Using Bayesian Spatial Smoothing and Extreme Value Theory to Develop Area-Yield Crop Insurance Rating

Eunchun Park, B. Wade Brorsen, and Ardian Harri

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2

Eunchun Park is a graduate research assistant, B. Wade Brorsen is a regents professor and A.J. and Susan Jacques Chair, Department of Agricultural Economics, Oklahoma State University, Stillwater, Oklahoma, Ardian Harri is an associate professor, in the Department of Agricultural Economics at Mississippi State University. The research was funded by the A.J. & Susan Jacques Chair as well as the Oklahoma Agricultural Experiment Station and USDA National Institute of Food and Agriculture, Hatch Project number OKL02939.

Abstract

Rating of insurance premiums depends on the probability of events in the tail of the distribution. Extreme value theory provides a promising way to assess tail risk. We assume that crop yield follows a Generalized Pareto Distribution (GPD), which is a family of extreme value distributions that has advantages for modeling rare events. GPD parameters are fitted using county-level historical winter wheat yield (1970-2014). Spatial smoothing with Kriging parameters is used within a Bayesian hierarchical framework that helps overcome a lack of data due to the rarity of extreme events. We assume that the spatial correlation of crop yield is embedded in the parameters of the GPD. To obtain the posterior distribution, we use Metropolis-Hastings (MH) steps within a Gibbs sampler. Maximum likelihood estimates of the GPD parameters are used for candidate density in the MH step. In the process, MCMC chains are run for 100,000 iterations and burn-in for the first 20,000 observations. We use Deviance Information Criterion (DIC) and out of sample performance to evaluate the quality of the model. From the estimated results, we verify spatial correlation in crop yield, which substantially affects estimates of posterior distributions of GPD parameters. We further simulate spatial random effect based on posterior values of Kriging parameters (range and sill) to visualize and verify the form of spatial correlation. Estimated premiums from an existing method from which current premiums are based, tend to underestimate premium rates compared to our new proposed method.

Key words: Rating crop insurance contract, Bayesian spatial smoothing, Spatial correlation, Bayesian hierarchical model, Extreme value theory.

Introduction

Agricultural producers purchase crop yield insurance to protect against the loss of crops from unexpected events. Area-yield insurance can avoid the problems of moral hazard and transactions cost arising from individual crop insurance policies since the indemnity payment is based on average yields of the county rather than the actual yields of the individual insured. A poorly rated product, however, does not protect against adverse selection and regional inequities since a producer who has sufficient information about the county's crop yield distribution could avoid overpriced area-yield insurance. Federally managed area-yield insurance has been offered by Risk Management Agency (RMA) for major crops in the United States. The Group Risk Plan (GRP), which was established in 1993, currently provides area-yield insurance based on the county level yield data from National Agricultural Statistics Service (NASS).

Several studies have developed methods to calculate premium rates for area yield crop insurance. Skees, Black, and Barnett (1997) proposed a premium rating method that is used as a base model in current U.S. design for area crop insurance. They employed a two-knot, linear spline to fit historical yield data and then calculate premium rates using residuals of the regression. Goodwin and Ker (1998), Ker and Goodwin (2000), and Ker and Coble (2003) proposed alternative methods. These proposed methods regress deterministic trend using historical yield and then adjust assumptions of heteroscedasticity from the estimated residuals to determine premium rates and expected indemnities. Current policies are rated following a method similar to Harri, et al. (2011). They use a hybrid approach where some parameters are estimated at the state or district level and some use only the data for each county. Crop yield distribution has been modeled with many different approaches in past studies. Some studies model crop yield with parametric distributions, such as the beta distribution (Nelson and Preckel; Tirupattur, Hauser, and Chaherli) or the log-normal distribution (Tirupattur, Hauser, and Chaherli; Jung and Ramezani; Stokes). Others have suggested a nonparametric framework (Ker and Goodwin). Although a wide range of crop yield models have been used in previous studies, it is generally agreed that crop yields are not normally distributed.

Crop yield of a given county tends to be spatially dependent with that of nearby counties because weather, geological features, and other hidden features that could potentially affect crop yields in one county are likely to be similar in neighboring counties (Annan, Tack, and Harri 2014; Du et al. 2015). Therefore, considering this spatial correlation can potentially lead to obtaining more precise yield distribution and thus more accurately rated premium rates. Several studies have proposed various distributions and dealt with heteroskedasticity, yet assumptions about the spatial correlation of the distribution are not fully adjusted. For instance, although the method suggested by Harri, et al. (2011) estimates some parameters at the state or district level and gives weight for county level estimation, when estimating the density in one county, only the historical yield data from the county is used. There is considerable interest in using spatial autocorrelation to improve crop insurance rating as reflected in Ker, Tolhurst, and Liu (2015) suggesting Bayesian Model Averaging (BMA) across space. They first estimate a posterior density for each county using its own data. Then they impose weights to posterior densities of other counties based on the fit of these densities to the observations of the target county. The posterior density of the target county is then a weighted average of its own posterior density and densities from other counties. This method does not require any knowledge in density similarity, and can be applied with both parametric and non-parametric estimators. What Ker, Tolhurst, and

Liu do is very different[1] than what we do, but their work does indicate that others are working toward the same goal.

A typical difficulty in the analysis of crop yield distribution is that the number of time-series observations is very limited. Particularly, estimating the probability of rare events is difficult since rare events usually have limited observations. Thus, we choose to face the challenge of directly considering spatial correlation as part of the estimation method. We address the problem by using a Bayesian hierarchical model. Unlike BMA method, the Bayesian hierarchical model directly incorporates the spatial correlation into the model using a spatial random effect. Therefore, posterior densities are estimated jointly using data for all counties being considered. We estimate parameters of the yield distribution for each region with a specified functional form of the spatial covariance matrix in the process layer of the hierarchical model. From this approach, spatial correlation of parameters is used to get more precise crop yield distributions and therefore more accurate premium rates.

The objective of the study is to develop a method for determining more accurate area yield insurance rates based on extreme value theory and Bayesian spatial smoothing. Crop yields are assumed to follow a Generalized Pareto Distribution (GPD), which is a member of the extreme value distribution family. GPD suggests that the focus should be on estimating the tail of the

---

[1] While we of course like our approach better, there may be instances where their approach could be preferred. Our approach has an explicit functional form for spatial correlation. Our approach yields estimates of the spatial correlation not provided by Bayesian Model Averaging. We use Bayesian methods because it lets us estimate the GPD distribution with more precision. The Ker, Tolhurst, and Liu (2015) approach is more like a nonparametric approach to spatial smoothing. Their approach is less restrictive than ours, but also less precise if our restrictions are at least close to being true. We would also argue that our approach may be more intuitive and easier to explain to RMA and to producers.

distribution that generates indemnity payments. The county-level historical wheat yields (1970-2014) from NASS are used for the estimation. GPD and Kriging parameters (sill and range) for the spatial smoothing are estimated under the Bayesian hierarchical framework. We assume that GPD parameters of the counties are spatially dependent and assumed to follow a multivariate normal distribution with non-zero spatial covariance across the counties. We employ Metropolis-Hastings (MH) steps within a Gibbs sampler to update posterior densities. Maximum likelihood estimates are used for candidates for the MH step so as to increase acceptance rate in the Markov Chain Monte Carlo (MCMC) procedure. We verify spatial correlation in crop yield distributions from the posterior density, and simulate spatial random process using the Kriging parameters to visualize and verify the form of spatial correlation across counties. We find that the estimated premiums from the RMA model[2] tend to underestimate premium rates compared to our proposed method. We compare out of sample performance of two approaches by assuming a representative insurance company that chooses whether to retain or cede polices under the Standard Reinsurance Agreement (SRA). We verify statistical significance of performance through bootstrapping. Our results show that the new model outperforms the RMA model.

The remainder of the paper is structured as follows. In the following section, we introduce an overview of the Generalized Pareto Distribution (GPD). We then describe the Bayesian hierarchical approach for spatial smoothing, which is used to obtain the posterior distribution of the spatially adjusted GPD parameters. Next, we briefly explain about MCMC structure of the

---

[2] We refer to the Harri et al. (2011) model as the RMA model. The exact RMA model is proprietary and not known. We compare our model to the Harri et al. (2011) model on which the RMA model is based. The RMA is believed to do some heuristic adjustments including some spatial smoothing.

study. The subsequent section describes the testing out of sample performance of the new proposed method and RMA method. The final section provides the results and conclusions.

Generalized Pareto Distribution

Extreme value theory is used for analysis of low probability events. The theory states that the tail of a loss distribution can be approximated by a Generalized Pareto Distribution (GPD), which is a member of the extreme value family. The extreme value family can be written in a simple form involving three parameters: location (λ), scale ($\sigma > 0$), and shape (ξ). Positive shape parameter ($\xi > 0,$ Fréchet type) implies a heavy tail distribution and a negative shape parameter ($\xi < 0,$ Weibull type) implies a bounded upper tail, and zero-value shape parameter ($\xi < 0,$ Gumbel type) implies a light tail distribution. Suppose X is a random variable. Now consider the conditional distribution of X given that it exceeds u. If F is a cumulative density function (CDF) for X, the probability of X exceeding x given X is greater than the threshold u can be given as

(1)
$$\mathrm{P}(X > x|\, X > u) = \frac{1 - F(x)}{1 - F(u)} = \begin{cases} (1 + \dfrac{\xi(x - \lambda)}{\sigma_u})^{-1/\xi} & \xi \neq 0 \\ \exp^{-(x-u)/\sigma} & \xi = 0 \end{cases}$$

where scale parameter $\sigma > 0$, and shape parameter $-\infty < \xi < \infty$.

One advantage of using a GPD model for the tail of the distribution is that the quantiles have closed form. Using equation (1) above, the quantiles can be defined as $X_q = F^{-1}(q)$,

(2) $$X_q = \begin{cases} u + \frac{\sigma}{\xi}\left[(\zeta_u/q)^{\xi} - 1\right] & \xi \neq 0 \\ u + \sigma \ln(\frac{\zeta_u}{q}) & \xi = 0 \end{cases}$$

where $\zeta_u = 1 - F(u)$ denotes the probability of exceeding the threshold u. Therefore, once we obtain the parameters of the distribution, a probability that a variable exceeds some specific threshold can be directly computed.

Suppose crop yields of the counties are spatially correlated. Then the parameters for the distribution should also be spatially correlated. Our focus in the paper is on how the distribution of crop yield spatially varies across counties and obtains the crop yield distribution of each county while considering the spatial correlation. Although the statistical tool for modeling univariate extremes are well-defined, extending these tools to model spatial data (i.e. crop yield data) with multivariate specification requires a more complex and advanced approach. We use a Bayesian hierarchical model to reflect spatial correlation via parameter random effects. Metropolis-Hastings(MH) steps within a Gibbs sampler are used to update the parameters of the model. Similar to Cooley, Nychka, and Naveau (2007)[3], we use maximum likelihood estimators of the GPD parameters as candidate densities for the posterior distribution.

The conceptual steps for the estimation are as follows. Given the GPD for crop yield, we add a spatial process considering σ_i and ξ_i to be functions of characteristics of each location i such as

---

[3] They use Bayesian hierarchical model with spatial smoothing to fit precipitation extremes for flood planning purpose. Our basic MCMC procedure (Metropolis-Hasting within Gibbs sampler) is identical to their work. However, they assume uniform prior distribution both for Kriging parameters (sill and range), whereas we assume inverse gamma prior for Kriging parameters to get more stable posterior distributions.

average yield, longitude, or latitude. In a Bayesian setting, the functions σ_i and ξ_i are the latent processes (in the process layer) of the hierarchical framework under the assumption of Gaussian spatial random processes[4]. We must choose a threshold level u in order to fit the GPD. The threshold selection is one difficulty in fitting the GPD data and thus finding an optimal approach to select the threshold is still an active research area. One may think that it is natural to obtain the threshold u by maximum likelihood along with the other parameters. However, this approach will not be stable because the number of observations is changed as u is changed. This fact will lead to a discontinuous or unbounded likelihood function. We focus on the estimation of each county's expected indemnity and premium rates under the predicted yield $\hat{y}_i$ and coverage rate λ. We set the average of historical yield $u_i = \bar{y}_i$ as the threshold to estimate the premium rates along with different coverage level. Therefore, observations of each county that are below the average of historical yields of the county are included to fit the GPD parameters.

Bayesian Hierarchical Model for Spatial Smoothing

The general Bayesian model assumes that the observations are independent. For example, basic Bayesian estimation for estimating crop yield distribution regards all observations in each county as independent of each other. In reality, however, there are many situations in which the assumption of independence does not hold. A Bayesian hierarchical model is a popular method

---

[4] A random vector consists of random variables, $\boldsymbol{X} = (X_1, \ldots X_N)^T$, is said to be Gaussian random process if the random variables $X_1, \ldots X_N$ are jointly normal. In our case, we regard GPD parameters as random variables, and they are jointly normally distributed. The covariance between two counties in the process can be specified as a function of the distance between these two counties.

to take the spatial correlation into consideration by adding one more prior layer (process layer) between the likelihood density and the prior density from the basic Bayesian model. Therefore, a Bayesian hierarchical model can be defined when a prior distribution of the general Bayesian model is also assigned to the additional prior parameters, say hyper priors, associated with the likelihood density parameters (GPD parameters). Recent statistical literature (Cooley, Nychka, and Naveau 2007, Gelman et al. 2004, Woodworth 2004, Nozoufras 2011) provide various adjustments of Bayesian hierarchical models for spatial modeling, such as spatial effects in extreme weather events, disease incidence, and mortality rates prediction.

The model has three layers. First, in the likelihood layer, we assume that each county's yield distribution follows a GPD. Therefore, this layer fits the crop yields at each county with GPD. Second, the process layer models the spatial process for the GPD parameters. This second layer determines the level and form of the spatial correlation across the counties and then the GPD parameters of each county in the layer are spatially correlated. Hence, the GPD parameters of each county are determined by a set of covariates that reflect the spatial characteristics of the county and the parameters are spatially correlated. The third layer consists of the prior distributions, called hyper priors, for the covariates of the process layer and Kriging parameters (sill and range). The Bayesian hierarchical model is structured as

$$\text{Likelihood layer: } \boldsymbol{Y}|\boldsymbol{\Omega_1},\boldsymbol{\Omega_2} \sim p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1},\boldsymbol{\Omega_2})$$

(3)

$$\text{Process layer: } \boldsymbol{\Omega_1}|\boldsymbol{\Omega_2} \sim p_2(\boldsymbol{\Omega_1}|\boldsymbol{\Omega_2})$$

$$\text{Prior layer: } \boldsymbol{\Omega_2} \sim p_3(\boldsymbol{\Omega_2})$$

where p_j is the density associated with each layer of the hierarchical model, $\boldsymbol{Y}$ is a matrix of yearly crop yields that spans all counties ($i = 1, \dots, K$) and years ($t = 1, \dots, T$), $\boldsymbol{\Omega_1}$ is a matrix of the GPD parameters that spans all counties ($i = 1, \dots, K$) so that $\boldsymbol{\Omega_1} = [\boldsymbol{\sigma}, \boldsymbol{\xi}]$ **,** and $\boldsymbol{\Omega_2}$ is a matrix of hyper parameters (coefficients for covariates and Kriging parameters) $\boldsymbol{\Omega}_2 = [\boldsymbol{\beta}_0, \boldsymbol{\beta}_h, \boldsymbol{\rho}, \boldsymbol{\theta}, \boldsymbol{\delta}]$ that spans all counties ($i = 1, \dots, K$). Note that each county has different values for the coefficients (β_0, β_h), but has identical Kriging parameters of sill and range (ρ, θ).

When we factorize the conditional density of the likelihood layer $p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$ using Bayes' theorem, the prior distribution for the likelihood layer density, say $p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$, can be separated into two components from

(4) $$p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) = p_2(\boldsymbol{\Omega_1}|\boldsymbol{\Omega_2})p_3(\boldsymbol{\Omega_2})$$

with the conditional distribution of GPD parameters given hyper parameters $p_2(\boldsymbol{\Omega_1}|\boldsymbol{\Omega_2})$ and prior distribution of hyper parameters $p_3(\boldsymbol{\Omega_2})$.

Therefore, the posterior distribution of the Bayesian hierarchical model can be expressed as,

(5) $$p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}|\, \boldsymbol{Y}) = \frac{p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}, \boldsymbol{Y})}{p(\boldsymbol{Y})} = \frac{p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})}{\int_{\boldsymbol{\Omega_1}}\int_{\boldsymbol{\Omega_2}} p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})d\boldsymbol{\Omega_2}d\boldsymbol{\Omega_1}}$$

$$p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}|\, \boldsymbol{Y}) \propto p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}).$$

Next, we plug equation (4) into the right-hand side of equation (5), and then we have

(6) $$p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}|\, \boldsymbol{Y}) \propto p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})p_2(\boldsymbol{\Omega_1}|\boldsymbol{\Omega_2})p_3(\boldsymbol{\Omega_2}).$$

Therefore, the posterior density of the Bayesian hierarchical model $p(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2} | \boldsymbol{Y})$ is proportional to the multiplication of three layers. As mentioned, the core difference between a general Bayesian model and a Bayesian hierarchical model is that the prior of the general Bayesian model can also be structured as additional priors (hyper priors). In a standard Bayesian model there would only be a prior for the distribution of the data, but in a hierarchical model, there is also a prior for the parameters. The spatial correlation of crop yield is captured in the "process layer" density $p_2(\boldsymbol{\Omega_1}|\boldsymbol{\Omega_2})$, so it is the parameters of the GPD density that are spatially autocorrelated.

Likelihood layer

A GPD likelihood function forms the likelihood layer of our hierarchical model. Let z_{it} denote the yearly crop yield at county $i = 1, \dots, I$ at year $t = 1, \dots, T$, and is assumed to follow a GPD. Since our interest is in minima of crop yield rather than maxima, we transform the likelihood using the fact that $min\{z_{i1}, \dots, z_{iT}\} = -\max\{-z_{i1}, \dots, -z_{iT}\} = -\max\{y_{i1}, \dots, y_{iT}\}$, where $y_{it} = -z_{it}$ for any i and t. We denote $\phi = \log \sigma$, which allows the parameter ϕ to take on both positive and negative values. Given that y_{it} exceeds the threshold u_i for county i, we assume that the yield of each county can be fitted by a GPD whose parameters depend on the location of the county.

Let ϕ_i and ξ_i represent the log-transformed scale and shape parameters of county i. By differentiating the cumulative distribution function (1) and replace the data, we get the probability density function (or likelihood) of the likelihood layer,

(7) $$p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1},\boldsymbol{\Omega_2}) = \prod_{i=1}^{I}\prod_{k=1}^{K}\frac{1}{\exp\phi_i}\left[1+\frac{\xi_i y_{ik}}{\exp\phi_i}\right]^{-\frac{1}{\xi_i}-1}$$

where subscript k represents the yield data of county i that exceeds the threshold u_i, and $\boldsymbol{\Omega}_1 = [\boldsymbol{\phi},\ \boldsymbol{\xi}]$ and $\boldsymbol{\Omega}_2 = [\boldsymbol{\beta}_0, \boldsymbol{\beta}_h, \boldsymbol{\theta}_{\boldsymbol{\phi}}, \boldsymbol{\rho}_{\boldsymbol{\phi}}, \boldsymbol{\theta}_{\xi}, \boldsymbol{\rho}_{\xi}]$, which are matrices for GPD parameters and hyper-parameters, respectively.

Process layer

The primary interest of the hierarchical model is the process layer since it determines the parameters of the likelihood density (GPD parameters) from a certain model structure and adjusts spatial correlation of these GPD parameters through the spatial random effect. In the process layer, we characterize the spatial process for the two GPD parameters (i.e., scale and shape) and form the spatial correlation function with Kriging parameters (i.e., sill and range). Since we are in the Bayesian framework, we treat the GPD parameters $\phi_i$and ξ_i as random variables and choose a prior distribution which allows us to model the latent spatial process. Therefore, we put independent priors on ϕ_i and ξ_i. We assume that the log-transformed scale parameter, $\phi_i = \log \sigma_i$, and shape parameter, ξ_i, of GPD are spatially sensitive and thus model their spatial correlation using a spatial random process.

First, we model the process of log-transformed scale parameters ϕ_i that follows a Gaussian spatial process of the form

$$\phi_i = \mu_i + \omega_i + \varepsilon_i$$
$$\mu_i = \beta_0 + \sum_{h=1}^{H} \beta_h Z_h$$
(8)
$$\omega_i \sim MVGP\left(0,\ \psi\left(D_{ij}; \theta_\phi, \rho_\phi\right)\right),$$
$$\psi\left(D_{ij}; \theta_\phi, \rho_\phi\right) = \rho_\phi e^{-D/\theta_\phi},$$
$$\varepsilon_i \sim MVN(0,\ \Lambda),$$

where μ_i is the mean equation of the log-transformed scale parameters ϕ_i at county i, Z_h is covariates of the scale parameters process, ω_i is spatial random effect for location i that follows multivariate Gaussian process[5], $\psi\left(D_{ij}; \theta_\phi, \rho_\phi\right)$ is spatial covariance matrix form with the Euclidean distance(D_{ij}) between counties i and j, sill parameters ρ, and range parameter θ, and ε_i is non-spatial error component that follows $\varepsilon_i \sim MVN(0, \Lambda)$ where Λ is a diagonal matrix where the diagonal elements are σ^2 and all other elements are zero. The spatial smoothing of the GPD parameters is due to the ω_i. The MCMC process generates spatially correlated values of ω_i and thus the GPD parameters are spatially correlated.

Similarly, the latent process of the shape parameters of GPD, ξ_i, is assumed to have a form,

$$\xi_i = \delta_i + \omega_i + \epsilon_i$$
(9)
$$\omega_i \sim MVGP\left(0,\ \psi\left(D_{ij}; \theta_\xi, \rho_\xi\right)\right),$$
$$\psi\left(D_{ij}; \theta_\xi, \rho_\xi\right) = \rho_\xi e^{-D/\theta_\xi},$$
$$\epsilon_i \sim MVN\left(0,\ \Psi\right),$$

[5] In the Gaussian spatial process, any subset of the field locations has a multivariate normal distribution. The covariance between any two locations (i.e., county) is determined by a covariance function (or kernel) of the Gaussian process evaluated at the spatial points of two locations.

where δ is the constant term for the spatial process, ω_i is the spatial random effect for location i that follows a multivariate Gaussian process, $\psi\left(D_{ij}; \theta_\xi, \rho_\xi\right)$ is spatial covariance matrix form with the Euclidean distance(D_{ij}) between counties i and j, sill parameters ρ, and range parameter θ, and ϵ_i non-spatial error component that follows $\epsilon_i \sim MVN(0, \Psi)$ where Ψ is a diagonal matrix where diagonal elements are σ^2 and all other elements are zero.

From the specification for GPD parameters, the vector of log-transformed scale parameter $\boldsymbol{\phi}$ and shape parameter $\boldsymbol{\xi}$ given the vector of parameters of counties $\boldsymbol{\delta}, \boldsymbol{\beta_0}, \boldsymbol{\beta_h}, \boldsymbol{\rho}$, and $\boldsymbol{\theta}$ follows

$$\boldsymbol{\phi} \mid \boldsymbol{\beta_0}, \boldsymbol{\beta_h}, \boldsymbol{\rho_\phi}, \boldsymbol{\theta_\phi} \sim MVN\left(\boldsymbol{\mu}, \ \Sigma_\phi\right)$$

(10)

$$\boldsymbol{\xi} \mid \boldsymbol{\delta}, \boldsymbol{\rho_\xi}, \boldsymbol{\theta_\xi} \sim MVN\left(\boldsymbol{\delta}, \ \Sigma_\xi\right)$$

where $\boldsymbol{\mu} = \boldsymbol{\beta_0} + \sum_{h=1}^{H} \boldsymbol{\beta_h X_h}$, $\Sigma_\phi = \psi\left(D_{ij}; \theta_\phi, \rho_\phi\right) \otimes \Lambda$ and $\Sigma_\xi = \psi\left(D_{ij}; \theta_\xi, \rho_\xi\right) \otimes \Psi$.

Modeling the GPD parameters ϕ_i and ξ_i as above, data at the county locations provide information about the latent spatial process that characterizes these parameters. Therefore, the second part of equation (7) is

$$p_2(\boldsymbol{\Omega_1} | \boldsymbol{\Omega_2}) = \frac{1}{\sqrt{(2\pi)^K \left|\Sigma_\phi\right|}} \exp\left[-\frac{1}{2}(\boldsymbol{\phi} - \mathrm{E}[\boldsymbol{\phi}])^T {\Sigma_\phi}^{-1} (\boldsymbol{\phi} - \mathrm{E}[\boldsymbol{\phi}])\right]$$

(11)

$$\times \frac{1}{\sqrt{(2\pi)^K \left|\Sigma_\xi\right|}} \exp\left[-\frac{1}{2}(\boldsymbol{\xi} - \mathrm{E}[\boldsymbol{\xi}])^T {\Sigma_\xi}^{-1} (\boldsymbol{\xi} - \mathrm{E}[\boldsymbol{\xi}])\right]$$

where Σ is the variance-covariance matrix for the process of GPD parameters ϕ_i and ξ_i, and $\mathbf{\Omega}_2$ is matrix for the hyper parameters $\mathbf{\Omega}_2 = [\boldsymbol{\beta}_0, \boldsymbol{\beta}_h, \boldsymbol{\theta}_{\boldsymbol{\phi}}, \boldsymbol{\rho}_{\boldsymbol{\phi}}, \boldsymbol{\theta}_{\boldsymbol{\xi}}, \boldsymbol{\rho}_{\boldsymbol{\xi}}]$.

To develop a crop insurance rating model for crops with temporal trend (i.e., technological progress), we would need to account for a trend variable in the process of the Bayesian hierarchical model. However, including a trend variable brings substantial technical challenges since we must consider both a spatial and a temporal random process, and correlation between two processes. One simple way to avoid the complexity in modeling is to estimate trend of the crop yield outside from the Bayesian hierarchical procedure in a manner similar to Harri et al. (2011). Since there is no significant trend in the wheat yields, we do not explore the issue of trend here. But, if the model were extended to corn or cotton that have strong yield trends, the issue of trend would need to be addressed.

Prior layer

In the prior layer, we impose a prior (i.e., hyper-prior) to the matrix of hyper-parameters $\mathbf{\Omega}_2$**,** which characterizes the GPD parameters in the process layer. We assume that each parameter in the layer is independent of the others. Since we do not have any prior information about a relationship between GPD parameters (scale and shape) and covariates (i.e., meanyield, longitude, and latitude) in the process layer, we choose uninformative priors (sufficiently large ranges) for the covariates parameters β_0 and β_h. For all the models, we choose all covariates parameters β_h follow $Uniform(-10000, 10000)$. However, the Kriging parameters of sill (ρ) and range (θ), which describe the spatial structure of the scale and shape parameter of the GPD, are more difficult to set priors. Bayesian statistics literature (Berger, DeOivelira, and Sanso; Banerjee, Carlin, and Gelfand; Cooley, Nychka, and Naveau) points out that improper priors for the Kriging parameters may result in improper posterior distributions. Banerjee, Carlin, and Gelfand(2004) suggested that choosing informative priors for Kriging parameters can be the safest way to avoid improper posteriors. Therefore, we use empirical information to construct proper priors for the Kriging parameters as well from the maximum likelihood estimation. We start with prior distributions of sill parameter. We first estimate scale and shape parameters for each county independently using maximum likelihood. A histogram of obtained scale ($\phi_{i,MLE}$) and shape ($\xi_{i,MLE}$) parameters for each county is illustrated in Figure 1. We then fit an empirical variogram[6], $\hat{\gamma}(h)$, suggested by Cressie (1993) using the obtained MLE parameters for each county to find the proper range of prior distribution of the sill parameter. We finally impose a prior of $IG(0.01, 5)$ for the sill parameter and $IG(0.01, 2)$ for the shape parameter. In order to find the prior distributions for range parameter θ, we use prior geographical knowledge of empirical data. Since spatial effect is measured using Euclidean distance based on longitude / latitude coordinate space, the nearest and farthest distance between the pair of locations in the empirical dataset are used for the prior distribution of the range parameter. Therefore, we impose prior of $gamma(0.23, 7{,}71)$ for the range parameters both in scale and shape parameters (θ_{ϕ} and θ_{ξ}). With the priors as above, the third layer in equation (6) can be expressed as

---

[6] For observation Z_i in location $i = 1, \dots, I$, empirical variogram can be defined as

$$\hat{\gamma}(h) \coloneqq \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} (Z_i - Z_j)^2$$

where $N(h)$ is number of possible pairs of observation i and j, h is geographical distance between two observations.

(12) $$p_3(\boldsymbol{\Omega_2}) = p(\boldsymbol{\beta_0})p(\boldsymbol{\beta_h})p(\boldsymbol{\theta_\phi})p(\boldsymbol{\theta_\xi})p(\boldsymbol{\rho_\phi})p(\boldsymbol{\rho_\xi}).$$

Since $p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$, $p_2(\boldsymbol{\Omega_1}|\boldsymbol{\Omega_2})$, and $p_3(\boldsymbol{\Omega_2})$ in equation (6) are obtained from the three layers, we now have posterior distributions for the parameters by multiplying these three layers.

Markov Chain Monte Carlo Procedure

We employ Metropolis-Hastings (MH) steps within a Gibbs sampler to update each parameter of the model. In the MH algorithm, we draw random values under a candidate density, then accept or reject each draw and the accepted values are included in the posterior density. Similar to Cooley, Nychka, and Naveau (2007), we use maximum likelihood estimates of the GPD parameters as the candidate density for the posterior distribution. Let $\widehat{\boldsymbol{\phi}}$ be MLE estimates for the set of GPD scale parameters $\boldsymbol{\phi}$ and let $\boldsymbol{\mu}$ be a vector of expectations for $\boldsymbol{\phi}$. From the asymptotic property of MLE's, we have

(13) $$\sqrt{\boldsymbol{N}}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \overset{\boldsymbol{d}}{\to} MVN\left(0, \lim_{N\to\infty}\left[\frac{1}{N}I(\boldsymbol{\phi})\right]^{-1}\right)$$

where $I(\boldsymbol{\phi})$ is the information matrix. Given the process layer density $\boldsymbol{\phi} \sim MVN(\boldsymbol{\mu}, \Sigma)$ in equation (11), where $\boldsymbol{\mu} = \boldsymbol{\beta_0} + \sum_{h=1}^{H} \boldsymbol{\beta_h X_h}$ and $\Sigma = \psi(D_{ij}; \theta, \rho) \otimes \Lambda$, we obtain the joint distribution of $\widehat{\boldsymbol{\phi}}$ and $\boldsymbol{\phi}$ as,

$$(14) \qquad \begin{pmatrix} \hat{\boldsymbol{\phi}} \\ \boldsymbol{\phi} \end{pmatrix} = MVN\left(\begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \end{pmatrix}, \begin{bmatrix} \Sigma + I(\boldsymbol{\phi})^{-1} & \Sigma \\ \Sigma & \Sigma \end{bmatrix} \right)$$

We then construct the conditional distribution

$$(15) \qquad \boldsymbol{\phi} \mid \hat{\boldsymbol{\phi}} \sim MVN\left(\boldsymbol{\mu} + \Sigma(I(\boldsymbol{\phi})^{-1} + \Sigma)^{-1}(\hat{\boldsymbol{\phi}} - \boldsymbol{\mu}), \quad \Sigma - \Sigma(I(\boldsymbol{\phi})^{-1} + \Sigma)^{-1}\,\Sigma\right),$$

that is used as the candidate density in the MH step. By the central limit theorem, the MLE estimates will be close to the Bayesian posterior, and thus the sampling distribution for the MLE should provide a good candidate distribution for this part of the posterior. According to Cooley, Nychka, and Naveau (2007), this approach has a significant advantage to improve the acceptance rate of MH steps. After updating the GPD parameters, we repeat the process of Gibbs sampling to update the other covariates parameters and spatial Kriging parameters. Currently, a variety of R-packages provide MCMC updating algorithms for Bayesian hierarchical models. We mainly employ SpatialExtremes,extRemes, and spBayes packages to construct MCMC procedure. Since these packages do not provide a function that exactly matched with our model estimation, we combine these packages and re-modified these packages to estimate our model.

Model selection

We model to county-level winter wheat yield data from National Agricultural Statistics Service (NASS). The data contain 1970-2014 yearly yields (bushels per acre) for 77 counties in Oklahoma. Counties with missing observations are discarded from the dataset. Therefore, 39 counties' yields are included in the final dataset. A single state is considered because RMA has generally wanted to tell producers that no data from another state affects their premium. Also, the algorithm is sufficiently slow enough that estimating the model for the whole United States at once is impractical with current computer resources. Since the RMA would estimate separate models for each state, it would be practical for the RMA to use the method proposed here. The description of the dataset is presented in Tables 4 and 5 in the appendix.

In the previous section, we have assumed that the parameters in the process layer ($\beta_0, \beta_i, \gamma_0$, and γ_1) determine the mean and covariance structure of the Gaussian process for ϕ_i. Therefore, we can draw from the conditional distribution for ϕ_i given the current value of ϕ in each iteration using the MCMC process. Deviance Information Criterion (DIC) suggested by Spiegelhalter et al. (2002) is used to evaluate the goodness of fit of each model. However, we do not only rely on the DIC to choose the most appropriate model but also evaluate the models using out of sample performance. DIC has substantial advantages for Gaussian likelihoods and is particularly convenient to compute from posterior samples (Finley, Banerjee, and Carlin 2007). This criterion is the sum of the Bayesian deviance and the effective number of parameters (a penalty for model complexity). The deviance is the negative of twice the log-likelihood,

(16) $$D(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) = -2\log p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$$

where the $p_1(\boldsymbol{Y}|\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})$ is likelihood function from the likelihood layer. The Bayesian deviance is the posterior mean, $\overline{D} = E[D(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2})]$, and an effective number of parameters is given by $P_D = \overline{D}(\boldsymbol{\Omega_1}, \boldsymbol{\Omega_2}) - D(\overline{\boldsymbol{\Omega}}_1, \overline{\boldsymbol{\Omega}}_2)$ where $\overline{\boldsymbol{\Omega}}_1$ and $\boldsymbol{\Omega_2}$ are the posterior means of the model parameters. Note that lower DIC values indicate preferred models.

The quality of several types of model specifications with two different spatial covariance function forms, Model 0 to Model 9, are tested. In the process, we draw 100,000 iterations for MCMC chains and burn-in the first 20,000 observations to get the posterior distribution of each parameter. Table 1 provides the models tested and their corresponding DIC values. We start with a model with no spatial covariance function (Model 0), and expand the model with several covariates in the scale parameter equation, including average yield ($\bar{y}_i$), longitude (lon_i), and latitude (lat_i) of the county. We find that the model 5 with a constant term and powered exponential covariance dominates other models with covariates. Therefore, adding other covariates does not improve the quality of the model. We select model 5 as the main model.

To measure and illustrate the degree of spatial correlation of the GPD parameters, we further obtain F-Madogram (Cooley, Naveau, and Poncet, 2006) and simulate spatial random process for scale and shape parameters, ω_i, on the longitude-latitude space. Generally, when observations y_i follow a stationary Gaussian random process with correlation function ψ, variogram (γ) suggested by Cressi (1993) is commonly used to illustrate the degree of spatial correlation, which is structured as $\gamma_{ij} = \frac{1}{2} var(y_i - y_j) = \sigma^2\{1 - \psi(D_{ij})\}$, where γ_{ij} is variogram between two region i and j, $y_i$and y_j are observations in region i and j, σ^2 is variance, and D_{ij} is Euclidean distance between two regions. However, if our interest is in extreme values, the variogram cannot be a useful tool. Cooley, Naveau, and Poncet (2006) propose a modified madogram called the F-madogram $v_F(h)$, which is a useful quantity to evaluate the degree of spatial correlation in a spatial random process with an extreme value distribution. Figure 2 illustrates the F-Madogram for the scale and shape parameters. It presents a measure of how distance between two locations will affect the value of parameters between two locations. X-axis

h is Euclidean distance between two locations (km), and Y-axis $v_F(h)$ represents F-madogram, which is the degree of spatial correlation. Therefore, an effect of spatial correlation on the two parameters (scale and shape) remains approximately 200km distance, and the maximum degree of the spatial effect is approximately 0.17 degree of F-Madogram.

We also simulate the spatial random process ω_i of the main model (model 5) to visualize the form of spatial correlation on the longitude-latitude space. This simulation is conducted using the posterior values for Kriging parameters. These obtained Kriging parameters determine the form and the degree of spatial correlation on the space. R-package ‘SpatialExtremes’ provides computational tools to obtain both F-Madogram and spatial random process simulation. In figure 3, the western region of Oklahoma has a relatively higher value of scale parameters compared to the eastern region, and northwestern region of Oklahoma shows a higher value of shape parameters than southeastern region. These spatial correlations of GPD parameters make higher premium rates of the western region of Oklahoma and mitigate regional inequalities of loss ratio in Oklahoma.

Out of sample comparison

The next step is calculating premium rates using the selected model and evaluating out of sample performance. We evaluate an out of sample performance of our model compared to the model suggested by Harri et al. (2011) from 2000 to 2014. To do that, we first calculate the premium rates of each county. The selected model provides different parameter values for each county and therefore premium rates differ by county. The break-even premium rate for county i, $prem_i$, for the area yield insurance contract that guarantees the expected yield, say $\lambda\hat{y}_i$, which is suggested by Barnett and Black (1997) and Ker and Coble (2003) is

$$prem_i = \frac{P(y_i, < \lambda\hat{y}_i)(\lambda\hat{y}_i - E[(y_i|y_i < \lambda\hat{y}_i)])}{\lambda\hat{y}_i}, \quad (17)$$

where λ is coverage rate, $\hat{y}_i$ is expected yield level in county i, and the expectation values and probability measure are taken from the posterior distribution of GPD parameters.

We calculate the 90% premium rates of our model and compare it with the current rating method that is suggested by Harri et al. (2011). As mentioned earlier, we refer to the Harri et al. (2011) model as the RMA model. Figure 4 illustrates interpolation of 90% premium rates from each model onto longitude-latitude space. The left figure has premium rates from the proposed model, and the right figure has premium rates from RMA model. The figure demonstrates that the shape of premium rates from the new model shows a smoother surface than that of the rates from RMA method. The crop yield distributions for each county capture spatial correlation due to the spatially correlated parameters. Therefore, the premium rates of the new model have a smoother surface than the Harri, et al. (2011) model.

We calculate the loss ratio of each model to measure the performance of the model. Given the historical yields from 2000 to 2014, we obtain the loss ratio for each model and compare the relative performance of the two models. The loss ratio can be given by

(18) $$lossratio_i = \frac{\sum_{t=1}^{T} \max[\lambda \hat{y}_{it} - y_{it},\ 0]}{\sum_{t=1}^{T} prem_{it}\hat{y}_{it}},$$

where λ is coverage level, $\hat{y}_{it}$ is predicted yield of county i at year t, y_{it} is actual yield for county i at year t, and $prem_{it}$ is premium rate of county i at year t, which is obtained from equation (15).

The premium gains (denominator) and indemnity losses (numerator) of equation (18), under the proposed and RMA (Harri et al, 2011) model, are obtained using actual yields of each county. Average, variance, maximum, and minimum loss ratio across counties of each model are presented in Table 2. The loss ratio of the new model is 1.19 and is closer to one than the RMA model (1.43). Since the loss ratio of fairly rated premium should equal one, a loss ratio closer to one implies the higher quality of out of sample performance. The last two columns in Table 5 show that the new model has a smaller variation of loss ratio across counties and therefore has less regional inequalities compared to the RMA model. Specifically, counties with high loss ratio under the RMA model, such as Alfalfa and Pottawatomie counties, become closer to one. Figure 5 illustrates 2D and 3D interpolation of loss ratio from the new model and the RMA model onto longitude and latitude space. The last column of Table 5 in appendix presents county level loss ratio in Oklahoma. In Figure 5, once again, we verify that the range of loss ratio across the counties of the new model is smaller than RMA model. RMA model shows significantly large loss ratio in the east central region in the Oklahoma, whereas the new model presents relatively equally distributed loss ratio.

Next, Similar to Harri et al. (2011), we assume a representative insurance company that can choose whether to retain or cede risks to Federal Crop Insurance Corporation (FCIC) under the Standard Reinsurance Agreement (Coble, Dismukes, and Glauber 2007). Then the premium rates of the company are assumed to be estimated using the proposed model and compared with RMA rates[7]. If the rates are higher than the RMA, then we cede risks to the FCIC since we consider that the RMA rates are underestimated, and therefore we expect a loss. Whereas, if the rates are smaller than the RMA, we retain risks since we believe that the RMA rates are overestimated, and therefore, we can expect a profit by retaining the policy. We repeat the procedure over the fifteen years from 2000 to 2014 and calculate the loss ratio of the retained and ceded policies. We then take a statistical test under the null hypothesis that the loss ratios of the two policies are equal. The non-parametric bootstrapping method is used to calculate the statistical significance of the hypothesis test. Under the null hypothesis of the test, the ratio of loss ratio between two policies should equal one unless the RMA rates are not fairly rated. The "loss ratio of ceded policy" in Table 3 is loss ratio of RMA rates when the representative insurance company chooses to cede a risk in the repeated comparison with the new model's rates. The "loss ratio of retained policy" is the loss ratio of RMA rates when the company chooses to retain a risk when one believes RMA rates are overpriced. The "cede to retained ratio" is calculated by dividing the loss ratio of ceded and the loss ratio of retained. The "percent of retained policy" is the percentage to choose retained policy over fifteen years with 39 counties. The "p-value" in Table 3 is the type 1 error (probability of rejecting the null hypothesis that the loss ratios of ceded and retained policy

[7] We assume that RMA calculates the premium rates using the method (empirical rates) suggested by Harri et al. (2011)

are indifferent when the null hypothesis is true) estimated from the bootstrapping method under the null hypothesis.

The results in Table 3 corroborate that the incorrect premium rate calculation may have a significant economic loss under the Standard Reinsurance Agreement (SRA). The ceded to retained ratio is greater than one meaning that the loss ratio of ceded policy is much higher than that of retained policy. The p-value of the test confirms the statistical significance of the test. The “percent of retained policy” is less than 50%. Therefore RMA model from which current premiums are based, tends to underestimate premiums compared to the suggested new model.

Conclusion

In this article, we suggest a new approach for the area-yield crop insurance rating based on extreme value theory and Bayesian spatial smoothing. Several studies have contributed to apply Bayesian models in various topics related to agricultural economics. However, there are few examples of Bayesian hierarchical models in the agricultural economics literature, and to our knowledge, this study is the first study that uses such a model to develop a crop insurance rating model. The main contributions of the study are a method to adjust spatial correlation into the crop insurance rating model and using a density function that focuses on the tail of the distribution that matters in pricing crop insurance. Our process differs substantially from the BMA method suggested by Ker, Tolhurst, and Liu (2015). We use a Bayesian hierarchical model to combine observations from all counties rather than using pre-obtained posterior distributions for each county using observations only in the county to develop a model that reflects all the information from the different counties. The weight of other counties' spatial influence is measured by the Euclidean distance between two counties.

Our results show spatial correlation of the crop yield distribution and show the form of spatial correlation across regions. Further, we measure the degree and the effective distance of the spatial correlation using F-madogram. Several model specifications are examined to identify the quality of the candidate models, and compare the performance between selected new model and RMA model. The empirical estimation results show that premium rates without considering the spatial correlation of the yield distribution may have a significant economic loss under the SRA. Our work suggests that considering spatial correlation in premium rating could lead to more precise risk measurement and therefore reduce inequalities in loss ratio across counties.

An important extension of our study is to make a comprehensive model for all kinds of crops, including crops with a temporal trend (i.e., technical changes). Adjusting trend variable into the procedure of Bayesian hierarchical model brings a substantial increase in the scope of our work since we should consider both spatial correlation and temporal correlation, and how these two correlations interact in our observations. Of course, this extensive modeling will lead to a significant increase in computational complexity. Therefore, Bayesian hierarchical modeling regarding both spatial and temporal correlation has received substantially increased attention in the current statistics literature. While our model does not directly include the temporal trend in our model structure, we can adjust the trend of crop yield outside of the model procedure using several types of trend estimation methods. However, for more technically sound modeling, future research should consider both spatial and temporal correlation in the procedure of the Bayesian hierarchical model. The model we proposed is the first model that directly considers the spatial correlation in the procedure of crop insurance rating. In the sense, this model might be a cornerstone to constructing an appropriate model for calculating crop insurance premium rates based on Bayesian spatial smoothing approach.

Table 1. Deviance Information Criterion (DIC) For Models Of Oklahoma County Wheat Yield

Spatial Effect	Model	Specification	$\bar{D}$	p_D	DIC
No spatial effect	Model 0	$\phi_{it} = \beta_0 + \varepsilon_{it}$	11441	57	11498
Matern	Model 1	$\phi_{it} = \beta_0 + \omega_i + \varepsilon_{it}$	11440	41	11481
	Model 2	$\phi_{it} = \beta_0 + \beta_1 \bar{y}_i + \omega_i + \varepsilon_{it}$	11440	43	11483
	Model 3	$\phi_{it} = \beta_0 + \beta_1 lon_i + \beta_2 lat_i + \omega_i + \varepsilon_{it}$	11441	45	11486
	Model 4	$\phi_{it} = \beta_0 + \beta_1 \bar{y}_i + \beta_2 lon_i + \beta_3 lat_i + \omega_i + \varepsilon_{it}$	11439	48	11487
Powered exponential	Model 5	$\phi_{it} = \beta_0 + \omega_i + \varepsilon_{it}$	11441	40	11481
	Model 6	$\phi_{it} = \beta_0 + \beta_1 \bar{y}_i + \omega_i + \varepsilon_{it}$	11439	44	11483
	Model 7	$\phi_{it} = \beta_0 + \beta_1 lon_i + \beta_2 lat_i + \omega_i + \varepsilon_{it}$	11438	45	11483
	Model 8	$\phi_{it} = \beta_0 + \beta_1 \bar{y}_i + \beta_2 lon_i + \beta_3 lat_i + \omega_i + \varepsilon_{it}$	11439	48	11487

Table 2. Loss Ratio Of New Model And RMA (Harri Et Al, 2011) Model

Model	Mean loss ratio	Variance of loss ratio	Max loss ratio	Min loss ratio
RMA	1.43	0.36	3.07	0.47
New model	1.19	0.21	2.21	0.35

Table 3. Loss Ratio Under Ceded And Retained Policy

Loss ratio of Ceded Policy	Loss ratio of Retained Policy	Ceded to Retained Ratio	Percent of Retained policy	P-value
2.44	1.12	2.18	0.26	0.046

Table 4. County Level Wheat Yield In Oklahoma (1970 – 2014)

County name	Average Yield (acre/bushel)	S.D	Minimum Yield	Maximum Yield
Alfalfa	32.1	8.0	14.4	52.0
Beaver	22.7	7.7	9.3	38.5
Blaine	28.4	5.6	16.7	40.6
Caddo	32.0	6.6	12.8	45.3
Canadian	30.8	5.6	20.0	43.5
Cimarron	25.2	7.6	9.3	41.1
Cleveland	28.9	5.7	16.5	45.0
Comanche	25.9	6.6	8.6	39.8
Cotton	27.5	7.0	10.0	40.9
Craig	32.4	9.0	15.0	54.7
Custer	30.3	7.2	13.2	46.9
Dewey	28.2	7.0	10.6	40.1
Garfield	31.7	7.1	15.1	45.0
Garvin	30.7	7.1	16.8	47.5
Grady	30.8	6.0	18.1	41.5
Grant	31.1	7.2	12.7	43.0
Greer	24.6	6.0	10.9	35.0
Harper	23.3	7.5	9.0	37.9
Kay	31.8	6.8	16.0	47.0
Kingfisher	28.7	6.2	17.5	44.5
Kiowa	26.9	7.2	10.1	41.9
Logan	29.2	7.0	13.1	43.5
Major	28.2	6.1	15.0	42.0
Mayes	30.8	6.6	20.0	47.3
Mcclain	30.9	6.3	17.5	48.8
Noble	31.2	6.9	15.0	42.5
Oklahoma	28.8	6.0	18.0	42.9
Osage	29.9	6.4	16.6	43.5
Ottawa	32.7	8.4	14.5	53.0
Pawnee	31.2	9.1	16.0	57.5
Payne	29.7	7.3	15.0	55.8
Pottawatomie	29.4	5.7	19.5	43.0
Roger Mills	25.6	7.6	12.2	43.2
Stephens	27.6	7.0	15.0	57.5
Tillman	27.1	6.8	10.4	40.7
Wagoner	29.4	7.5	12.3	42.5
Washita	28.7	6.6	14.1	41.8
Woods	29.4	8.5	13.7	53.1
Woodward	26.1	8.1	10.5	42.9

Data Source: www.nass.usda.gov/ok

Table 5. Premium Rates And Loss Ratio Of Oklahoma (2014)

County name	90% Premium rates (New model)	90% Premium rate (RMA model)	Loss ratio (New model)	Loss ratio (RMA model)
Alfalfa	4.97%	4.31%	1.25	2.47
Beaver	7.94%	7.68%	0.68	0.90
Blaine	4.28%	3.78%	1.29	1.48
Caddo	5.31%	4.74%	0.83	0.75
Canadian	4.10%	2.87%	1.08	1.67
Cimarron	7.63%	5.79%	0.82	2.39
Cleveland	4.12%	3.34%	1.34	2.07
Comanche	5.72%	7.18%	1.65	1.03
Cotton	6.18%	7.88%	1.90	1.27
Craig	5.36%	3.55%	0.35	0.71
Custer	5.37%	4.25%	1.22	1.45
Dewey	5.33%	5.34%	1.75	1.84
Garfield	4.40%	4.42%	1.48	1.41
Garvin	4.42%	3.66%	0.56	0.87
Grady	4.43%	4.55%	1.43	1.21
Grant	4.87%	5.38%	1.44	1.07
Greer	5.76%	5.49%	1.00	1.16
Harper	7.07%	7.91%	1.27	1.02
Kay	5.07%	4.50%	1.84	1.77
Kingfisher	4.02%	2.55%	0.88	1.76
Kiowa	5.67%	5.62%	1.08	1.06
Logan	4.82%	4.47%	1.71	1.65
Major	4.67%	4.14%	0.84	0.69
Mayes	4.99%	4.35%	0.52	1.10
McClain	4.07%	3.88%	0.61	0.48
Noble	4.79%	4.73%	2.21	1.94
Oklahoma	4.02%	2.83%	0.67	1.77
Osage	5.35%	4.30%	1.04	1.21
Ottawa	5.58%	6.17%	1.18	1.77
Pawnee	5.25%	6.81%	1.77	0.74
Payne	4.86%	4.20%	1.66	2.13
Pottawatomie	4.18%	4.07%	0.94	3.08
Roger Mills	6.04%	3.53%	0.80	1.84
Stephens	4.95%	4.44%	0.72	0.62
Tillman	6.08%	6.33%	1.64	1.42
Wagoner	5.75%	7.16%	1.82	2.56
Washita	5.12%	2.93%	0.66	1.25
Woods	5.82%	6.11%	1.54	1.49
Woodward	6.35%	5.56%	0.81	0.78

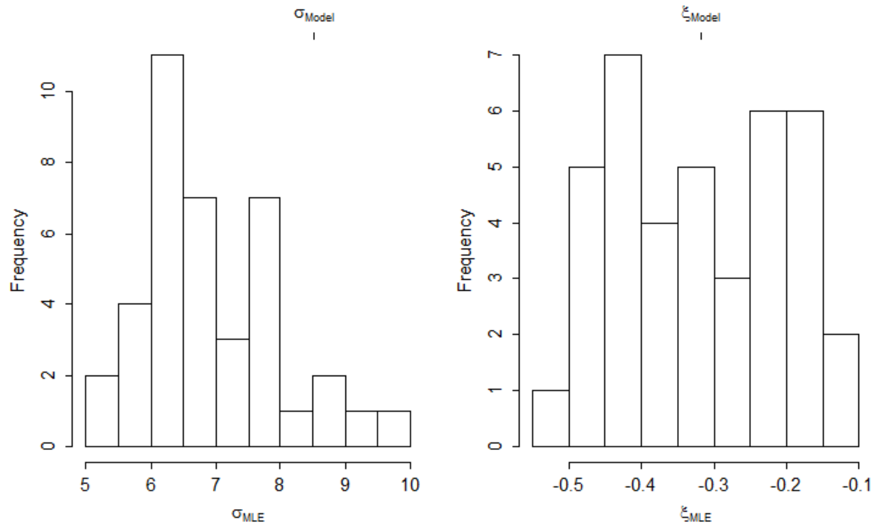

Figure 1. Histogram Of MLE Estimator For Each County

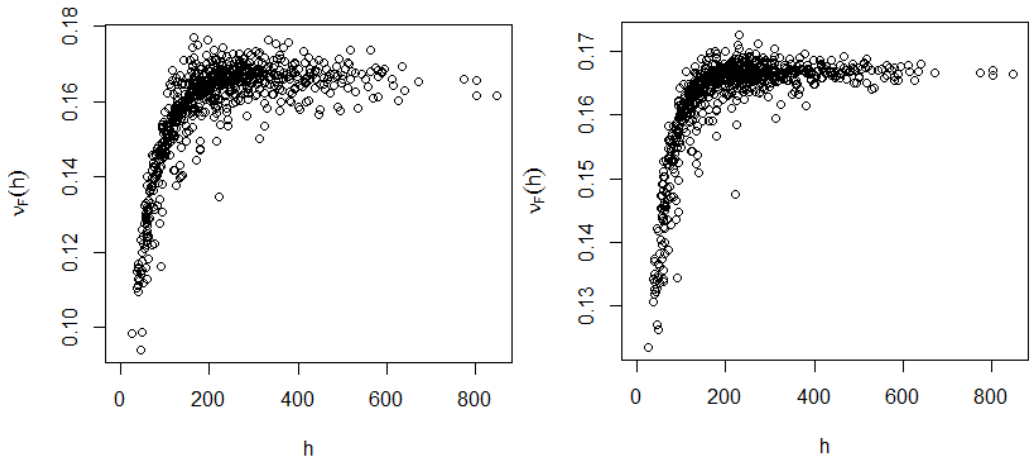

Figure 2. F-Madogram Of Scale(Left) And Shape Parameters(Right)

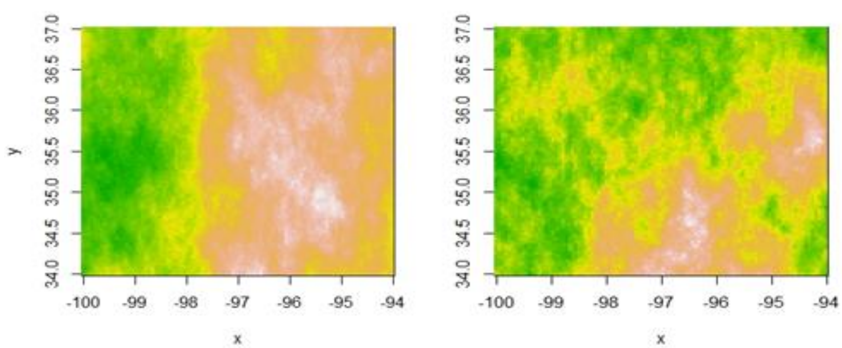

Figure 3. Spatial Random Field Simulation For Scale And Shape Parameters

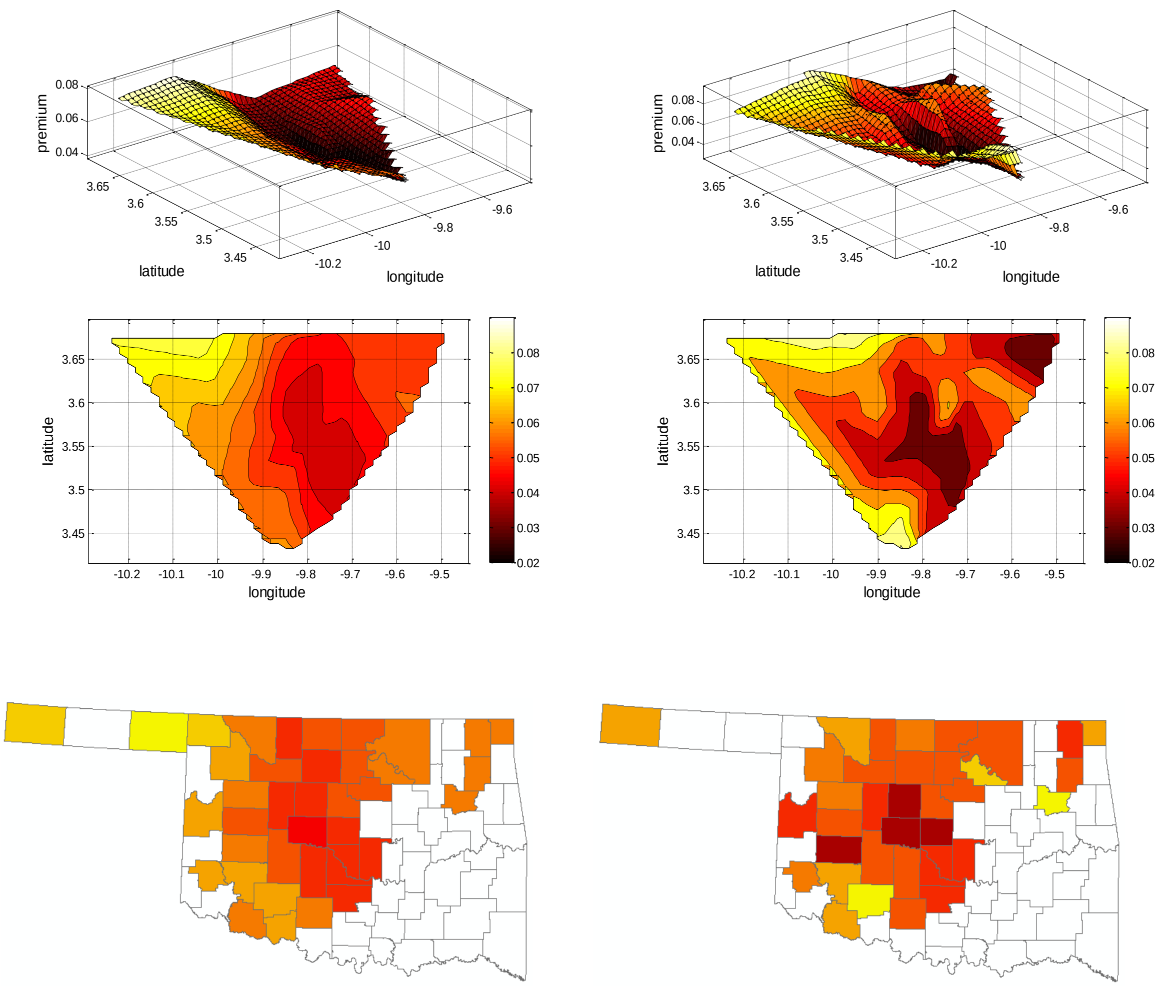

Figure 4. 90% Premium Rates From New Model (Left) And RMA (Harri Et Al, 2011) Model (Right)

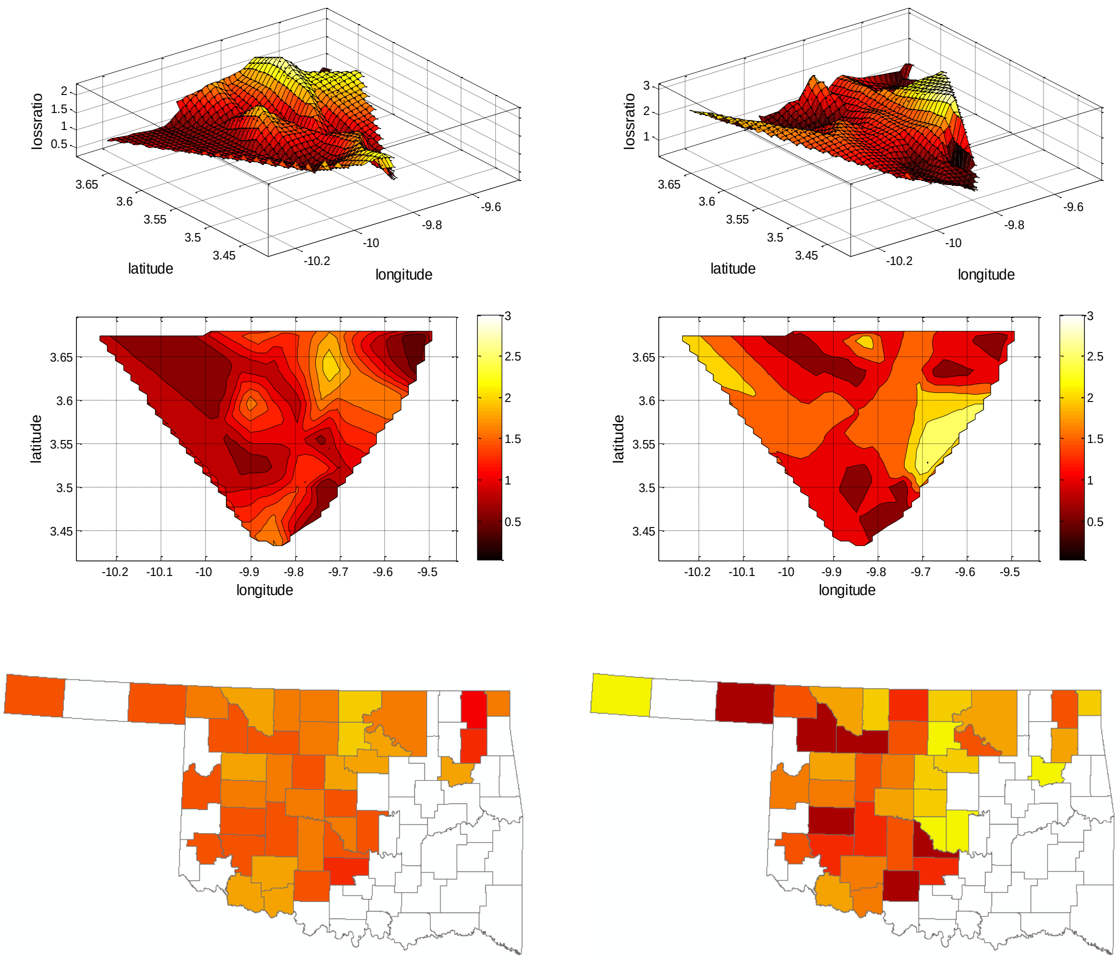

Figure 5. Loss Ratio Of New Model (Left) And RMA(Harri Et Al, 2011) Model (Right).

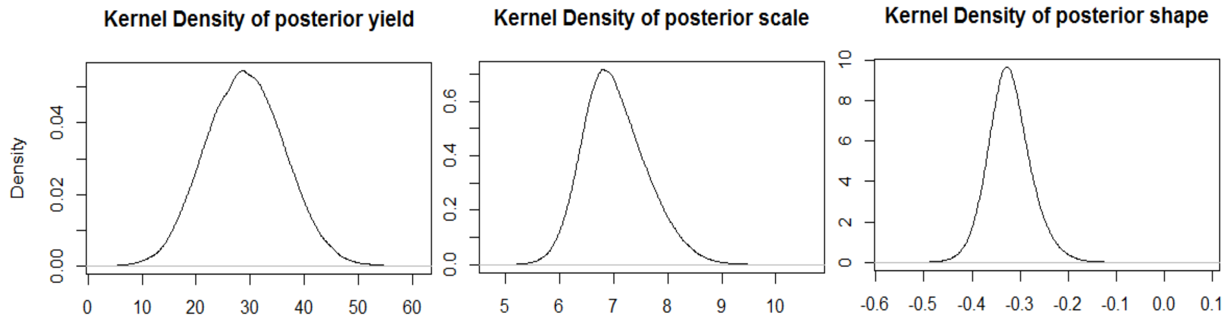

Figure 6. Posterior Density Of Yield, Scale, And Shape Parameter

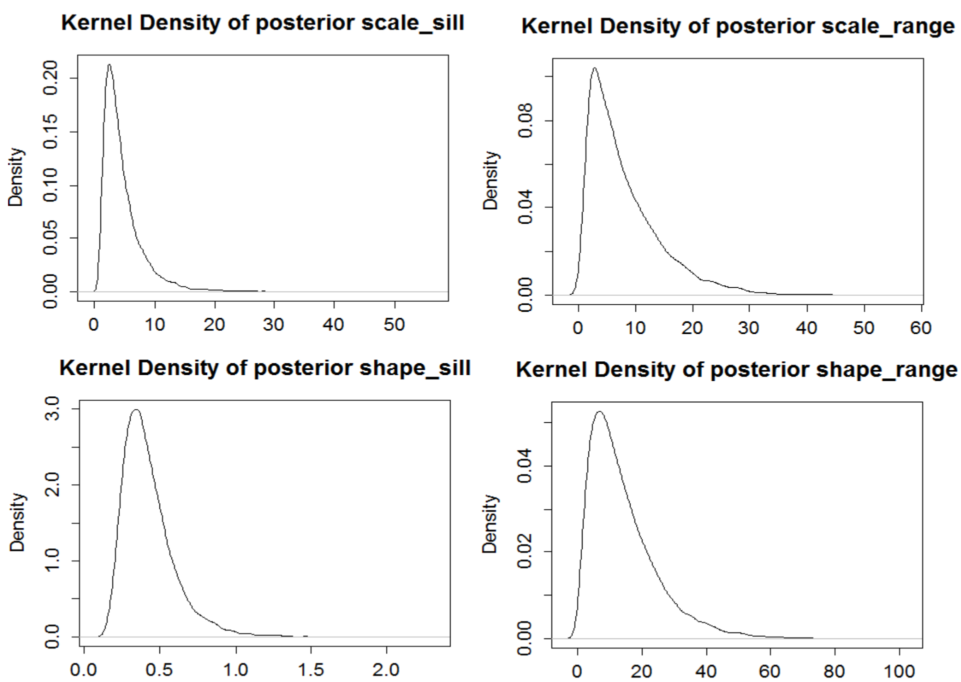

Figure 7. Posterior Density Of Kriging Parameters

References

Abramson, I.S. 1982. “On Bandwidth Variation in Kernel Estimates—A Square Root Law.” *Annals of Statistics* 10:1217–23.

Atwood. J., S. Shaik, and M. Watts. 2002. “Can Normality of Yields Be Assumed for Crop Insurance?” *Canadian Journal of Agricultural Economics* 50:171–84.

Babcock, B.A, and D.A. Hennessy. 1996. “Input Demand Under Yield and Revenue Insurance.” *American Journal of Agricultural Economics* 78:416–27.

Banerjee, S., B. Carlin., and A. Gelfand. 2004. Hierarchical Modeling and Analysis for Spatial Data, Boca Raton, FL: Chapman & Hall/CRC Press.

Berger, J., V. DeOiveira., and B. Sanso. 2001. “Objective Bayesian Analysis of Spatially Correlated Data.” *Journal of the American Statistical Association* 96: 1361–74.

Bessler, D. 1994. “Aggregate Personalistic Beliefs on Yields of Selected Crops Estimated Using ARIMA Processes.” *American Journal of Agricultural Economics* 76:801–9.

Borges, R., and W. Thurman. 1994. “Marketing Quotas and Subsidies in Peanuts.” *American Journal of Agricultural Economics* 76:801–17.

Coble, K.H., T.O. Knight, R.D. Pope., and J.R. Williams. 1996. “Modeling Farm-Level Crop Insurance Demand with Panel Data.” *American Journal of Agricultural Economics* 78:439–47.

Cooley, D., D. Nychka, and P. Naveau. 2007. "Bayesian Spatial Modeling of Extreme Precipitation Return Levels." *Journal of the American Statistical Association* 102:824-840.

Cooley, D., P. Naveau, and P. Poncet. 2006. “Variograms for Spatial Max-stable Random Fields.” *Dependence in Probability and Statistics* 187:373-390.

El-Sayyad, G.M., M. Samiuddin., and A.A. Abdel-Ghaly. 1984. “A New Kernel Density Estimate.” *Journal of Nonparametric Statistics* 3:1– 11.

Gelman, A., J. H. S. Carlin., and D. Rubin. 2014. Bayesian Data Analysis. Vol. 2. Boca Raton, FL, USA: Chapman & Hall/CRC.

Goodwin, B.K., and A.P. Ker. 1998. “Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk (GRP) Crop Insurance Contracts.” *American Journal of Agricultural Economics* 80:139–53.

Harri, A., K. H. Coble, C. Erdem., and T. O. Knight. 2009. “Crop Yield Normality: A Reconciliation of Previous Research.” *Review of Agricultural Economics* 31: 163–182.

Harri, A., K. H. Coble., A. P. Ker., and B. J. Goodwin, 2011. “Relaxing heteroscedasticity assumptions in area-yield crop insurance rating”. *American Journal of Agricultural Economics*, 93 (3): 707–17.

Hennessy, D.A., B.A. Babcock., and D.J. Hayes.1997. ″Budgetary and Producer Welfare Effects of Revenue Insurance.” *American Journal of Agricultural Economics* 79:1024–34.

Just, R. E., and R. D. Pope. 1978. “Stochastic Specification of Production Functions and Economic Implications.” *Journal of Econometrics* 7: 67–86.

Just, R. E., and Q. Weninger. 1999. “Are Crop Yields Normally Distributed?” *American Journal of Agricultural Economics* 81(May): 287–304.

Ker, A. P., Tolhurst, T. N., and Liu, Y. 2015. “Bayesian estimation of possibly similar yield densities: implications for rating crop insurance contracts” *American Journal of Agricultural Economics* 98(2):360–82

Ker, A. P., and K. H. Coble. 2003. “Modeling Conditional Yield Distributions.” *American Journal of Agricultural Economics* 85: 291–304.

Ker, A, P., and B. K Goodwin. 2000. “Nonparametric Estimation of Crop Insurance Rates Revisited.” *American Journal of Agricultural Economics* 82(2): 463–478.

Miranda, M. J. 1991. “Area-Yield Crop Insurance Reconsidered.” *American Journal of Agricultural Economics* 73(May):233–242.

Miranda, M. J., and J. W. Glauber. 1997. “Systemic Risk, Reinsurance and the Failure of Crop Insurance Markets.” *American Journal of Agricultural Economics* 79: 206–215.

Moss, C. B., and J. S. Shonkwiler. 1993. “Estimating Yield Distribution with a Stochastic Trend and Nonnormal Errors.” *American Journal of Agricultural Economics* 75(November): 1056–1062.

Ntzoufras. I. 2011. Bayesian modeling using WinBUGS. Vol. 698. John Wiley & Sons.

Norwood, B., M. C. Roberts., and J. L. Lusk. 2004. “Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions.” *American Journal of Agricultural Economics* 86(November): 1032–1043.

Ozaki,V., B. K. Goodwin., and R. Shirota. 2008. “Parametric and Nonparametric Statistical Modeling of Crop Yield: Implications for Pricing Crop Insurance Contracts.” *Applied Economics* 40: 1151–1164.

Ramirez, O. A. 1997. “Estimation and Use of a Multivariate Parametric Model for Simulating, Heteroskedastic, Correlated, Nonnormal Random Variables: The Case of Corn Belt Corn, Soybean and Wheat Yields.” *American Journal of Agricultural Economics* 79(February): 191–205.

Skees, J.R., J.R. Black., and B.J. Barnett. 1997. “Designing and Rating an Area Yield Insurance Contract.” *American Journal of Agricultural Economics* 79:430–38.