm plot of a sample $\mathbf{h} = L_{\kappa}^{-1}\mathbf{y}$ from the predictive distribution; (d) average posterior variogram of weed counts (—) and empirical variograms for various binnings

Table 1. Mean-square prediction error for subsamples of the Bjertorp data set[†]

| <i>Spatial sampling design</i> | <i>Mean-square prediction errors for the following processes:</i> | | |
|--------------------------------|---|-------------|---------------------|
| | <i>TGCP</i> | <i>TGRF</i> | <i>(Co-)kriging</i> |
| 99 counts, no image | 2710 (2510) | 2417 (2404) | 2373 |
| 30 counts, no image | 3409 (3255) | 2932 (2862) | 2878 |
| 10 counts, no image | 5522 (4439) | 3831 (3561) | 3743 |
| 30 counts, 99 images | 2426 | 2211 | 1998 |
| 10 counts, 99 images | 4340 | 3240 | 2358 |

[†]Numbers in italics are mean-squared errors obtained when parameters estimated off line are plugged into the predictor.

4.2.2. Setting for kriging and co-kriging

We also implemented a linear spatial predictor (kriging or co-kriging). The variant that was used was ordinary kriging. It assumes stationarity of the mean and variance but accounts for unknown mean. It was performed following the usual practice in geostatistics (Chilès and Delfiner, 1999) consisting in using plugged-in parameters. We assumed an exponential covariance function for the exact counts. For combinations of data using image estimates, we implemented co-kriging with a linear co-regionalization model (Wackernagel, 1995). This was implemented with the R package RandomFields (Schlather, 2006). We also tried kriging on the log-transformed and cubic transformed data (and back-transform), a technique introducing a bias in the prediction but potentially reducing the mean-squared error. It appeared that the mean-squared error was actually slightly larger than that obtained with plain kriging. Hence we do not report numerical results for that. The results in terms of mean-square prediction error are reported in Table 1.

5. Discussion

5.1. Use of image-derived estimates

The first conclusion of this study is that using image estimates clearly improves the accuracy of predictions compared with those obtained with exact counts only. We assumed throughout this study that some of the image data were collected at the same sites as exact count data. All our computations are based on such data and it has been observed on simulated data where $\mathcal{S} \cap \mathcal{T} = \emptyset$ that the predictions were less accurate than those where $\mathcal{S} \subset \mathcal{T}$ in the TGRF and TGCP models, in particular because of poor estimates of the parameters γ and δ in the Poisson regression. However, the requirement that $\mathcal{S} \subset \mathcal{T}$ is not very constraining in practice and this should not be considered as a drawback of our study. We also note that the Poisson regression model linking images and counts assumes the relationship

$$\text{var}(i) = \gamma E[c^\delta] + \gamma^2 \text{var}(c^\delta). \quad (6)$$

In particular, for γ and δ close to 1, we have $\text{var}(i) \approx E[c] + \text{var}(c)$, implying that $\text{var}(i) > \text{var}(c)$. In the present data set, the empirical variances of i and c rank in the opposite way, which suggests that the manual counting process involves some kind of smoothing. Our model is therefore not perfect and perhaps a little pessimistic and further improvement could probably be obtained with a model enforcing a tighter relationship between i and c .

5.2. Prediction under complex models with small data sets

The second result of our analysis is that, for the sample sizes that were considered (within a range from small to moderate), the simple models (kriging and TGRF) give more accurate prediction than the most complex (and potentially more realistic) TGCP model. The difference that was observed between kriging and TGRF on one hand and TGCP on the other hand is limited when many data points are available and increases when the size of the data set decreases. The largest difference is observed with a small number of exact counts, where kriging and TGRF still perform well whereas TGCP brings large prediction errors. Figs 5(e) and 5(j) reveal that the posterior distribution of the scale parameter κ is flatter for the TGCP model than for the TGRF, giving a non-negligible weight to large values of κ . This complies well with the modelling assumptions. Indeed, under the TGRF model, c (data) and y (Gaussian field) are related through a deterministic relationship. In contrast, under the TGCP model, the data do not relate directly to the Gaussian random field but only through a conditional probability dependence: hence the increased difficulty to learn from the data and to transfer information across space appropriately.

5.3. Influence of predictor

Although the existence of a first-order moment in the TGCP and TGRF models is not guaranteed, to check that the results were not due to the use of the median as predictor, we also made prediction by computing numerically the posterior mean on the simulated samples. We observed a small difference with $|c_{\text{mean}}^* - c_{\text{median}}^*|/c_{\text{median}}^*$ consistently below 1% both for TGRF and for TGCP predictions.

5.4. Full Bayes versus plug-in prediction

Kriging prediction was performed with plug-in parameters. This contrasts with the method that we used for the Bayesian models consisting in estimating parameters and making predictions jointly. The discrepancy between TGRF and kriging seems to be largely explained by this fact as the use of plug-in parameters in TGRF prediction decreases the mean-squared error to a value that is comparable with those obtained with kriging. This is surprising and the reasons for that are not fully understood. We also note that inference under the TGCP model requires Markov chain Monte Carlo computations as the intensity field w is not observed. In contrast, the off-line (Markov chain Monte Carlo free) inference of parameters under the TGRF model is possible as long as there is some overlap between exact weed counts locations and image locations. Indeed, if weed counts and images are observed at a set of common sites, then γ and δ can be estimated easily. Then parameters of the random field c can be estimated following standard geostatistical methods. This is fortunate for the TGRF model as this allows us to reduce the prediction error, which is often the main goal of weed studies based on such models.

5.5. Continuous versus discontinuous covariance function

In the case of kriging, we assumed that the exact counts have an exponential covariance function which is continuous on \mathbb{R} . For the TGRF model, the assumption of an exponential covariance function refers to the underlying Gaussian random field y . In that case, the covariance of the exact weed count c cannot be obtained analytically, and Hermite polynomials expansions are used to derive theoretical properties (Chilès and Delfiner, 1999). But, here, the continuity of the transform ϕ guarantees the continuity of the covariance function of c . In contrast, under the TGCP model the covariance of the weed count is discontinuous (see Section 3.1.2). The nugget effect (a jump at 0) is equal to the mean of the intensity field, whose estimated value is

80.5. This value is small but might explain partly the lower performance of the TGCP model for prediction.

5.6. Conclusion

The factors affecting the performances of the various approaches are

- (a) a better robustness of kriging due to parsimony,
- (b) a slight (TGRF) to moderate (TGCP) sensitivity of Bayesian approaches to parameter uncertainty,
- (c) the weak identifiability of parameters under the TGCP model (in particular the scale parameter κ of the hidden Gaussian random field) and
- (d) the better performance for prediction of models based on a continuous covariance function (TGRF, kriging).

In terms of practical recommendations, the conclusions to draw from our work seem to be that, with data sets that are similar to the one which we studied in terms of sample size, spatial design and statistical characteristics, it is good to use image estimates. Prediction should then preferably be done with a simple model such as kriging or a TGRF model. We also noted that kriging seems to be more robust than TGRF, in particular in the case where the parameters are unknown. However, making prediction under a TGRF model allows us to derive not only point predictions but also probability distributions of parameters, which can be useful even when the focus is on prediction rather than on inference, e.g. for combining information from various data sets in a meta-analysis. The interest of more sophisticated Bayesian models such as TGCP models lies potentially on the side of inference rather than on the side of prediction, for instance if a quantification of a mean intensity is required which can be interpreted in terms of local weed growth potential of the field.

5.7. Computer program

The program that was developed during this work is available from the Comprehensive R Archive Network as an R package named ([WeedMap](http://cran.r-project.org/packages/WeedMap)<http://cran.r-project.org/packages/WeedMap>.)

Acknowledgements

This work has been carried out within the project ‘Image analysis as a tool for site specific weed control’ funded by the Swedish Farmers Foundation for Agricultural Research. Niklas Lorén acknowledges the Foundation for financial support during this work. Mats Rudemo was supported by the Swedish Foundation for Strategic Research through the Gothenburg Mathematical Modelling Centre project and the Swedish Research Council through the Gothenburg Stochastic Centre. Anders Lars-Olle (Swedish Agricultural University), Mats Söderström (Lantmännen–Sweco) and Thomas Börjesson (Lantmännen) are thanked for the preparation of the geographic information systems data and valuable discussions. We also thank our colleagues Anastassia Baxevani, Jacques de Maré, Antti Penttinen, Petter Mostad, Aila Särkkä and Martin Schlather for comments, in particular during the Gothenburg Stochastic Centre workshop on ‘Spatial modelling’ in 2006 in Smögen.

Appendix A: Detail of Markov chain Monte Carlo computations

In the TGRF model, the vector of parameters and variables to be simulated is $(\mathbf{c}_{T \setminus S}, \mathbf{c}_U, \theta)$ with $\theta =$

$(\mu, \sigma, \kappa, \gamma, \delta)$, where μ is the mean of the field $\psi(y)$ and σ its standard deviation. In the TGCP model, there is an extra layer and the vector of parameters and variables to be simulated is $(\mathbf{c}_{T \setminus S}, \mathbf{c}_U, \mathbf{y}_S, \mathbf{y}_{T \setminus S}, \mathbf{y}_U, \theta)$. The steps that are involved in both algorithms are similar and we only describe below the updates for the TGCP model.

A.1. Updating exact counts c at image data sites in $T \setminus S$

We iterate in a deterministic order single-component updates $\mathbf{c}_{T \setminus S}$. The acceptance ratio for such a move is

$$a(\mathbf{c}_{T \setminus S}, \mathbf{c}_{T \setminus S}^*) = \mathbf{1} \wedge \frac{\pi(\mathbf{c}_{T \setminus S}^* | \mathbf{w}_{T \setminus S})}{\pi(\mathbf{c}_{T \setminus S} | \mathbf{w}_{T \setminus S})} \frac{\pi(\mathbf{i}_{T \setminus S} | \mathbf{c}_{T \setminus S}^*, \theta)}{\pi(\mathbf{i}_{T \setminus S} | \mathbf{c}_{T \setminus S}, \theta)} \quad (7)$$

where all the densities are Poisson distributions with independent components.

A.2. Updating exact counts c at prediction sites U

We iterate in a deterministic order single-component updates of \mathbf{c}_U . The acceptance ratio for such a move is

$$a(\mathbf{c}_U, \mathbf{c}_U^*) = \mathbf{1} \wedge \frac{\pi(\mathbf{c}_U^* | \mathbf{w}_U)}{\pi(\mathbf{c}_U | \mathbf{w}_U)}. \quad (8)$$

A.3. Updating the components of the vector $(\mathbf{y}_S, \mathbf{y}_{T \setminus S}, \mathbf{y}_U)$ of the hidden Gaussian field y

Denoting $\mathbf{y}_U = (\mathbf{y}_S, \mathbf{y}_{T \setminus S}, \mathbf{y}_U)$, $\mathbf{c}_U = (\mathbf{c}_S, \mathbf{c}_{T \setminus S}, \mathbf{c}_U)$ and $Q(\mathbf{y}_U) = \pi(\mathbf{c}_{T \setminus S}, \mathbf{c}_U, \mathbf{y}_U, \theta | \mathbf{c}_S, \mathbf{i}_T)$, we make Langevin proposals for \mathbf{y}_U , namely

$$\mathbf{y}_U^* = \mathbf{y}_U + \frac{\eta^2}{2} \partial \ln \{Q(\mathbf{y}_U)\} + \delta_{\mathbf{y}_U} \quad (9)$$

where $\partial \ln(Q)$ denotes the gradient of $\ln(Q)$ and $\delta_{\mathbf{y}_U}$ is a vector with independent Gaussian components with zero mean and variance η^2 .

The gradient is

$$\partial \ln \{Q(\mathbf{y}_U)\} = \partial \ln \{\pi(\mathbf{y}_U | \theta)\} + \partial \ln \{\pi(\mathbf{c}_U | \theta, \mathbf{y}_U)\}. \quad (10)$$

The first term is $-\Sigma y$ where Σ is the covariance matrix of \mathbf{y}_U . The second term is a vector whose generic term is $\{\mathbf{c}_k / \psi(\mathbf{y}_k) - 1\} \psi'(\mathbf{y}_k)$.

The move from \mathbf{y}_U to \mathbf{y}_U^* is accepted with probability

$$a(\mathbf{y}_U, \mathbf{y}_U^*) = \mathbf{1} \wedge \frac{\pi(\mathbf{y}_U^* | \theta) \pi(\mathbf{c}_U | \theta, \mathbf{y}_U^*) q(\mathbf{y}_U | \mathbf{y}_U^*)}{\pi(\mathbf{y}_U | \theta) \pi(\mathbf{c}_U | \theta, \mathbf{y}_U) q(\mathbf{y}_U^* | \mathbf{y}_U)} \quad (11)$$

where $\pi(\mathbf{y}_U | \theta)$ is a multivariate Gaussian density, $\pi(\mathbf{c}_U | \theta, \mathbf{y}_U)$ is a product of Poisson densities and

$$q(\mathbf{y}_U^* | \mathbf{y}_U) = g\left[\mathbf{y}_U^* - \mathbf{y}_U - \frac{\eta^2}{2} \partial \ln \{Q(\mathbf{y}_U)\}\right] \quad (12)$$

where g denotes the independent centred multivariate Gaussian density.

A.4. Updating the mean μ of the intensity field w

We propose a move from μ to μ^* by the addition of a small Gaussian increment. Since we let the values of the Gaussian random field y be unchanged, the values of the intensity field w are implicitly changed. The move is accepted with probability

$$a(\mu, \mu^*) = \mathbf{1} \wedge \frac{\pi(\mathbf{c}_U | \mathbf{y}_U, \theta^*)}{\pi(\mathbf{c}_U | \mathbf{y}_U, \theta)} \frac{\pi(\mu^*)}{\pi(\mu)}. \quad (13)$$

A.5. Updating the variance σ^2 of the intensity field w and the scale parameter κ of the hidden Gaussian field y

We propose a move from σ to σ^* by the addition of a small Gaussian increment and simultaneously a move from κ to κ^* by the addition of a small uniform increment on the discretized range of values that are spanned by κ . Here again, the values of the Gaussian random field y are unchanged; the values of the intensity field w are implicitly changed. The move is accepted with probability

$$a\{(\sigma, \kappa), (\sigma^*, \kappa^*)\} = \mathbf{1} \wedge \frac{\pi(y_{\cup}|\kappa^*)}{\pi(y_{\cup}|\kappa)} \frac{\pi(c_{\cup}|y_{\cup}, \theta^*)}{\pi(c_{\cup}|y_{\cup}, \theta)} \frac{\pi(\kappa^*)}{\pi(\kappa)} \frac{\pi(\sigma^*)}{\pi(\sigma)}. \quad (14)$$

A.6. Updating γ and δ

We update γ and δ jointly by addition of small Gaussian increments. It would be natural to compute the acceptance ratio accounting for all the variables that are available in the current state. However, this implies evaluation of a ratio of probabilities $\pi(\mathbf{i}|\mathbf{c}, \theta^*)/\pi(\mathbf{i}|\mathbf{c}, \theta)$ where some of the sites involved do not include image or count data, but values that are simulated in the current state of the chain. We observed that this could bring some numerical instabilities, in particular in the case where $|\mathcal{S}|$ is small with respect to $|\mathcal{T}|$. Hence we chose to compute the ratio disregarding the count and image values at sites $\mathcal{T} \setminus \mathcal{S}$ (brought by the data augmentation), and we used i - and c -values at sites in $\mathcal{S} \cap \mathcal{T}$ only. We found this strategy more efficient. It avoids numerical instabilities while it preserves the convergence to the desired target distribution of the global algorithm.

The move from (γ, δ) to (γ^*, δ^*) is therefore accepted with probability

$$a\{(\gamma, \delta), (\gamma^*, \delta^*)\} = \mathbf{1} \wedge \frac{\pi(\mathbf{i}_{\mathcal{T} \cap \mathcal{S}}|\mathbf{c}_{\mathcal{T} \cap \mathcal{S}}, \theta^*)}{\pi(\mathbf{i}_{\mathcal{T} \cap \mathcal{S}}|\mathbf{c}_{\mathcal{T} \cap \mathcal{S}}, \theta)}. \quad (15)$$

References

- Allcroft, D. J. and Glasbey, C. A. (2003) A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *Appl. Statist.*, **52**, 487–498.
- Andreasen, C., Rudemo, M. and Sevestre, S. (1997) Assessment of weed density at an early stage by use of image processing. *Weed Res.*, **37**, 5–18.
- Åstrand, B. and Baerveldt, A.-J. (2002) An agricultural mobile robot with vision-based perception for mechanical weed control. *Auton. Robots*, **13**, 21–35.
- Berger, J. O., De Oliveira, V. and Sansó, B. (2001) Objective Bayesian analysis of spatially correlated data. *J. Am. Statist. Ass.*, **96**, 1361–1374.
- Brivot, R. and Marchant, J. (1996) Segmentation of plants and weeds for a precision crop protection robot using infrared images. *IEEE Proc. Visn Image Signal Process.*, **143**, 118–124.
- Brivot, R. and Marchant, J. (1998) Real-time segmentation of plants and weeds. *Realtime Imagng*, **4**, 243–253.
- Chilès, J. and Delfiner, P. (1999) *Geostatistics*. New York: Wiley.
- Christensen, O. and Waagepetersen, R. (2002) Bayesian prediction of spatial count data using generalised linear mixed models. *Biometrics*, **58**, 280–286.
- De Oliveira, V. (2003) A note on the correlation structure of transformed Gaussian random fields. *Aust. New Zeal. J. Statist.*, **45**, 353–366.
- De Oliveira, V., Kedem, B. and Short, D. (1997) Bayesian prediction of transformed Gaussian random fields. *J. Am. Statist. Ass.*, **92**, 1422–1433.
- Diggle, P., Ribeiro, P. and Christensen, O. (2003) Spatial statistics and computational methods. *Lect. Notes Statist.*, **173**, 43–86.
- Diggle, P. J., Tawn, J. A. and Moyeed, R. A. (1998) Model-based geostatistics (with discussion). *Appl. Statist.*, **47**, 299–350.
- Dille, J., Milner, M., Groetke, J., Mortensen, D. and Williams II, M. M. (2002) How good is your weed map?: a comparison of spatial interpolators. *Weed Sci.*, **51**, 44–55.
- Eidsvik, J., Martino, S. and Rue, H. (2006) Approximate Bayesian inference in spatial generalized linear mixed models. *Technical Report*. Norwegian University of Science and Technology, Trondheim.
- Gerhards, R. and Christensen, S. (2003) Real-time weed detection, decision making and patch spraying in maize, sugarbeet, winter wheat and winter barley. *Weed Res.*, **43**, 385–392.
- Heisel, T., Andreasen, C. and Ersbøll, A. (1996) Annual weed distributions can be mapped with kriging. *Weed Res.*, **36**, 325–337.
- Heisel, T., Ersbøll, A. and Andreasen, C. (1999) Weed mapping with co-kriging using soil properties. *Precsn Agric.*, **1**, 39–52.

- Lamb, D. W., Weedon, M. and Rew, L. J. (1999) Evaluating the accuracy of mapping weeds in seedling crops using airborne digital imaging: *Avena spp.* in seedling triticale. *Weed Res.*, **39**, 481–492.
- Marchant, J. (1996) Tracking of row structure in three crops using image analysis. *Comput. Electron. Agric.*, **15**, 161–179.
- Marchant, J. and Brivot, R. (1995) Real-time tracking of plant rows using a Hough transform. *Realtime Imagng*, **1**, 363–371.
- Onyango, C. and Marchant, J. (2003) Segmentation of row crop plant from weeds using colour and morphology. *Comput. Electron. Agric.*, **39**, 141–155.
- Otsu, N. (1979) A threshold selection method from grey-level histograms. *IEEE Trans. Syst. Man Cybernet.*, **9**, 62–66.
- Perez, A., López, F., Benloch, J. and Christensen, S. (2000) Colour and shape analysis techniques for weed detection in cereal fields. *Comput. Electron. Agric.*, **25**, 197–212.
- Rew, L. and Cousins, R. (2001) Spatial distribution of weeds in arable crops: are current sampling and analytical methods appropriate? *Weed Res.*, **41**, 1–18.
- Rew, L., Whelan, B. and McBratney, A. B. (2001) Does kriging predict weed distribution accurately enough for site-specific weed control? *Weed Res.*, **41**, 245–263.
- Russ, J. (1994) *The Image Processing Handbook*, 2nd edn. Boca Raton: CRC Press.
- Schlather, M. (2006) RandomFields: simulation and analysis of random fields. *R Package Version 1.3.29*. (Available from <http://www2.hsu-hh.de/schlath/index.html>.)
- Soille, P. (2000) Morphological image analysis applied to crop field mapping. *Image Visn Computng*, **18**, 1025–1032.
- Tillet, N., Hague, T. and Miles, S. (2001) A field assessment of potential method for weed and crop mapping on the basis of crop planting geometry. *Comput. Electron. Agric.*, **32**, 229–246.
- Waagepetersen, R. (2006) A simulation-based goodness-of-fit test for random effects in generalized linear mixed models. *Scand. J. Statist.*, **33**, 721–731.
- Wackernagel, H. (1995) *Multivariate Geostatistics: an Introduction with Applications*. Berlin: Springer.
- Woebbecke, D., Meyer, G., von Bargen, K. and Mortensen, D. (1995) Color indices for weed identification under various soil, residue, and lighting conditions. *Trans. Am. Soc. Agric. Engrs*, **38**, 259–269.