

Bayesian analysis of within-field variability of corn yield using a spatial hierarchical model

Pingping Jiang · Zhuoqiong He · Newell R. Kitchen · Kenneth A. Sudduth

Published online: 9 July 2008
© Springer Science+Business Media, LLC 2008

Abstract Understanding relationships of soil and field topography to crop yield within a field is critical in site-specific management systems. Challenges for efficiently assessing these relationships include spatially correlated yield data and interrelated soil and topographic properties. The objective of this analysis was to apply a spatial Bayesian hierarchical model to examine the effects of soil, topographic and climate variables on corn yield. The model included a mean structure of spatial and temporal co-variates and an explicit random spatial effect. The spatial co-variates included elevation, slope and apparent soil electrical conductivity, temporal co-variates included mean maximum daily temperature, mean daily temperature range and cumulative precipitation in July and August. A conditional auto-regressive (CAR) model was used to model the spatial association in yield. Mapped corn yield data from 1997, 1999, 2001 and 2003 for a 36-ha Missouri claypan soil field were used in the analysis. The model building and computation were performed using a free Bayesian modeling software package, WinBUGS. The relationships of co-variates to corn yield generally agreed with the literature. The CAR model successfully captured the spatial association in yield. Model standard deviation decreased about 50% with spatial effect accounted for. Further, the approach was able to assess the effects of temporal climate co-variates on corn yield with a small number of site-years. The spatial Bayesian model appeared to be a useful tool to gain insights into yield spatial and temporal variability related to soil, topography and growing season weather conditions.

Keywords Crop yield spatial variability · Bayesian statistics · Conditional auto-regressive model · WinBUGS

P. Jiang (✉)
Department of Environmental Sciences, University of California, Riverside, CA 92521, USA
e-mail: pingping.jiang@ucr.edu

Z. He
Department of Statistics, University of Missouri, Columbia, MO 65211, USA

N. R. Kitchen · K. A. Sudduth
USDA-ARS Cropping Systems and Water Quality Research Unit, 269 Agricultural Engineering Bldg., Columbia, MO 65211, USA

Abbreviations

CAR	Conditional auto-regressive model
WinBUGS	Windows version of Bayesian Updating using Gibbs Sampler
EC _a	Apparent electrical conductivity
Max.temp	Mean maximum daily temperature in July and August
Temp.range	Means daily temperature range of July and August
cPREC	Cumulative precipitation in July and August

Introduction

Understanding within-field yield variability as affected by soil properties and field topography is important in site-specific management practices. Topographic variables such as elevation and slope have been shown to relate to crop yield (Yang et al. 1998; Kravchenko and Bullock 2000; Jiang and Thelen 2004). Sometimes, these variables can account for as much as 60% of the within-field variability (Yang et al. 1998). However, these relationships vary across years depending on weather conditions. In site-specific management, soil apparent electrical conductivity (EC_a) has also become an important tool in assessing soil suitability and productivity because it relates to a wide range of soil chemical and physical properties that affect crop yield (McNeill 1992; Lund et al. 2000; Kitchen et al. 2003; Sudduth et al. 2005). Applications of mapped EC_a have included characterizing soil spatial variability (Corwin and Lesch 2005), delineating management zones (Kitchen et al. 2005), and estimating topsoil depth for claypan soils (Kitchen et al. 1999; Sudduth et al. 2003). Together, these soil and topographic variables can provide guidance for site-specific management decision-making, partly because high resolution datasets can be obtained at reasonable costs using on-the-go sensing technologies.

Yield monitoring and mapping technologies that enable measuring, geo-referencing and recording grain yield with high precision and spatial resolution have created a great opportunity to study the spatial and temporal variability observed in crop yield. Yield monitors collect dense datasets in a pointwise fashion as the combine harvester moves through the field, and these data can be imported to a GIS system for display, processing and analysis. A common procedure for processing yield monitor data is to aggregate point yield data to a network of grids with an appropriate cell size for analysis as areal data (Birrell et al. 1996; Perez-Quezada et al. 2003).

Data analysis relating yield and soil and/or topographic variables has mostly adopted classical multiple linear regression techniques. As non-linear alternatives, projection pursuit regression (Sudduth et al. 1996), multiple quadratic regression (Kitchen et al. 1999), and neural network methods (Drummond et al. 2003) have also been employed. However, none of these approaches were able to explicitly model the spatial association of yield information. Further challenges for these analysis techniques include soil and topographic variables that are often interrelated, causing over-fitting of the model, and the general inability of these methods to assess the effects of temporal climatic conditions for the number of site-years commonly available (Drummond et al. 2003).

Recent development of the spatial Bayesian hierarchical framework has gained increasing popularity in ecological studies that often deal with spatial data (Wikle 2003; Oleson and He 2004). One of the important advantages of spatial Bayesian models is that the spatial effect can be explicitly expressed in the model by assuming a prior distribution

with its parameters specified from the information of spatial neighbors. In soil science, spatial Bayesian models have been applied to estimate the correlation coefficients of socio-ecological variables most strongly associated with soil $\text{NO}_3\text{-N}$ and C, while accounting for varying spatial associations among the two dependent soil variables caused by different land use types (Oleson et al. 2006).

Texts devoted to the theory and applications of Bayesian hierarchical models include, e.g., Gelman et al. (2005) and Banerjee et al. (2004), the latter focusing on spatial data analysis applications. For the reader's convenience, basic background information is presented here. The Bayesian hierarchical framework is developed from Bayes' rule of probability which provides a relationship between joint probability distribution and conditional distributions of two variables. For example, assuming there are two variables Y and θ , Bayes' theorem states that the joint probability distribution of Y and θ , $p(Y, \theta)$, can be obtained by the product of the conditional distributions and marginal distributions of Y and θ . Mathematically, Bayes' rule is expressed as $p(Y, \theta) = p(Y|\theta)p(\theta) = p(\theta|Y)p(Y)$. Thus, the conditional probability of Y can be obtained by $p(Y|\theta) = p(\theta|Y)p(Y)/p(\theta)$, and likewise, the conditional probability of θ by $p(\theta|Y) = p(Y|\theta)p(\theta)/p(Y)$.

As summarized in Banerjee et al. (2004), given observed data $\mathbf{y} = (y_1, \dots, y_n)$ with probability density function $f(y|\theta)$, and a vector of unknown parameters $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$, the fundamental difference of Bayesian approach from the classical statistical models is the former assumes that $\boldsymbol{\theta}$ is a random quantity sampled from a *prior* distribution $\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})$ based on previous knowledge, where $\boldsymbol{\lambda}$ is a vector of *hyper-parameters* (i.e., parameters for the prior distribution). If $\boldsymbol{\lambda}$ is known, inference concerning $\boldsymbol{\theta}$ is based on the *posterior* distribution of $\boldsymbol{\theta}$,

$$p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\lambda}) = \frac{p(\mathbf{y}, \boldsymbol{\theta}|\boldsymbol{\lambda})}{p(\mathbf{y}|\boldsymbol{\lambda})} = \frac{p(\mathbf{y}, \boldsymbol{\theta}|\boldsymbol{\lambda})}{\int p(\mathbf{y}, \boldsymbol{\theta}|\boldsymbol{\lambda})d\boldsymbol{\theta}} = \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})}{\int f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})d\boldsymbol{\theta}} \quad (1)$$

It can be seen in this relationship that both the data (f) and the prior knowledge (π) contribute to the posterior. In reality, however, $\boldsymbol{\lambda}$ is often not known. Thus, a second-stage (or a *hyper-prior*) distribution $h(\boldsymbol{\lambda})$ is required, and Eq. 1 becomes

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}, \boldsymbol{\theta})}{p(\mathbf{y})} = \frac{\int f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})h(\boldsymbol{\lambda})d\boldsymbol{\lambda}}{\int \int f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})h(\boldsymbol{\lambda})d\boldsymbol{\theta}d\boldsymbol{\lambda}} \quad (2)$$

Usually, $\boldsymbol{\lambda}$ can be replaced by its estimator, $\hat{\boldsymbol{\lambda}}$, which maximizes the marginal distribution $p(\mathbf{y}|\boldsymbol{\lambda})$ with respect to $\boldsymbol{\theta}$. Then inference about $\boldsymbol{\theta}$ can be based on the estimated posterior distribution $p(\boldsymbol{\theta}|\mathbf{y}, \hat{\boldsymbol{\lambda}})$.

One obstacle to adopting Bayesian analysis applications for scientists working in fields of ecology, agriculture and natural resources has been the programming skills required to evaluate the full conditionals required by the Bayesian models. The development of the free computer program "Windows version of Bayesian Updating using Gibbs Sampler" (WinBUGS, <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>), which has a built-in Gibbs sampler, an algorithm commonly used for generating random samples in Bayesian analysis, greatly facilitated Bayesian model formulation and computation. The WinBUGS program eliminated the need for specifying the full conditional distributions of the model, allowing the user to create their own models by providing only model distribution and prior distributions of model parameters. White and Sun (2006) illustrated the use and efficacy of this program using an ecological case study and reported that WinBUGS produced results compatible with those using the same model programmed in FORTRAN.

The objective of this project was to apply a spatial Bayesian hierarchical approach, implemented using the WinBUGS program, to examine the effects of soil, topographic and climatic variables on corn yield for a Missouri field.

Materials and methods

The dataset

The study site was a 36-ha corn-soybean field located near Centralia, Missouri. Elevation of the field ranged from 262 to 266 m. Soil types found in the field were claypan soils and included Mexico (fine, smectitic, mesic Aeric Vertic Epiaqualfs), Adco (fine, smectitic mesic Aeric Vertic Epiaqualfs), and Leonard (fine smectitic, mesic, Vertic Epiaqualfs). These soils were somewhat poorly or poorly drained and have a restrictive layer with high clay content (the claypan) occurring below topsoil at a highly variable depth. Corn yield data for 1997, 1999, 2001 and 2003 were used in the analyses. These yield data were collected using commercial yield monitors mounted on combine harvesters. During the harvest, the combine usually traveled at approximately 5–8 km h⁻¹ and yield data were recorded every second. Thus, depending on swath width, a single yield data point represented an average yield for an area approximately 6–10 m². An automatic yield data processing program—Yield Editor (Sudduth and Drummond 2007)—was used to remove questionable and unrealistic yield data points caused by operating errors such as abrupt changes of speed, partial swath and combine stops and starts. Descriptive statistics of corn grain yields are given in Table 1. Growing season precipitation records showed that droughty conditions prevailed in July, 1999 and July, 2003. In August 2003, monthly precipitation was above the long-term average, but the timing of the rain events was too late to alleviate the water stress of the corn plants (Fig. 1). Thus, corn yield was severely reduced in these 2 years. Continuous surfaces of yield data were generated using ordinary kriging with an exponential semivariogram using the ArcGIS Spatial Analyst (ESRI, Redlands, CA, USA) and then were output to a 30 m by 30 m cell resolution, resulting in a total of 390 data cells. The interpolated yield maps are shown in Fig. 2. The choice of 30 m cell size was mainly a consideration of data capacity of the WinBugs program, even though others have also used a 30-m resolution to evaluate the relationship between crop yield and soil information for the same field (Sudduth et al. 1996; Drummond et al. 2003). At this resolution some fine spatial structures maybe lost, but by visual examination of kriged maps the general spatial pattern was largely preserved.

Table 1 Descriptive statistics of corn grain yield and growing-season weather conditions

Year	Corn yield			Weather variables ^a		
	Mean (Mg ha ⁻¹)	SD (Mg ha ⁻¹)	CV (%)	Max.temp (°C)	Temp.range (°C)	cPREC (mm)
1997	7.2	1.1	0.15	29.7	16.3	150
1999	2.6	0.68	0.26	31.3	15.8	128
2001	6.1	0.92	0.15	30.3	13.0	116
2003	2.1	1.1	0.66	31.2	19.1	87

^a Max.temp, mean maximum daily temperature; Temp.range, mean daily temperature range; cPREC, cumulative precipitation

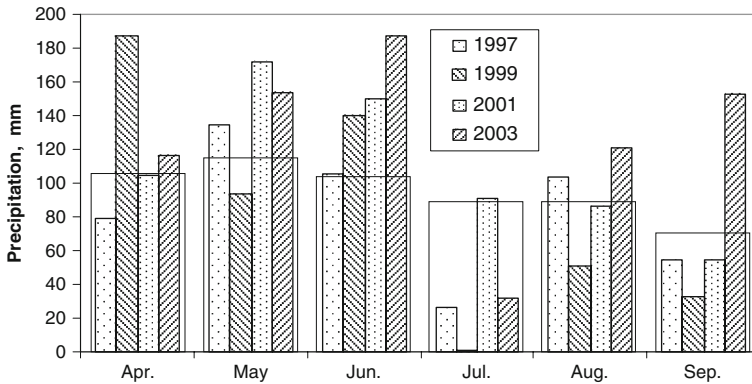


Fig. 1 Growing season (April–September) precipitation (mm). The cross bar of the hollow boxes indicates 30-year average for the month (i.e., average from 1971 to 2000)

Elevation of the field was mapped using a total station surveying instrument and standard mapping procedures. Soil EC_a was measured using an on-the-go Dualem-2S sensor (Dualem Inc., Milton, Ontario, Canada)¹ every second on a 10-m transect spacing. The Dualem-2S sensor had two effective sensing depths of 1.2 and 3.0 m, and the 1.2-m sensing depth was used. The speed of travel of the sensor was approximately 7.2 km h⁻¹. Both elevation and EC_a data were kriged to the same spatial extent as the yield data. Maximum slope in degrees was derived from the interpolated elevation data using Spatial Analyst in ArcGIS (ESRI 2007). The Spatial Analyst calculates slope values along the rate of maximum change in elevation for each cell to its neighbors using a 3 by 3 search window. The kriged elevation and EC_a maps are presented in Fig. 3.

Three weather variables for the period of July and August were included as temporal co-variates in the model. They were mean maximum daily temperature (Max.temp), mean daily temperature range (Temp.range) and cumulative precipitation (cPREC; Table 1). These particular climatic variables were chosen because they have been shown to significantly affect corn yield using a dataset collected over 100 years in Central Missouri (Hu and Buyanovsky 2003).

The model

The baseline model

The baseline model was a linear regression model including a mean structure of co-variates (elevation, slope, soil EC_a , Max.temp, Temp.range, and cPREC), and a random spatial effect.

Let $Y_{i,j}$ denote the observed yield at cell i in year j , and assume

$$Y_{i,j} = \mathbf{X}_{i,j}\boldsymbol{\beta} + Z_i + e_{i,j} \quad (3)$$

where $e_{i,j}$ represents the measurement errors which are assumed to be independent and identically distributed (iid) as $N(0, \sigma_e^2)$, Z_i is the spatial effect of cell i . $\boldsymbol{\beta}$ is the $n \times 1$

¹ Mention of trade names or commercial products is solely for the purpose of providing specific information and does not imply recommendation or endorsement by University of Missouri or the USDA.

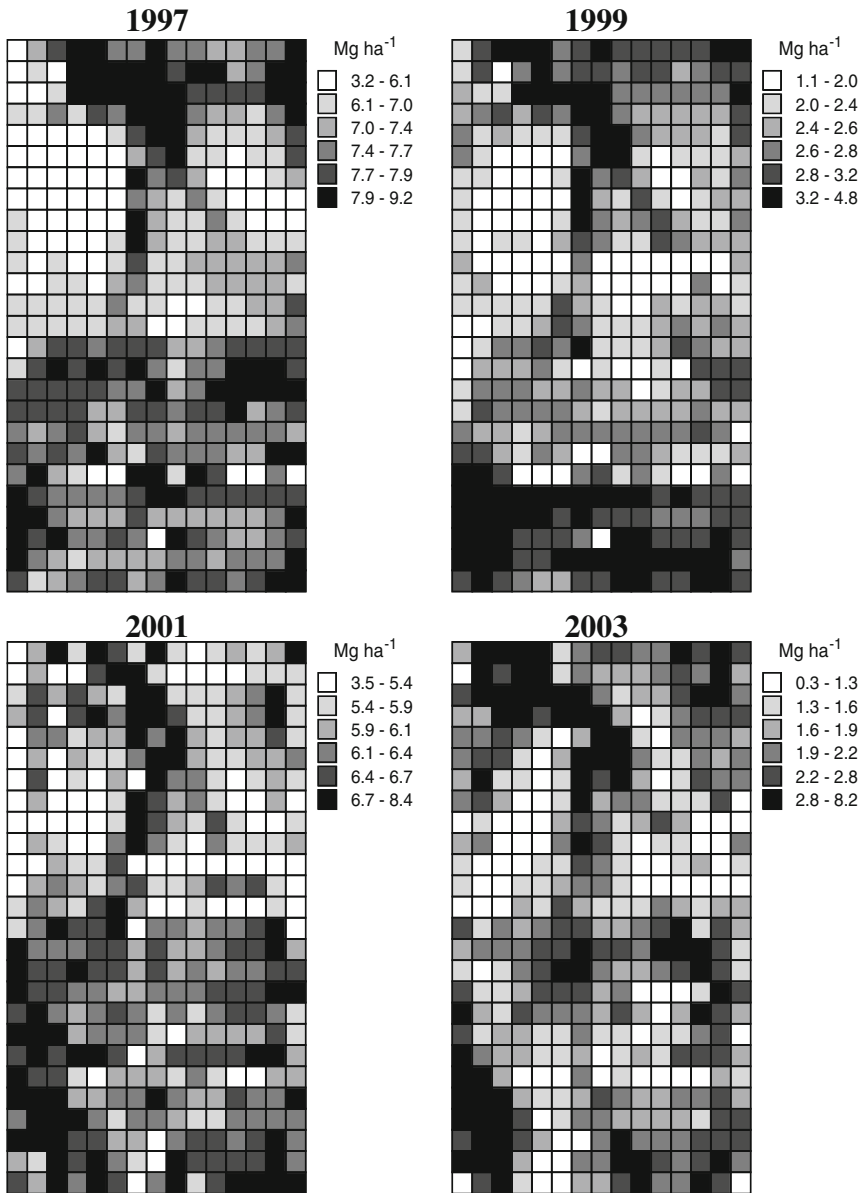


Fig. 2 Yield maps for 1997, 1999, 2001, and 2003. The dimension of each map cell is 30 m by 30 m

parameter coefficient vector of co-variates and $\mathbf{X}_{i,j}$ is the $1 \times n$ covariate vector. The spatial co-variates were considered fixed in this case.

Spatial effect modeling

A conditional auto-regressive (CAR) model was used to account for the spatial association among yield (i.e., Z_i in Eq. 3). In general, a CAR model can be expressed as follows:

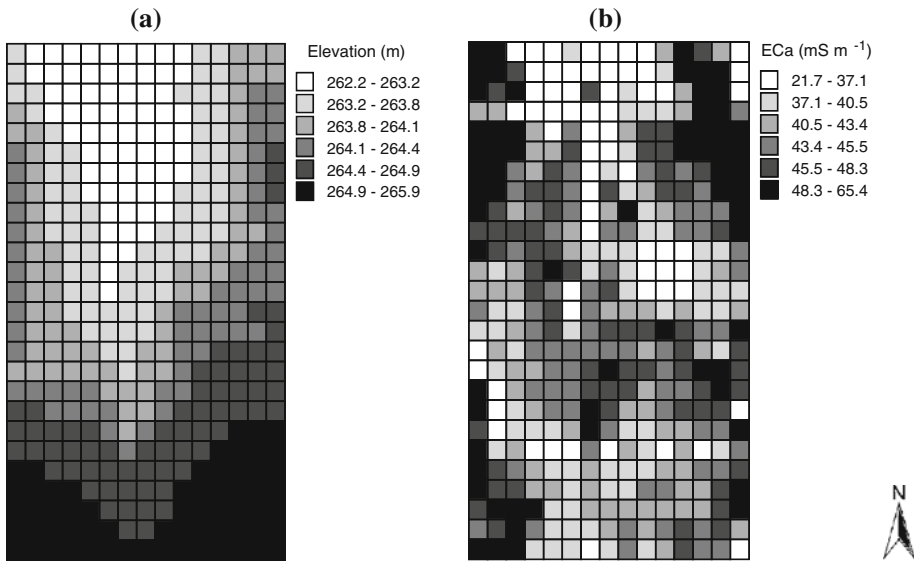


Fig. 3 Elevation (a) and EC_a (b) maps of the field. The dimension of each map cell is 30 m by 30 m

$$(Y_i|y_m, m \neq i) \sim N\left(\sum_m b_{im}y_m, \tau^2\right), \quad i = 1, \dots, n. \tag{4}$$

where m denotes the set of “neighbor cells” around cell i . It is assumed that Y_i has a normal distribution with conditional mean given by the average of its neighbors, y_m , and a common variance component, τ^2 . b_{im} are elements of a symmetric matrix \mathbf{B} , which is called an adjacency matrix, where b_{im} are equal to 1 if cell i and cell m are adjacent, and 0 otherwise. The CAR model allows for the “borrowing” of information from adjacent cells in estimating parameters for each individual cell.

Specification of prior distributions and hyper-parameters

As mentioned above, each model parameter needed a prior distribution, and the parameters of the prior distribution (i.e., hyper-parameters) were also needed to complete the model. In our dataset, all the coefficients of co-variates in Eq. 3 were given a normal distribution $N(0, \sigma_0^2)$, where σ_0^2 is a positive constant. The model parameters, variance (σ_e^2), and the spatial variance (τ^2), were assumed to follow the inverse gamma distribution $IG(a, b)$, which is a common prior distribution specification for error variance (Gelman et al. 2005). Therefore, the two variance terms were specified as:

$$\sigma_e^2 \sim IG(a_e, b_e) \quad \text{and} \quad \tau^2 \sim IG(a_z, b_z) \tag{5}$$

where the hyper-parameter values of (a_e, b_e) and (a_z, b_z) are positive constants.

All hyper-parameters (i.e., $\sigma_0^2, a_e, b_e, a_z,$ and b_z) were given such that prior distributions were non-informative. Non-informative priors have large or infinite variance so parameter estimations are largely driven by data and less by the prior-distributions specified.

Table 2 Prior distributions and hyper-parameter values for co-variables (β), model variance (σ_e^2) and spatial variance (τ^2)

Parameter ^a	Prior distribution with hyper-parameters (same as in the text)	Hyper-parameter values	Prior distribution specification used in WinBUGS	Hyper-parameter values in WinBUGS coding syntax
EC _a	$N(0, \sigma_0^2)$	$N(0, 1000)$	$N(0, 1/\sigma_0^2)$	$N(0, 0.001)$
Slope	$N(0, \sigma_0^2)$	$N(0, 1000)$	$N(0, 1/\sigma_0^2)$	$N(0, 0.001)$
Elevation	$N(0, \sigma_0^2)$	$N(0, 1000)$	$N(0, 1/\sigma_0^2)$	$N(0, 0.001)$
Tamp.range	$N(0, \sigma_0^2)$	$N(0, 1000)$	$N(0, 1/\sigma_0^2)$	$N(0, 0.001)$
Max.temp	$N(0, \sigma_0^2)$	$N(0, 1000)$	$N(0, 1/\sigma_0^2)$	$N(0, 0.001)$
cPREC	$N(0, \sigma_0^2)$	$N(0, 1000)$	$N(0, 1/\sigma_0^2)$	$N(0, 0.001)$
τ^2	$IG(a_e, b_e)$	$IG(2, 1)$	$Gamma(a_e, 1/b_e)$	$Gamma(2, 1)$
σ_e^2	$IG(a_e, b_e)$	$IG(2, 1)$	$Gamma(a_e, 1/b_e)$	$Gamma(2, 1)$

^a Temp.range, mean daily temperature range in July and August; Max.temp, mean maximum daily temperature in July and August; cPREC, cumulative precipitation in July and August

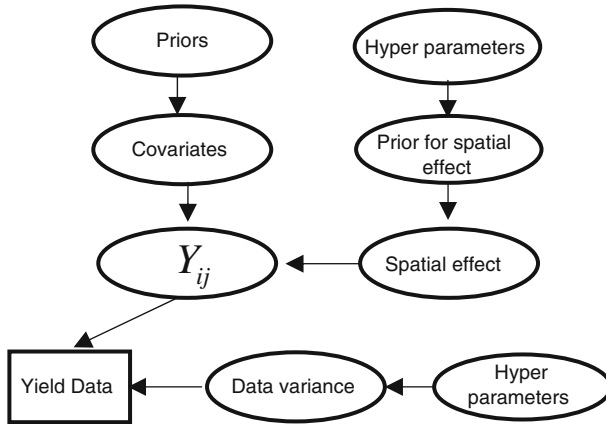


Fig. 4 Diagram of the spatial Bayesian hierarchical model described by Eqs. 3–5

A complete list of prior distributions and hyper-parameters are summarized in Table 2. Further, a diagram illustrating the inter-relationships between model components (Eqs. 3–5) is given in Fig. 4.

Bayesian computation using WinBUGS

Posterior distributions are estimated using the Gibbs sampler in WinBUGS. If the model has k parameters, $\theta = (\theta_1, \dots, \theta_k)'$, the Gibbs sampler generates random samples from each of the full conditional distributions $\{p(\theta_i | \theta_{j \neq i}, \mathbf{y}), i = 1, \dots, k\}$ in the model. Given initial values of $\{\theta_2^{(0)}, \dots, \theta_k^{(0)}\}$, The Gibbs sampler iterates as follows (Banerjee et al. 2004):

- Step 1: Draw $\theta_1^{(t)}$ from $p(\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \mathbf{y})$
- Step 2: Draw $\theta_2^{(t)}$ from $p(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_k^{(t-1)}, \mathbf{y})$

Step k : Draw $\theta_k^{(t)}$ from $p(\theta_1|\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{k-1}^{(t)}, \mathbf{y})$

Thus, when t is sufficiently large, $\{\theta_1^{(t)}, \dots, \theta_k^{(t)}\}$ can be shown to converge to the true joint posterior distribution $p(\theta_1, \dots, \theta_k|\mathbf{y})$.

Model implementation

Three versions of the baseline model (Eq. 3) were implemented. The first version, Model (a), was run for each year with neither temporal co-variates nor the spatial effect Z_i in the model; the second version, Model (b), was run for each year with the spatial effect Z_i , but without temporal co-variates; and the last version, Model (c), was run for the 4-year combined dataset with the three temporal co-variates and the spatial effect Z_i .

For each version of the baseline model, posterior densities for each model parameter were obtained after 100,000 iterations, with the first 10,000 iterations discarded as “burn-in”. Model runtime for all model versions was less than 10 min using a Pentium IV computer.

Results and discussion

By-year analysis without spatial effect

Parameter posterior means, standard deviations (SD) and 95% credible intervals obtained using Model (a) are given in Table 3. Soil EC_a was negatively related to corn yield in three (1997, 1999 and 2001) out of the four site-years, with statistical significance in 1997 and 1999 (i.e., the 95% credible intervals excluded zero). For claypan soils, negative correlations between EC_a and crop yield under droughty growing seasons has been previously reported, especially when rainfall was below normal in July and August (Kitchen et al. 1999). In agreement with these findings, 1997 and 1999 had well-below-normal precipitation during July or August (Fig. 1). Slope was negatively and

Table 3 Posterior means, standard deviations, and 95% credible intervals obtained using Model (a)

Parameter	Mean (SD)	(95% Credible interval)	Parameter	Mean (SD)	(95% Credible interval)
<i>1997</i>			<i>2001</i>		
Soil EC _a	-0.038 (0.007)	(-0.053, -0.024)	Soil EC _a	-0.008 (0.006)	(-0.019, 0.003)
Slope	-1.283 (0.181)	(-1.638, -0.927)	Slope	-0.504 (0.142)	(-0.782, -0.226)
Elevation	0.028 (0.059)	(-0.088, 0.144)	Elevation	0.074 (0.048)	(-0.020, 0.169)
σ_e^a	0.895 (0.032)	(0.835, 0.960)	σ_e	0.694 (0.025)	(0.647, 0.744)
<i>1999</i>			<i>2003</i>		
Soil EC _a	-0.016 (0.005)	(-0.026, -0.007)	Soil EC _a	0.005 (0.008)	(-0.011, 0.020)
Slope	-0.538 (0.117)	(-0.766, -0.310)	Slope	-0.668 (0.194)	(-1.046, -0.288)
Elevation	0.050 (0.041)	(-0.029, 0.130)	Elevation	-0.301 (0.063)	(-0.422, -0.178)
σ_e	0.567 (0.020)	(0.529, 0.608)	σ_e	0.957 (0.034)	(0.892, 1.027)

^a σ_e , model standard deviation, in Mg ha⁻¹

significantly related to yield for all four site-years. The relationships between slope and crop yield are generally negative (Yang et al. 1998; Kravchenko and Bullock 2000; Jiang and Thelen 2004) because steep slopes usually result in severe erosion characterized by thinner topsoil, higher clay content, lower infiltration rate, and greater runoff and, hence, lower soil productivity. Another topographic variable, elevation, was negatively significant only in 2003. Relationships of elevation to crop yield varied, depending on location and year. As Kravchenko and Bullock (2000) pointed out, the effect of elevation on yield is reflected through water availability, and this effect is more readily observed under extreme weather conditions and field topography. The elevation change was gradual in our field and thus elevation did not affect crop yield in most of the site-years. The negative relationship in 2003 was probably because more soil water was stored in the low-elevation areas of the field. The low-elevation areas (depositional areas) had greater topsoil (silt loam) thickness, and topsoil usually provides about twice as much plant available water capacity as the underlying claypan. On the other hand, erosional areas, especially the highly eroded side-slopes where claypan was near the soil surface, often show deficiencies in soil water storage due to slow recharge of the clay material (Jiang et al. 2007). This effect was particularly an issue in 2003 because below-normal precipitation was also recorded for the fall/winter recharge period before the 2003 growing season (data not shown). This resulted in very low level of soil water, except for the depositional areas in the low elevations.

Model residuals of yield (predicted yield–measured yield) are mapped in Fig. 5. Spatial patterns in the residuals were similar to those seen in the yield maps (Fig. 2). Clearly, these spatial patterns in the yield were preserved in the residuals without a spatial component in the model. Large positive residuals (over-estimation) were mostly in the most-eroded side-slope areas, whereas large negative residuals (under-estimation) occurred along the lower drainage-way and the outlet on the north. Another distinctive pattern can be found around the south-west corner of the field. The south end of the field had not been in crop production until 1980, and the south-west portion of the field was the location of an old farmhouse and livestock housing area. Therefore, animal manures, as well as less eroded soil, contributed to the higher yield in that part of the field (under-estimation).

By-year analysis with spatial effect

Parameter posterior means, standard deviations and 95% credible intervals obtained using Model (b) are given in Table 4. In general, consistent relationships were found for soil and topographic co-variates with those obtained by Model (a). For example, soil EC_a was negatively significant in 1997 and 1999 and slope was negatively significant for all four site-years. The standard deviations of the soil EC_a and slope estimates were comparable between Models (a) and (b). The standard deviation of elevation, however, increased for all four years compared with Model (a). For example, in 2003 the standard deviation increased to 0.107 Mg ha^{-1} in Model (b) from 0.063 Mg ha^{-1} in Model (a) and as a result, elevation was no longer significant. This result was possibly because the spatial association of yield was related to elevation changes in parts of the field. With the spatial effect added, the model standard deviations (σ_e) decreased for all 4 years, from an average of 0.778 Mg ha^{-1} for Model (a) to 0.382 Mg ha^{-1} for Model (b). The standard deviations of the spatial effect (τ) were lower for the dry and low-yielding years (1999 and 2003) and higher for the high-yielding years (1997 and 2001). The correlation coefficient was 0.7 (not shown).

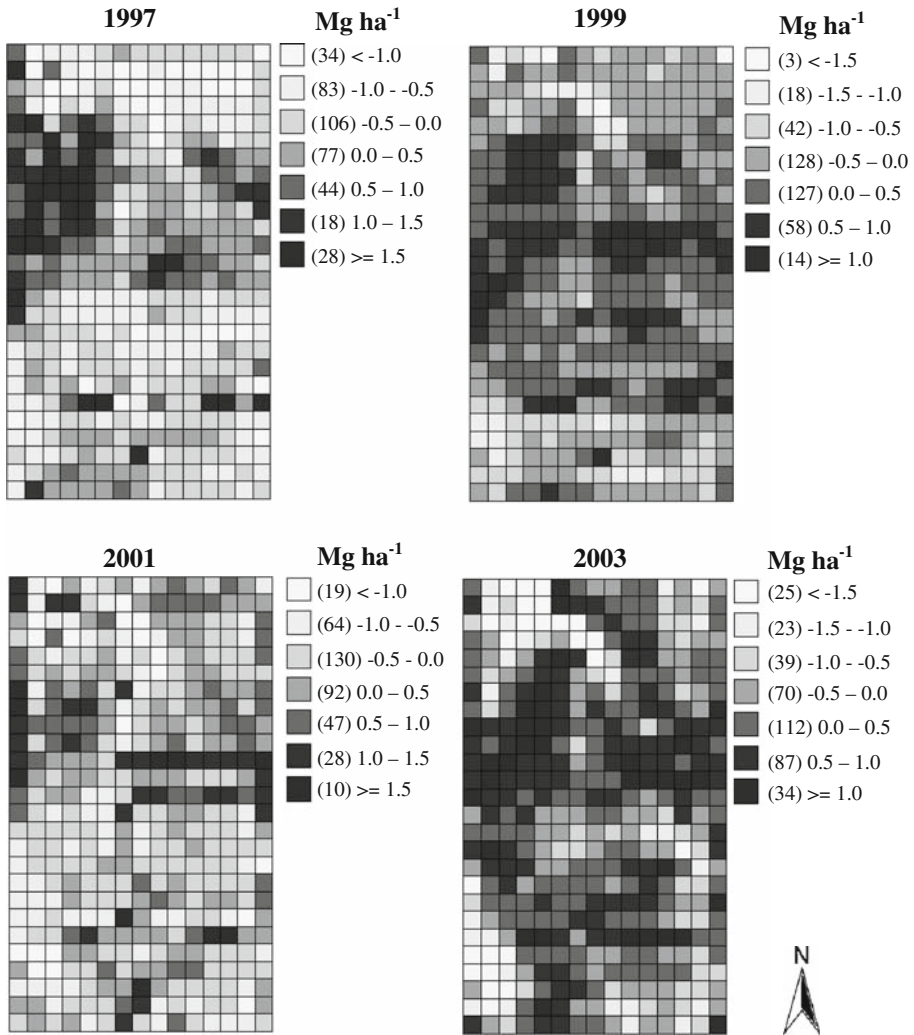


Fig. 5 Residual (predicted yield–observed yield) maps of corn yield obtained using Model (a), without a spatial component. Numbers in parentheses are number of cells in each yield range. The dimension of each map cell is 30 m by 30 m

Spatial effect maps are presented in Fig. 6. The Model (b) spatial maps were, in large, opposite of the residual maps using Model (a) (Fig. 5). For the spatial effect, patterns of negative values were observed in the eroded side-slope areas and patterns of positive values were found along the drainage outlet areas and in the south-west corner of the field. Visually, more resemblance seemed to exist between 1999 and 2003, the two dry years, and between 1997 and 2001, the two moderately-dry years. Allowing for the small number of site-years, this result suggests that the spatial patterns of yield can be temporally consistent under similar weather patterns. This result agreed with Sadler (1998) and Jiang et al. (2007), who found that spatial pattern of corn yield was temporally consistent when soil water was limited.

Table 4 Posterior means and 95% credible intervals for regression coefficients, and model and spatial error, using Model (b)

Parameter	Mean (SD)	(95% Credible interval)	Parameter	Mean (SD)	(95% Credible interval)
<i>1997</i>			<i>2001</i>		
Soil EC _a	-0.020 (0.007)	(-0.034, -0.005)	Soil EC _a	0.003 (0.006)	(-0.009, 0.016)
Slope	-1.001 (0.201)	(-1.397, -0.611)	Slope	-0.218 (0.171)	(-0.553, 0.118)
Elevation	-0.131 (0.103)	(-0.332, 0.074)	Elevation	-0.067 (0.096)	(-0.252, 0.122)
τ^a	1.392 (0.10)	(1.20, 1.577)	τ	1.018 (0.111)	(0.803, 1.237)
σ_e^b	0.384 (0.045)	(0.302, 0.475)	σ_e	0.441 (0.041)	(0.359, 0.520)
<i>1999</i>			<i>2003</i>		
Soil EC _a	-0.011 (0.004)	(-0.020, -0.002)	Soil EC _a	0.008 (0.008)	(-0.008, 0.023)
Slope	-0.569 (0.121)	(-0.809, -0.333)	Slope	-0.885 (0.222)	(-0.884, -0.454)
Elevation	-0.124 (0.084)	(-0.287, 0.043)	Elevation	-0.184 (0.107)	(-0.392, 0.028)
τ	0.794 (0.054)	(0.689, 0.899)	τ	1.612 (0.089)	(1.435, 1.785)
σ_e	0.266 (0.022)	(0.225, 0.311)	σ_e	0.356 (0.043)	(0.278, 0.447)

^a τ , spatial effect standard deviation in Mg ha⁻¹

^b σ_e , model standard deviation in Mg ha⁻¹

Residual maps obtained using Model (b) are presented in Fig. 7. With yield spatial association accounted for by the CAR model, residuals were greatly reduced, the spatial patterns observed in the residual maps from Model (a) were resolved and the residuals began to look more or less random. This indicated that the spatial CAR model successfully explained the spatial association of yield.

Analysis for four-year data combined with spatial effect

Parameter posterior means, standard deviations and 95% credible intervals obtained using Model (c) are given in Table 5. The Temp.range and Max.temp were negatively related to corn yield, similar to the findings in Hu and Buyanovsky (2003), who reported high corn yields were related to cooler and stable daily temperatures from June through August. Surprisingly, cPREC was not significant, emphasizing that the timing of precipitation during the critical development stages corn plants can be more important than the total amount (Hu and Buyanovsky 2003). When data were combined, the model standard deviation was substantially increased to 0.606 Mg ha⁻¹, compared with the average SD value (0.382 Mg ha⁻¹) of Model (b). This increase was because overall yield variance increased when data were combined. The standard deviation for the spatial effect, however, decreased to 0.900 Mg ha⁻¹ from an average of 1.204 Mg ha⁻¹ using Model (b). These results demonstrated the spatial Bayesian model was able to fit the additional complexities introduced by the inclusion of climate co-variates, and was able to detect the effect of weather conditions on corn yield with a small number of site-years, fewer than generally needed in more traditional methodologies such as multiple regression and some non-linear fitting techniques (Drummond et al. 2003).

The spatial effect map for the 4-year combined yield is presented in Fig. 8. The combined map seemed to be smoother than the spatial effects mapped for individual years using Model (b) (Fig. 6). Nonetheless, the most distinctive patterns were retained. This map indicated how strongly and persistently the corn yield was spatially associated within the study period.

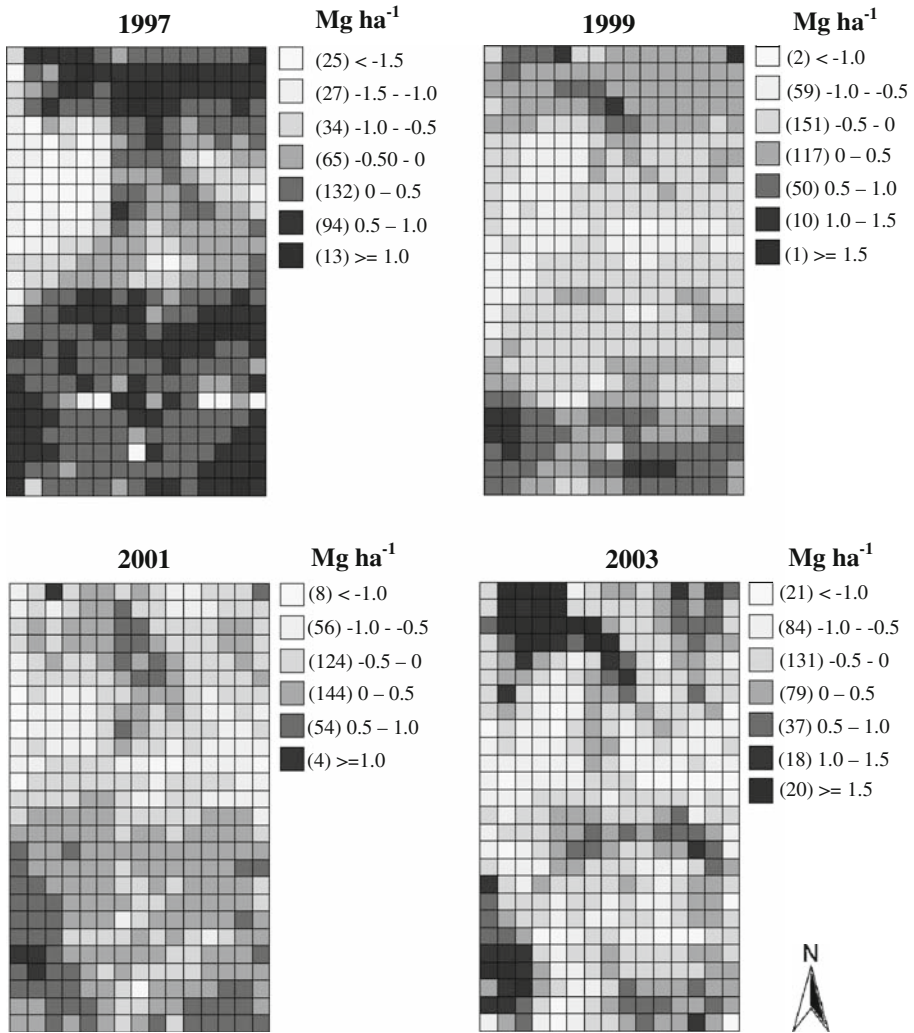


Fig. 6 Spatial effect maps for corn grain yield using Model (b). Numbers in parentheses are number of cells in each yield range. The dimension of each map cell is 30 m by 30 m

Conclusions

A spatial Bayesian hierarchical approach was employed to model the relationships of soil, topographic and climate variables to corn yield. Our results generally agreed with previous research findings in the literature. Compared with other methodologies, the Bayesian hierarchical approach distinguishes itself by being able to explicitly model the spatial association of yield data. This is an advantage to the classical linear regression technique, which assumes independence of the dependent variables in the model. Thus, all data points could be utilized in the analysis, which greatly increased the efficiency of the data and the power of statistical inference. The CAR model successfully captured the structures of spatial association of corn yield both for individual years and for 4 years combined. Model

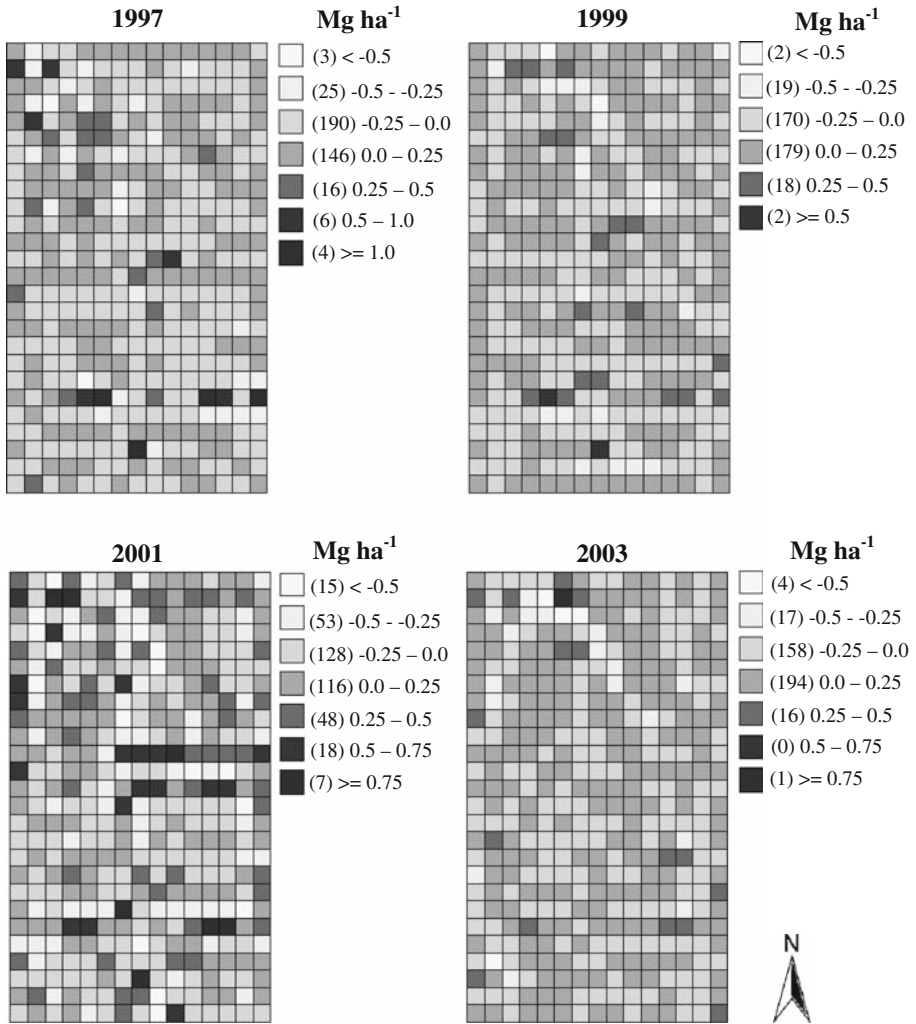


Fig. 7 Residual (predicted yield–observed yield) maps of corn yield obtained using Model (b), with a spatial component. Numbers in parentheses are number of cells in each yield range. The dimension of each map cell is 30 m by 30 m

errors were reduced by about 50% when adding the spatial effect component to the model. Further, the approach was able to assess the effects of temporal variables (i.e., climatic variables) on corn yield with a small number of site-years. When more site-years become available, it is conceivable and reasonable to consider a temporal random effect to capture the time dependency in the corn yield over time. The Bayesian model provides the flexibility to accommodate that possibility. The flexible nature of the model also makes it possible to consider the inclusion of additional spatial (e.g., soil or topographic) variables when appropriate.

The WinBUGS software and its spatial module, GEOBUGS, proved to be useful tools in constructing spatial Bayesian models. However, some limitations need to be noted for the

Table 5 Posterior means and 95% credible intervals for regression coefficients, and model and spatial effect standard deviations, using Model (c)

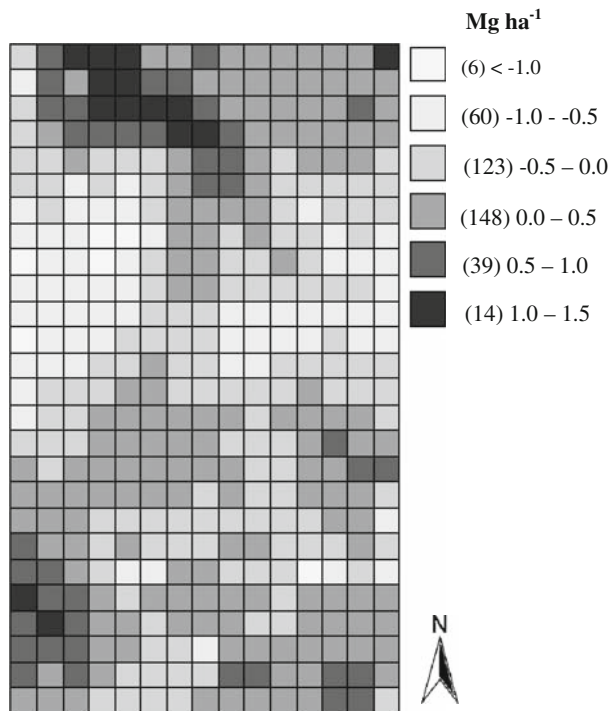
Parameter ^a	Posterior mean (SD)	(95% Credible interval)
Soil EC _a	-0.006 (0.005)	(-0.016, 0.004)
Slope	-0.660 (0.136)	(-0.926, -0.395)
Elevation	-0.005 (0.092)	(-0.184, 0.174)
Temp.range	-0.227 (0.008)	(-0.243, -0.211)
Max.temp	-2.876 (0.032)	(-2.938, -2.814)
cPREC	0.001 (0.001)	(-0.001, 0.003)
τ^b	0.900 (0.052)	(0.799, 1.003)
σ_e^c	0.606 (0.012)	(0.583, 0.630)

^a Temp.range, mean daily temperature range of July and August; Max.temp, mean maximum daily temperature in July and August; cPREC, cumulative precipitation in July and August

^b τ , spatial effect standard deviation

^c σ_e , model standard deviation

Fig. 8 Strength of spatial association, demonstrated using Model (c), for the combined corn yield of 1997, 1999, 2001, and 2003. Numbers in parentheses are number of cells in each yield range. The dimension of each map cell is 30 m by 30 m



sake of future development of the software. First, the ability of the software to handle large datasets is limited. For example, we attempted to run a 10 m by 10 m cell-size resolution (which would give 3634 data points) for yield and each covariate; WinBUGS was unable to handle this. Second, WinBUGS has primitive color schemes for displaying spatial results. Also, it lacks an exporting function by which spatial results can be output to a data file, which then can be imported to a GIS or other package for display and further analysis.

As a result, subtle changes in spatial patterns may be challenging to discern. Overall, the WinBUGS software was able to facilitate the adoption of the spatial Bayesian approach by providing relatively easy access to the full Bayesian formulation and computation to scientists working in soil science and other natural resource fields where spatial data are commonplace.

Acknowledgment We thank Scott Drummond for his assistance in handling and graphing the data.

References

- Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2004). *Hierarchical modeling and analysis for spatial data*. New York, USA: Chapman & Hall.
- Birrell, S. J., Sudduth, K. A., & Borgelt, S. C. (1996). Comparison of sensors and techniques for crop yield mapping. *Computers and Electronics in Agriculture*, *14*, 215–233. doi:10.1016/0168-1699(95)00049-6
- Corwin, D. L., & Lesch, S. M. (2005). Characterizing soil spatial variability with apparent soil electrical conductivity. I. Survey protocols. *Computers and Electronics in Agriculture*, *46*, 135–152. doi:10.1016/j.compag.2004.11.003.
- Drummond, S. T., Sudduth, K. A., Joshi, A., Birrell, S. J., & Kitchen, N. R. (2003). Statistical and neural methods for site-specific yield prediction. *Transactions of the ASAE*, *46*(1), 5–14.
- RI, E. S. (2007). *ArcGIS 9.2 Desktop help*. Redlands, CA: ESRI.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2005). *Bayesian data analysis*. Boca Raton, Florida, USA: Chapman and Hall/CRC.
- Hu, Q., & Buyanovsky, G. (2003). Climate effect on corn yield in Missouri. *Journal of Applied Meteorology*, *42*, 1626–1635. doi:10.1175/1520-0450(2003)042<1626:CEOCYI>2.0.CO;2.
- Jiang, P., Anderson, S. H., Kitchen, N. R., Sudduth, K. A., & Sadler, E. J. (2007). Estimating plant-available water capacity for claypan landscapes using apparent electrical conductivity. *Soil Science Society of America Journal*, *71*, 1902–1908. doi:10.2136/sssaj2007.0011.
- Jiang, P., & Thelen, K. D. (2004). Effect of soil and topographic properties on crop yield in a North-Central corn-soybean cropping system. *Agronomy Journal*, *96*, 252–258.
- Kitchen, N. R., Drummond, S. T., Lund, E. D., Sudduth, K. A., & Buchleiter, G. W. (2003). Soil electrical conductivity and topography related to yield for three contrasting soil crop systems. *Agronomy Journal*, *95*, 483–495.
- Kitchen, N. R., Sudduth, K. A., & Drummond, S. T. (1999). Soil electrical conductivity as a crop productivity measure for claypan soils. *Journal of Production Agriculture*, *12*, 607–617.
- Kitchen, N. R., Sudduth, K. A., Myers, D. B., Drummond, S. T., & Hong, S. Y. (2005). Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. *Computers and Electronics in Agriculture*, *46*, 285–308. doi:10.1016/j.compag.2004.11.012.
- Kravchenko, A. N., & Bullock, D. G. (2000). Correlation of corn and soybean grain yield with topography and soil properties. *Agronomy Journal*, *92*, 75–83. doi:10.1007/s100870050010.
- Lund, E. D., Christy, C. D., & Drummond, P. E. (2000). Using yield and soil electrical conductivity (EC) maps to derive crop production performance information. In P. C. Robert, R. H. Rust, & W.E. Larson (Eds.), *Proceedings Of the 5th International Conference on Precision Agriculture, ASA, CSSA, and SSSA, Madison, WI, USA* [CD-ROM].
- McNeill, J. D. (1992). Rapid accurate mapping of soil salinity by electromagnetic ground conductivity meters. In *Advances in measurement of soil physical properties: Bringing theory into practices* (pp. 209–229). Soil Science Society of America Special Publication 30, SSSA, Madison, WI, USA.
- Oleson, J., & He, C. Z. (2004). Space-time modeling for the Missouri turkey hunting survey. *Environmental and Ecological Statistics*, *11*, 85–101. doi:10.1023/B:EEST.0000011366.68489.d8.
- Oleson, J. J., Hope, D., Gries, C., & Kaye, J. (2006). Estimating soil properties in heterogeneous land-use patches: A Bayesian approach. *Environmetrics*, *17*, 517–525. doi:10.1002/env.789.
- Perez-Quezada, J. F., Pettygrove, G. S., & Plant, R. E. (2003). Spatial-temporal analysis of yield and the influence of soil factors in two-four-crop-rotation fields in the Sacramento Valley, California. *Agronomy Journal*, *95*, 676–687.
- Sadler, E. J. (1998). Spatial scale requirements for precision farming: A case study in the Southeastern USA. *Agronomy Journal*, *90*, 191–197.

- Sudduth, K. A., & Drummond, S. T. (2007). Yield editor: Software for removing errors from crop yield map. *Agronomy Journal*, *99*, 1471–1482.
- Sudduth, K. A., Drummond, S. T., Birrell, S. J., & Kitchen, N. R. (1996). Analysis of spatial factors influencing crop yield. In P. C. Robert, R. H. Rust, & W. E. Larson (Eds.), *Proceedings of the 3rd International Conference on Precision Agriculture, ASA, CSA, and SSSA* (pp. 129–140), Madison, WI, USA.
- Sudduth, K. A., Kitchen, N. R., Bollero, G. A., Bullock, D. G., & Wiebold, W. J. (2003). Comparison of electromagnetic induction and direct sensing of soil electrical conductivity. *Agronomy Journal*, *95*, 472–482.
- Sudduth, K. A., Kitchen, N. R., Weibold, W. J., Batchelor, W. D., Bollero, G. A., Bullock, D. G., et al. (2005). Relating EC_a to soil properties across the north-central USA. *Computers and Electronics in Agriculture*, *46*, 263–283. doi:[10.1016/j.compag.2004.11.010](https://doi.org/10.1016/j.compag.2004.11.010).
- White, G., & Sun, D. (2006). Simultaneous estimation of hunting pressure, harvest and hunter success rates using WinBUGS. *Far East Journal of Theoretical Statistics*, *19*, 91–116.
- Wikle, C. K. (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology*, *84*, 1382–1394. doi:[10.1890/0012-9658\(2003\)084\[1382:HBMFPT\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[1382:HBMFPT]2.0.CO;2).
- Yang, C., Peterson, C. L., Shropshire, G. J., & Otawa, T. (1998). Spatial variability of field topography and wheat yield in the Palouse region of the Pacific Northwest. *Transactions of the ASAE*, *41*, 17–27.