Rating Crop Insurance Contracts with Gaussian Processes and Model Stacking

Weifang Liang

Department of Agricultural Economics, Texas A&M University

wf.liang@tamu.edu

Yong Liu

Department of Agricultural Economics, Texas A&M University.

yongliu@tamu.edu

Bart Fischer

Department of Agricultural Economics, Texas A&M University

bart.fischer@ag.tamu.edu

Henry Bryant

Department of Agricultural Economics, Texas A&M University

henry.bryant@ag.tamu.edu

Selected Paper prepared for presentation at the 2023 Agricultural & Applied Economics Association Annual Meeting, Washington DC; July 23-25, 2023

Rating Crop Insurance Contracts with Gaussian Processes and Model Stacking

Weifang Liang[1], Yong Liu[2], Bart Fischer[3], and Henry Bryant[4]

Abstract

It is crucial to predict crop yield distribution accurately to obtain an actuarially fair premium rate when designing crop insurance contracts considering the importance of crop insurance in U.S. agricultural policy. We proposed a new rating methodology that combines two machine learning techniques: Gaussian Process and stacking. Gaussian Process consolidates the conventional two-stage estimation into one and is capable of capturing nonlinear temporal effects and time-varying distributions of crop yields. Stacking pools similar counties together and prioritizes the ability to forecast future crop yields over fitting existing data points when determining optimal weights on model averaging. Heteroscedasticity adjustment is also added which further improves the predictive performance. A repeated out-of-sample rating game is conducted to demonstrate the economic importance of the proposed method. Based on the game results, stacking models outperform individual Gaussian Process models by having smaller average loss ratios. Also, the stacking model benefits from enlarging the candidate pool size. Compared to the Risk Management Agency (RMA) rating methodology, the stacking model after heteroscedasticity adjustment generates more accurate premium rates.

Keywords: Crop insurance, Gaussian Process, stacking, premium rates

[1]Weifang Liang, Ph.D. Candidate, Texas A&M University.
[2]Yong Liu, Assistant Professor, Department of Agricultural Economics, Texas A&M University. Email: yongliu@tamu.edu
[3]Bart Fischer, Research Assistant Professor, Department of Agricultural Economics, Texas A&M University.
[4]Henry Bryant, Research Professor, Department of Agricultural Economics, Texas A&M University.

Introduction

Crop insurance is a critical component of U.S. agricultural policy, with over $193 billion in total liability in 2022 (U.S. Department of Agriculture Office of Inspector General, 2022). Accurately forecasting crop yield distribution is essential in designing a crop insurance policy to obtain an actuarially fair premium rate for insurance contracts. A more precise rating methodology for crop insurance improves the overall soundness of the program and helps to reduce adverse selection and moral hazard in crop insurance markets caused by asymmetric information.

Rating area-type crop insurance contracts relies on county-level yield data, which depend on various inputs, such as weather, seed technology, cropping intensity, soil quality, and farming practices. Therefore, crop yield data exhibit nonlinear trends and time-varying distributions. Previous research commonly uses a two-stage approach to predict crop yield distribution. In the first stage, the trend of crop yields is estimated and removed to obtain the demeaned yields. In the second stage, the distribution of demeaned yields is estimated using a parametric or nonparametric approach. This approach has two potential limitations (Wu et al., 2021). First, if the trend is not accurately estimated in the first stage, an incorrect result may be concluded in the second stage. Second, while the actual yield distribution can be changing over time, a stationary yield distribution is generally assumed in the second stage. Currently, the Risk Management Agency (RMA) of the U.S. Department of Agriculture (USDA) uses a two-knot linear spline model to estimate the trend function, which lacks flexibility due to its simplicity. Therefore, to capture the nonlinear temporal function and time-varying distributions of crop yields more accurately, we propose a Gaussian Process (GP) model for each individual county. GP is a probabilistic machine learning method that combines the two stages of yield density prediction into one and provides a more accurate prediction that incorporates uncertainty that can vary over time.

On the other hand, county-level crop yields often have short histories of 60-70 years and exhibit spatial dependence due to similar weather, soil conditions, and other geographical features. Pooling or averaging individual county models can lead to more accurate estimates

by borrowing information from similar counties. Thus, identifying the "similarity" between neighboring counties is crucial for implementing this averaging scheme. Some literature incorporates the geographic location of counties for this purpose, as neighboring counties tend to be similar in crop production (Du et al., 2015; Wu and Zhang, 2020). Bayesian Model Averaging (BMA) (Ker, Tolhurst, and Liu, 2016) offers an alternative approach by identifying similar counties without specifying geographic information, which allows candidate counties far away from the target county to contribute information on predicting the yield density as long as they are "similar." However, BMA has the flaw that it converges to the single candidate model in the pool closest to the true model when the true model is not in the candidate model pool, which is almost always the case in forecasting crop yield density. In contrast, model stacking, the method proposed in this paper, overcomes this problem by averaging individual models to generate a distribution closest to the data-generating process under the chosen scoring rule. The weights for averaging are derived by maximizing the overall predictive log scores, which measure the forecast ability of each candidate model when predicting the crop yields of the target county (Geweke and Amisano, 2011; Yao et al., 2018).

In this study, we evaluate the proposed estimation strategy using county-level U.S. corn yield data from 1960 to 2020 for seven major corn-producing states. By incorporating information from similar counties, the stacking densities demonstrate better predictive performance, as measured by yearly average log scores. A repeated out-of-sample rating game is also conducted to evaluate the economic importance of the proposed method. The rating game results show that stacking models perform better than individual models by showing a substantial performance gain from enlarging the candidate pool size. Heteroscedasticity adjustment also improves rating game results in general. Furthermore, stacking models outperform the conventional two-stage process used by RMA for capturing economical rents and generating more accurate premium rates.

Tradition Approaches on Crop Insurance Rating

Many approaches have been applied to estimate crop yield density in the literature. Most of these approaches contain two stages. In the first stage, the trend of the yield distribution is estimated. Conceptually, the trend represents technological advancements and climate change over time. Polynomial (Just and Weninger, 1999), univariate ARIMA (Goodwin and Ker, 1998), and mixtures of trend functions (Tolhurst and Ker, 2015) are employed in previous studies for detrending. In the second stage, residuals taken from the first stage are used to model the yield distribution after adjustment of potential heteroscedasticity. Methods here can be classified into two classes: parametric and non-parametric. Parametric methods are estimated by assuming a specific functional form of the yield distribution. Among them, the normal distribution is used by Botts and Boles (1958) and considered to be not unreasonable by Just and Weninger (1999). Other commonly used specifications include Gamma (Gallagher, 1987), Beta (Nelson and Preckel, 1989), and Logistic (Atwood, Shaik, and Watts, 2003) distributions. However, if the parametric distribution is specified incorrectly with respect to the underlying crop yield distribution, biases may arise. Non-parametric methods on the other hand provide a more flexible way to avoid specification error and estimate yield density but require more effective samples to ensure convergence. More examples include inverse sine transformation method (Moss and Shonkwiler, 1993), kernel density estimation (Ker and Goodwin, 2000; Liu and Ker, 2020a; Wen, Wu, and Leatham, 2021), maximum entropy (Wu and Zhang, 2012; Tack, Harri, and Coble, 2012) and normal mixtures (Goodwin, Roberts, and Coble, 2000; Woodard and Sherrick, 2011; Tolhurst and Ker, 2015).

In estimating crop yield density, the major difficulty is the lack of data. Commonly the county-level crop yield only has a history of 60-70 years. On the other hand, county-level crop yield data is featured with spatial dependencies among counties due to the similar weather, farming practice, and soil conditions. To exploit this feature, one can pool or 'borrow' information from similar counties to achieve more accurate estimates. A variety of approaches have been applied here. For example, Annan et al. (2013) conduct a test for

distributional equivalence to identify similar densities and find pooled model results in more accurate rates. Bayesian model average is used by Ker, Tolhurst, and Liu (2016) to identify and pool similar densities together and further employed by Liu and Ker (2020b) to smooth across both space and time. Zhang (2017) proposes a density-ratio estimator which first sets a common baseline density and then estimates individual distributions as deviations from the baseline. This method is further extended by incorporating spatial dependence among counties, where the weights of pooled counties are assigned by a local maximum likelihood estimator proposed by Wu and Zhang (2020).

Gaussian Process

This study estimates the individual county model by Gaussian Process regression. Gaussian Process is a powerful nonparametric method that can produce the entire predictive distribution instead of only point estimates. In past literature for crop insurance, Wu et al. (2021) employ a mixture of GPs to estimate crop yield density and it outperforms the traditional two-stage estimators, especially with nonstationary underlying distributions. For more technical details of the Gaussian Process, we refer to Williams and Rasmussen (2006) for a thorough overview.

Preliminaries

A Gaussian Process can be considered a multivariate Gaussian distribution with infinitely many variables. Therefore, to completely specify a GP, we only need a mean function $m(\mathbf{x})$ and covariance function $K(\mathbf{x}, \mathbf{x}')$

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), K(\mathbf{x}, \mathbf{x}')) \tag{1}$$

By the marginalization property, we can partition the data points we care about and only work with this subset. Then the infinite dimension GP reduces to a finite multivariate Gaussian distribution which is feasible for calculation.

Given a training data set with input vector $\mathbf{x} = [x_1, x_2, ..., x_n]'$ and corresponding noisy observations $\mathbf{y} = [y_1, y_2, ..., y_n]'$, where $\mathbf{y} = f(\mathbf{x}) + \epsilon, \epsilon \sim N(0, \sigma_n^2)$, a GP regression model is formulated by placing a GP prior over the non-linear function f. Here we assume f to be a zero-mean GP with covariance function $K(\mathbf{x}, \mathbf{x})$. There is a variety of choices for this covariance function.

To predict the outcomes at test location $\mathbf{x}^*$, the joint distribution of $f(\mathbf{x}^*)$ and $\mathbf{y}$ need to be evaluated. Denote $\mathbf{f}^* = f(\mathbf{x}^*)$, we have:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(\mathbf{x}, \mathbf{x}) + \sigma_n^2 I & K(\mathbf{x}, \mathbf{x}^*) \\ K(\mathbf{x}^*, \mathbf{x}) & k(\mathbf{x}^*, \mathbf{x}^*) \end{bmatrix}\right) \tag{2}$$

With n training points and n^* testing points, $K(\mathbf{x}, \mathbf{x}^*)$ is an $n \times n^*$ matrix where the $(i, j)^{th}$ entry is the covariance between the training point x_i and testing point x_j^*, and similarly for $K(\mathbf{x}, \mathbf{x}), K(\mathbf{x}^*, \mathbf{x})$ and $K(\mathbf{x}^*, \mathbf{x}^*)$

The predictive distribution of $\mathbf{f}^*$ given the data is:

$$\mathbf{f}^* | \mathbf{x}, \mathbf{y}, \mathbf{x}^* \sim N(\bar{\mathbf{f}}^*, cov(\mathbf{f}^*)) \tag{3}$$

with predictive mean and predictive covariance:

$$\bar{\mathbf{f}}^* = K(\mathbf{x}^*, \mathbf{x})[K(\mathbf{x}, \mathbf{x}) + \sigma_y^2 I]^{-1}\mathbf{y} \tag{4}$$

$$cov(\mathbf{f}^*) = K(\mathbf{x}^*, \mathbf{x}^*) - K(\mathbf{x}^*, \mathbf{x})[K(\mathbf{x}, \mathbf{x}) + \sigma_y^2 I]^{-1} K(\mathbf{x}, \mathbf{x}^*) \tag{5}$$

The predictive uncertainty can be interpreted as prior uncertainty minus the reduction in uncertainty from seeing the training data. Therefore, with more training data points observed and more correlation between training data and predictive points, the predictions will be more confident than prior.

Application to Crop Yields

This study is using 1960 - 2020 United States county-level corn yield data obtained from National Agricultural Statistics Service. Seven major corn-producing states are selected: Illinois, Iowa, Michigan, Minnesota, Missouri, Ohio, and Wisconsin. After removing counties with missing values, a total of 300 counties are left in the final data set.

Let county-level corn yields be output y and each year as input x. For each county i ($i = 1, ..., 300$), to predict the corn yield distribution in each year of x^* ($x^* = 1990, 1991, ...2021$) we use all data from that county before that year to fit a Gaussian Process. In total, 32 times 300 different GP models are built in this step. A zero mean GP with Squared Exponential (SE) covariance function is used to fit the model. The SE kernel takes the form:

$$k(x_i, x_j) = \sigma_f^2 \exp(-\frac{1}{2l^2}(x_i - x_j)^2) \tag{6}$$

where the length-scale l defines how quickly these correlations fall away, and signal variance σ_f^2 defines the scale of the function. Since corn yields are noisy observations, there is another parameter σ_n^2 that defines how noisy the data are at each point. For $y_i = f(x_i) + \varepsilon_i$, the error term takes the form:

$$\varepsilon_i \sim N(0, \sigma_n^2) \tag{7}$$

The covariance function for y becomes

$$k(x_i, x_j) = \sigma_f^2 \exp(-\frac{1}{2l^2}(x_i - x_j)^2) + \sigma_n^2 \delta_{ij} \tag{8}$$

where δ_{ij} is the Kronecker delta function.

For each GP model, the three hyperparameters (l, σ_f, σ_n) are estimated through maximum a posteriori (MAP), which is known as the penalized maximum likelihood estimate. It obtains a posterior mode estimate, which is also the mean of the posterior distribution since we are dealing with Gaussian posterior.

Finally, with the estimated hyperparameters, we can predict the corn yield distribution as a Normal distribution centered at predictive mean with predictive variance calculated by equations (4) and (5).

Heteroscedasticity Adjustment

Possible heteroscedasticity in crop yields is studied in past research (Harri et al., 2011; Tolhurst and Ker, 2015; Ker and Tolhurst, 2019). Due to technological change, corn yield data is showing signs of nonconstant variance with respect to time. Tolhurst and Ker (2015) find yield density becomes more dispersed with bimodal shape over time. To adjust for possible heteroscedasticity of crop yield, we assume:

$$\epsilon_i | x_i \sim N(0, \sigma^2_{n,i}) \tag{9}$$

where

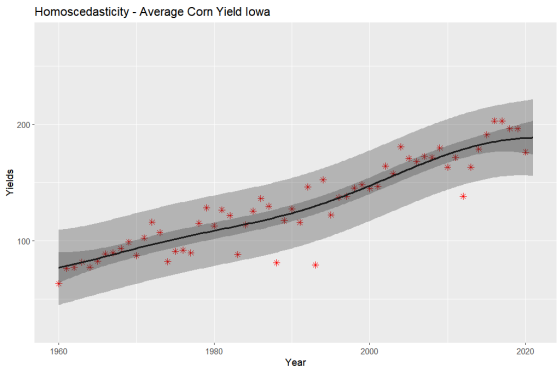
$$\ln \sigma^2_{n,i} = \alpha + \beta x_i \tag{10}$$

The covariance function for $y_i = f(x_i) + \varepsilon_i$ is now

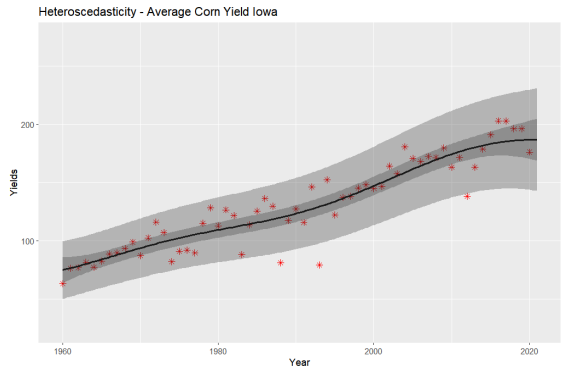
$$k(x_i, x_j) = \sigma_f^2 \exp(-\frac{1}{2l^2}(x_i - x_j)^2) + \sigma^2_{n,i}\delta_{ij} \tag{11}$$

Instead of three, four hyperparameters $(l, \sigma_f, \alpha, \beta)$ are estimated through MAP when accounting for heteroscedasticity.

To demonstrate the overall pattern and show the changes when incorporating heteroscedasticity, Figure 1a and 1b show the fitted GP model of the average corn yields from all counties in Iowa using historical data from 1960 to 2020 and predict the mean of the yield distribution in 2021. The darker shaded area is the GP predicted mean plus and minus 1.96 predicted signal standard deviation. The lighter shaded area is the GP predicted mean plus and minus 1.96 predicted standard deviation of signal with noise. With heteroscedasticity incorporated, variance is now increasing over time.



(A) Homoscedasticity - Iowa



(B) Heteroscedasticity - Iowa

Stacking

Crop yield data is having a short history which is restricting the performance of the prediction from the GP model. We proposed stacking as the approach to identify similar counties and pool information together to overcome this difficulty.

Preliminaries

Model stacking is an ensemble machine learning technique that trains multiple models and combines them in a way that maximizes their collective predictive performance. Generally, it contains two steps: First, Given a set of data $((x_i, y_i), i = 1, ..., n)$ and the model list $M = (M_1, ..., M_m)$, each model M_k is fitted separately to obtain the leave-one-out (LOO) predictor for each data point i. Then the weight for each model is obtained by minimizing the LOO mean squared error:

$$\hat{w} = \arg\min_{w} \sum_{i=1}^{n} (y_i - \sum_{k} w_k \hat{f}_k^{(-i)}(x_i))^2 \tag{12}$$

where $\hat{f}_k^{(-i)}(x_i)$ is the LOO predictor from model k and data point i (**?**).

However, minimizing LOO mean squared error is focusing on point predictions. To extend this approach to predictive distributions, Yao et al. (2018) suggest finding stacking weights by maximizing scoring rules. There is a variety of choices for the scoring rules, but under regularity conditions, the log score is the only proper local score. Geweke and Amisano (2011)

also suggest finding the optimal weight of the linear combination of prediction models by maximizing the predictive log score.

When stacking predictive distributions, we first obtain the LOO predictive densities by fitting model k without data point (x_i, y_i) for each i and k. Then the log score is the logarithm of the probability that an actual data point y_i is generated from the LOO predictive distribution estimated without using the i^{th} data point.

$$\log S_{ik} = \log(p(y_i|y_{-i}, M_k)) \tag{13}$$

The weight for each model is derived from maximizing the summation of the log score:

$$\hat{w_k} = \underset{w}{\operatorname{argmax}} \sum_{i=1}^{n} \log \sum_{k=1}^{m} (w_k p(y_i|y_{-i}, M_k)) \tag{14}$$

Finally, the optimal predictive density is the weighted average of predictive distributions:

$$\hat{p}(\tilde{y}|y) = \sum_{k=1}^{m} \hat{w}_k p(\tilde{y}|y, M_k) \tag{15}$$

Instead of focusing on fitting existing data points more accurately, stacking is more focused on enhancing the predictive ability by utilizing the leave-one-out approach. It generates a distribution that is closest to the data-generating process under the selected scoring rule (Yao et al., 2018).

Application to Crop Yields

After estimating the individual county predictive distribution from 1990 to 2021 by Gaussian Process, we stack candidates counties together and derive the optimal weighted average predictive distribution for each county. The predictive distribution of each county is averaged in different candidate pools: using counties in the same States, or counties in the same crop-reporting-district (CRD) in each State.

Assume the corn yield data from each county is generated from a different underlying model in the model list $M = (M_1, ..., M_{300})$. To get the stacking predictive distribution of target county i in the year T $(1990 \leq T \leq 2021)$, we will need the predictive distribution

p_{kj} generated in the previous section, where k denotes each candidate county and $j = 1, ..., (T - 1990 + 1)$ represents each forecast year x_j^*. This approach can be considered a special case of the LOO method by only leaving one data point out instead of repeating this procedure for all historical data points. It mimics the decision environment of real-world insurers: predict the crop yield distribution for the next year using all historical data, which is increasing in sample size as time goes by.

Denote the pool size as m. Then, for each target county i in each forecast year x_j^* ($j \neq 32$), we put the actual yield y_{ij}^* into the GP predictive distribution p_{kj} obtained from each candidate county k in the pool ($k = 1, 2, ..., m$) to calculate the log score:

$$S_{ikj} = \log(p_{kj}(y_{ij}^*)),\ p_{kj} \sim N(\bar{f}_{kj}^*, cov(f_{kj}^*) + \sigma_{nkj}^2) \tag{16}$$

where $\bar{f}_{kj}^*$ and $cov(f_{kj}^*) + \sigma_{nkj}^2$ are the predictive mean and variance obtained from the individual GP model of county k when forecasting year x_j^*

Then, for each target county i, by maximizing the summation of the log score before year T, the stacking weights of the candidate models can be derived:

$$\hat{w}_{ik} = \underset{w}{\operatorname{argmax}} \sum_{j=1}^{T-1990} \log \sum_{k=1}^{m} (w_{ik} p_{kj}(y_{ij}^*)) \tag{17}$$

The predicted density in year T of target county i from the stacking model is:

$$p_i^* = \sum_{k=1}^{m} \hat{w}_{ik} p_{kj}^*,\ j = T - 1990 + 1 \tag{18}$$

Model stacking combines the predictions of multiple candidate models to improve predictive performance, rather than focusing solely on fitting existing data points. This feature is especially important for our study, as accurate predictions of crop yield density one or two years in advance are necessary for rating crop insurance contracts. Unlike other methods, the yield densities of candidate counties assigned with non-negative weights in stacking are not necessarily similar to the target county's densities, and these candidate counties are not

restricted to the geographical closeness to the target county. However, they still contribute to identifying the underlying data-generating process of the target county.

By definition, a higher log score represents higher forecasting performance. Figure 2 and Figure 3 present the average log score in two example states: Illinois and Wisconsin, from the year 2001 to 2020. The solid lines are from the homoscedasticity GP model and the dotted lines are the GP models after heteroscedasticity adjustment. By incorporating heteroscedasticity, we found a large gain in predictive performance. Especially when crop yield experienced large shocks, like the 2012 crop losses due to severe drought, the heteroscedasticity models are affected less than the models without adjustments.

Furthermore, the black lines indicate individual GP models, while the blue and red lines are stacking models with candidate pools using counties within CRD and counties within states, respectively. Overall, by incorporating more "similar" models, the stacking model outperforms individual models, with performance gains increasing as the candidate pool size grows.

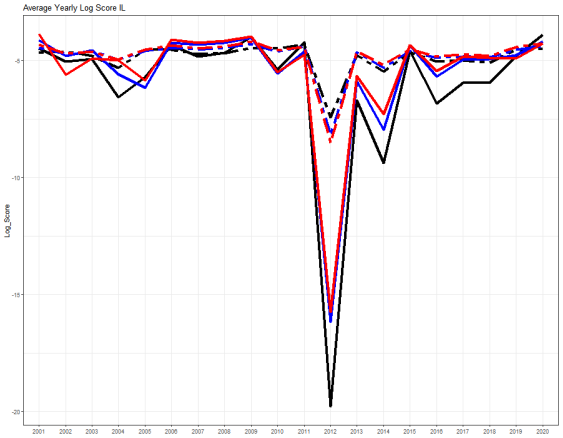


FIGURE 2. Illinois

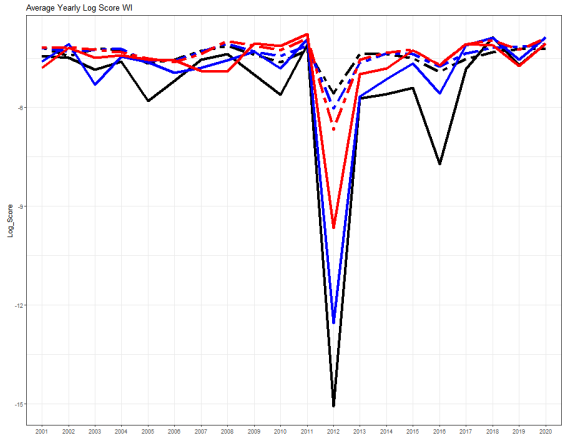


FIGURE 3. Wisconsin

Figure 4 and Figure 5 show the predictive distribution in the year 2021 using models after heteroscedasticity adjustment from two example counties: Effingham County in Illinois and Burnett County in Wisconsin. The individual predictive distributions are normal distributions with means and variances estimated by GP. The stacking distribution is no longer normal distributions but rather a mixture of normal distributions, which is more flexible

and appropriate in terms of crop yield distribution. By borrowing information from similar counties, stacking densities tend to have quite different moments than the individual ones.

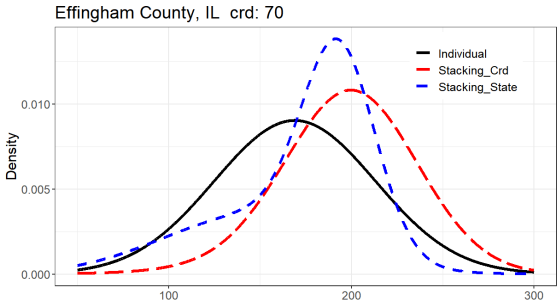


FIGURE 4. Illinois, Ford County

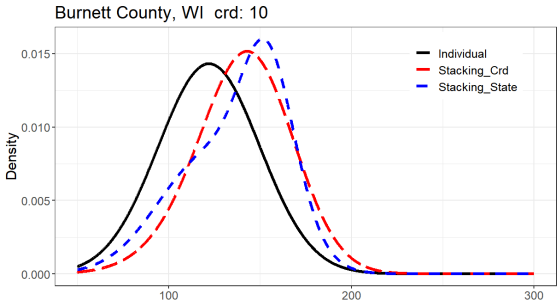


FIGURE 5. Wisconsin, Barron County

Out-of-sample Rating Games

To demonstrate the economic importance of the proposed method, we conduct a repeated out-of-sample rating game to compare the performance of the individual method and the stacking methods (**?**). The rating game is repeated for the recent 20 years ($t = 2001, ..., 2020$). First, we calculate the log scores in the years 1990-2020 using the method described in the previous section. For each t in the rating game, the optimal weights are calculated by maximizing the summation of the log score within each pool from all previous years.

Then, we estimate and compare the premium rate in year t using the individual predictive distribution as well as the weighted average of the predictive distributions. We assume government uses the individual premium rate $\hat{\pi}_G$ while private insurance company uses the stacking premium rate $\hat{\pi}_P$ (both in bushels per acre). If $\hat{\pi}_P$ is larger than $\hat{\pi}_G$ then the insurance company considers government underestimates the risk and cedes the contract. If $\hat{\pi}_P$ is smaller than $\hat{\pi}_G$ then the insurance company thinks government overestimates the risk and retains the contract. Using the actual yield y_i we can obtain the loss ratio of a set of policies Ω in that year:

$$\text{LossRatio}_\Omega = \frac{\sum_{i\in\Omega} \max(0, \hat{y}_{G_i} - y_i)}{\sum_{i\in\Omega} \hat{\pi}_G} \tag{19}$$

where $\hat{y}_{G_i}$ is the yield guarantee under 90% coverage level. It is equal to the expected yield times the coverage level, where the expected yield is the predicted yield calculated by the government. After the yearly loss ratios are calculated, two statistical tests are employed to examine the performance of the proposed methods.

To show the incremental gain from increasing the size of the candidate pool, rating games are further implemented by assigning the government a stacking premium rate obtained from a smaller candidate pool, while insurance companies use a stacking premium rate calculated from a larger pool. We demonstrate the gain under both homoscedasticity and heteroscedasticity assumptions in Table 1. All the rating games are conducted under a 90% coverage level. The loss ratios from the individual models are larger than the stacking models. With more counties added to the pool of stacking, the loss ratios decrease in general. The null hypothesis of Game 1 is that ceding or obtaining policies using the estimated premium rates is as good as ceding or obtaining the same number of policies randomly. The result shows all methods are significantly better than randomly choosing the policies.

Under the setup of the rating games, the insurance company could take advantage by always selecting ahead of the government (Ker, Tolhurst, and Liu, 2016). Therefore, we use Game 2 to nullify the advantage by letting government also select ahead of the insurance company. The metric is calculated as:

$$D = \frac{LR_p/LR_g}{LR_g^*/LR_p^*} \tag{20}$$

LR_p and LR_g are the loss ratio of the insurance company and government when the insurance company selects against the government. LR_p^* and LR_g^* are the loss ratios of the insurance company and government when they switch roles. Denote C^* as the number of $D \leq 1$ in the 20 years rating game. Under the null, $C^* \sim \mathrm{B}(20, 0.5)$, which indicates in terms of variance, the rates are equally efficient. Small p-values are obtained in the result of Table 1 indicating the stacking models are more efficient.

Table 2 on the other hand focus on the improvement of the heteroscedasticity adjustment. We found that without incorporating heteroscedasticity, even after stacking using

TABLE 1. Corn Rating Game Results: 90% Coverage Level

Method-State	Number of Counties	Retained by Private (%)	Loss Ratio Government	Loss Ratio Private	Game 1 p-value	Game 2 p-value
Homoscedasticity: Individual vs Within CRD						
Illinois	56	47.6	6.084	2.977	0.0000	0.0577
Iowa	75	46.2	6.931	2.388	0.0002	0.1316
Michigan	23	42.4	14.871	6.935	0.0000	0.1316
Minnesota	46	37.5	6.215	2.364	0.0000	0.0577
Missouri	12	68.3	4.683	2.220	0.0002	0.0207
Ohio	46	36.2	8.612	4.389	0.0000	0.0207
Wisconsin	42	40.7	7.165	5.356	0.0207	0.5881
Homoscedasticity: Within CRD vs Within State						
Illinois	56	41.2	2.874	2.274	0.0013	0.0013
Iowa	75	32.3	3.534	2.826	0.0059	0.2517
Michigan	23	34.8	4.498	3.096	0.0000	0.0059
Minnesota	46	15.3	2.328	2.062	0.0207	0.0577
Missouri	12	35.8	1.880	2.110	0.0059	0.1316
Ohio	46	24.6	3.255	2.517	0.0002	0.0577
Wisconsin	42	32.4	5.022	1.283	0.0000	0.1316
Heteroscedasticity: Individual vs Within CRD						
Illinois	56	74.6	0.729	0.396	0.0000	0.0207
Iowa	75	72.2	0.510	0.228	0.0000	0.0059
Michigan	23	69.1	0.388	0.256	0.0000	0.0059
Minnesota	46	75.2	0.248	0.138	0.0000	0.0577
Missouri	12	82.9	0.596	0.569	0.0002	0.0059
Ohio	46	75.8	0.282	0.296	0.0002	0.0577
Wisconsin	42	67.6	0.435	0.360	0.0059	0.0577
Heteroscedasticity: Within CRD vs Within State						
Illinois	56	66.5	0.849	0.648	0.0000	0.0013
Iowa	75	86.5	0.486	0.441	0.0013	0.0002
Michigan	23	63.0	1.085	0.537	0.0000	0.0059
Minnesota	46	70.7	0.415	0.352	0.0013	0.0059
Missouri	12	67.9	0.414	0.794	0.0013	0.0013
Ohio	46	77.6	0.666	0.534	0.0000	0.0002
Wisconsin	42	71.4	1.133	0.476	0.0059	0.2517

all counties within states, the loss ratios are still higher than individual GP models with heteroscedasticity adjustment. p-values from both Game 1 and Game 2 are small enough to indicate heteroscedasticity adjustment leads to a better and more efficient method.

TABLE 2. Corn Rating Game Results: 90% Coverage Level

Method-State	Number of Counties	Retained by Private (%)	Loss Ratio Government	Loss Ratio Private	Game 1 p-value	Game 2 p-value
Homoscedasticity Within State vs Heteroscedasticity Individual						
Illinois	56	29.6	0.589	0.422	0.0013	0.0577
Iowa	75	35.1	0.381	0.331	0.0000	0.0577
Michigan	23	50.0	0.482	0.213	0.0002	0.0000
Minnesota	46	57.8	0.293	0.128	0.0002	0.0002
Missouri	12	42.9	0.789	0.431	0.0207	0.0059
Ohio	46	58.7	0.412	0.216	0.0059	0.0577
Wisconsin	42	35.8	0.516	0.365	0.0002	0.0013

Lastly, Table 3 compares the stacking model under heteroscedasticity adjustment using all counties within states as the candidate pool with the conventional RMA rating methodology. All states except Iowa are having a smaller average loss ratio under the proposed method. Game 1 and Game 2 are both having small p-values for every state in the data set. Compared to the current RMA approach, our proposed method is generating more accurate premium rates.

TABLE 3. Corn Rating Game Results: 90% Coverage Level

Method-State	Number of Counties	Retained by Private (%)	Loss Ratio Government	Loss Ratio Private	Game 1 p-value	Game 2 p-value
Heteroscedasticity: RMA vs Within State						
Illinois	56	75.7	0.679	0.485	0.0000	0.0207
Iowa	75	85.1	0.128	0.332	0.0000	0.0207
Michigan	23	72.6	0.670	0.315	0.0059	0.0059
Minnesota	46	79.8	0.227	0.205	0.0002	0.1316
Missouri	12	78.8	0.723	0.627	0.0002	0.0013
Ohio	46	65.4	0.649	0.581	0.0059	0.1316
Wisconsin	42	81.8	0.650	0.467	0.0013	0.0207

Conclusions

In this study, we combine two machine learning techniques: Gaussian Process and stacking to predict county-level crop yield distribution. Gaussian Process merges the conventional two-step approach into one and is capable of capturing the nonlinear trends and time-varying distributions of crop yields accurately. Stacking on the other hand pools similar counties together to allow "borrowing" information among counties. By using a mixture model of normals, the stacking model can capture non-normal or asymmetric distributions, which may be more appropriate for some crop yield density data that exhibit complex and diverse patterns across different counties. The proposed methodology has shown to be promising in obtaining more accurate premium rates by incorporating model uncertainty, dealing with situations where the true model is not in the candidate pool, and focusing more on predictive performance compared to existing approaches.

By performing a repeated out-of-sample rating game, we demonstrate the economic importance of our proposed method. The stacking models outperform individual models by showing incremental gains as the candidate pool size increases. There is also a great improvement in rating game results using GP models with heteroscedasticity adjustment. Furthermore, after incorporating heteroscedasticity, stacking using all counties within states outperforms the traditional RMA method by generating more accurate premium rates. The crop insurance policymakers may benefit from the proposed methodology with a smaller loss ratio compared to the current RMA method. The proposed method can be further enhanced by using the leave-one-out method for calculating mixing weights when the data are cross-sectional, and more GP kernels can be considered to capture different temporal effects. The proposed framework has the potential to contribute not only to yield forecasting and crop insurance rating but also to more general applied works where the focus is distributional prediction. Additionally, the proposed method can be easily adapted to a big data setup.

References

Annan, F., J. Tack, A. Harri, and K. Coble. 2013. "Spatial Pattern of Yield Distributions: Implications for Crop Insurance." *American Journal of Agricultural Economics* 96:253–268.

Atwood, J., S. Shaik, and M. Watts. 2003. "Are Crop Yields Normally Distributed? A Reexamination." *American Journal of Agricultural Economics* 85:888–901.

Botts, R.R., and J.N. Boles. 1958. "Use of Normal-Curve Theory in Crop Insurance Ratemaking." *Journal of Farm Economics* 40:733–740.

Du, X., C.L. Yu, D.A. Hennessy, and R. Miao. 2015. "Geography of crop yield skewness." *Agricultural Economics* 46:463–473.

Gallagher, P. 1987. "U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances." *American Journal of Agricultural Economics* 69:796–803.

Geweke, J., and G. Amisano. 2011. "Optimal prediction pools." *Journal of Econometrics* 164:130–141.

Goodwin, B.K., and A.P. Ker. 1998. "Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts." *American Journal of Agricultural Economics* 80:139–153.

Goodwin, B.K., M.C. Roberts, and K.H. Coble. 2000. "Measurement of Price Risk in Revenue Insurance: Implications of Distributional Assumptions." *Journal of Agricultural and Resource Economics* 25:195–214.

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* 93:707–717.

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

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

Ker, A.P., and T.N. Tolhurst. 2019. "On the Treatment of Heteroscedasticity in Crop Yield Data." *American Journal of Agricultural Economics* 101:1247–1261.

Ker, A.P., T.N. Tolhurst, and Y. Liu. 2016. "Bayesian Estimation of Possibly Similar Yield Densities: Implications for Rating Crop Insurance Contracts." *American Journal of Agricultural Economics* 98:360–382.

Liu, Y., and A.P. Ker. 2020a. "Rating Crop Insurance Contracts with Nonparametric Bayesian Model Averaging." *Journal of Agricultural and Resource Economics* 45:244–264.

—. 2020b. "Simultaneous borrowing of information across space and time for pricing insurance contracts: An application to rating crop insurance policies." *Journal of Risk and Insurance* 88:231–257.

Moss, C.B., and J.S. Shonkwiler. 1993. "Estimating Yield Distributions with a Stochastic Trend and Nonnormal Errors." *American Journal of Agricultural Economics* 75:1056–1062.

Nelson, C.H., and P.V. Preckel. 1989. "The Conditional Beta Distribution as a Stochastic Production Function." *American Journal of Agricultural Economics* 71:370–378.

Tack, J., A. Harri, and K. Coble. 2012. "More than Mean Effects: Modeling the Effect of Climate on the Higher Order Moments of Crop Yields." *American Journal of Agricultural Economics* 94:1037–1054.

Tolhurst, T.N., and A.P. Ker. 2015. "On Technological Change in Crop Yields." *American Journal of Agricultural Economics* 97:137–158.

U.S. Department of Agriculture Office of Inspector General. 2022. "Federal Crop Insurance Corporation/Risk Management Agency's Financial Statements for Fiscal Years 2022 and 2021." Report No. 05401-0014-11, U.S. Department of Agriculture.

Wen, K., X. Wu, and D.J. Leatham. 2021. "Spatially Smoothed Kernel Densities with Application to Crop Yield Distributions." *Journal of Agricultural, Biological and Environmental Statistics* 26:349–366.

Williams, C.K., and C.E. Rasmussen. 2006. *Gaussian processes for machine learning*. MIT press Cambridge, MA.

Woodard, J.D., and B.J. Sherrick. 2011. "Estimation of Mixture Models using Cross-Validation Optimization: Implications for Crop Yield Distribution Modeling." *American Journal of Agricultural Economics* 93:968–982.

Wu, W., X. Wu, Y.Y. Zhang, and D. Leatham. 2021. "Gaussian process modeling of nonstationary crop yield distributions with applications to crop insurance." *Agricultural Finance Review* 81:767–783.

Wu, X., and Y.Y. Zhang. 2020. "A local maximum likelihood model of crop yield distributions." *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie* 68:117–125.

—. 2012. "Nonparametric estimation of crop yield distributions: A panel data approach." Working paper.

Yao, Y., A. Vehtari, D. Simpson, and A. Gelman. 2018. "Using Stacking to Average Bayesian Predictive Distributions (with Discussion)." *Bayesian Analysis* 13.

Zhang, Y.Y. 2017. "A Density-Ratio Model of Crop Yield Distributions." *American Journal of Agricultural Economics* 99:1327–1343.