SPATIALLY SMOOTHED CROP YIELD DENSITY ESTIMATION : PHYSICAL DISTANCE VS CLIMATE SIMILARITY

Eunchun Park, B. Wade Brorsen, and Ardian Harri

Selected Poster prepared for presentation at the 2017 Agricultural & Applied Economics Association Annual Meeting, Chicago, Illinois, July 30-August 1

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

Crop yield density tends to be spatially correlated because nearby areas share similar climate and agronomic characteristics. Many crop insurance studies have pointed out that the spatial yield correlation should be considered to provide more precise premium rating. Bayesian Kriging for spatial smoothing offers a promising way to use such spatial correlation when estimating crop yield densities. This article contributes to agricultural economics literature by providing a spatial smoothing method based on a climate space, which is composed of climatological measures. We compare the spatial smoothing from the climate space and a general physical space (longitude-latitude space) to evaluate the performance of each method. We use loss ratio of crop insurance to test the performance for county level yearly corn yield data from six U.S. states. Spatial smoothing from climate space dominates the results from the physical space in out-of-sample prediction and mitigates regional inequalities in crop insurance loss ratios. The climate space notably outperforms the physical space in Colorado that has varying climate due to its varying topography.

Key words: Bayesian hierarchical structure, Bayesian spatial smoothing, Bayesian Kriging, climate space, crop insurance, crop yield similarity, physical space, spatial correlation.

Introduction

Yields of a given crop tend to be spatially correlated due to agronomical and climatological similarity of nearby areas. A risk inherent in crop yield, therefore, can also be spatially correlated due to such spatial correlation in crop yields. As a result, shortfalls of crop yield in a particular region, such as a county or district, usually tend to be correlated with shortfalls of yield in the neighboring regions. Previous literature (Annan et al. 2014; Du et al. 2015; Goodwin 2015; Ker, Tolhurst, and Liu 2015; Woodard 2016) has discussed this issue and pointed out that considering spatial correlation could more accurately measure downside yield risk and thus reduce adverse selection in crop insurance.

Several statistical methods have been suggested in agricultural economics literature to reflect the spatial correlation in estimating crop yield density. For example, Goodwin and Ker (1998) use pooled observations from surrounding counties. Ozaki et al. (2008) use a spatial weighting matrix in an attempt to consider spatial correlation in crop yield. They impose uniform weights on parameter estimates from surrounding counties but impose zero weights beyond the surrounding counties. Ozaki and Silva (2009) propose a skewed normal multivariate conditional yield distribution for spatial smoothing. However, similar to Ozaki et al. (2008), they do not consider correlation beyond surrounding counties. Current area-based crop insurance programs are rated with a model suggested by Harri et al. (2011). To reflect the spatial correlation, the model imposes a district level restriction on the county level parameters. Other studies consider spatial correlation using Bayesian Model Averaging (BMA). Ker, Tolhurst, and Liu (2015) estimate a posterior density for each county using observations of each county and then take Bayesian averaging of its own posterior density and densities from other counties. Woodard (2016) employs BMA to get a weighted average of county and district level parameters. More recently, Park, Brorsen, and Harri (2016) suggest Bayesian Kriging as a method for spatial smoothing. Their Bayesian Kriging approach produces spatially smoothed parameter estimates that vary smoothly over space. Their weight for smoothing is determined by a physical distance in longitude-latitude space (i.e., physical space).

In addition to the Bayesian Kriging method for spatial smoothing under the traditional way of using physical space, we offer an alternative spatial smoothing under a space with climatological coordinates, which is a climate space. The climate space uses temperature and precipitation as coordinates rather than latitude and longitude. We use an average number of days in months (July and August) with maximum temperature greater than or equal to 90°F (DT90) and a total average precipitation amount (mm) for the months from May to August (TPCP) as the two axes. In this application, our focus is on how the distribution of crop yield varies over two different types of spaces (physical and climate space). We then compare the performances of the estimates from these two spaces using a loss ratio under a crop insurance program.

The primary goal of the article is to suggest a new method for actuarially accurate crop insurance rating considering the spatial correlation of the crop yield densities. We extend the Bayesian Kriging spatial smoothing method to use climate space instead of physical distance. We are not aware of any literature in agricultural economics that uses spatial smoothing based on climate space. We evaluate and compare the performances of the spatial smoothing from physical space and from climate space.

We choose corn as a crop for evaluating the performance of each model. We utilize annual county level yield data from the National Agricultural Statistics Service (NASS) for Iowa, Illinois, Nebraska, Minnesota, Indiana, and Colorado from 1955 to 2014. We find that climate space performs better or at least similar to physical space in every state in the dataset. Specifically, in Colorado and Nebraska, climate space substantially mitigates regional inequalities of loss ratio for crop yield densities.

In the following section, we discuss a theoretical framework for Bayesian hierarchical structure of our Bayesian Kriging method. In the empirical application section, we explain the dataset used for empirical estimation and describe two different types of smoothing spaces. We then introduce premium calculating procedures for evaluating the performance of the models from the different smoothing spaces. The last section has conclusions.

Theoretical Framework

We use a Kriging method for spatial smoothing. Kriging is a geostatistical spatial interpolation method that has been actively employed in a broad variety of disciplines. The method assumes that spatial correlation (i.e., density similarity) varies smoothly and decreases with the distance between locations. Note that regardless of which space (physical or climate) is used for the spatial smoothing, the estimation procedure is identical.

Overview of the Bayesian hierarchical structure

The Kriging method here is estimated under a Bayesian hierarchical structure. A

Bayesian hierarchical model can be specified when Bayesian modeling structure can be written in multiple levels (i.e., hierarchies). In a Bayesian hierarchical framework, therefore, a prior distribution of the general Bayesian model can also be structured as additional prior parameters, called hyper-priors.

We consider two types of specifications for the process layer: a Gaussian spatial process type (GP) and an auto-regressive type with Gaussian spatial process[1] (AR). GP only considers spatial correlation of the crop yield distribution based on the Gaussian spatial process whereas AR takes into consideration both spatial and temporal correlation (spatio-temporal) using the Gaussian spatial process and the auto-regressive process. Both GP and AR can be represented in the Bayesian hierarchical structure with three layers: likelihood layer, process layer, and prior layer. In the likelihood layer of the hierarchy, the crop yield distribution for each county is assumed to follow a normal distribution. Second, the process layer models the spatial and temporal structure for parameters of the crop yield distribution. In this layer, we only model the hierarchical structure of mean parameter μ of the crop yield distribution[2]. The process layer has both deterministic and stochastic effects. The deterministic part of the process at each county is determined by a set of explanatory variables of the county and the stochastic part will operate the spatial and temporal smoothing process. The third layer of the hierarchy consists of the prior density for the coefficients of the explanatory variables and Kriging parameters to conduct spatial smoothing, which are called hyper priors. The hierarchy we use can be structured as,

$$\begin{aligned} \boldsymbol{Y} \mid \boldsymbol{\mu}_t, \boldsymbol{\Theta} &\sim p_1(\boldsymbol{Y} \mid \boldsymbol{\mu}_t, \boldsymbol{\Theta}) \\ \boldsymbol{\mu}_t \mid \boldsymbol{\Theta} &\sim p_2(\boldsymbol{\mu}_t \mid \boldsymbol{\Theta}) \end{aligned} \quad (1)$$

$$\mathbf{\Theta} \sim p_3(\mathbf{\Theta})$$

where p_1, p_2, and p_3 are the density associated with each layer of the hierarchy, likelihood layer, process layer, and prior layer, respectively, $\boldsymbol{Y}$ is a matrix of crop yields that spans all counties $(n = 1, \dots, N)$ and all years $(t = 1, \dots, T)$, $\boldsymbol{\mu}_t$ is a vector of the mean parameters of the likelihood function at year t that contains all counties, where $\boldsymbol{\mu}_t = [\mu_{1t}, \dots, \mu_{Nt}]'$, and $\mathbf{\Theta}$ is a vector of hyper parameters, where $\mathbf{\Theta} = [\beta_1, \dots, \beta_K, \omega, \theta, \rho, \sigma^2]'$.

By Bayes' theorem, the joint posterior distribution of the model is

(2)
$$p(\boldsymbol{\mu}_t, \mathbf{\Theta} \mid \boldsymbol{Y}) \propto p_1(\boldsymbol{Y} \mid \boldsymbol{\mu}_t, \mathbf{\Theta}) p_2(\boldsymbol{\mu}_t \mid \mathbf{\Theta}) p_3(\mathbf{\Theta}).$$

Therefore, the joint posterior density of the model $p(\boldsymbol{\mu}_t, \mathbf{\Theta} \mid \boldsymbol{Y})$ is proportional to the multiplication of the three layers of the hierarchy, which will be specified in the following subsections.

Likelihood layer

A likelihood function of the crop yield distribution forms the first layer of the model. Both GP and AR assume that the crop yield of each county follows a normal distribution. Then, the first layer of the model, the likelihood layer is

(3)
$$p_1(\boldsymbol{Y} \mid \boldsymbol{\mu}_t, \mathbf{\Theta}) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(\boldsymbol{y}_t - \boldsymbol{\mu}_t)'(\boldsymbol{y}_t - \boldsymbol{\mu}_t)}{2\sigma^2}$$

where $\boldsymbol{y}_t$ denotes a vector of crop yield at year t that spans all counties, $\boldsymbol{y}_t = [y_{1t}, \dots, y_{Nt}]'$, $\boldsymbol{\mu}_t$ is a vector of the mean parameter at year t that includes all counties, and $\boldsymbol{\Theta}$ is a vector of hyper parameters, $\boldsymbol{\Theta} = [\beta_1, \ldots, \beta_K, \omega, \theta, \rho, \sigma^2]'$.

Process layer

In the second layer of the hierarchy, we model the spatial process of mean parameters μ for each GP and AR accounts for spatial / spatial and temporal correlations relevant to crop yield distribution. Since we assume a Gaussian spatial process, mean parameters of all counties are assumed to be multivariate normally distributed[3]. Spatial and temporal smoothing for the parameter $\boldsymbol{\mu}_t$ is conducted from the stochastic part of the process via Gaussian spatial process and first order autoregressive, AR(1), process. The level of spatial dependence is measured from a spatial covariance matrix with Kriging parameters, which captures the detailed spatial structure for the mean parameters $\boldsymbol{\mu}_t$.

First, the Gaussian spatial process (GP) allows spatial correlation of crop yield. The GP is specified as,

$$
\begin{aligned}
\boldsymbol{\mu}_t &= \boldsymbol{\psi}_t + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_t, \\
\boldsymbol{\psi}_t &= \boldsymbol{X}_t \boldsymbol{\beta} \\
\boldsymbol{\eta} &\sim MVGP(\Sigma), \\
\Sigma_\eta &= \rho_\eta e^{-D_{ij}/\theta_\eta}, \\
\boldsymbol{\varepsilon}_t &\sim MVN(\mathbf{0}, \sigma^2 \boldsymbol{I}),
\end{aligned}
\tag{4}
$$

where $\boldsymbol{\mu}_t$ is a vector of mean parameter for crop yield distribution at year t that contains all counties, $\boldsymbol{\mu}_t = [\mu_{1t}, \ldots, \mu_{Nt}]'$, $\boldsymbol{\psi}_t$ deterministic part of the mean structure, $\boldsymbol{X}_t$ is a T by K matrix of explanatory variables that determine the mean structure at time t such as historical moving average yield level of each county and trend variable, $\boldsymbol{\beta}$ is a K by

1 vector of coefficients of explanatory variables $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_K]'$, $\boldsymbol{\eta}$ is the spatial random effects, $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_N]'$ that is assumed to follow a multivariate Gaussian process with exponential type spatial covariance matrix, $\Sigma = \rho e^{-D_{ij}/\theta}$, which is a function of Euclidean distance (D_{ij}) between counties i and j, sill parameter ρ, and range parameter θ, and $\boldsymbol{\varepsilon}_t$ is a non-spatial error component. In empirical part, we obtain the parameter estimates under the two different types of spatial smoothing structure: traditional physical space and climate space. Therefore, the distance D_{ij} between two counties in the spatial covariance matrix will be differently measured in accordance with which spatial space is used.

The second type of specification is an auto-regressive model with Gaussian spatial process (AR). The AR is defined as,

$$\begin{aligned}
&\boldsymbol{\mu}_t = \omega \boldsymbol{\mu}_{t-1} + \boldsymbol{\psi}_t + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_t, \\
&\boldsymbol{\psi}_t = \boldsymbol{X}_t \boldsymbol{\beta} \\
&\boldsymbol{\eta} \sim MVGP(\Sigma), \\
&\Sigma = \rho e^{-D_{ij}/\theta}, \\
&\boldsymbol{\varepsilon}_t \sim MVN(\mathbf{0}, \sigma^2 \boldsymbol{I}),
\end{aligned} \tag{5}$$

where $\boldsymbol{\mu}_{t-1}$ is a vector of lagged mean parameter for crop yield distribution at year $t-1$ that contains all counties, $\boldsymbol{\mu}_t = [\mu_{1t-1}, \ldots, \mu_{Nt-1}]'$, ω denotes the temporal correlation parameter under the first order auto regressive process that is assumed to be in the interval, $-1 < \omega < 1$, and all other parameters are identical to the GP model. Obviously, for $\omega = 0$, AR is exactly the same as the GP model.

The AR addresses temporal correlation of the crop yield together with its spatial correlation. Some factors that affect crop yield realizations (i.e., climate) tend to be correlated, both spatially and temporally, and thus adjacent counties would experiences spatial correlations of crop yields over multiple periods of time. The AR specification reflects these spatio-temporal aspects of the crop yield densities.

From the process for the mean parameters, the vector of the parameter $\boldsymbol{\mu}_t$ given the parameters $, \omega, \theta,\ \rho,$ and σ^2 follows

(6) $$\boldsymbol{\mu}_t \mid \beta, \omega, \theta, \rho, \sigma^2 \sim MVGP(\boldsymbol{\psi}_t,\ \Sigma\,).$$

Then the process layer densities for GP and AR model can be specified as equation (7) and (8), respectively,

(7) $$p_2(\boldsymbol{\mu}_t|\boldsymbol{\Theta}) = \frac{1}{\sqrt{(2\pi)^N|\Sigma|}}\exp\left[-\frac{1}{2}(\boldsymbol{\mu}_t-\boldsymbol{\psi}_t)'\Sigma^{-1}(\boldsymbol{\mu}_t-\boldsymbol{\psi}_t)\right]$$

(8) $$p_2(\boldsymbol{\mu}_t|\boldsymbol{\Theta}) = \frac{1}{\sqrt{(2\pi)^N|\Sigma|}}\exp\left[-\frac{1}{2}(\boldsymbol{\mu}_t-\omega\boldsymbol{\mu}_{t-1}-\boldsymbol{\psi}_t)'\Sigma^{-1}(\boldsymbol{\mu}_t-\omega\boldsymbol{\mu}_{t-1}-\boldsymbol{\psi}_t)\right]$$

where $\boldsymbol{\psi}_t$ is a vector of the deterministic part of the mean process at time t defined by equations (4) and (5), and $\boldsymbol{\Theta} = [\beta, \omega, \theta, \rho, \sigma^2]'$ is a vector of hyper-parameters, Σ is spatial covariance parameter, where $\Sigma = \rho e^{-D_{ij}/\theta}$. Since we are in the Bayesian framework, the parameters of the crop yield process are treated as random variables. Therefore, we impose independent priors for the hyper parameters $(\beta, \omega, \theta,\ \rho,\ \sigma^2)$ in the following prior layer.

Prior layer

The third layer of the hierarchy has the priors for the hyper-parameters $\boldsymbol{\Theta}$, which are parameters for explanatory variables, Kriging parameters, and variance parameter in the process layer. Since the model assumes the parameters in the prior layer are independent, a multiplication of each prior for the hyper parameters forms the prior layer. For convenience, we group the hyper parameters into three different types depending on their role in the process layer: coefficient, variance, and Kriging parameters. First, all coefficient parameters in the process layer $\beta_1, \ldots, \beta_k$ and ω are given normal priors. We impose $N(0, 10^4)$ priors for each of the coefficient parameters. For the variance parameter σ^2, we impose general inverse gamma priors $IG(0.1, 0.1)$ same as Ozaki et al. (2008). However, imposing priors for the Kriging parameters (ρ, θ), which describe the spatial structure of the Gaussian spatial process, is more problematic than the other priors. There is a large Bayesian statistics literature (Berger, DeOivelira, and Sanso 2001; Banerjee, Carlin, and Gelfand 2004; Cooley, Nychka, and Naveau 2007) regarding consistency of proper priors for the Kriging parameters that argues improper priors for such parameters may induce significant improper posteriors. Some statistics literature (Banerjee, Carlin, and Gelfand 2004; Sahu, Gelfand, and Holland 2006; Cooley, Nychka, and Naveau 2007) suggest an empirical Bayes method in which the Kriging priors are estimated from the empirical data to avoid improper priors. In this regards, we use the empirical information to find the priors of the Kriging parameters. Since the sill parameter ρ determines the maximum level of the variogram, which is a function describing the degree of spatial correlation of a stochastic spatial process, an empirical variogram is a general way to collect prior information about the sill parameter. Therefore, we first estimate the mean parameter of each county using maximum likelihood. Then using the estimated MLE parameters for each county, we fit the empirical variogram.[4] The results of the empirical variogram are used to impose inverse gamma prior for the sill parameter ρ since the value of sill parameter determines maximum of variogram. Two parameters of the inverse gamma prior are obtained from maximum likelihood by using the empirical variogram values.

The next step is to find the prior distributions for the range parameter θ. We use prior empirical distance information of the empirical data to impose prior for the range parameter since the range parameter θ determines maximum distance of the spatial effect. Two parameters of gamma prior for the range parameter θ are imposed based on the previous empirical distance information and maximum likelihood estimation.

With the priors as above, the third layer in equation (2) can be expressed as

$$p_3(\mathbf{\Theta}) = p(\beta_k)p(\omega)p(\rho_\eta)p(\theta_\eta)p(\sigma^2). \tag{9}$$

Joint posterior distribution

We now have densities for each hierarchy, $p_1(\boldsymbol{Y} \mid \boldsymbol{\mu}_t, \mathbf{\Theta})$, $p_2(\boldsymbol{\mu}_t \mid \mathbf{\Theta})$, and $p_3(\mathbf{\Theta})$ from the previous sections. The joint posterior distributions for our model can be obtained by multiplying these three layers. The logarithm of the joint posterior distributions of the GP is

$$\begin{aligned} \log p(\boldsymbol{\mu}_t, \mathbf{\Theta} \mid \boldsymbol{Y}) \propto & -\frac{NT}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(\boldsymbol{y}_t - \boldsymbol{\mu}_t)'(\boldsymbol{y}_t - \boldsymbol{\mu}_t) - \frac{\sum_{t=1}^{T}\log|\Sigma|}{2} \\ & -\tfrac{1}{2}\textstyle\sum_{t=1}^{T}(\boldsymbol{\mu}_t - \boldsymbol{\psi}_t)'\Sigma^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\psi}_t) + \log(p_3(\mathbf{\Theta})). \end{aligned} \tag{10}$$

Likewise, the logarithm of the joint posterior distributions of the AR is

$$\log p(\boldsymbol{\mu}_t, \mathbf{\Theta} \mid \boldsymbol{Y}) \propto -\frac{NT}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(\boldsymbol{y}_t - \boldsymbol{\mu}_t)'(\boldsymbol{y}_t - \boldsymbol{\mu}_t) - \frac{\sum_{t=1}^{T}\log|\Sigma|}{2} \tag{11}$$

$$-\frac{1}{2}(\boldsymbol{\mu}_t - \omega\boldsymbol{\mu}_{t-1} - \boldsymbol{\psi}_t)'\Sigma^{-1}(\boldsymbol{\mu}_t - \omega\boldsymbol{\mu}_{t-1} - \boldsymbol{\psi}_t) + \log\big(p_3(\boldsymbol{\Theta})\big).$$

Empirical Application

Our study uses county-level yearly corn yield data from NASS. The data contains 1955-2014 annual yields (bushels per acre) for Iowa, Illinois, Nebraska, Minnesota, Indiana, and 1963-2009 for Colorado. Counties with missing observations are discarded. Therefore, the final dataset includes 99 counties for Iowa, 77 for Illinois and Nebraska, 68 for Minnesota, 75 for Indiana, and 18 counties for Colorado.

The states of Iowa, Illinois, and Nebraska have been the first, second and third largest corn producers in the United States. In 2015, Iowa produced around 2.5 billion bushels of corn, Illinois produced 2.01 billion bushels of corn, and Nebraska produced 1.7 billion bushels of corn (2015 state agriculture overview, NASS). Minnesota and Indiana are fourth and fifth largest corn producer in the United States, respectively. Colorado is the fourteenth largest corn producer in the United States. Colorado is located where the Great Plains of North America connects with the Rocky Mountains. Colorado is included because it has more varying climatic conditions than the other states.

A coordinate of the physical space consists of longitude and latitude of each county and thus the spatial smoothing from the physical space uses Euclidean distance between the locations on the physical space (i.e., physical distance). On the other hand, a coordinate of each point in the climate space is given by its climatological quantities. Therefore, spatial smoothing based on climate space uses the Euclidean distance between the locations on the climate space, which reflects climatological similarity between the locations[5]. From this climate space, we construct the spatial structure that relates the parameters of the crop yield density to the climate characteristics of the locations.

Several empirical studies have tried to determine the relationship between climate factors and crop yields. Schlenker and Roberts (2009) estimate the impact of climate change on agricultural output using panel data. They find that temperature above a threshold level has a negative impact on crop yields since it increases heat exposure and water stress. They use temperature and precipitation as the main climate variables for their research. Other related studies (Hendricks and Peterson 2012; Lobell et al. 2013; Dell, Jones, and Olken 2014) employ temperature and precipitation as explanatory variables. In accordance with the previous literature, we choose temperature and precipitation of each location as the climate coordinates.

The data for the climate space are collected from the Global Historical Climatology Network Database (GHCND) under the National Oceanic and Atmospheric Administration (NOAA). GHCND includes 18 meteorological variables including temperature (monthly means, extremes, and number of days that exceed a threshold), precipitation (total, mean, extremes), and snowfall, snow depth, and some other elements for each weather station. The climatological quantities for our climate space are a set of collected data over an extended period of time. Therefore, an average number of days in months (July and August) with maximum temperature greater than or equal to 90°F (DT90) and an average precipitation amount (mm) for the months from May to August (TPCP) from 1955 to 2014 are used as the climate coordinates. We take an average of the climate quantities from the counties with multiple weather stations. Counties with no weather stations (5 counties in all dataset) are discarded from our dataset.

In the climate space, counties with similar climate features are grouped together even when their locations are physically distant. Colorado has diverse geographical features, including mountainous terrain, vast plains, desert canyons, and mesas. For that reason, we expect that Colorado has diverse climate conditions as well so that county locations in Colorado on the two different spaces are substantially different. Figure 1 translates county locations in Colorado from the physical space (longitude/latitude) to the climate space (DT90/TPCP). The x-axis of the climate space is DT90 and the y-axis is TPCP. In contrast to Colorado, counties in corn-belt states such as Iowa tend to be grouped together in the climate space as well as in the physical space. The reason may be that the counties in Iowa have similar geographical features and thus nearly located counties have similar climate characteristics as well. Therefore, we expect that distances among the counties in Iowa on the physical space and the climate space are closely related.

To verify the difference between these two states, we calculate correlations between the distances from two different spaces[6]. As with our expectation for the two states, the correlations between two types of distances are 0.47 in Iowa and -0.09 in Colorado. This fact may result in considerable differences in the spatial smoothing estimation from the two different spaces, and yield more substantial difference in Colorado than in Iowa.

Posterior predictive distribution

As suggested by Ozaki et al. (2008), we compute premium rates from posterior predictive values. The posterior predictive distribution $p(\boldsymbol{y}^{*} | \boldsymbol{Y})$ is obtained by integrating over the parameters with respect to the joint posterior distributions,

$$p(\boldsymbol{y}^*|\,\boldsymbol{Y}) = \int_{\boldsymbol{\mu}}\int_{\boldsymbol{\Theta}} p_1(\boldsymbol{y}^*|\,\boldsymbol{\mu},\boldsymbol{\Theta})p_2(\boldsymbol{\mu}^*\mid\boldsymbol{\mu},\boldsymbol{\Theta})p(\boldsymbol{\mu},\boldsymbol{\Theta}\mid\boldsymbol{Y})d\boldsymbol{\Theta}d\boldsymbol{\mu}, \tag{12}$$

where $p_1(\boldsymbol{y}^*|\,\boldsymbol{\mu},\boldsymbol{\Theta})$ is the density of the likelihood layer, $p_2(\boldsymbol{\mu}^*\,|\boldsymbol{\mu},\boldsymbol{\Theta})$ is the density of the process layer, $p(\boldsymbol{\mu},\boldsymbol{\Theta}\mid\boldsymbol{Y})$ is the posterior density of the model obtained from Markov Chain Monte Carlo (MCMC) procedure[7], $\boldsymbol{\mu}^*$ and $\boldsymbol{y}^*$ are vector of posterior predicted mean parameters and crop yields, respectively. Note that $\boldsymbol{\mu}^*$ and $\boldsymbol{y}^*$ can denote either or both of a new location (county) and a new time point (year). Conceptual steps for prediction are as follows. First, vectors of random samples $\boldsymbol{\mu}$ and $\boldsymbol{\Theta}$ are drawn from the posterior density $p(\boldsymbol{\mu},\boldsymbol{\Theta}\mid\boldsymbol{Y})$. Then the Bayesian spatial smoothing (Kriging) for $\boldsymbol{\mu}^*$ is applied from the process layer $p_2(\boldsymbol{\mu}^*\mid\boldsymbol{\mu},\boldsymbol{\Theta})$ by updating conditional distribution for $\boldsymbol{\mu}^*$ given the current values of $\boldsymbol{\mu},\boldsymbol{\Theta}$ in each iteration of the MCMC algorithm. Finally, the vector of predicted yields $\boldsymbol{y}^*$ is drawn from the likelihood layer density $p_1(\boldsymbol{y}^*|\,\boldsymbol{\mu},\boldsymbol{\Theta})$.

One important advantage of using the Kriging method is to get densities of counties with insufficient historical observations (i.e., missing observation) or even no observations. Since Kriging identifies a spatial structure for explaining a variation of densities across space, we can predict densities of locations with no historical observations using the Kriging parameters and posterior predictive formula in equation (12)[8].

The prediction quality of the model is evaluated by calculating the Predictive Model Choice Criteria (PMCC) suggested by Gelfand and Ghosh (1998), which is defined as,

(13) $$PMCC = \sum_{t=1}^{T} E[(\boldsymbol{y}_t^* - \boldsymbol{y}_t)'(\boldsymbol{y}_t^* - \boldsymbol{y}_t)] + var(\boldsymbol{y}_t^*),$$

where $\boldsymbol{y}_t^*$ denotes a vector of predicted yield level at year t, and $\boldsymbol{y}_t$ is a vector of actual yields at year t. The first term of the PMCC represents the goodness of fit of the model, and the second term represents a penalty of model complexity. Several specifications for deterministic parts of GP and AR are tested as potential candidates for the empirical analysis. A model with lower PMCC is chosen as the preferred model. We test multiple different specifications for the process layer structure with a moving average term for historical yields (from last five years to ten years), simple linear trend, and quadratic linear trend. Among these alternatives, the model including five years of moving average and simple linear trend minimizes PMCC and thus is selected as our main model. The main model for the GP can be expressed according to the following structure for the process layer,

(14) $$\boldsymbol{\mu}_t = \beta_0 + \beta_1 \overline{\boldsymbol{y}}_t + \beta_2 \boldsymbol{t} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_t,$$

and for the AR model,

(15) $$\boldsymbol{\mu}_t = \omega \boldsymbol{\mu}_{t-1} + \beta_0 + \beta_1 \overline{\boldsymbol{y}} + \beta_2 \boldsymbol{t} + \boldsymbol{\eta} + \boldsymbol{\varepsilon}_t,$$

where $\overline{\boldsymbol{y}}_t$ is a vector of current five years average of yields, $\overline{\boldsymbol{y}} = [\bar{y}_1, \dots, \bar{y}_N]'$, and $\boldsymbol{t}$ is a vector of trend variable $\boldsymbol{t} = [t_1, \dots, t_N]'$.

Several R-packages provide MCMC algorithms for Bayesian Kriging estimation. We mainly use spTimer, spBayes, and SpatialExtremes packages. These packages are used to estimate posterior distributions, posterior predictive distributions, and PMCC of the models. We run 20,000 iterations for MCMC chains and burn-in the first 5,000 observations to avoid an autocorrelation problem of the posterior values. We check for all parameters the graphical diagnostics of convergence using trace plots and autocorrelation plots. All posterior densities achieve fast convergence with no significant autocorrelation. To save space, Table 1 presents only the averages and standard deviations of the posterior parameter values for the four states in the dataset. Two leading states in the corn-belt area, Iowa and Illinois, show similar posterior parameter values. Both states have very small levels of temporal correlation parameter ω (0.01 for Iowa and 0.02 for Illinois) and thus the parameter of the current five year average yields (β_1) and trend parameter (β_2) differ little between the GP and the AR. In contrast, we find a notable level of temporal correlations (ω) in Nebraska and Colorado (0.37 for Nebraska and 0.57 for Colorado). The results indicate that the two leading states in the corn-belt area (Iowa and Illinois) have a more steadily increasing trend than Nebraska and Colorado.

Out of sample performance

Providing actuarially sound premiums is an essential task to RMA. Crop producers will turn down insurance contracts when the premiums are overrated and may result in an adverse selection problem. Likewise, underrated premiums may induce insurance losses to agencies. The premium rate of crop insurance represents expected payouts as a proportion of total liability. The premium rate can be structured as

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

where λ is the coverage level, $0 < \lambda < 1$, $\hat{y}_i$ is an expected crop yield at county i.

The premium rates can be estimated using the posterior predictive distributions and the premium rate formula equation (16). The posterior predictive distribution for each county can be obtained from the formula in equation (12). We then calculate the 90 percent coverage premium rates from equation (16). The expected yield $\hat{y}_i$ in county i is to be the posterior mean of the predictive distribution for each county. The premium rates differ across the alternative model specification and use of the spatial smoothing spaces. For example, each state has premium rates from AR and GP model estimated under both the physical and the climate space. Table 2 presents each state's average premium rates across counties from 2000 to 2014. One interesting result is that Colorado shows notable difference in average premium rates between using the physical space and the climate space compared to the other states. Figure 2 and 3 illustrate estimated premium rates of Iowa and Colorado from the two types of spaces, respectively. The left map presents the premiums from the physical space, and the right map presents the premiums from climate space. In accordance with our expectation, premium estimates in Iowa from the two spaces show no substantial difference, which reflects Iowa having similar spatial structure both climate and physical distance. In contrast to Iowa, premiums in Colorado show meaningful differences in spatial structure. Specifically, we find that premiums of western region counties (Delta, Mesa, and Montrose) and southern region counties (Baca and Otero) are increased when we use climate space.

We use a loss ratio under the corresponding crop insurance program as a tool for evaluating out of sample performance of the models under the two different smoothing spaces. The loss ratio is given by

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

where λ is coverage level, $\hat{y}_{it}$ is a predicted yield of county i at year t, y_{it} is an actual yield for county i at year t, and $prem_{it}$ is the premium rate of county i at year

t, which is obtained from equation (16).

The premium gains and indemnity losses from equation (17) are calculated using actual yields and estimated premiums of each county from 2000 to 2014. Average, variance, maximum, and minimum loss ratio across counties for the six states are presented in Table 3. The loss ratio of fairly rated crop insurance should equal one. Thus, a model with average loss ratio close to one and with a small variance across counties (i.e., regional equality) might be a preferred model. Our results demonstrate that considering the temporal correlation of crop yield (AR model) results in notable improvement in measuring the premiums in every state in our dataset. The average loss ratios from the AR model are close to one and have smaller variance in every state compared to the GP model.

Climate space performs better or equal to physical space in every state in our dataset. In Colorado and Nebraska, the climate space conspicuously provides greater performance in out of sample prediction. Both of the average loss ratios from the climate space are closer to one than the physical space in Colorado (4.75 and 4.17) and Nebraska (1.97 and 1.67), which indicates the premiums from the climate space are more fairly rated than the physical space. Perhaps most significant is the finding that climate space smoothing resolves the serious regional inequality problem in Colorado. Our results also show that climate space has a smaller variation of loss ratio across counties. The reduction in the variance of the loss ratio of Colorado is from 36.31 to 18.80 and Nebraska is from 12.34 to 7.46. Specifically, for counties in Colorado with a high loss ratio under physical space, such as Adams and Washington counties, the loss ratio becomes closer to one under climate space smoothing. Our results demonstrate that climate space smoothing more pertinently describes spatial structure of the crop yield density compared to the physical space smoothing especially for a state like Colorado where the climate is diverse. In case of Colorado, a closeness of physical distance between locations would not be a successful factor to explain crop yield similarity across space.

Conclusion

Federal crop insurance programs have been solidified by the Agricultural Act of 2014. In this regard, providing accurate premiums for insurance contracts has become again the utmost important role of RMA. However, the current RMA model does not fully attempt to use historical data from other areas in estimating the premium for an area of interest. Therefore, the current model has problems both in reflecting spatial correlation and retaining enough observations to properly estimate the premiums. The Bayesian Kriging model proposed here suggests a promising way to solve such problems.

There are only a few examples of Bayesian Kriging models in the agricultural economics literature, and this article is the first to use climate space smoothing. Our results show that crop yield similarity across locations tends to be more affected by climate similarity than locational closeness. By conducting climate space smoothing, we can provide a better way to measure the crop yield densities especially for a geographically diverse region such as Colorado. Future research could consider additional measures of similarity such as soil type and slope of agricultural land.

The Kriging method could be adopted in any research area that involves spatial correlation. A prominent extension of the model, for example, would be for precision agriculture. Today the availability of accurate and abundant field monitoring data allows developing a crop yield response model that parameters are smoothed by site specific agrological characteristics such as soil type, water, etc. This application could provide better site specific Variable Rate Application (VRA) fertilizer prescriptions.

One of great strengths of the Kriging model would be density estimation for the counties where historical yield data is limited or there are no observations. Bayesian Kriging method defines and describes a spatial structure for a variation of densities across space. The method allows a density prediction of any locations (counties) with no historical observations by using a posterior predictive distribution. In this regard, the method suggests a useful way to produce density measures for counties with no yield reported.

Because we are focusing on introducing and comparing the performance of the climate space smoothing method compared to the physical distance smoothing, our model treats crop yield density in a simple manner using the normality assumption. Although the normality assumption has the advantage of simplifying the MCMC structure and could easily include trend variable into our model specification, the assumption still has a shortcoming to adjust higher moment characteristics of crop yield density such as asymmetric skewness. Therefore, while it may require developing techniques that are not yet available in the statistics literature, future research should attempt to relax this distributional assumption.

The Kriging method for climate space smoothing proposed here clearly has the potential to offer significant efficiency gains in crop yield density estimation, where historical observations is limited and has varying climate conditions.

Table 1. Average Posterior Parameter Values for Selected States

State / Smoothing Space		Physical Space		Climate Space	
Model Structure		GP (S.D)	AR (S.D)	GP (S.D)	AR (S.D)
Iowa	β_1	1.01	1.02	1.01	1.02
		(0.02)	(0.03)	(0.02)	(0.03)
	β_2	1.75	1.78	1.76	1.78
		(0.02)	(0.03)	(0.02)	(0.02)
	ω	-	0.01	-	0.01
			(0.00)		(0.00)
	ρ	40.35	40.15	51.31	50.23
		(3.25)	(3.25)	(4.12)	(4.19)
	θ	18.23	18.42	21.83	22.43
		(2.65)	(2.95)	(3.65)	(3.11)
Illinois	β_1	1.03	1.01	1.02	1.01
		(0.03)	(0.03)	(0.02)	(0.02)
	β_2	1.72	1.71	1.72	1.72
		(0.02)	(0.03)	(0.02)	(0.02)
	ω	-	0.02	-	0.02
			(0.00)		(0.00)
	ρ	36.35	36.15	45.31	45.23
		(3.95)	(3.99)	(5.22)	(5.23)
	θ	19.38	19.22	24.13	24.15
		(3.26)	(3.15)	(4.82)	(5.11)
Nebraska	β_1	0.78	0.63	0.99	0.63
		(0.03)	(0.03)	(0.02)	(0.03)
	β_2	2.07	1.34	2.13	1.34
		(0.02)	(0.02)	(0.02)	(0.02)
	ω	-	0.37		0.36
			(0.01)		(0.02)
	ρ	42.11	42.19	61.31	62.23
		(6.51)	(6.52)	(4.12)	(4.19)
	θ	15.12	14.21	20.69	20.66
		(1.11)	(1.12)	(2.18)	(2.26)
Colorado	β_1	0.30	0.14	0.30	0.15
		(0.02)	(0.02)	(0.02)	(0.02)
	β_2	1.83	0.74	1.81	0.74
		(0.05)	(0.01)	(0.04)	(0.01)
	ω	-	0.56	-	0.58
			(0.01)		(0.01)
	ρ	77.25	78.10	98.31	98.44
		(7.42)	(7.44)	(11.12)	(11.19)
	θ	12.13	12.12	19.13	19.12
		(1.00)	(1.00)	(3.65)	(3.61)

Table 2. Average of 90 Percent Coverage Premium Rates across Counties

State / Smoothing Space		Physical Space		Climate Space	
Model Structure		GP	AR	GP	AR
Iowa	Premium Rate (%)	1.73	1.79	1.57	1.59
Illinois	Premium Rate (%)	1.52	1.54	1.54	1.55
Nebraska	Premium Rate (%)	1.38	1.27	1.28	1.31
Minnesota	Premium Rate (%)	1.59	1.57	1.63	1.59
Indiana	Premium Rate (%)	1.65	1.69	1.64	1.75
Colorado	Premium Rate (%)	1.48	1.62	2.41	2.53

Table 3. Estimated Loss Ratio under the Physical Space and the Climate Space

State / Smoothing Space		Physical Space		Climate Space	
Model Structure		GP	AR	GP	AR
Iowa	Mean	0.94	0.94	1.05	1.00
	Variance	1.26	1.30	1.23	0.96
	Max	5.06	4.88	4.26	3.62
	Min	0.00	0.55	0.00	0.56
Illinois	Mean	2.06	2.03	2.03	2.00
	Variance	1.08	1.10	1.07	1.01
	Max	4.28	4.35	4.26	4.32
	Min	0.00	0.00	0.00	0.13
Nebraska	Mean	2.22	1.97	2.36	1.67
	Variance	7.55	4.30	9.54	3.15
	Max	14.39	10.28	12.34	7.46
	Min	0.00	0.00	0.05	0.31
Minnesota	Mean	0.37	0.41	0.36	0.42
	Variance	0.21	0.20	0.18	0.19
	Max	1.67	1.49	1.45	1.47
	Min	0.00	0.00	0.00	0.00
Indiana	Mean	0.86	0.92	0.91	0.98
	Variance	0.58	1.74	0.36	0.12
	Max	5.36	4.89	5.34	4.98
	Min	0.05	0.09	0.06	0.23
Colorado	Mean	6.97	4.75	5.86	4.17
	Variance	59.90	26.01	36.31	18.80
	Max	23.19	15.78	17.65	13.25
	Min	0.17	0.05	0.17	0.35

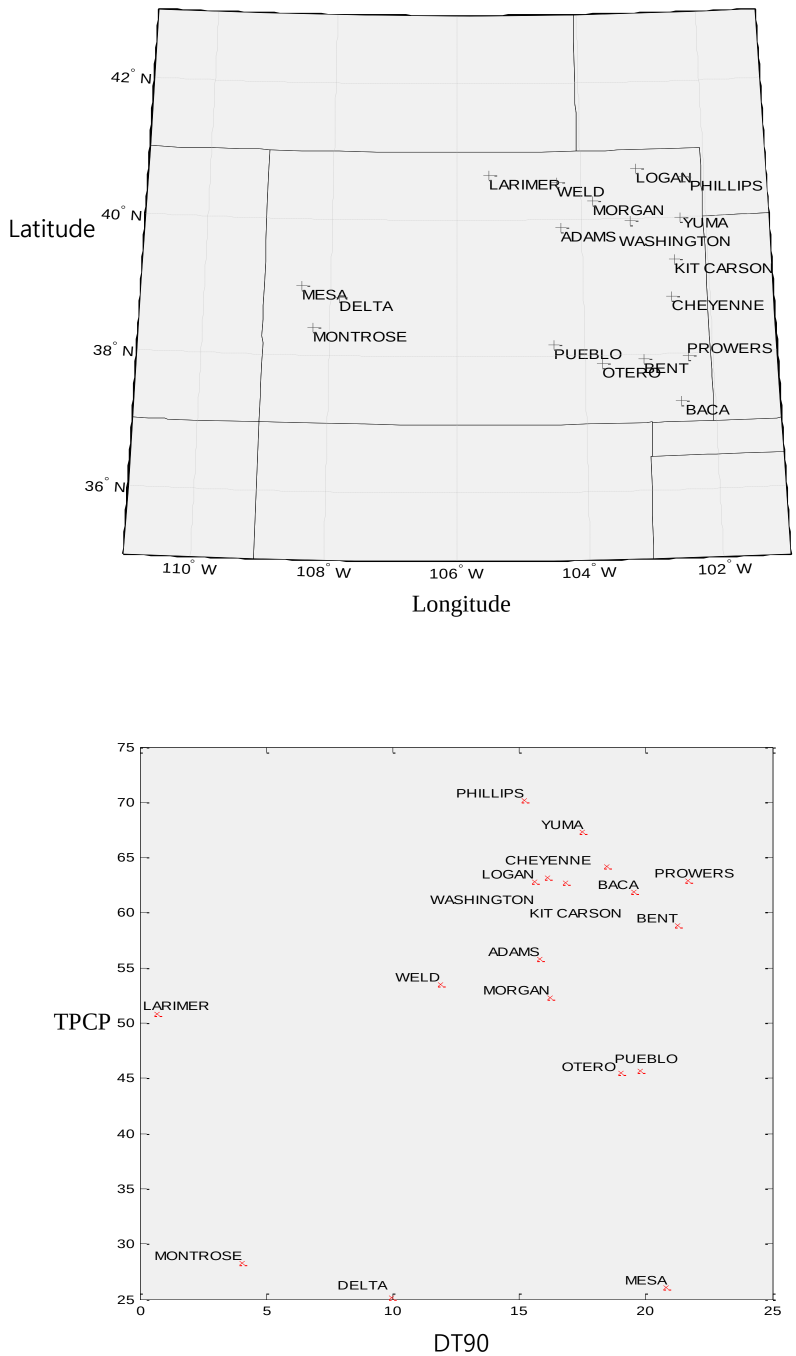

Figure 1. Translation of counties of Colorado in physical space (above) to climate space (below)

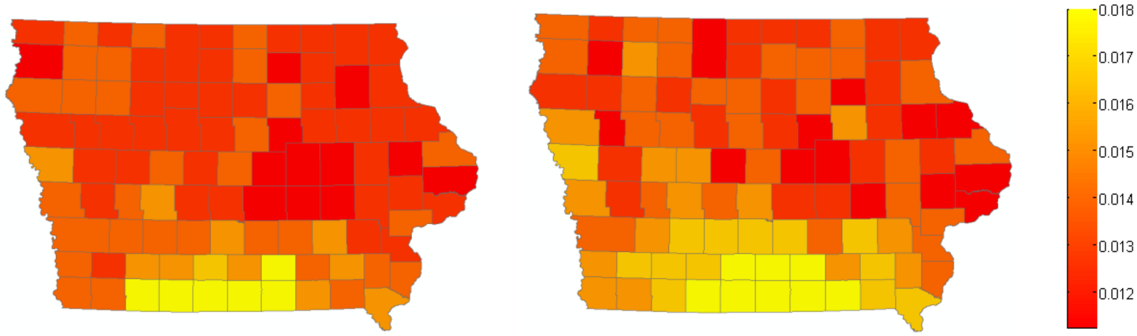

Figure 2. Premiums of Iowa from physical space (left) and from climate space (right)

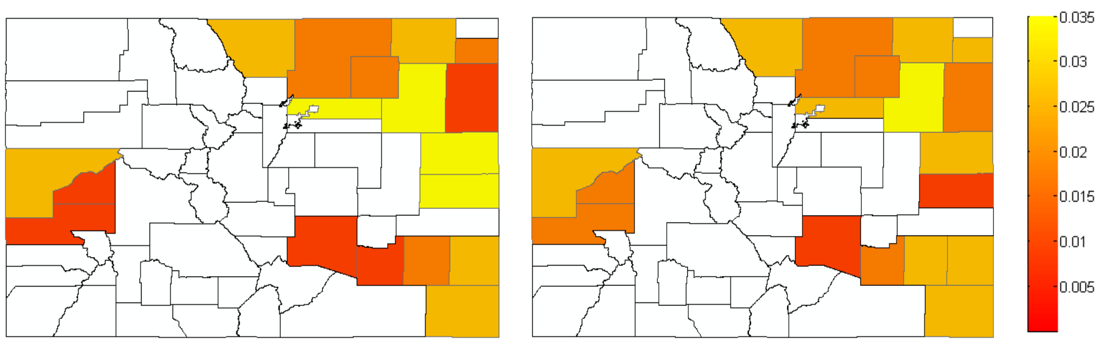

Figure 3. Premiums of Colorado from physical space (left) and from climate space (right)

REFERENCES

Annan, F., J. Tack, A. Harri, and K.H. Coble. 2013. Spatial Pattern of Yield Distributions: Implications for Crop Insurance. *American Journal of Agricultural Economics* 96(1):253–68.

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.

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.

Coble, K.H., R. Dismukes., and J.W. Glauber. 2007. Private Crop Insurers and the Reinsurance Fund Allocation Decision. *American Journal of Agricultural Economics* 89:582–95.

Cooley, D., D. Nychka., and P. Naveau. 2007. Bayesian Spatial Modeling of Extreme Precipitation Return Levels. *Journal of the American Statistical Association* 102:824-40.

Cooley, D., P. Naveau., and P. Poncet. 2006. Variograms for Spatial Max-stable Random Fields. *Dependence in Probability and Statistics* 187:373-90.

Dell, M., B.F. Jones, and B.A. Olken. 2014. What Do We Learn from the Weather? The New Climate-Economy Literature. *Journal of Economic Literature* 52:740–98.

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.

Goodwin, B.K., and A. Hungerford. 2015. Copula-Based Models of Systemic Risk in U.S. Agriculture: Implications for Crop Insurance and Reinsurance Contracts. *American Journal of Agricultural Economics* 97(3):879–96.

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.

Hendricks, N. P., and J. M. Peterson. 2012. Fixed Effects Estimation of the Intensive and Extensive Margins of Irrigation Water Demand. *Journal of Agricultural and Resource Economics* 37(1): 1-19.

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

Lobell, D.B., G.L. Hammer, G. McLean, C. Messina, M.J. Roberts, and W. Schlenker. 2013. The Critical Role of Extreme Heat for Maize Production in the United States. *Nature Climate Change* 3:497–501.

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: 1151–64.

Ozaki, V.A., and R.S. Silva. 2009. Bayesian Ratemaking Procedure of Crop Insurance Contracts with Skewed Distribution. *Journal of Applied Statistics* 36 (4): 443–52.

Park, E., B.W Brorsen., and A. Harri., 2016. Using Bayesian Spatial Smoothing and Extreme Value Theory to Develop Area-Yield Crop Insurance Rating. Paper presented at the 2016 Agricultural and Applied Economics Association Annual Meeting in Boston, Massachusetts, July 31 – Aug 2.

Schlenker, W., and M.J. Roberts. 2009. Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change. *Proceedings of the National Academy of Sciences* 106:15594–15598.

Woodard, J. D. 2016. Determining Optimal Data Aggregation: An Application of Out-of-Sample Mixture Models. Paper presented at the 2016 Agricultural and Applied Economics Association Annual Meeting in Boston, Massachusetts, July 31 – Aug 2.

---

[1] The Gaussian spatial process is a stochastic process where any finite subcollection of random variable (i.e., mean parameter of any county) is multivariate normally distributed. The covariance of those random variables between any two locations (i.e., covariance between μ_i and μ_j) is determined by a Euclidean distance between two locations and spatial covariance matrixΣ.

[2] To avoid complexity of the Markov Chain Monte Carlo (MCMC) structure, we only assume the hierarchical structure for the mean parameter. Therefore, we directly impose general inverse-gamma type prior distribution to the variance parameter of the likelihood layer σ^2.

[3] Under the Bayesian framework, we regard the mean parameter μ as a random variable. When a random vector of mean equation, $\boldsymbol{\mu} = (\mu_1, \dots \mu_N)^T$ is assumed to follow a Gaussian spatial process, the mean parameters of the counties $\mu_1, \dots \mu_N$ are jointly normally distributed.

[4] We use the empirical variogram to impose proper priors for the sill parameter ρ. For MLE estimates of mean parameter $\hat{\mu}_i$ for county location $i = 1, \dots, N,$ empirical variogram can be defined as

$$\hat{\gamma}(D_{ij}) \coloneqq \frac{1}{2M} \sum_{(i,j) \in M} (\hat{\mu}_i - \hat{\mu}_j)^2$$

where M is the number of all possible pairs of counties, D_{ij} is Euclidean distance between two counties i and j, $\hat{\mu}_i$ and $\hat{\mu}_j$ are MLE estimates for the mean parameter of county i and j, respectively.

[5] In the climate space, locations (counties) that have similar climate features are closely located, even though their locations may be distant on the traditional physical space.

[6] Since our dataset includes 99 counties for Iowa and 18 counties for Colorado, all possible pairs of distances among the counties for each state are 4,851 and 153, respectively. We first obtain the distances from two different types of space (physical and climate). We then normalize these distances and calculate correlations between the two types of distance.

[7] All models in this article are fitted using a Metropolis-Hastings (MH) within Gibbs sampling algorithm. As mentioned in the prior layer section, standard conjugate priors are assumed for all coefficients (normal) and variance (inverse gamma) parameters. However, prior densities for the Kriging parameters (sill and range) are non-standard. Hence the MH algorithm is employed to draw samples from the Kriging parameters.

[8] Kriging defines a spatial interpolation function that determines spatial correlation structure across density parameters of locations. Therefore, we can generate an interpolated density for any specific point using the

Kriging parameters (sill and range) and spatial information of the point (coordinates). In this article, we do not aim to produce a specific county's density with no historical observations. However, Kriging can produce densities with no observations by using the posterior predictive distribution in equation (12).