Site-Specific Nitrogen Recommendation: Using Bayesian Kriging Method with Different Correlation Matrices

Davood Poursina[a]

B. Wade Brorsen[a]

[a]Department of Agricultural Economics, Oklahoma State University, Stillwater, OK, 74074-6026, United States

Site-Specific Nitrogen Recommendation: Using Bayesian Kriging Method with Different Correlation Matrices

Abstract

Estimating a production function and finding the optimal value of a fertilizer is of significant interest in any agricultural system. Precision fertilizer application essentially requires finding a production function for each piece of the field and then finding the optimal value based on this model. In finding the optimal nitrogen value, the spatial behavior of data is often assumed on the error term, and the covariates' coefficients are considered constant over space. However, this assumption might be questionable, especially in agricultural data. So, finding a fast and accurate model with spatially random coefficient is essential. In addition, solving the profit function for the mentioned model needs additional efforts of computation. An analytical approach to finding the optimal value of nitrogen could help the computational burden for non-linear mode. A comparison among models in the accuracy and computational burden shows that the restrictions only lose a little accuracy while significantly reducing the computational burden.

Keywords: fertilizer, Gaussian spatial process, linear plateau, optimal nitrogen, spatial smoothing

Abbreviations

Conditional autoregressive (CAR)

Geographically weighted regression (GWR)

Hamiltonian Monte Carlo (HMC)

Simultaneous autoregressive (SAR)

1. *Introduction*

Precision fertilizer application essentially requires finding a production function for each piece of the field. There are several interesting papers in this field of research (Anselin et al. 2004, Evans et al. 2020, Griffin et al. 2008, Hurley et al. 2005, Park et al. 2018, Rodrigues et al. 2013). Several methods exist to handle the data's spatial behavior. There is a tradeoff between the accuracy and complexity of a model and the ability to estimate it.

In this paper, we address the problem of finding a specific optimal value with good precision for each part of the field and at the same time find an efficient computational way to make these models feasible for data with a large number of locations. An analytical solution for obtaining the optimal nitrogen value is provided. This method applies to the linear plateau model.

There is rich literature in this field of study. Early approaches sought to estimate a production function with the parameters constant for clusters in the data set. A dummy variable is then added for each cluster (Lambert et al. 2004, Liu et al. 2006). The dummy variable approach has drawbacks. It requires knowledge about how to form the clusters. It assumes parameters vary discretely rather than smoothly across a field. Finally, this method could suffer from a lack of degrees of freedom or multicollinearity if the number of variables that should be considered dummies increases.

The second approach, which is widely used, is Geographically weighted regression (GWR). GWR usually uses a contiguity matrix (with a different number of neighbors) to find the model's spatial weights (Evans et al. 2020, Trevisan et al. 2020). Although this model can fit the

data well, it suffers from the lack of statistical theory for optimality behavior (Dambon et al. 2020).

The third approach, which is employed here is Bayesian Kriging (Park et al. 2016. Several papers have been published to compare the GWR and spatially random coefficient models. Wheeler and Calder (Wheeler and Calder 2007) showed that the spatially random coefficient model provides more accurate parameter estimates than GWR through a simulation study. Wheeler and Waller (Wheeler and Waller 2009) used a public health data set and showed that spatially random coefficient models provide more robust regression coefficients in the moderate to high multicollinearity situation. Finley (Finley 2011) compared these two models with several criteria. He concluded that although the GWR is faster and useful in fitting the data, the spatially random coefficient model has a significantly smaller prediction mean square error. Besides, in the GWR, the weight is fixed (a grid search across weights can be done), while in Bayesian Kriging, the optimal weight is estimated simultaneously. The methods from first to last become more complicated and time-consuming. The Bayesian Kriging method has mostly been used with a dense continuous correlation matrix. Park et al. (Park et al. 2018) used Bayesian Kriging to find optimal nitrogen recommendations. However, they only estimate the plateau spatially and restrict the number of locations to 160 to reduce computational time. When the number of random coefficients and sites increases, the model is very time-consuming. Hence, finding a method that is not only feasible for any data but also accurate is essential if Bayesian Kriging is to be competive with GWR. In this situation, sparsity in the precision matrices (covariance inverse) could make the code faster. Firstly, the code does not need to compute the inverse of a large covariance matrix. In addition, the sparsity could reduce the number of computations.

We use Bayesian Kriging to estimate a random coefficient linear plateau model and precisely estimate each location's optimal nitrogen level. In this model, intercept and plateau parameters vary across the field with a spatial correlation matrix. Calculations involving the spatial correlation matrix are a major reason for the slow computation. We estimate models with different spatial correlation functions and then compare computer time and estimation accuracy. Sacrificing a little precision to gain computation power might be needed to make such approaches commercially viable.

We estimate a linear plateau model on the 486 out of 1738 locations, corn (*Zea mays L.*) nitrogen response data from Bongiovanni and Lowenberg-DeBoer (Bongiovanni and Lowenberg-DeBoer 2000). While there is only one observation for each location, the estimated parameters differ for each location. The Hamiltonian Monte Carlo (HMC) algorithm, provided by Rstan, estimates the posterior density function. The optimal N value at each site is obtained by maximizing the expected profile using the posterior density.

2. *Bayesian Linear Plateau model*

The end goal is to find the optimum amount of nitrogen at each location. A common and effective data generating process for this purpose is a linear plateau model (Llewelyn and Featherstone 1997, Tembo et al. 2008. The innovation is that we assume that the parameters in these models vary by location. The proposed model is

$$y_i = \min(a_i + bN_i, Plateau_i) + \epsilon_i \qquad (1)$$

where y_i and N_i are the yield and the amount of nitrogen input in location i; $a_i$is the intercept and $Plateau_i$ is the plateau parameter. The effect of nitrogen (b) is fixed over space to reduce

the computational burden and $\epsilon_i \sim N(0, \sigma_\epsilon^2)$. Let $\boldsymbol{a} = (a_1, a_2, \cdots, a_n)^T$ and $\boldsymbol{Plateau} = (plateau_1, plateau_2, \cdots, plateau_n)^T$ are the $n \times 1$vector of parameters follow a Gaussian random process with spatial correlation matrices of $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_p$. Hence,

$$\begin{aligned} \boldsymbol{a} &\sim MVGP(\alpha \mathbf{1}, \boldsymbol{\Sigma}_0) \\ b &\sim N(\beta, \sigma_b^2) \\ \boldsymbol{Plateau} &\sim MVGP(p\mathbf{1}, \boldsymbol{\Sigma}_p) \end{aligned} \quad (2)$$

where α, β, p are the mean parameters, $\mathbf{1}$ is an $n \times 1$vector with all elements equal to one, $\sigma_b^2$is the variance component for the slope. Besides, we assume that the parameters a_i and $Plateau_i$ vary across locations, and parameters are spatially autocorrelated. Hence, $\boldsymbol{\Sigma}_0$, and $\boldsymbol{\Sigma}_p$are the $n \times n$ covariance matrices of the multivariate Gaussian process (MVGP) that depicts this behavior in the parameters. The covariance matrices in the MVGP could have varied structures.

3. *Optimal nitrogen level recommendation*

Assume that all other inputs are fixed, the optimal level of input nitrogen is selected to maximize expected profit:

$$\max_{N_i} E(\pi_i | N_i) = \max_{N_i} \int [Price(\min(a_i + bN_i, Plateau_i)) - rN_i] f(\boldsymbol{\Psi}) \boldsymbol{d\Psi} \quad (3)$$

where the $\boldsymbol{\Psi}$ contains all the parameters which should be estimated, and *f* is the posterior distribution function of parameters. Since the price and cost do not depend on the parameters, the integration is calculated only on the profit equation's production function.

Tembo et al. (Tembo et al. 2008) consider a plug-in method to find nitrogen's economic optimal value for a stochastic linear plateau. This method is not applicable in the current situation due to uncertainty in both the linear plateau model parts.

Ouedraogo et al. (Ouedraogo et al. 2020) used a grid search and found the expectation using the Monte Carlo sample of the posterior distribution. This method could be used to find the optimal value for each location; however, the grid search method in large data sets would be time-consuming.

The posterior distribution of the parameters converges in limit to the multivariate normal distribution (Van der Vaart 2000). We find the optimal value's analytical solution based on the normality assumption of the posterior distribution.

In the problem at hand, we want to calculate

$$E(\min(a_i + bN, plateau_i))$$

Nadarajah and Kotz (Nadarajah and Kotz 2008) provide the distribution and moment generating function of the minimum and maximum of two jointly normal random variables.

Let

$$(X_1, X_2)^T \sim N_2\left((\mu_1, \mu_2)^T, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right) \quad (4)$$

and $Y = \min(X_1, X_2)$ then

$$f_Y(y) = f_1(y) + f_2(y)$$

where

$$f_1(y) = \frac{1}{\sigma_1}\phi\left(\frac{y-\mu_1}{\sigma_1}\right)\Phi\left(\frac{\rho(y-\mu_1)}{\sigma_1\sqrt{(1-\rho^2)}} - \frac{(y-\mu_2)}{\sigma_2\sqrt{(1-\rho^2)}}\right)$$

$$f_2(y) = \frac{1}{\sigma_2}\phi\left(\frac{y-\mu_2}{\sigma_2}\right)\Phi(\frac{\rho(y-\mu_2)}{\sigma_2\sqrt{(1-\rho^2}} - \frac{(y-\mu_1)}{\sigma_1\sqrt{(1-\rho^2}})$$

and the mean of Y is

$$E(Y) = \mu_1\Phi\left(\frac{\mu_2-\mu_1}{\theta}\right) + \mu_2\Phi(\frac{\mu_1-\mu_2}{\theta}) - \theta\mu_1\phi(\frac{\mu_2-\mu_1}{\theta}) \quad (5)$$

where ϕ, and Φ are the PDF and CDF of normal distribution respectively, and $\theta = \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$.

If we assume that the $a_i, b,$ and $plateau_i$ are independent, and the posterior distribution of the parameters is

$$a_i \sim N\left(\bar{a_i}, \sigma_{a_i}^2\right)$$

$$b \sim N\left(\bar{b}, \sigma_b^2\right)$$

$$plateau_i \sim N\left(\bar{p_i}, \sigma_{p_i}^2\right)$$

then

$$a_i + bN \sim N\left(\bar{a}_i + bN, \sigma_{a_i}^2 + N^2\sigma_b^2\right)$$

$$plateau_i \sim N\left(\bar{p_i}, \sigma_{p_i}^2\right)$$

so, the expected value of $Y = \min(a_i + bN, plateau_i)$ is

$$E(Y) = (\bar{a} + \bar{b}N)\Phi\left(\frac{\bar{p}-\bar{a}-\bar{b}N}{\theta}\right) + \bar{p}\Phi(\frac{\bar{a}+\bar{b}N-\bar{p}}{\theta}) - \theta(\bar{a} + \bar{b}N)\phi(\frac{\bar{p}-\bar{a}-\bar{b}N}{\theta}) \quad (6)$$

$$\theta = \sqrt{\sigma_a^2 + N^2\sigma_b^2 + \sigma_p^2}$$

where the index *i* is dropped, for simplicity. We are looking for the *N* value to maximize equation (3). The first-order condition for this profit function in every location is

$$\frac{\partial E\pi}{\partial N} = \bar{b}price\left(1 - \Phi\left(\frac{\bar{a} + \bar{b}N - \bar{p}}{\theta}\right)\right) - r - \phi\left(\frac{\bar{a} + \bar{b}N - \bar{p}}{\theta}\right) \qquad (7)$$

The root of equation (7) cannot be obtained analytically. Hence we use the "optimize" function in R (Team 2013) to find the optimal value.

Some might argue that the intercept, slope, and plateau part cannot be independent in a real situation. By increasing the intercept, the slope will decrease. The plateau part in equation (1) may depend on the linear model, and the independence is questionable. Equation (6) can be adjusted for the correlated circumstances. Suppose that the parameters are correlated and the

$$\boldsymbol{V} = \begin{pmatrix} a \\ b \\ plateau \end{pmatrix} \sim MVN(\begin{pmatrix} \bar{a} \\ \bar{b} \\ \bar{p} \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix})$$

then, the vector $(a + bN, plateau)$ is equal to $A^T V$ where A is

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ N & 0 \\ 0 & 1 \end{pmatrix}$$

and $A^T V$ follows a multivariate normal distribution with mean and variance equal to

$$\boldsymbol{\mu} = (\bar{a} + \bar{b}N, \bar{p})$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} + N^2\sigma_{22} + 2N\sigma_{12} & \sigma_{13} + N\sigma_{23} \\ \sigma_{13} + N\sigma_{23} & \sigma_{33} \end{pmatrix}$$

If we want to rewrite the covariance matrix as (4), we have

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11}^* & \rho\sigma_1^*\sigma_2^* \\ \rho\sigma_1^*\sigma_2^* & \sigma_{22}{}^* \end{pmatrix}$$

where $\rho = \frac{\rho_{13}\sigma_1\sigma_3 + \rho_{23}N\sigma_2\sigma_3}{\sigma_1\sigma_3 + N\sigma_2\sigma_3}$, $\sigma_{11}^* = (\sigma_1 + N\sigma_2)^2$, $\sigma_{22}^* = \sigma_{33}$. Hence the expectation of the linear plateau model is equal to (6) with

$$\theta = \sqrt{(\sigma_1 + N\sigma_2)^2 + \sigma_{33} - 2\frac{\rho_{13}\sigma_1\sigma_3 + \rho_{23}N\sigma_2\sigma_3}{\sigma_1\sigma_3 + N\sigma_2\sigma_3}(\sigma_1 + N\sigma_2)\sigma_3} \qquad (8)$$

The first-order condition calculation is complicated and unnecessary because we maximize the equation (6) with the new θ, given in equation (8), directly.

For the switching regression model of Paris (1992), $y_{it} = \min(\mathrm{a} + bN + \kappa_{it}, \mu_m + \omega_{it})$, the tow random variables, κ_{it} and ω_{it} have marginal normal distributions. They do not necessarily have a joint bivariate normal distribution (the copula for the joint distribution is unspecified). So to use this approach in the Paris stochastic linear plateau would require an additional assumption of joint normality that is not imposed in the estimation.

4. *Spatial correlation matrices and their behavior*

In section 2, we consider the linear plateau model as a data generation function. We assume that the coefficients have spatial behavior, and this behavior can be explained with a covariance function in the normal distribution. Two well-known autoregressive precision matrices (covariance inverse) are Simultaneous Auto-regressive (SAR) and Conditional Auto-regressive (CAR). The term conditional in the CAR structure shows conditional independence in the distribution of each element dependent on neighbors' values; however, the simultaneous form mostly emphasizes regressing the random part on themselves simultaneously (Hooten et al.

2014). Conditional independence between the element i and j can be easily seen in the precision matrix ($q_{ij} = 0$).

The CAR model is usually presented as a conditional distribution

$$\beta_i|\boldsymbol{\beta}_{-i} \sim N(\sum_{j=1}^{n} c_{i,j}\beta_j, m_{i,i})$$

where $\boldsymbol{\beta}_{-i}$ is the vector of all elements of vector $\boldsymbol{\beta}$ except β_i, $c_{i,j}$ are the *i*th and *j*th element of spatial weight matrix $\boldsymbol{C}$ and $\boldsymbol{M}$ is a diagonal matrix with positive diagonal elements of $m_{i,i}$. Following Besag (Besag 1974), if $(\boldsymbol{I} - \boldsymbol{C})^{-1}\boldsymbol{M}$ is a positive definite matrix then the CAR model can be written as

$$\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{CAR}) \quad (9)$$

where $\boldsymbol{\Sigma}_{CAR}$, the covariance matrix is (Ver Hoef et al. 2018a)

$$\boldsymbol{\Sigma}_{CAR} = (\boldsymbol{I} - \boldsymbol{C})^{-1}\boldsymbol{M}.$$

In practice, usually $\boldsymbol{\Sigma}_{CAR} = \frac{1}{\tau^2}(diag(\boldsymbol{W}\mathbf{1}) - \rho_c \boldsymbol{W})^{-1}$ is used where $\boldsymbol{W}$ is contiguity matrix**,** $\mathbf{1}$ is the vector of ones, and ρ_c shows the amount of dependency between neighbors (Ver Hoef et al. 2018b).

In the SAR model, an $n \times n$ weight matrix, B, relates the vector of parameters to themselves. In contrast to the CAR model, the SAR model can directly define the vector complete distribution. If we define

$$\boldsymbol{\beta} = \boldsymbol{B}\boldsymbol{\beta} + \boldsymbol{\vartheta}$$

where the matrix $\boldsymbol{B}$ is a $n \times n$ spatial weight matrix and $\boldsymbol{\vartheta} \sim MVN(\mathbf{0}, \boldsymbol{\Omega})$, then

$$\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{SAR}). \quad (10)$$

In the SAR model, the $\boldsymbol{\Sigma}_{SAR} = (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\Omega}(\boldsymbol{I} - \boldsymbol{B}^t)^{-1}$, where $\boldsymbol{B}$ is not necessarily a symmetric matrix since $\boldsymbol{\Sigma}_{SAR}$ is symmetric even if $\boldsymbol{B}$ is not symmetric. In the SAR model, it is enough for $(\boldsymbol{I} - \boldsymbol{B})$ to be a non-singular matrix, $\boldsymbol{\Omega}$ be a diagonal matrix with positive values and $b_{ii} = 0$. In practice, usually consider $\boldsymbol{B}$ as a row standardized non-symmetric contiguity matrix. So the covariance matrix is $\boldsymbol{\Sigma}_{SAR} = (\tau(\boldsymbol{I} - \rho\, \boldsymbol{W}^*)(\boldsymbol{I} - \rho\, \boldsymbol{W}^{*\prime}))^{-1}$ where $\boldsymbol{W}^*$ is the row standardized contiguity matrix.

Another common framework in geostatistics modeling is considering the correlation matrix as an elementwise decreasing function of distance among locations. Suppose that

$$\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma}_{ij} = cov\left(\beta_i, \beta_j\right) = \sigma^2 e^{-\frac{d_{ij}}{\rho}}$ is a positive definite covariance matrix, d_{ij} is the Euclidean distance between location i and j, ρ is the effective range and σ^2 is the sill. The exponential covariance matrix implies that the observations near each other are highly correlated while the far observations are nearly independent. Although this model uses the correlation matrix directly and is straightforward to interpret, the researcher needs to specify a point to represent each unit, and for an extensive data set, fitting this model could be time-consuming.

Although the precision matrices in the CAR and SAR models are sparse, which leads to faster computing, the related covariance matrix is dense. It shows that although a structure is assumed on the covariance matrices, no extra independence is considered between the locations. Besides, in both CAR and SAR covariance matrices, the correlation between them decreases by increasing the distance between two places.

5. *Model fitting and layer specification in the Bayesian framework*

The spatial behavior parameters need to be estimated, and there is uncertainty about their actual value. A hierarchical Bayesian perspective implies that some uncertainty may exist in the mean and correlation structures of the prior in the data generating process. So the proposed model in the previous section contains three layers: likelihood, process (priors), and hyper prior level. We assume that the response variable follows a linear plateau model in the likelihood layer, a non-linear model. Also, we assume that the parameters in this model follow a multivariate Gaussian process. The dependency between the parameters in this model handles the Gaussian process's correlation structures. The third layer contains the hyperparameters priors that can assure that the covariance matrix is positive definite. Based on the Bayesian framework, we can express the posterior distribution of the parameters as

$$f(\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \boldsymbol{\Theta}_3 | \boldsymbol{Y}) \propto f(\boldsymbol{Y} | \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) \times f(\boldsymbol{\Theta_2} | \boldsymbol{\Theta_3}) \times f(\boldsymbol{\Theta}_3)$$

where $f(\boldsymbol{Y} | \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2)$, $f(\boldsymbol{\Theta_2} | \boldsymbol{\Theta_3})$, and $f(\boldsymbol{\Theta}_3)$ are the likelihood layer, process layer, and hyper prior layer, respectively. $\boldsymbol{\Theta}_1 = (a_1, a_2, \cdots, a_n, b, plateau_1, plateau_2, \cdots, plateau_n, \sigma_\epsilon)$ is the set of parameters for the likelihood layer, $\boldsymbol{\Theta}_2 = (\alpha, \beta, p, \rho_0, \rho_p, \tau_0, \tau_p)$, and $\boldsymbol{\Theta}_3$ is the set of all hyperparameters in the distributions of the $\boldsymbol{\Theta}_2$. Likelihood layer is

$$f(\boldsymbol{Y} | \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2) = \left(\frac{1}{\sqrt{2\pi\sigma_\epsilon^2}}\right)^n \exp \frac{(\boldsymbol{y} - \boldsymbol{\mu})'(\boldsymbol{y} - \boldsymbol{\mu})}{2\sigma_\epsilon^2}$$

where $\boldsymbol{y}$ is the vector of yield data, $\boldsymbol{\mu}$ is the vector with the elements equal to $\mathrm{E}(\min(a_i + bN_i, Plateau_i) + \epsilon_i)$, n is the number of observations, and σ_ϵ^2 is the variance component of ϵ.

The process layer deals with the model's spatial structure and finds a specific estimate for each location. The correlation matrix plays a vital role in the spatial structure of the data. Different

parameters have been defined in the three mentioned methods, which should be determined in this layer. In the CAR and SAR model, the parameters are τ, and ρ, and in the exponential model, the parameters are σ, and ρ.

The stochastic spatial process in this model has distribution

$$f(\boldsymbol{\Theta}_2|\boldsymbol{\Theta}_3) = \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}_p|}}\exp\left[-\frac{1}{2}(\boldsymbol{plateau} - \bar{\boldsymbol{P}})'\boldsymbol{\Sigma}_p^{-1}(\boldsymbol{plateau} - \bar{\boldsymbol{P}})\right] \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}_0|}}\exp\left[-\frac{1}{2}(\boldsymbol{a} - \bar{\boldsymbol{a}})'\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{a} - \bar{\boldsymbol{a}})\right] \frac{1}{\sqrt{(2\pi)^n\sigma_b^2}}\exp\left[-\frac{1}{2\sigma_b}(b - \bar{b})^2\right]$$

where $\bar{\boldsymbol{P}}$, and $\bar{\boldsymbol{a}}$, are $\bar{p}\mathbf{1}$, $\bar{a}\mathbf{1}$; respectively. The covariance matrix in this layer could be any of the covariances defined in the previous section.

The hyper prior layer contains the priors for all the parameters in the process layer and some from the likelihood layer. The priors for p, α, and β are normal, with large variances to be non-informative. The priors for the variance components are inverse gamma with parameters $\alpha = 2$, $\beta = 1$ for the τ. Obviously, the covariance matrix in the normal distribution must be positive definite. Since the value of ρ could affect the positive definity of the covariance matrix, a careful selection of prior seems necessary. In practice, the restriction $\frac{1}{\lambda_1} < \rho < \frac{1}{\lambda_n}$ should be imposed in the prior where the λ_i is the eigenvalue of the W, however, the restriction turns to $\frac{1}{\lambda_1} < \rho < 1$ when the row standardized form of W is used (Haining, 1993). We consider $0 < \rho < 1$, standard uniform prior, because only positive autocorrelation is expected. The improper prior proportional to the inverse of standard error for the variance component of σ_ϵ is considered in all three models. The prior for the sill and range parameter in the exponential correlation function

consider being an improper distribution of $f(\rho, \sigma) \propto \frac{1}{\sigma}$. Fuglstad et al. 2015 showed that this improper prior has stable results and can be used widely.

6. *Data and Data Analysis*

We consider the corn yield response to nitrogen from Bongiovanni and Lowenberg-DeBoer 2000. The data were collected from a strip plot design in "Las Rosas" farm in Cordoba's southwestern corner of Argentina. The data are for the single year 1999. Six different levels of nitrogen, namely 0,19,53,66,106, and 131.5 kg ha^{-1}, were applied to the farm based on a strip plot design. The highest nitrogen rate was higher than the value of nitrogen that was expected to maximize the response. The yield data and the selected nitrogen levels are given in Fig.1 and Fig.2, respectively. The original data contain 1738 locations that were digitalized as polygons. The centroid point is generated and considered as a data point in each area. 486 plots of the fields are selected from the data set such that all six levels of nitrogen were chosen, and we face unbalanced data for each level.

To estimate the linear plateau model in the Bayesian framework, the HMC algorithm is employed through the Rstan package in R. HMC algorithm in Stan uses a dynamic Hamiltonian Markov Chain to reduce the time of calculation and increase the chance of convergence. Different iteration and warmup values are employed for models to meet the convergence criteria for each model. The number of iterations and warmup for each model are given in Table1 ($number\ of\ iteration \times number\ of\ chains$). Different convergence criteria such as Gelman-Rubin statistics (Rhat) showed the ratio for the variance of parameters when the chain's data are pooled, and the number of effective samples is considered. Trace and Trunk plot, which shows

the Markov property of the data and mixing chain property of chains, respectively, for the model also monitored to ensure convergence of the estimators and the results' reliability in the model. Models with SAR and exponential correlation matrix converge nicely. However, these criteria in the CAR correlation matrix model are not met entirely like the two other models. The time for running this model is given in Table 1. The site-specific estimated results of $a_i, plateau_i,$ and y_i are given in figure 3 to figure 5.

Table 1 shows the estimated values for the parameters and their related Gelman-Rubin statistics (Rhat) for all three models. The estimated correlation parameters emphasize the spatial behavior's existence in the model's parameters that should be considered in the data analysis. We can compare the posterior likelihood value for the models since they have the same number of parameters. The time for getting an effective sample indicates that although the CAR model is far faster than the SAR model for creating an iteration, it needs more iterations and hence more time to converge. Also, the Rhat statistics show that the CAR model does not converge as well as the SAR model even with more iterations. Based on the likelihood criterion, it is clear that the model with exponential correlation model is the best and the model with CAR correlation matrix is not as good as the SAR model. However, the exponential correlation matrix model is not feasible for a more extensive data set due to the computational burden. The SAR model fits well on the data and simultaneously met the convergence criteria. Besides, this model is faster than the two other models in simulating an effective sample.

The optimal nitrogen value is calculated based on the posterior distribution of the parameters for every specific part of the field. The posterior distribution of each a_i, b and $plateau_i$ is estimated, then the results plug into equation (6), and the optimal value for nitrogen is obtained.

The results are given in figure 6. The exponential differs more than the other two because it was estimated using a smaller dataset.

7. *Conclusions*

Recently, Park, et al. (2020) provide a stochastic linear plateau model with site-specific parameters; however, they consider only the plateau part as a site-specific parameter. In this paper, intercept and plateau parameters in the plateau model can be determined specifically for each location. Three different correlation matrices are considered.

In this application, all three models perform well in fitting the data set. In the CAR and SAR model, the neighbors' covariance is considered equal without attention to the distance between these two points. In the example in hand, the neighbors have a similar distance. However, in general, the distance between every neighbor could be far different. Table 1 shows that the CAR and SAR models are far faster than the exponential covariance model. They can be more easily used for large datasets due to the precision matrix's sparsity. If the number of locations is large and the data have some well-defined equally distant regions, the CAR and SAR models are feasible. Simultaneously, the exponential correlation function cannot be used in large data sets due to the computational burden of calculating the inverse and determinant of an extensive dense matrix. In the problem at hand, it seems that the model with SAR correlation matrix is preferred to the CAR based on the likelihood and the computational time to simulate an effective sample. The site-specific optimal nitrogen value depends on the mean of the posterior distribution and the variance of estimated parameters in each site. So, the accuracy of the estimation procedure could significantly affect the optimal recommendation value.

Acknowledgements

The work benefited directly or indirectly from the work of Eunchun Park, Dayton Lambert, Whoi Cho, and Stephen Wild.

Funding

This work was supported by the A.J. and Susan Jacques Chair, the Oklahoma Agricultural Experiment Station, and USDA National Institute of Food and Agriculture [Hatch Project number OKL03170].

References

Anselin, L., Bongiovanni, R., Lowenberg-DeBoer, J., 2004. A spatial econometric approach to the economics of site-specific nitrogen management in corn production. American Journal of Agricultural Economics 86, 675-687.
Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society: Series B (Methodological) 36, 192-225.
Bongiovanni, R., Lowenberg-DeBoer, J., 2000. Nitrogen management in corn using site-specific crop response estimates from a spatial regression model, Proceedings of the Fifth International Conference on Precision Agriculture.
Dambon, J.A., Sigrist, F., Furrer, R., 2020. Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. arXiv preprint arXiv:2001.08089.
Evans, F.H., Recalde Salas, A., Rakshit, S., Scanlan, C.A., Cook, S.E., 2020. Assessment of the use of geographically weighted regression for analysis of large On-Farm experiments and implications for practical application. Agronomy 10, 1720.
Finley, A.O., 2011. Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. Methods in Ecology and Evolution 2, 143-154.
Fuglstad, G.-A., Simpson, D., Lindgren, F., Rue, H., 2015. Interpretable priors for hyperparameters for Gaussian random fields. arXiv preprint arXiv:1503.00256.
Griffin, T.W., Dobbins, C.L., Vyn, T.J., Florax, R.J., Lowenberg-DeBoer, J.M., 2008. Spatial analysis of yield monitor data: case studies of on-farm trials and farm management decision making. Precision Agriculture 9, 269-283.
Haining, R., 1993. Spatial data analysis in the social and environmental sciences. Cambridge University Press.
Hooten, M.B., Ver Hoef, J.M., Hanks, E.M., 2014. Simultaneous Autoregressive (SAR) Model. Wiley StatsRef: Statistics Reference Online, 1-10.
Hurley, T.M., Oishi, K., Malzer, G.L., 2005. Estimating the potential value of variable rate nitrogen applications: A comparison of spatial econometric and geostatistical models. Journal of Agricultural and Resource Economics, 231-249.

Lambert, D.M., Lowenberg-Deboer, J., Bongiovanni, R.J.P.A., 2004. A comparison of four spatial regression models for yield monitor data: A case study from Argentina. 5, 579-600.
Liu, Y., Swinton, S.M., Miller, N.R., 2006. Is site-specific yield response consistent over time? Does it pay? American Journal of Agricultural Economics 88, 471-483.
Llewelyn, R.V., Featherstone, A.M., 1997. A comparison of crop production functions using simulated data for irrigated corn in western Kansas. Agricultural Systems 54, 521-538.
Nadarajah, S., Kotz, S., 2008. Exact distribution of the max/min of two Gaussian random variables. IEEE Transactions on very large scale integration (VLSI) systems 16, 210-212.
Ouedraogo, F.B., Brorsen, B.W., Biermacher, J.T., Rohla, C.T., 2020. Effects of Pruning at Planting on Pecan Trunk Development and Total Shoot Growth. HortTechnology 30, 248-250.
Park, E., Brorsen, B., Harri, A., 2016. Using Bayesian Spatial Smoothing and Extreme Value Theory to Develop Area-Yield Crop Insurance Rating. Selected paper, 2016 Annual Meeting, Boston, MA, USA.
Park, E., Brorsen, W., Li, X., 2018. How to Use Yield Monitor Data to Determine Nitrogen Recommendations: Bayesian Kriging for Location Specific Parameter Estimates. Agricultural and Applied Economics Association.
Rodrigues, M.S., Corá, J.E., Castrignanò, A., Mueller, T.G., Rienzi, E., 2013. A spatial and temporal prediction model of corn grain yield as a function of soil attributes. Agronomy Journal 105, 1878-1887.
Team, R.C., 2013. R: A language and environment for statistical computing. Vienna, Austria.
Tembo, G., Brorsen, B.W., Epplin, F.M., Tostão, E., 2008. Crop input response functions with stochastic plateaus. American Journal of Agricultural Economics 90, 424-434.
Trevisan, R., Bullock, D., Martin, N., 2020. Spatial variability of crop responses to agronomic inputs in on-farm precision experimentation. Precision Agriculture.
Van der Vaart, A.W., 2000. Asymptotic statistics. Cambridge university press.
Ver Hoef, J.M., Hanks, E.M., Hooten, M.B., 2018a. On the relationship between conditional (CAR) and simultaneous (SAR) autoregressive models. Spatial statistics 25, 68-85.
Ver Hoef, J.M., Peterson, E.E., Hooten, M.B., Hanks, E.M., Fortin, M.J., 2018b. Spatial autoregressive models for statistical inference from ecological data. Ecological Monographs 88, 36-59.
Wheeler, D.C., Calder, C.A., 2007. An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. Journal of Geographical Systems 9, 145-166.
Wheeler, D.C., Waller, L.A., 2009. Comparing spatially varying coefficient models: a case study examining violent crime rates and their relationships to alcohol outlets and illegal drug arrests. Journal of Geographical Systems 11, 1-22.

Table 1. Parameters estimation and the Rhat value in three models for the yield

Parameters	CAR (Rhat)	SAR (Rhat)	Exponential (Rhat)
$\bar{a}$	60.86 (2.01)	58.85 (1.01)	58.69(1.01)
ρ_{int}	0.97 (1.00)	0.97 (1.00)	0.04(1.00)
τ_{int}	0.35 (1.01)	0.25(1.00)	10.17(1.00)
$\bar{b}$	0.11(1.10)	0.13(1.00)	0.11(1.01)
$\bar{p}$	69.35(2.25)	68.17(1.01)	70.15 (1.01)
ρ_p	0.89 (1.91)	0.96 (1.00)	1.29 (1.00)
τ_p	0.35 (1.18)	0.28(1.00)	0.60 (1.01)
Log posterior likelihood	-2427.81(1.81)	-1811.05(1.00)	794.18(1.14)
Max time for an effective sample (parameter)	21481 ($\bar{p}$)	348.52 ($\bar{a}$)	108952 (lp)
Time(hours)[a]	30.96	25.58	268.55
Iteration	600000	435000	60000
Warmup	300000	120000	35000

[a] Desktop PC with Intel Core i5-9500 CPU @ 3.00 GHz and 32 GB DDR4

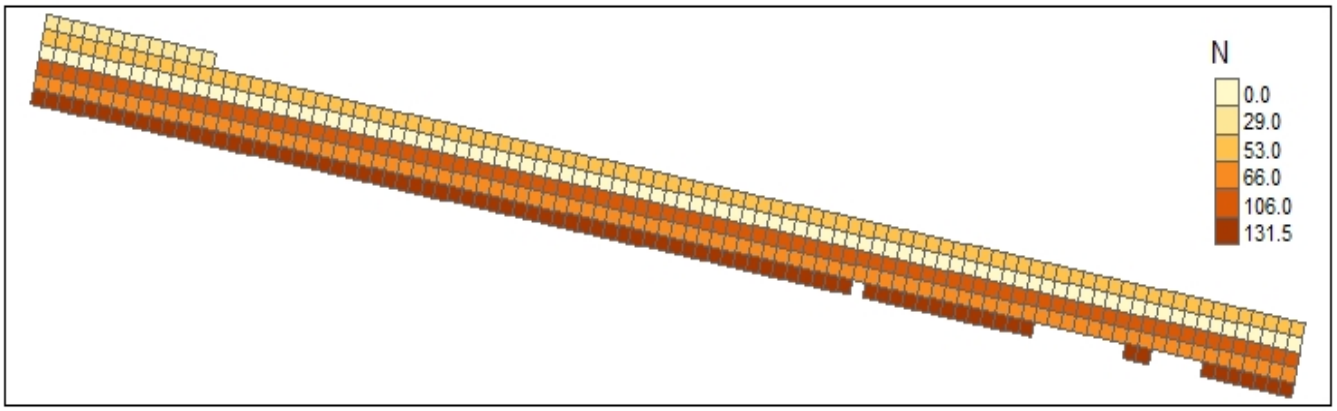

Figure 1. The amount of applied nitrogen

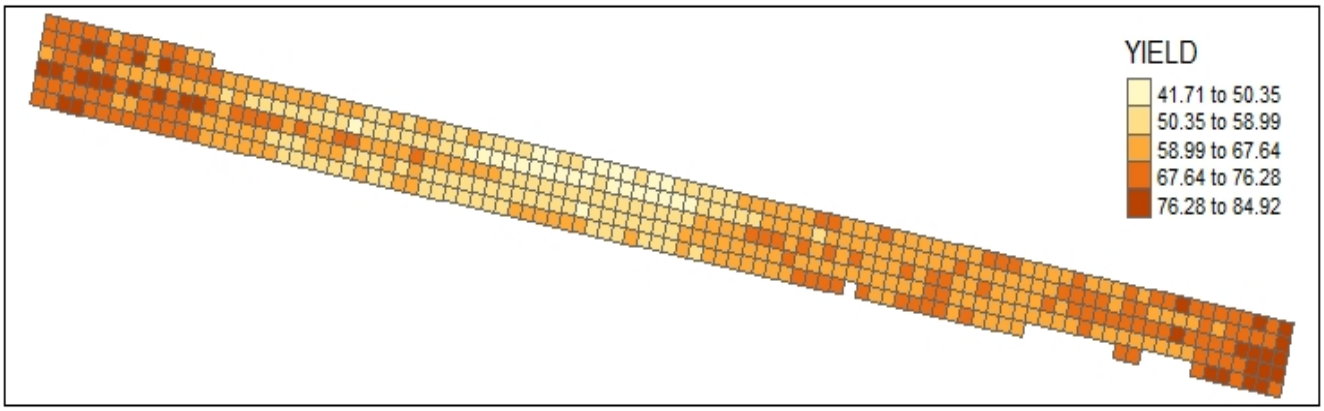

Figure 2. Value of actual yield

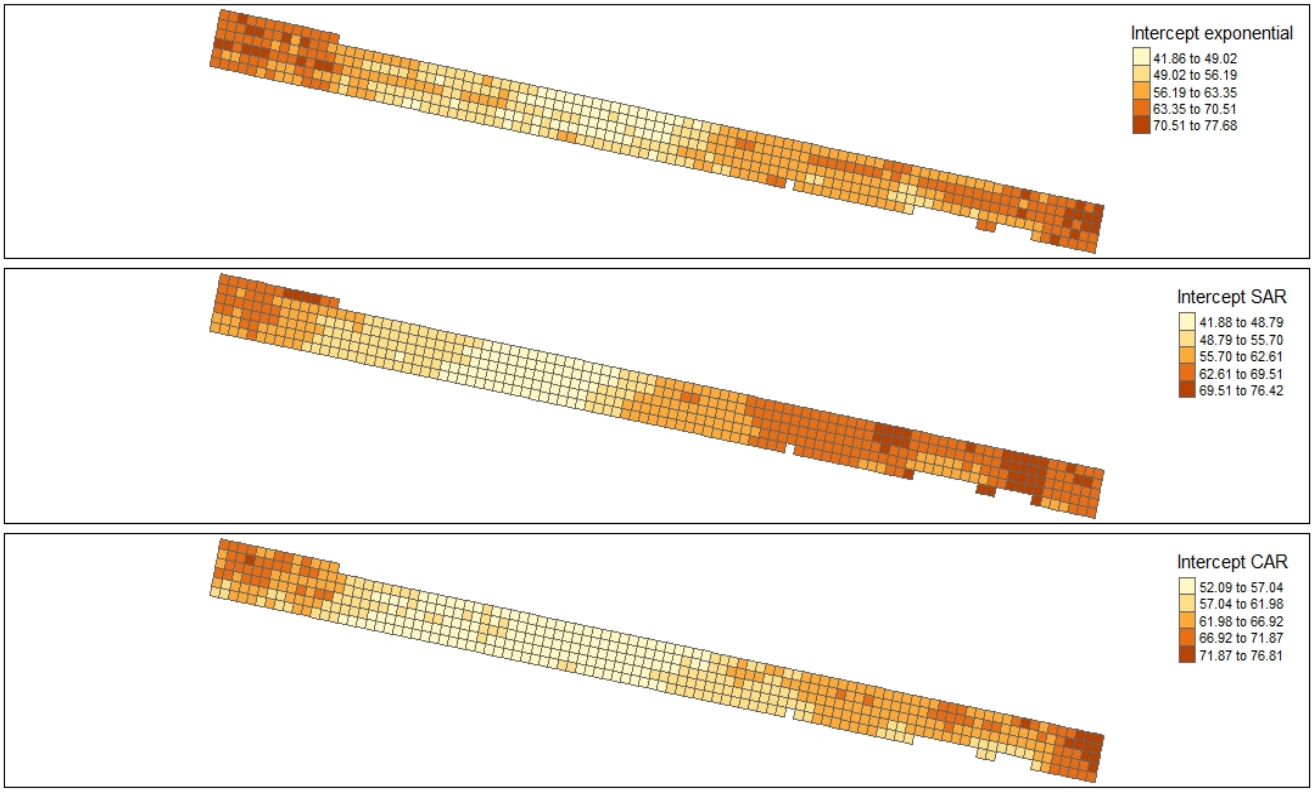

Figure 3. The intercept value for exponential, SAR, and CAR model

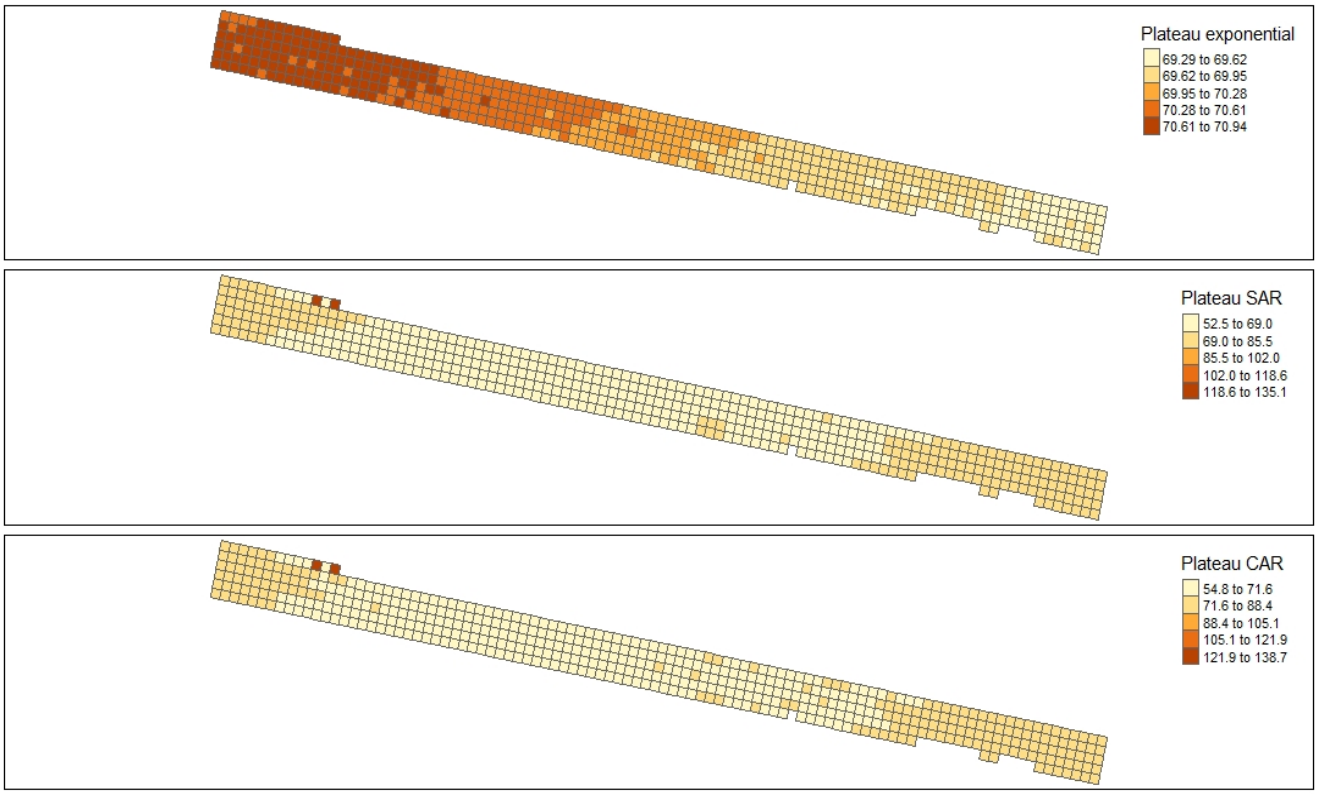

Figure 4. The plateau value for exponential, SAR, and CAR model

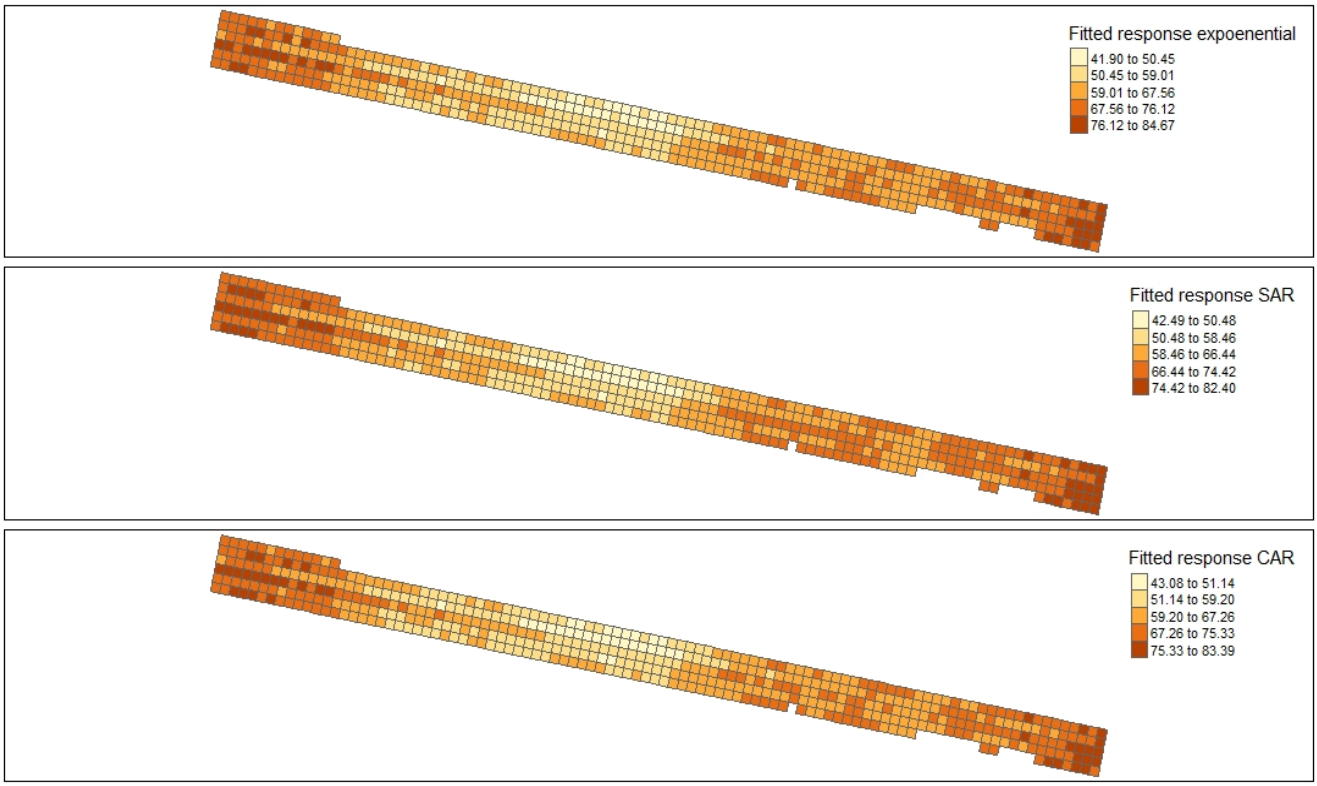

Figure 5.The fitted value for exponential, SAR, and CAR model

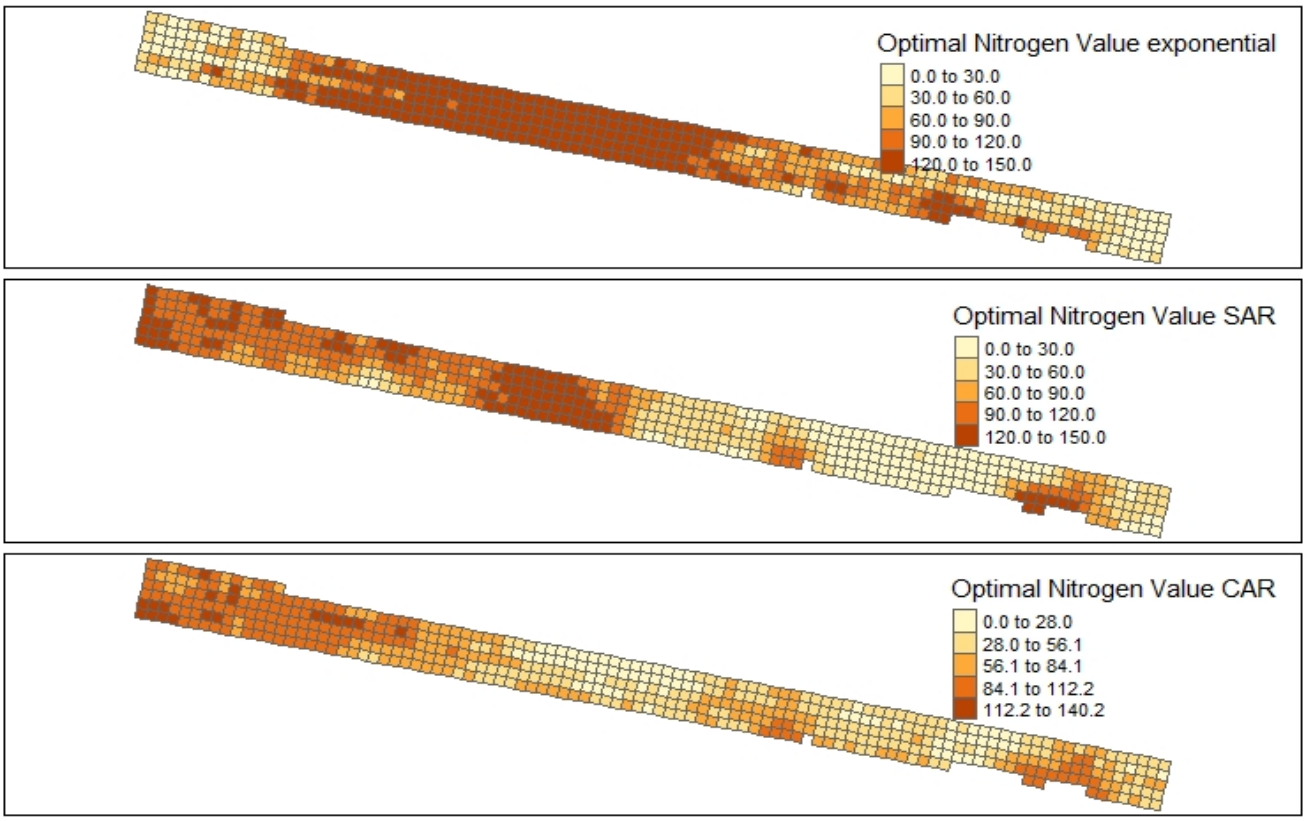

Figure 6. The optimal nitrogen value for exponential, SAR, and CAR model