

# Defining and Characterizing Aflatoxin Contamination Risk Areas for Corn in Georgia, USA: Adjusting for Collinearity and Spatial Correlation

EunHye Yoo, Ruth Kerry, Ben Ingram, Brenda Ortiz, Brian Scully

---

## Abstract

Aflatoxin is a carcinogenic toxin to humans and animals produced by mold fungi in staple crops. It is expensive to measure and is usually measured at harvest leaving little scope for implementing mitigation strategies within season. If high and low risk areas and year can be identified, then applying such strategies should be possible and the cost of measuring levels could be reduced by targeting the high risk years and counties. Aflatoxin contamination is driven by drought conditions. Kerry et al. (2017) developed a risk factors approach for determining county level Aflatoxin contamination risk which they assessed using Aflatoxin data for 1977-2004. The approach was not able to combine Aflatoxin and risk factor data in the same analysis so a regression approach was proposed. As the risk factor data are highly correlated amongst themselves, the Aflatoxin data approach a Poisson distribution and exhibit spatial autocorrelation, traditional multiple regression approaches would be flawed. In this paper, spatial Poisson profile regression is applied to Aflatoxin and risk factor data including remotely sensed data for the entire study period and high risk years. The approach identifies clusters of counties which have similar Aflatoxin risk and risk factor profiles whilst taking into account multicollinearity in the risk factors and spatial autocorrelation in the Aflatoxin data. The results identify plausible clusters compared to previous work but also give the important risk factors associated with those clusters. The approach also helps show that some factors like well drained soil behaves differently from expectations and irrigation is not useful.

*Keywords:* Spatial Autocorrelation, Profile Regression, Aflatoxin Contamination Risk, Risk Factors

---

## 1. Introduction

Aflatoxin is a carcinogenic toxin produced by the mold fungi *Aspergillus flavus* and *Aspergillus parasiticus*. As these fungi can infect crops that are some of the dominant staples in different parts of the world, infection is considered a worldwide problem (Brenneman et al., 1993; Liu and Wu, 2010; Wang et al., 2010). Ingestion of infected grain/food can cause esophageal and liver cancer in humans, livestock and poultry (Ghasemi-Kebria et al., 2013; Liu and Wu, 2010). Consequently, the U.S. Food and Drug Administration (FDA) has set a limit of 20 ppb, total Aflatoxin, for use of corn, peanut, cottonseed meal, and other animal feeds/feed ingredients intended for animal consumption, particularly by immature animals. There is also a limit of 100 ppb restricting use of corn and peanut products intended for breeding beef cattle, swine, or mature poultry (FDA, 2015). However, in less developed countries where such administrative standards do not exist, there are 16 to 31 times more deaths from liver cancer and this has been attributed, at least in part, to Aflatoxin contamination of food (Liu and Wu, 2010).

Even in areas where governmental standards for maximum contamination levels exist, like the USA, food scares still occur (Garland and Reagor, 2001; Newman et al., 2007) because testing is limited as it is time-consuming, expensive and requires several grain samples (Papadoyannis et al., 1990). There is also the problem that different methods of determining Aflatoxin have differing accuracies, detection limits and advantages and disadvantages (Wacoo et al., 2014). These challenges of accurate Aflatoxin assessment make regular monitoring infeasible with current technology, which results in the scarce availability of data in the study region, southern Georgia (GA), as well as inconsistent data availability over the study period.

An important way to reduce the risk of contaminated food going undetected whilst also keeping the cost of Aflatoxin measurement to a minimum, is to identify years and areas with different levels of Aflatoxin contamination risk. This means that less samples would need to be collected in low risk years and areas and more in high risk years and areas. Also, as Aflatoxin levels are usually determined at harvest, no in-season mitigation strategies are possible. However, if years and areas with different levels of risk are identified, strategies that prevent, reduce or manage Aflatoxin levels in crops can be employed at key periods in the growing season. During planting, the seeding rate can be varied between different risk zones, or more resistant corn hybrids can be planted in the high risk areas. During crop growth, irriga-

tion, pest and nutrient management strategies could be altered between the different risk zones. Finally, at harvest, areas with different levels of risk can be harvested separately.

Previous studies have shown that infection of corn by *A. flavus* or *A. parasiticus* is linked with high temperatures, drought and high net evaporation (Guo et al., 2008; Horn et al., 2014; Payne and Widstrom, 1992) so it is associated with particular climatic areas (Abbas et al., 2007) and soil types (Palumbo et al., 2010). In the southern states of the USA, summer crop corn is highly susceptible to Aflatoxin contamination (Widstrom et al., 1996) due to high temperatures and rainfall variability, along with light textured soils that compound crop water stress. Lack of irrigation infrastructure in some areas also means that crop water stress cannot be easily relieved (Brenneman et al., 1993). Using logistic regression, Salvacion et al. (2011) showed that in southern GA, Aflatoxin risk level changed based on deviations of June maximum temperature (TMax) and rainfall (RF) levels from climatic normals. Damianidis et al. (2015) also found that the risk of contamination changes with corn hybrid, soil type and the weather conditions before and after the mid-silk growth stage (usually in June in south eastern USA). In addition, when the Agricultural Reference Index for Drought (ARID) (Fraisie et al., 2006) was included in determining Aflatoxin risk, a 0.1 increase of in-field drought, as quantified by ARID, around the mid-silk period, increased the probability of Aflatoxin exceeding the 20 ppb FDA threshold (Damianidis et al., 2015).

Recently, Kerry et al. (2017) used a risk factors approach to identify years and areas at high risk of Aflatoxin contamination. They used an additive indicator approach based on key thresholds in variables that previous research (Abbas et al., 2007; Brenneman et al., 1993; Palumbo et al., 2010; Salvacion et al., 2011; Widstrom et al., 1996) had identified as important to Aflatoxin contamination risk. This approach, however, had the limiting assumption that each variable has equal weight in determining overall risk. Here, we consider regression models as an alternative methodology to assess the relative importance of different risk factors (explanatory variables or co-variates) for predicting Aflatoxin levels (response variable). While climate information available from weather stations are clearly helpful in determining drought prone areas and years, it is only through the use of interpolated data that information on drought can be extracted for areas that are not close to weather stations.

One way to gain more complete spatial and temporal information on

actual drought stress in crops is from remotely sensed imagery. Normalized difference vegetation index (NDVI) data from images indicate the degree of crop greenness and have been used to indicate variations in crop health and yield within fields in response to differences in water supply and nutrients (Hatfield and Prueger, 2010; Wang et al., 2016). Thermal InfraRed (IR) data have also been used to indicate drought stress (Jones et al., 2009; Sepúlveda-Reyes et al., 2016). Navarro et al. (2017) found for one year (2011), with a high risk of Aflatoxin contamination, an 82% agreement in the cells that were identified as high and low risk when NDVI and Thermal IR were used as opposed to rainfall and maximum temperature data. It is important to further determine if these variables can be used to indicate drought stress in other years and thus, Aflatoxin contamination risk level can be identified within growing seasons and at a finer spatial scale. Thus, our goals include determining if NDVI and Thermal IR data from remotely sensed imagery are useful predictors of Aflatoxin levels and could therefore be used to determine risk at finer scales.

A challenge associated with the representation of the risk of Aflatoxin contamination levels and building a regression model is that the Aflatoxin measurements were obtained from corn grain samples, but their specific location other than the county of origin is not known, so they are essentially aggregated data represented at the county level. Other information, such as weather, soil, or remotely sensed data are available for different representations, such as point locations for weather stations, polygon maps of soil type and raster data from imagery but all were aggregated to the county level. Regardless of the unit of analysis, aggregating point data to areas is likely to induce spatial autocorrelation. Also, the very fact that Kerry et al. (2017) were able to fit a reliable Poisson variogram to the Aflatoxin data and typical variograms to the weather data shows that spatial autocorrelation is present in these data. The presence of spatial autocorrelation in standard regression models needs to be accounted for, otherwise the uncertainty intervals of errors will be under-estimated.

The final challenge with applying a multiple regression approach to these data is that the known risk factors for Aflatoxin contamination in corn crops are highly correlated causing problems of multi-collinearity. For example, the correlation of irrigation and excessively drained soil class is as high as 0.60 and the two remotely sensed variables (NDVI and Thermal IR) were negatively correlated (-0.71) with one another. The recent development of spatial Profile regression (Molitor et al., 2010, 2011; Papathomas et al., 2011)

appears to be a promising alternative to address both the presence of spatial autocorrelation in Aflatoxin contamination risk and collinearity among the explanatory variables. Profile regression is a Bayesian statistical approach that assesses the link between potentially collinear variables and a response variable through cluster membership (Liverani et al., 2016). Its application in environmental and social epidemiology (Hastie et al., 2013; Coker et al., 2016) and infectious disease studies (Shekhar et al., 2017) has demonstrated its potential to be useful in the definition and characterization of Aflatoxin contamination risk areas.

In this paper, we aim to apply Bayesian spatial profile regression, which enables investigators to identify areas with different levels of Aflatoxin contamination risk based on both Aflatoxin measurements and risk factors, while accounting for problems of collinearity among risk factors and the spatially correlated structure present in Aflatoxin risk. We will perform profile regression for two time frames, high risk years identified by Kerry et al. (2017) and the entire study period, respectively. The latter allows comparison with the high risk regions of Aflatoxin contamination derived from Poisson Kriging (Kerry et al., 2017) purely based on the primary data from corn plant samples. Lastly, we estimated the uncertainty associated with contamination risk for each cluster using both Aflatoxin measurements and risk factors within a Bayesian framework by deriving the posterior distribution of random effects for each cluster.

## 2. Methods

### 2.1. Data Sources/Collection

The study area consists of 53 counties in southern GA, USA. County level Aflatoxin data were collected between 1977 and 2004, although data were not collected in every year. Specifically, data were collected in 1977-1978, 1981 and 1990-2004. Figure 1 shows the number of years that Aflatoxin data was collected for each county. Data were collected for most counties (45) in 1978 and least (23) in 1990. Corn grain samples were collected at harvest using a grab sampling technique where 10 ears were collected for each sampling and there was an average of 3 replications per county. Aflatoxin levels in ppb were measured by thin layer chromatography until the late 1990s and the most recent Aflatoxin measurements were made with the VICAM AflaTest (Wacoo et al., 2014). These measurements were made by the USDA-ARS Crop Protection and Management Research Unit and the University of Georgia,

Natural Products Laboratory in Tifton, GA. Some uncertainty about Aflatoxin levels is obviously introduced by the difference in measurement methods used during the study period.

[FIGURE 1]

Monthly weather data were obtained for each year 1977-2004 in the study period from the Georgia Weather Network (<http://georgiaweather.net>). The locations of weather stations (82 in total) are shown as black dots in Figure 1. It should be noted that weather data were not available for all counties in all years because not all counties have a weather station, but some have more than one with varying installation dates. In addition, the weather stations are not located at the center of the county. Therefore, monthly maximum temperatures for June (June TMax, °C) and June rainfall data (June RF, mm) (Salvacion et al., 2011), were kriged from the recording stations to counties.

The area (hectares) planted with corn per county and the proportion (%) of each county planted in corn were determined using the CropScape Cropland data layer produced by the National Agricultural Statistics Service (NASS, <http://nassgeodata.gmu.edu/CropScape>). It was also used to determine each 30m pixel in the study area where corn was grown so that soil and remotely sensed data could be isolated just for the corn growing area in each county for a given year. As the CropScape data layer was not available for the study period between 1977 and 2004, the proportion of land in each county planted as corn had to be determined from the 2008/2009 growing season which was the first growing season with full coverage in southern GA. Using the CropScape data for this year assumes that the corn growing regions in the study area have not changed markedly within or since the study period although it is known that crop rotation management strategies are used. The area of corn grown in each county each year (with exact locations unknown), was available from agricultural census data using the quick stats tool of USDA-NASS ([https://www.nass.usda.gov/Statistics\\_by\\_State/Georgia/Publications/County\\_Estimates/2016/GACorn14\\_15.pdf](https://www.nass.usda.gov/Statistics_by_State/Georgia/Publications/County_Estimates/2016/GACorn14_15.pdf)). There were strong, positive and significant ( $p < 0.05$ ) correlations in the county-specific corn growing areas between all years in the study period and 2008. This means that the counties growing the most corn have remained quite con-

sistent throughout the study period and that our assumption above about the CropScape data is reasonable. However, it was also found that the larger the temporal gap between a specific year and 2008, the lower the correlation coefficient, which indicates that there will be greater uncertainty in the corn planted area data for 1977.

A 1:250,000 geo-corrected map of soil associations (NRCS, 2006) was simplified and used to generate a map with 3 drainage classes: excessively, well and poorly drained soil. For the corn-growing pixels identified for each county using the 2008/2009 CropScape data layer mentioned above, the percentage of the area with soil in each drainage class was calculated. There will be some obvious uncertainty in this data related to mis-identification of corn-growing pixels for a given growing season. The uncertainty is likely to be greatest for the earlier study years and least for more recent years.

Data on the acreage of irrigated, harvested cropland on farms per county were available from an agricultural census (<http://agcensus.mannlib.cornell.edu/AgCensus/homepage.do>) at 5 yearly intervals during the study period (1978, 1982, 1987, 1992, 1997 and 2002). We calculated the proportion of irrigated cropland per county from the agricultural census data.

Landsat Thematic Mapper (TM) data for southern GA from the second half of June were collected for most years (1990-2004) in the study period from <https://earthexplorer.usgs.gov>. The only years for which remotely-sensed data were not collected were 1977, 1978 and 1981 as the pixel sizes and wavebands measured for these years were not consistent with other years and there was no quality assessment band available. Imagery were collected for the least cloud-covered collection day in the latter half of June. If, however, there was more than 50% cloud cover in all imagery available for the latter half of June, then an image from the first week in July was used. Images were geo-corrected and stitched together to create a large image covering southern GA. The NDVI was calculated and Thermal IR values extracted for each corn-growing pixel as indicated in the 2008/2009 CropScape data layer. Some corn-growing pixels were discarded due to false-positives in the CropScape data layer. To determine which falsely-labeled corn pixels to discard, it was assumed that the minimum size of a corn field should be at least 9 connected pixels which translates to a minimum field size of approximately 2 acres. Furthermore, the quality assessment band was used to discard cloud-covered pixels. Due to noise in the data based on just the 2008/2009 CropScape data layer it was decided to create a composite data layer of all pixels where corn was grown in the 2008-2016 period. With

this composite corn data layer, NDVI and Thermal IR pixel values were extracted for all locations where corn had been grown one or more times in the year range 2008-2016. This should help account, at least partially, for the fact that crop rotations exist. Finally, average NDVI and Thermal IR values of corn-growing (2008 to 2016) and cloud free pixels were calculated for each county.

## 2.2. Statistical Analyses

We characterized the Aflatoxin risk level for each county based on FDA thresholds of 20 ppb and 100 ppb, respectively, and performed statistical analyses for the two time frames. The risk of Aflatoxin at each county, denoted as  $Y_i^c$ , was quantified by the number of years with Aflatoxin measurements ( $Z_i$ ) greater than or equal to  $c$  thresholds ( $c = 20$  ppb or 100 ppb) as  $Y_i^c = \sum_{t=1}^{T_i} I(Z_i > c), i = 1, \dots, 53$ . Here  $T_i$  denotes the total number of years that Aflatoxin measurements were available at the  $i$ th county. The response variable was defined for each county over the entire study period (1977 to 2004) and within the high risk years (1977, 1981, 1998) identified by Kerry et al. (2017).

Given that the measured Aflatoxin values in southern GA between 1977 and 2004 approach a Poisson distribution, we assumed that the number of years with Aflatoxin measurements above the two FDA thresholds follows a Poisson distribution. As a base case, we fit a Poisson regression model for the observed Aflatoxin data  $Y_i^c$  with known risk factors by the logarithmic function. The county-specific risk factors are mixed with both continuous and dichotomous variables relevant for each time frame. More specifically, the continuous risk factors include the weather conditions, June TMax and June RF for each county per year and the proportion of years that June TMax is greater than 33 °C and June RF is less than 50 mm. Soil conditions are also included as percentage of the areas with different drainage levels (excessively and well drained), as well as the percent of areas that grows corns, the percentage of irrigated cropland, and the average NDVI and Thermal IR per county. Three dichotomous variables are derived from the percentage of the area with soil drainage types and the percentage of the area growing corn according to thresholds, which are given in Table 1. These thresholds for indicators were determined with regard to tails of the distribution in histograms.



[TABLE 1]

Given that both the continuous and dichotomous variables were derived from the same sources so the explanatory variables are likely to be collinear. This multi-collinearity in the regression affects the inference of the risk factors on the outcome making it less precise and resulting in a large standard error. In addition, the association between risk factors and the Aflatoxin contamination risk is likely to vary in the study area. The standard regression is likely to fail to capture non-stationary interaction between risk factors and Aflatoxin contamination risk. As an alternative to regression models, we developed profile clusters using a non-parametric dimension reduction technique, called Bayesian profile regression (Molitor et al., 2010; Coker et al., 2016). This allows us to investigate the joint effects of multiple risk factors on Aflatoxin contamination risk. Under the profile regression approach, both the response variable  $Y_i^c$  and a set of explanatory variables  $\mathbf{X}_i^c = (X_{i1}^c, \dots, X_{ip}^c, \dots, X_{i13}^c)$  are simultaneously considered to determine clusters following the joint probability model as

$$p(Y_i^c, \mathbf{X}_i^c | \theta_i^c = k) = \Theta_k \sum_{k=1}^{\infty} p(Y_i^c | \theta_i^c = k) p(\mathbf{X}_i^c | \theta_i^c = k) \quad (1)$$

where  $\theta_i^c = k$  denotes the allocation variable for the  $i$ th county with respect to the FDA threshold  $c$  belongs to cluster  $k$ , and  $\Theta_k$  represents the weight of the mixture component  $k$ . In the present paper, we used the stick-breaking priors for the mixture weights, that are constructive definitions of the Dirichlet process (Liverani et al., 2016). This prior implies that we believe that conditioning on the observed data for response and explanatory variables, the 53 counties have the same probability of being assigned to a particular cluster  $k$ . The covariate model  $p(\mathbf{X}_i^c | \theta_i^c = k)$  is defined as a mixture of 10 continuous and 3 discrete covariates. For the response model  $p(Y_i^c | \theta_i^c = k)$ , we specified the log-transformed mean Aflatoxin risk  $\log(\rho_i^c)$  as a linear regression model with known risk factors at each county, denoted by  $\mathbf{x}_i = [x_{i1}, \dots, x_{i13}]$  and their effects  $\boldsymbol{\beta}$  on the Aflatoxin contamination, as

$$\begin{aligned} Y_i^c &\sim \text{Poisson}(E_i \rho_i^c) \\ \log(\rho_i^c) &= \beta_0 + \mathbf{x}_i^T \boldsymbol{\beta} + S_i^c + \epsilon_i^c \end{aligned} \quad (2)$$

where  $E_i$  denotes the number of years of observations of Aflatoxin samples at each county.  $S_i^c$  and  $\epsilon_i^c$  denote the random effects for the spatially structured

residual error and independent residual error, respectively. The independent residual error term follows a zero mean Gaussian process as  $\epsilon_i^c \sim N(0, \sigma^2)$ , but the former is defined as a realization of a Gaussian Markov Random Field (Besag, 1974) to accommodate the spatial autocorrelation present in the response variable. More specifically, we include a spatially structured error model in the spatial profile regression using the conditional autoregressive term based on neighborhood specification (Liverani et al., 2016) and defined a county-specific relative risk as a function of the Aflatoxin risks of neighboring counties as

$$S_i^c \sim N\left(\frac{\sum_{i \neq j} w_{ij} S_j^c}{\sum_{i \neq j} w_{ij}}, \frac{\tau^2}{\sum_{i \neq j} w_{ij}}\right), \quad (3)$$

where the weights  $w_{ij}$  are elements of the zero-one neighborhood adjacency matrix defined to be equal to one when counties  $i$  and  $j$  are adjacent and zero otherwise. This approach implements the Besag-York-Molly (BYM) model (Besag et al., 1991). The standard deviation of the spatially structured error is denoted as  $\tau$ .

Bayesian profile regression uses covariate values to observe joint patterns within the data (Coker et al., 2016), which reduces the dimensionality of the covariates. This feature allows investigation of the Aflatoxin risk as it relates to the joint patterns of multiple risk factors, while accounting for their multicollinearity. Unlike other clustering algorithms, such as K-means, Bayesian profile regression does not call for the target number of clusters and it provides greater flexibility than other similar clustering methods. For a detailed discussion of Bayesian profile regression, we refer readers to recent works (Molitor et al., 2011; Papathomas et al., 2011).

We investigated the joint effects of collinear environmental conditions, while accounting for the spatially correlated structure in Aflatoxin risk levels. Uncertainty associated with the clustering was also incorporated through the Markov Chain Monte Carlo (MCMC) iteration and its subsequent post-processing (90,000 iterations and 10,000 burn-in period). The normal score transformation was applied for continuous risk factor variables for profile regression.

Based on the MCMC results, we examine the posterior distribution of risk factor profiles associated with each cluster and the spatial patterns of cluster-specific relative risk of Aflatoxin contamination. The Bayesian profile regression analysis was conducted using the PReMiuM package (v.3.1.4) in R (v. 3.3.2).

### 3. Results

The measured Aflatoxin levels for corn in the southern counties of GA had a mean of 60.18 ppb (standard deviation of 111.07) for the entire study period and a mean of 256.40 ppb (standard deviation of 460.98) when only the high risk years were considered. The differences in mean Aflatoxin levels between the two time frames are substantial and require two separate profile regressions. Our data also showed spatial autocorrelation at the county level with respect to the risk of Aflatoxin contamination as demonstrated by Kerry et al. (2017). This makes it necessary to explicitly model the spatially dependent structure in the profile regression.

#### 3.1. Risk Factor Correlation

As indicated in Figure 2, there is evidence of strong correlation among known risk factors for Aflatoxin contamination. The two variables from remotely sensed images, the NDVI and Thermal IR, show strong negative correlation (-0.71) with one another. As expected, strong positive correlation is found between indicator variables and original variables from which indicator variables were defined, such as well drained (0.71) and excessively drained (0.58) soil classes, and the corn production area (0.73). Similarly, the proportion variables derived from June TMax and June RF have strong positive correlations of 0.77 and 0.87, respectively with their original variables. It is also worth noting that the correlation between irrigation and excessively drained soil class is as high as 0.60 and the correlation between Thermal IR and June RF was -0.51.

[FIGURE 2]

#### 3.2. Poisson Regression

To allow proper assessment of the performance of profile regression, we first performed a Poisson linear regression with the full list of covariates. The results are summarized in Table 2. Based on AIC values, overall the model fit was better for the contamination risk defined with the 100 ppb threshold rather than the 20 ppb threshold and for high risk years rather than the entire study period. None of the indicator variables are statistically

significant. Due to missing values in observed covariates, only 48 counties were considered in all four Poisson regression models.

The model for contamination risk with the 20 ppb threshold for the high risk years has no statistically significant variables. For the entire study period, Thermal IR has a positive effect on Aflatoxin contamination risk with statistical significance regardless of the threshold value (0.32 for 20 ppb and 0.50 for 100 ppb). The proportion of June rainfall has a statistically significant negative effect (-2.14) on the contamination risk for the entire study period with the 20 ppb threshold. For the high risk years, only June TMax has a statistically significant positive effect (0.99) on the contamination risk with the 100 ppb threshold. When effects were identified they were positive or negative as one would expect given the theories about what conditions drive contamination. The only exception to this was the indicator for well-drained soils and the continuous variable for well-drained soil for the entire study which are negative when one would expect positive coefficients as well-drained soil should be relatively drought, and thus, Aflatoxin prone.

[TABLE 2]

### 3.3. *Optimal Profile Clusters*

Since we are interested in the joint distribution of risk factors and their impacts on Aflatoxin contamination risk, we examined the optimal clustering obtained from spatial profile Bayesian models with the number of years whose Aflatoxin level was greater than an FDA threshold of 20 ppb or 100 ppb. The clusters that best fit the data were identified for two different time frames, the entire study period and high risk years. As summarized in Table 3, the Bayesian profile regression identified two clusters for the threshold value of 20 ppb and two clusters for 100 ppb for both the high risk years and the entire study period.

[TABLE 3]

The spatial distribution of the optimal partition of counties for the entire

study period and the high risk years is visualized in Figures 3 and 4, respectively. The three maps in Figures 3(b), 4(a), and 4(b) show similar patterns in terms of which counties are assigned to the higher risk cluster (cluster 2), and these tend to be located in the south west of the study region. Particularly, the two maps for high risk years in Figures 4(a) and 4(b) are almost identical despite the 100 ppb threshold having far lower estimated relative risks. In contrast, Figure 3(a) shows the highest risk counties (cluster 2) being found in a north to south line in the centre of the study region. This pattern is similar to the finding of Kerry et al. (2017) with Poisson kriging of Aflatoxin data and kriging of the number of years that both June TMax and June RF breached thresholds relating to drought.

[FIGURE 3]

The two box-plots in Figures 3(c) and 3(d) show the posterior distribution of the probabilities of the response (“relative risk” of Aflatoxin contamination) estimated for the entire study period. The two hinges of each box plot correspond to the first and third quartiles of estimated risks for each cluster. Outliers beyond the first and third quartiles are denoted by points. The relative risk estimates per cluster suggest that the differences between clusters are not substantial for the high risk years but are quite marked for the entire study period. As depicted in Table 3, the 95 % posterior credible intervals of clusters for the entire study period are narrower (mean of 0.125) than those for the high risk years (mean of 0.37) regardless of their threshold values, which suggests that there is less uncertainty for the optimal clusters formed for the entire study period. This makes sense given that there are more years of data for the entire study period.

The spatial patterns of the optimal partition of counties for the high risk years are shown in Figures 4(a) and 4(b). The two optimal groups identified for 20 ppb and 100 ppb thresholds are almost identical to each other in terms of their spatial patterns. However, the mean of estimated relative risk associated with the 20 ppb threshold (0.73 and 0.75 for cluster 1 and 2) is more than double those defined with the 100 ppb threshold (0.31 and 0.35). For both thresholds, the differences between the two clusters are relatively small during high risk years compared to the entire study period. Lastly the uncertainty associated with the clusters is greater for high risk years for both

thresholds as the range of the 95% credible intervals for clusters 1 and 2 for the 20 ppb threshold is 0.12 and 0.21 for the entire study period and 0.39 and 0.60 for the high risk years, respectively. Only 52 counties have sample data for the high risk years unlike the entire study period where a total of 53 counties were available for the analysis.

For both time frames, the relative risk for each cluster with the 100 ppb threshold is smaller than for the 20 ppb threshold which is expected given that breaches of the 100 ppb threshold are a less common occurrence. Overall risk, however, of Aflatoxin contamination is substantially elevated in the high risk period regardless of FDA threshold value (see Table 3). It is worth noting that Bayesian profile regression makes inference using MCMC and the results may depend on the chain lengths and initial values (Liverani et al., 2016). To assure the stability of our findings, we performed a sensitivity analysis by conducting a number of trials with different random seed numbers (over 100 different values) and different combinations of iterations and burning periods. We noticed that the clusters for the entire study period with the 20 ppb threshold were the most sensitive among the four cases considered in the present study. We examined the spatial patterns of clusters obtained from multiple runs. Each run yielded slightly different results, while the spatial patterns of clusters reported above are rather consistent and dominant (51 %). We found that for the entire study period with the 20 ppb threshold, some clusters contained only one county (39 %) and other runs had three clusters (10 %). We discarded the latter cases and reported only case for two clusters in the present study. We also found that the results are more sensitive to the initial value (random seed number) than the chain lengths.

[FIGURE 4]

### 3.4. Risk Factor Profiles

Figures 5 and 6 show the risk factor profiles of the two clusters for the entire study period with the 20 ppb and 100 ppb thresholds, respectively. Similarly, the results for the high risk years are presented in Figures 7 and 8. They represent the cluster-specific probabilities for the three binary risk factors (see Table 1 for their definitions) and the posterior mean vectors of ten risk factor profiles for the continuous variables. The hinges of each box

plot correspond to the first and third quartiles of the posterior distribution of each parameter estimate and the black two dots outside the box in each cluster represent the lower and upper levels of the 95% credible intervals, respectively. The means across clusters (overall mean) are denoted by a horizontal line.

[FIGURE 5]

[FIGURE 6]

The dominant pattern in Figures 5, 6 and 8 is that the high risk cluster (cluster 2) is characterized by higher values than the overall mean for Thermal IR, June TMax and the proportion of June TMax and lower values for NDVI, June RF, the proportion of June RF, the corn growing area and its indicator. While these are the dominant patterns, there is still, in some cases, overlap in the distribution of risk values shown in the box plots between the two hinges for the two clusters. In contrast, the patterns for well drained and excessively drained soil classes and their indicators and for irrigation are inconsistent between clusters or cannot be distinguished as higher or lower in the two clusters. For Figure 7, patterns that are the reverse of those in the other figures are shown.

As shown in Figure 5 for the entire time period with the 20 ppb threshold, the risk factor profiles for the two clusters are most different for corn-growing area and its indicator, June TMax and the proportion of June TMax and for Thermal IR. The difference in the Thermal IR between the two clusters is consistent with the Poisson regression result (a statistically significant positive coefficient estimate for Thermal IR; see Table 2). While June RF, the proportion of June RF, NDVI and Irrigation show the expected patterns with on average lower values for the high risk cluster (cluster 2), the difference is less marked and sometimes only very slight. The well and excessively drained soil classes and their indicators show either very little difference between clusters or patterns that are the reverse of what one would expect. One would expect larger amounts of excessively and well drained soils in the high risk cluster.

The high risk cluster for the 100 ppb threshold in Figure 6 is characterized

by difference in all the variables in terms of the boxes for each cluster being found either side of the overall mean. The only variables where there is some overlap between the boxes for the two clusters are irrigation, excessively-drained soil class and June RF. For the latter two variables, however, the associated indicators and the proportion of June RF did not show overlap and showed the same pattern of values as one might expect with cluster 2 having lower June RF and higher amounts of excessively drained soil class. As with the 20 ppb threshold, however, well drained soil class showed lower values for cluster 2. The marked difference in the values of several variables between clusters 1 and 2 for the 100 ppb threshold is consistent with the difference observed in the relative risk between the two clusters in Figure 3(d).

[FIGURE 7]

[FIGURE 8]

For high risk years, the risk profiles for the 20 and 100 ppb thresholds show almost the reverse pattern to each other. The risk profiles for two clusters with the 100 ppb threshold for the high risk years show no overlap of boxes and boxes on different sides of the overall mean line for the percentage of corn-growing area and well drained soil class and their indicators, June RF and the proportion of June RF, June TMax, NDVI and the indicator for excessively drained soil class. Also for each of these variables the values show the patterns one would expect to be associated with higher risk of Aflatoxin contamination apart from for well-drained soil class. The boxplots for the 20 ppb threshold for the high risk years show the reverse patterns to those for the 100 ppb threshold and the general patterns for cluster 2 that one would not expect to be associated with higher risk of Aflatoxin contamination. This result however, is not completely unexplained or unexpected as the posterior distribution of relative risk in Figure 4(c) show that the median risk values for clusters 1 and 2 are similar to each other and both hinges of the box for cluster 1 lie inside the range of values for the hinges of the box for cluster 2. This shows that during high risk years, the 20 ppb threshold is very frequently breached and so it is difficult to distinguish the characteristics of



counties that are above and below the threshold.

In most box plots (Figures 5-8) there was slightly more difference between the two clusters in terms of June TMax than June RF, but the latter always showed the expected pattern with higher risk of Aflatoxin contamination being associated with lower June RF. This does suggest though, that heat stress may be an additional cause of contamination rather than just drought. The patterns displayed in Figures 5-8 for the indicator variables and the proportion variables compared with those for the original variables from which they were derived are very similar and there does not seem to be any particular benefit of deriving these new variables based on thresholds.

#### 4. Discussion

We characterized Aflatoxin contamination risk in southern Georgia using a Bayesian spatial profile regression approach. While there are similarities with existing literature (Kerry et al., 2017) in the counties that are identified as high risk for the 20 ppb threshold and the entire study period, our results are different from Poisson kriging for the other thresholds and time periods and in that both Aflatoxin levels from corn samples and risk factors are simultaneously taken into account. Understanding how, and the extent to which, the various risk factors influence Aflatoxin levels is important as the present study has shown that some of the theoretical assumptions about what factors influence Aflatoxin contamination made by Kerry et al. (2017) seem to be wrong and also it can help identify suitable inexpensive proxy variables like NDVI and Thermal IR that vary at finer scales.

While regression models can be used to assess the effects of explanatory variables on the Aflatoxin contamination risk, their application to Aflatoxin risk assessment imposes some serious methodological challenges. First, known risk factors and their impact on Aflatoxin contamination risk vary spatially, which is often ignored in standard regression models. In the proposed profile regression analysis, clusters were formed according to their levels of Aflatoxin contamination risk and their associations with risk factors. Within each cluster, the associations between the Aflatoxin contamination risk and risk factors were established. This allows the spatial variation in the estimates of the relationship between risk factors and Aflatoxin levels to be assessed and allows the model to make more use of spatial information and their relationships (Schabenberger and Gotway, 2005). Second, parameter estimation via a simple Poisson regression was not likely to to

effectively identify statistically significant risk factors in the current study due to the strong correlation among explanatory variables and the presence of spatial autocorrelation in the residual errors. The spatial profile regression overcomes multi-collinearity among explanatory variables and explicitly accounts for the spatial autocorrelation present in Aflatoxin contamination data by allowing the joint consideration of multiple risk factors through their risk profile. Lastly, corn sampling for Aflatoxin determination was not conducted across all counties with equal intensity as some counties have more observations than others (see Figure 1). When analyzing this small number of sample data, we often observe excessively high variability that can not be explained by a standard Poisson regression. This over-dispersion problem, essentially resulting from the small numbers problem, is taken care of by the Bayesian spatial profile regression approach. This is done by incorporating spatial random effects, which can absorb the extra variability, and also by conducting a regression within a cluster in which both response variable and explanatory variables share similar characteristics.

Profile regression enabled identification of regions with higher Aflatoxin contamination risk, while accounting for the uncertainty associated with the risk estimation. We assessed the uncertainty with respect to the random effects ( $S_i^c, \epsilon_i^c$ ) for each cluster on the Aflatoxin contamination risk. That is, the posterior probability that a specific profile cluster's posterior distribution of relative risk for Aflatoxin contamination is obtained and visualized. The greater certainty in the cluster formed, the narrower the intervals would be. Despite the optimal partitioning obtained from model averaging through the entire MCMC output, the higher uncertainty due to the different clusters formed from iteration to iteration of the MCMC sampler suggests that there is substantial uncertainty to account for in the cluster assignment and the number of clusters used. Nevertheless, the sensitivity analysis suggested that 2 clusters were optimal for the majority of runs. With regard to the number of clusters, it should be noted that Bayesian profile regression does not require analysts to pre-specify the target number of clusters but rather the number changes from iteration to iteration. This flexibility, however, leads to a rich output that requires careful interpretation (Molitor et al., 2011).

The risk factor profiles for each cluster showed patterns that were generally consistent with the theory of Aflatoxin contamination, but also showed some other key findings. First, irrigation was largely unimportant to consider as a risk factor. As shown in Figures 5-8, irrigation as a risk factor seems to have very little influence on Aflatoxin contamination as the pattern in rela-

tion to the high risk cluster changes between thresholds and study periods or it is just not markedly different between clusters. This could reflect the fact that only the hottest driest counties have the most irrigation infrastructure and these are the counties where you might expect most Aflatoxin contamination. So due to this positive association, the negative association, that one might expect, with higher Aflatoxin contamination occurring in areas that are not well irrigated is not always observed. Also this could reflect the fact that the irrigation figures were for agricultural land in general that was harvested rather than just the corn-growing area that was harvested. If more precise irrigation data were obtained the results might change. More precise irrigation data could be obtained from imagery in individual years by using shape indices to identify circular or semi-circular fields that grow corn and thus have a central pivot irrigation system installed.

Second, the pattern of results for well-drained soil as a risk factor in Figures 5-8 were generally the reverse of what one might expect. Showing that, including it in an additive risk factors approach like that of Kerry et al. (2017) was counter-productive. This is probably a result of the majority of soil in the study area being well drained and so it is the excessively drained soil that is different and more worthy to investigate. Excessively drained soil and its indicator tended to show the largest difference between classes and with cluster 2 being associated with higher amounts of excessively drained soil only for the 100 ppb thresholds (Figures 6 and 8). This suggests that this type of soil drainage exacerbates drought when weather is extreme and tends to cause breaching of the higher 100 ppb threshold but is not important in determining breaches of the lower 20 ppb threshold.

Third, corn-production and June TMax were the only two variables that consistently showed no overlap between the boxes for the two clusters and the boxes were found on opposite sides of the global mean for each cluster. This suggests that they are the most important factors in determining Aflatoxin contamination risk. The effects of high maximum temperatures in causing plant heat stress and increased risk of Aflatoxin contamination are clear. The effect of corn-production area is less clear. However, the decision to plant various crops in particular locations is usually based on some notion of temperature, rainfall and soil conditions in that area that are favourable for growth of a particular crop. Considerations such as the length of the growing season and the likely number of growing degree days until key growth stages in relation to climate normals are usually considered to avoid crop failure. Given this, greater risk of Aflatoxin contamination being found in counties

where less corn is grown would be expected as it is suggesting that those counties have more marginal conditions for the growth of corn, so less farmers risk growing it there.

Finally, the patterns of NDVI and Thermal IR were very similar to June RF and June TMax, respectively, but the differences between clusters were usually a little less well defined than for the weather variables. This was expected given that we were unable to determine the exact corn growing pixels to extract from imagery for each individual growing season in our study period. The smaller difference between clusters is also expected, particularly in the case of the high risk years, as there was no suitable remotely sensed imagery available for two of the three high risk years (1978 and 1981). Nevertheless, the results do suggest that NDVI and Thermal IR are promising surrogates for June TMax and June RF to determine areas at risk or Aflatoxin contamination at the sub-county scale. They are likely to better represent actual crop stress in future years rather than the past years used in this study because for years after 2008, the actual corn growing pixels are available via the CropScape data layer.

Regardless of the FDA thresholds used to define the level of Aflatoxin risk for each county, all dichotomous variables showed similar patterns to the original variables from which they were derived and showed marked differences in values between clusters. This contrasts with the results of standard Poisson regression where no coefficients for indicator variables were significant. This implies that the threshold values used to define the dichotomous variables were effective, but their impacts were not properly estimated when the covariates are correlated and the effects are constant throughout the study area. Nevertheless, as indicator variables showed similar patterns to the original variables there does not seem to be any real advantage to using these, but it was useful to be able to assess this by including them in profile regression where their strong correlation with the original variables would not be a problem.

Our work is not short of limitations. First, the stability of results was not guaranteed for our model despite the high computation and extra runs. Given the relatively low Aflatoxin relative risk estimates for both time frames and thresholds, we assumed that the mixing took a long time and thus the optimal partition was hard to reach. Second, both the Aflatoxin sample data and risk factors considered in the present study contain missing values in counties and specific years. Thus, our analysis was based on a relatively small number of observations in both Poisson regression and profile regression. Third,

our study was conducted based on the spatial scale of counties due to data availability. The scale effects on our findings are expected. In future work we will investigate the potential impacts of changes in environmental/weather conditions on Aflatoxin risk levels by estimating (or predicting) Aflatoxin risk associated with a hypothetical risk profile (Ceccato et al., under review). We will be able to better understand the role of a particular risk factor or group of risk factors on Aflatoxin contamination risk under a specified scenario, for example, by investigating how the Aflatoxin risk will change and which areas will be the most vulnerable under extreme drought. This will be useful for determining likely risks of contamination under different climate change scenarios. We will also conduct analysis at a finer spatial scale than counties by incorporating both interpolated Aflatoxin risk and risk factors. Lastly validation analysis for profile regression will be performed.

## 5. Conclusions

In the current work we assessed Aflatoxin risks at county levels using both corn sample measurements and risk factors using a Bayesian spatial profile regression approach. The application of spatial profile regression addressed methodological challenges that are often encountered in a regression, but also allowed investigation of the effects of a range of variables that we know are well correlated with one another to determine which are the best predictors of Aflatoxin contamination risk. The greatest difference between clusters was found for the entire study period and the 100 ppb threshold. For this scenario most variables were significant and all except well drained soil showed expected patterns. The data for this scenario were based on more years of observations and it seems like there is less uncertainty in defining breaches of the 100 ppb threshold. The results suggest that known risk factors have the potential to be useful for identifying areas with high Aflatoxin risk, but that well drained soil behaves in the opposite way to that expected and that the irrigation data used were not useful. They showed that June TMax is slightly more important than June RF, and that excessively drained soils are more important when considering breaches of the 100 ppb threshold. Analysis showed that exploring the use of NDVI and Thermal IR for finer scale analyses is worthwhile and such would be required for this analysis to be more practically useful to individual farmers.

## References

- Abbas, H. K., Shier, W. T. and Cartwright, R. D. (2007), ‘Effect of temperature, rainfall and planting date on Aflatoxin and fumonisin contamination in commercial Bt and non-Bt corn hybrids in Arkansas’, *Phytoprotection* **88**(2), 41–50.
- Besag, J. (1974), ‘Spatial interaction and the statistical analysis of lattice systems’, *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 192–236.
- Besag, J., York, J. and Mollié, A. (1991), ‘Bayesian image restoration, with two applications in spatial statistics’, *Annals of the Institute of Statistical Mathematics* **43**(1), 1–20.
- Brenneman, T., Wilson, D. and Beaver, R. (1993), ‘Effects of diniconazole on aspergillus populations and Aflatoxin formation in peanut under irrigated and nonirrigated conditions’, *Plant Disease* **77**(6), 608–612.
- Ceccato, V., Li, G. and Haining, R. (under review), ‘The ecology of outdoor rape: The Stockholm case study’, *European Journal of Criminology*.
- Coker, E., Liverani, S., Ghosh, J. K., Jerrett, M., Beckerman, B., Li, A., Ritz, B. and Molitor, J. (2016), ‘Multi-pollutant exposure profiles associated with term low birth weight in Los Angeles county’, *Environment International* **91**, 1–13.
- Damianidis, D., Ortiz, B., Windham, G., Scully, B. and Woli, P. (2015), Predicting pre-harvest aflatoxin corn contamination risk with a drought index, in ‘Precision agriculture’15’, Wageningen Academic Publishers, pp. 788–809.
- FDA (2015), ‘U.S. Food and Drug Administration’, online.  
**URL:** <http://www.fda.gov/ICECI/ComplianceManuals/CompliancePolicyGuidanceManual/ucm>
- Fraisse, C., Breuer, N., Zierden, D., Bellow, J., Paz, J., Cabrera, V., y Garcia, A. G., Ingram, K., Hatch, U., Hoogenboom, G. et al. (2006), ‘AgClimate: A climate forecast information system for agricultural risk management in the southeastern USA’, *Computers and Electronics in Agriculture* **53**(1), 13–27.

- Garland, T. and Reagor, J. (2001), ‘Chronic canine aflatoxicosis and management of an epidemic’, *Mycotoxins and phycotoxins in perspective at the turn of the millennium. Wageningen, The Netherlands: Ponsen and Looven* pp. 231–236.
- Ghasemi-Kebria, F., Joshaghani, H., Taheri, N. S., Semnani, S., Aarabi, M., Salamat, F. and Roshandel, G. (2013), ‘Aflatoxin contamination of wheat flour and the risk of esophageal cancer in a high risk area in Iran’, *Cancer Epidemiology* **37**(3), 290–293.
- Guo, B., Chen, Z.-Y., Lee, R. D. and Scully, B. T. (2008), ‘Drought stress and preharvest aflatoxin contamination in agricultural commodity: genetics, genomics and proteomics’, *Journal of Integrative Plant Biology* **50**(10), 1281–1291.
- Hastie, D. I., Liverani, S., Azizi, L., Richardson, S. and Stücker, I. (2013), ‘A semi-parametric approach to estimate risk functions associated with multi-dimensional exposure profiles: application to smoking and lung cancer’, *BMC Medical Research Methodology* **13**(1), 129.
- Hatfield, J. L. and Prueger, J. H. (2010), ‘Value of using different vegetative indices to quantify agricultural crop characteristics at different growth stages under varying management practices’, *Remote Sensing* **2**(2), 562–578.
- Horn, B. W., Sorensen, R. B., Lamb, M. C., Sobolev, V. S., Olarte, R. A., Worthington, C. J. and Carbone, I. (2014), ‘Sexual reproduction in aspergillus flavus sclerotia naturally produced in corn’, *Phytopathology* **104**(1), 75–85.
- Jones, H. G., Serraj, R., Loveys, B. R., Xiong, L., Wheaton, A. and Price, A. H. (2009), ‘Thermal infrared imaging of crop canopies for the remote diagnosis and quantification of plant responses to water stress in the field’, *Functional Plant Biology* **36**(11), 978–989.
- Kerry, R., Ortiz, B. V., Ingram, B. R. and Scully, B. T. (2017), ‘A spatio-temporal investigation of risk factors for aflatoxin contamination of corn in southern Georgia, USA using geostatistical methods’, *Crop Protection* **94**, 144–158.

- Liu, Y. and Wu, F. (2010), ‘Global burden of aflatoxin-induced hepatocellular carcinoma: A risk assessment’, *Environmental Health Perspectives* **118**(6), 818.
- Liverani, S., Lavigne, A. and Blangiardo, M. (2016), ‘Modelling collinear and spatially correlated data’, *Spatial and Spatio-temporal Epidemiology* **18**, 63–73.
- Molitor, J., Papathomas, M., Jerrett, M. and Richardson, S. (2010), ‘Bayesian profile regression with an application to the national survey of children’s health’, *Biostatistics* **11**(3), 484–498.
- Molitor, J., Su, J. G., Molitor, N.-T., Rubio, V. G., Richardson, S., Hastie, D., Morello-Frosch, R. and Jerrett, M. (2011), ‘Identifying vulnerable populations through an examination of the association between multipollutant profiles and poverty’, *Environmental Science & Technology* **45**(18), 7754–7760.
- Navarro, F., Ingram, B., Kerry, R., Ortiz, B. and Scully, B. (2017), ‘A web-based gis decision support tool for determining corn aflatoxin risk: A case study data from Southern Georgia’, *Advances in Animal Biosciences* **8**(2), 718–723.
- Newman, S. J., Smith, J. R., Stenske, K. A., Newman, L. B., Dunlap, J. R., Imerman, P. M. and Kirk, C. A. (2007), ‘Aflatoxicosis in nine dogs after exposure to contaminated commercial dog food’, *Journal of Veterinary Diagnostic Investigation* **19**(2), 168–175.
- Palumbo, J. D., O’Keeffe, T. L., Kattan, A., Abbas, H. K. and Johnson, B. J. (2010), ‘Inhibition of aspergillus flavus in soil by antagonistic pseudomonas strains reduces the potential for airborne spore dispersal’, *Phytopathology* **100**(6), 532–538.
- Papadoyannis, I. et al. (1990), *HPLC in Clinical Chemistry*, Dekker.
- Papathomas, M., Molitor, J., Richardson, S., Riboli, E. and Vineis, P. (2011), ‘Examining the joint effect of multiple risk factors using exposure risk profiles: lung cancer in nonsmokers’, *Environmental Health Perspectives* **119**(1), 84.



- Payne, G. A. and Widstrom, N. W. (1992), ‘Aflatoxin in maize’, *Critical Reviews in Plant Sciences* **10**(5), 423–440.
- Salvacion, A., Ortiz, B. V., Scully, B., Wilson, D. M., Hoogenboom, G. and Lee, D. (2011), Effect of rainfall and maximum temperature on corn aflatoxin in the southeastern U.S coastal plain, *in* ‘Proceedings of the Climate Information for Managing Risks’, Orlando, Florida.
- Schabenberger, O. and Gotway, G. (2005), *Statistical Methods for Spatial Data Analysis*, CRC Press, Boca Raton, FL.
- Sepúlveda-Reyes, D., Ingram, B., Bardeen, M., Zúñiga, M., Ortega-Farías, S. and Poblete-Echeverría, C. (2016), ‘Selecting canopy zones and thresholding approaches to assess grapevine water status by using aerial and ground-based thermal imaging’, *Remote Sensing* **8**(10), 822.
- Shekhar, S., Yoo, E., Ahmed, S., Haining, R. and Kadannolly, S. (2017), ‘Analysing malaria incidence at the small area level for developing a spatial decision support system: A case study in Kalaburagi, Karnataka, India’, *Spatial and Spatio-temporal Epidemiology* **20**, 9–25.
- Wacoo, A. P., Wendi, D., Vuzi, P. C. and Hawumba, J. F. (2014), ‘Methods for detection of aflatoxins in agricultural food crops’, *Journal of Applied Chemistry* **2014**.
- Wang, R., Cherkauer, K. and Bowling, L. (2016), ‘Corn response to climate stress detected with satellite-based NDVI time series’, *Remote Sensing* **8**(4), 269.
- Wang, T., Zhang, E., Chen, X., Li, L. and Liang, X. (2010), ‘Identification of seed proteins associated with resistance to pre-harvested aflatoxin contamination in peanut (*arachis hypogaea* l)’, *BMC Plant Biology* **10**(1), 267.
- Widstrom, N., Forster, M., Martin, W. and Wilson, D. (1996), ‘Agronomic performance in the southeastern united states of maize hybrids containing tropical germplasm’, *Maydica* **41**(1), 59–63.

Table 1: Threshold Values Used for Risk Factor Indicators

Binary Risk Factors	Description	Threshold
I(Excessively drained)	Percent of corn growing area in county with excessively drained soils (classes 1-2.5)	> 2.5%
I(Well drained)	Percent of corn growing area in county with well-drained soils (classes 1-4)	> 90%
I(Corn production)	Percent of county area growing corn	> 1.75%

Table 2: Poisson Regression Results

	<i>The Entire Study Period</i>		<i>High Risk Years</i>	
	$Y^{20}$	$Y^{100}$	$Y^{20}$	$Y^{100}$
I(Excessively drained)	0.13(-0.29,0.56)	0.01(-0.61,0.63)	-0.10(-0.76,0.56)	0.05(-0.91,1.02)
I(Well drained)	-0.11(-0.43,0.22)	-0.11(-0.63,0.41)	-0.25(-0.85,0.36)	-0.29(-1.18,0.60)
I(Corn Production)	0.03(-0.32,0.38)	0.25(-0.32,0.81)	-0.13(-0.70,0.44)	-0.05(-0.97,0.88)
Irrigation	-0.00(-0.00,0.00)	-0.00(-0.00,0.00)	-0.00(-0.00,0.00)	-0.00(-0.00,0.00)
NDVI	1.63(-3.74,6.99)	3.59(-5.07,12.25)	-0.69(-8.91,7.54)	-0.34(-12.90,12.22)
Thermal IR	0.32** (0.11,0.52)	0.50*** (0.19,0.81)	-0.03(-0.19,0.12)	-0.14(-0.40,0.12)
Excessively drained	0.02(-0.03,0.07)	0.05(-0.01,0.11)	0.02(-0.04,0.08)	0.06(-0.02,0.13)
Well drained	-0.003(-0.01,0.004)	-0.01(-0.02,0.001)	0.01(-0.004,0.02)	0.01(-0.004,0.03)
Percent of corn growing	0.03(-0.07,0.13)	0.04(-0.12,0.21)	0.10(-0.12,0.33)	0.001(-0.41,0.41)
Prop. June RF	-2.14* (-4.28,-0.001)	-1.96(-5.20,1.29)	-1.09(-4.32,2.14)	0.90(-4.10,5.91)
Prop. June TMax	-0.09(-3.09,2.92)	-2.44(-6.98,2.10)	-0.55(-4.05,2.95)	-2.29(-7.90,3.31)
June RF (mm)	0.03(-0.01,0.07)	0.01(-0.06,0.07)	0.01(-0.05,0.06)	-0.01(-0.10,0.08)
June TMax (°C)	0.14(-0.45,0.73)	0.86(-0.07,1.79)	0.20(-0.39,0.79)	0.99* (0.03,1.96)
Constant	-50.75** (-90.38,-11.13)	-99.23*** (-160.99,-37.47)	-2.62(-30.68,25.43)	-14.09(-55.78,27.59)
Observations	48	48	48	48
Log Likelihood	-88.20	-68.69	-61.09	-44.53
Akaike Inf. Crit.	204.41	165.38	150.17	117.06

*Note:*

\*p&lt;0.1; \*\*p&lt;0.05; \*\*\*p&lt;0.01

Table 3: Membership of Georgia Counties under the Optimal Partition. Profile group names are given followed by the number of counties belonging to each cluster in parenthesis and the mean relative risk of Aflatoxin contamination is given with the 95% posterior credible intervals in parenthesis.

Time Frame	Threshold Value	Profile Groups	Relative Risk
Entire Study Period	20 ppb	Cluster 1 (38)	0.36 (0.30, 0.42)
		Cluster 2 (15)	0.44 (0.35, 0.56)
	100 ppb	Cluster 1 (31)	0.11 (0.08, 0.15)
		Cluster 2 (22)	0.18 (0.14, 0.24)
High Risk Period	20 ppb	Cluster 1 (35)	0.73 (0.56, 0.95)
		Cluster 2 (17)	0.75 (0.53, 1.03)
	100 ppb	Cluster 1 (36)	0.31 (0.16, 0.50)
		Cluster 2 (16)	0.35 (0.23, 0.48)

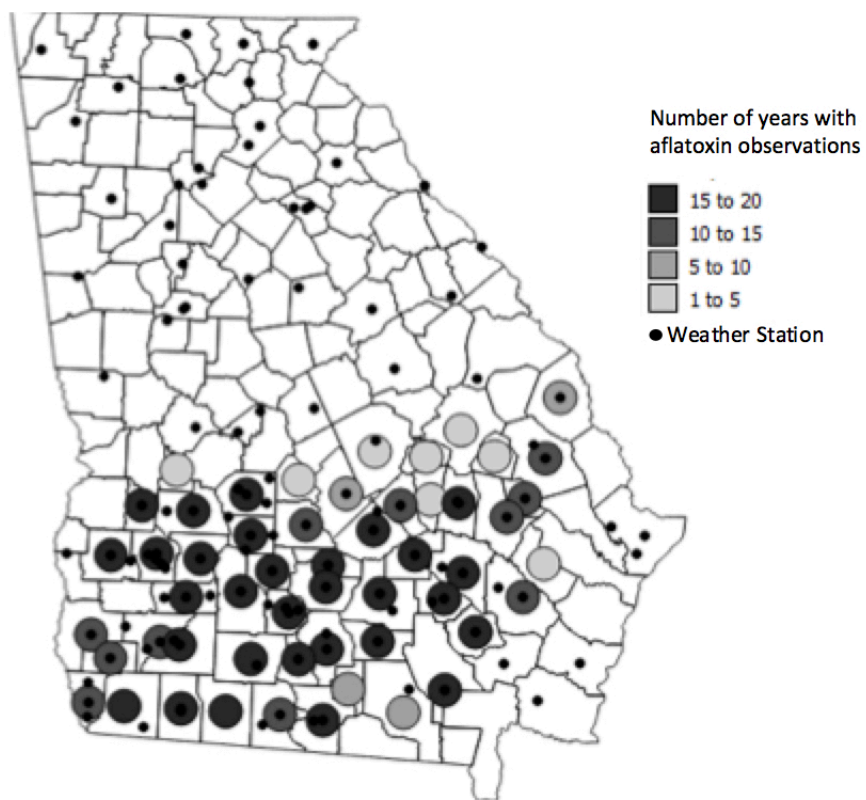


Figure 1: Location of weather stations and number of years with Aflatoxin data for each county

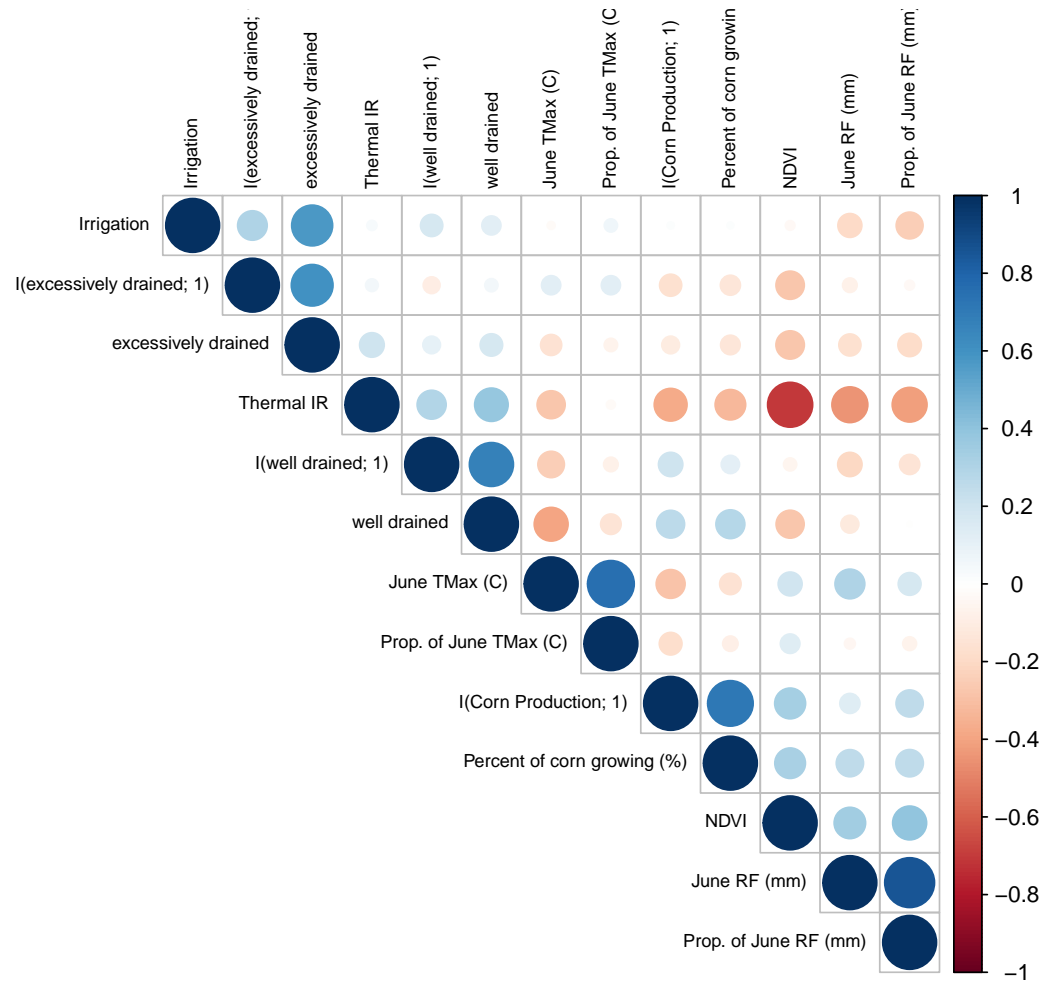


Figure 2: Pearson Correlation among Risk Factors (Explanatory Variables)

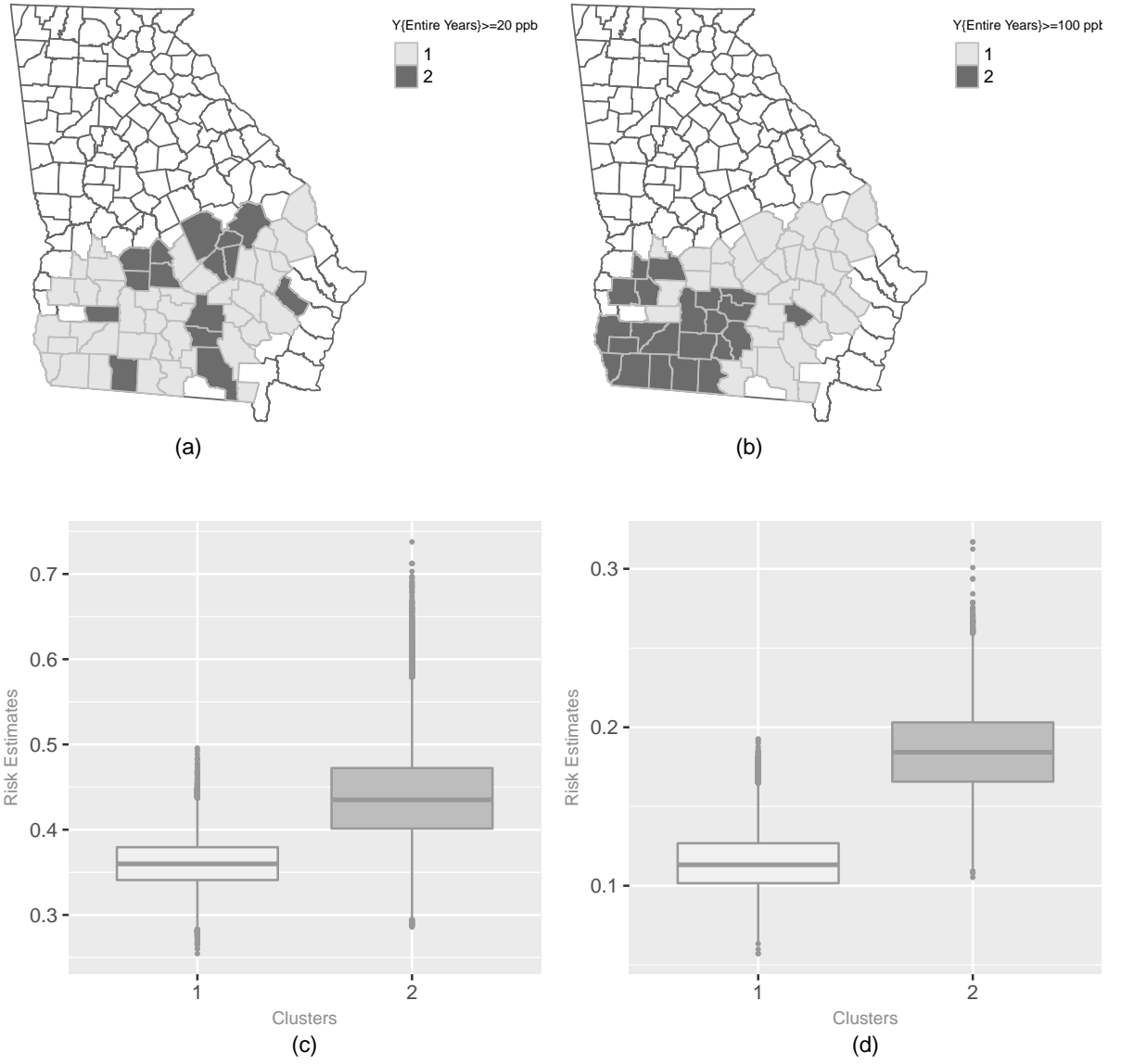


Figure 3: Spatial Distributions of Aflatoxin Risk Clusters and Posterior Distribution of Relative Risk with the 20 ppb threshold (a and c) and the 100 ppb threshold (b and d) for the Entire Study Period.

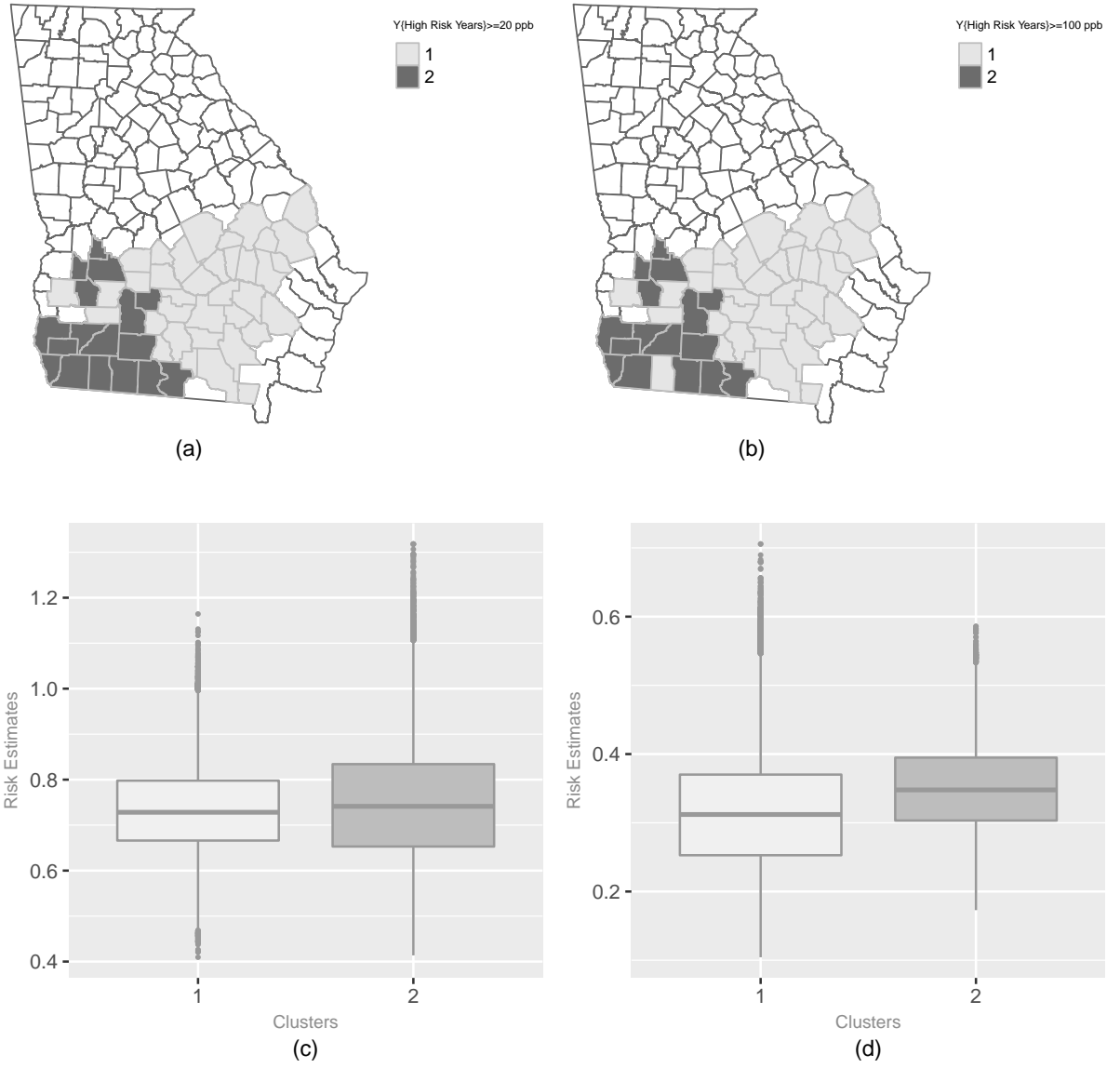


Figure 4: Spatial Distributions of Aflatoxin Risk Clusters and Posterior Distribution of Relative Risk with the 20 ppb threshold (a and c) and the 100 ppb threshold (b and d) for High Risk Years.



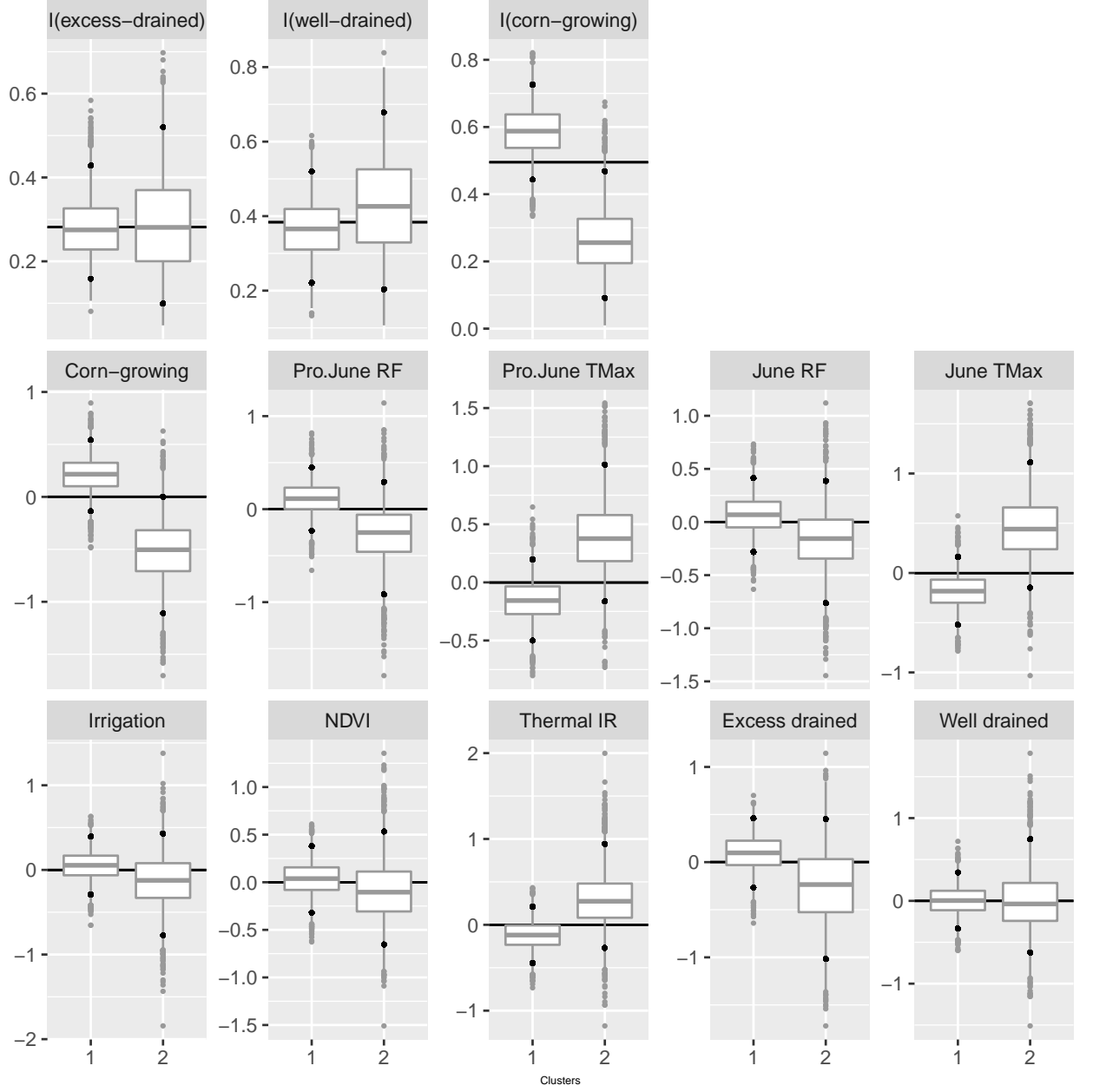


Figure 5: Posterior Distribution of Risk Factor Profile for Each of the Two Clusters of  $Y_i^{20}$  for the Entire Study Period

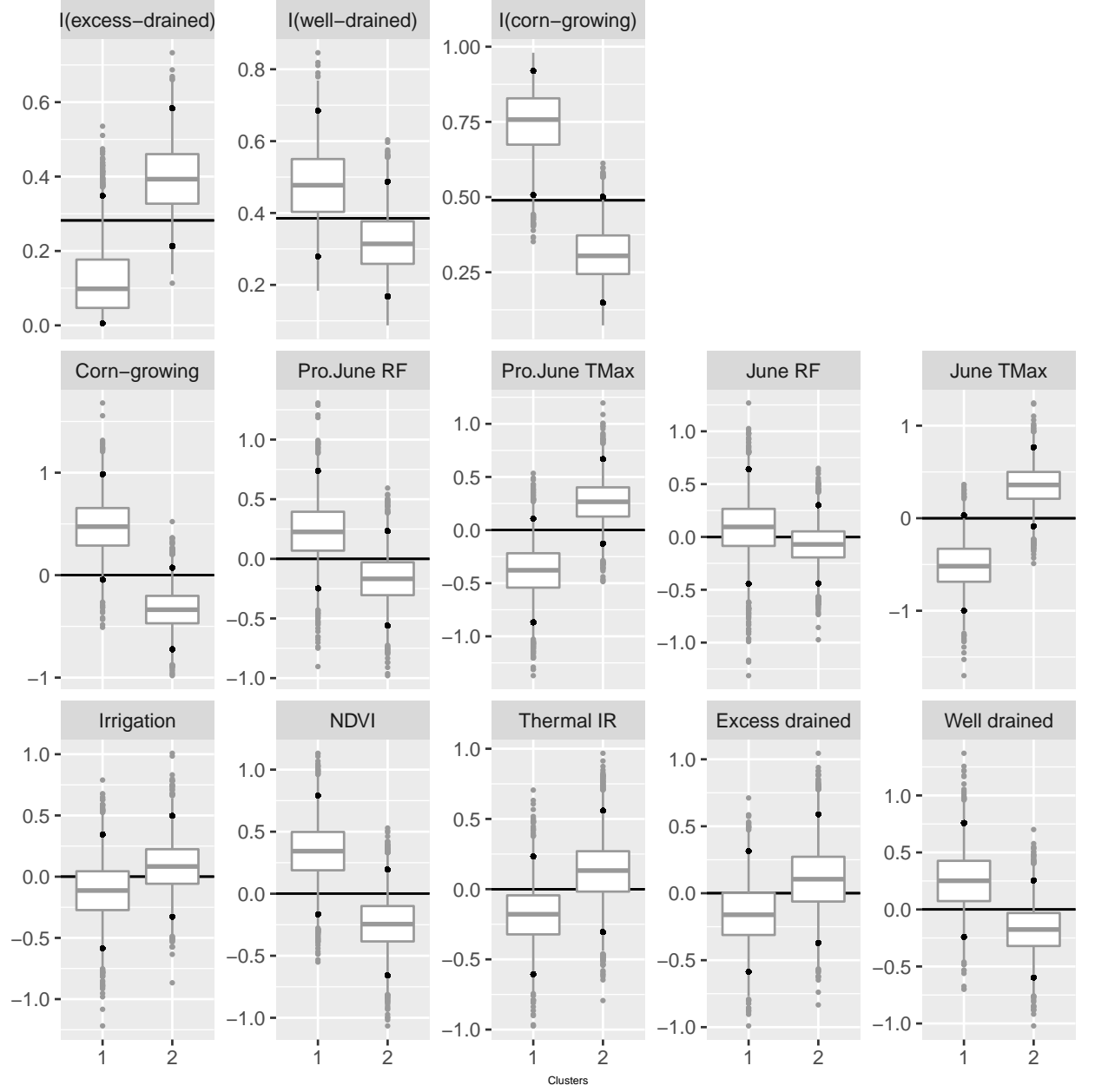


Figure 6: Posterior Distribution of Risk Factor Profile for Each of the Two Clusters of  $Y_i^{100}$  for the Entire Study Period

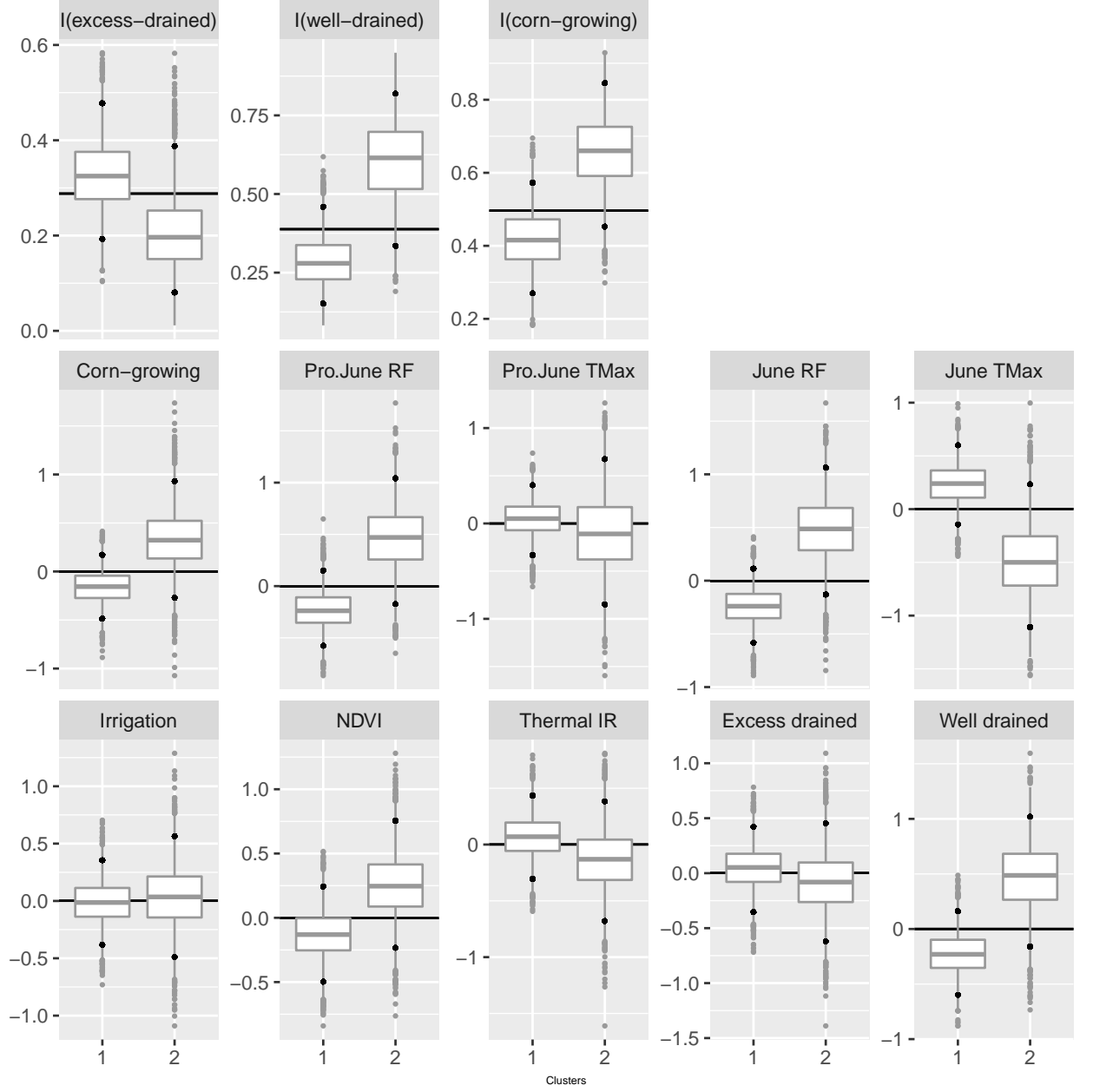


Figure 7: Posterior Distribution of Risk Factor Profile for Each of the Two Clusters of  $Y_i^{20}$  for the High Risk Years

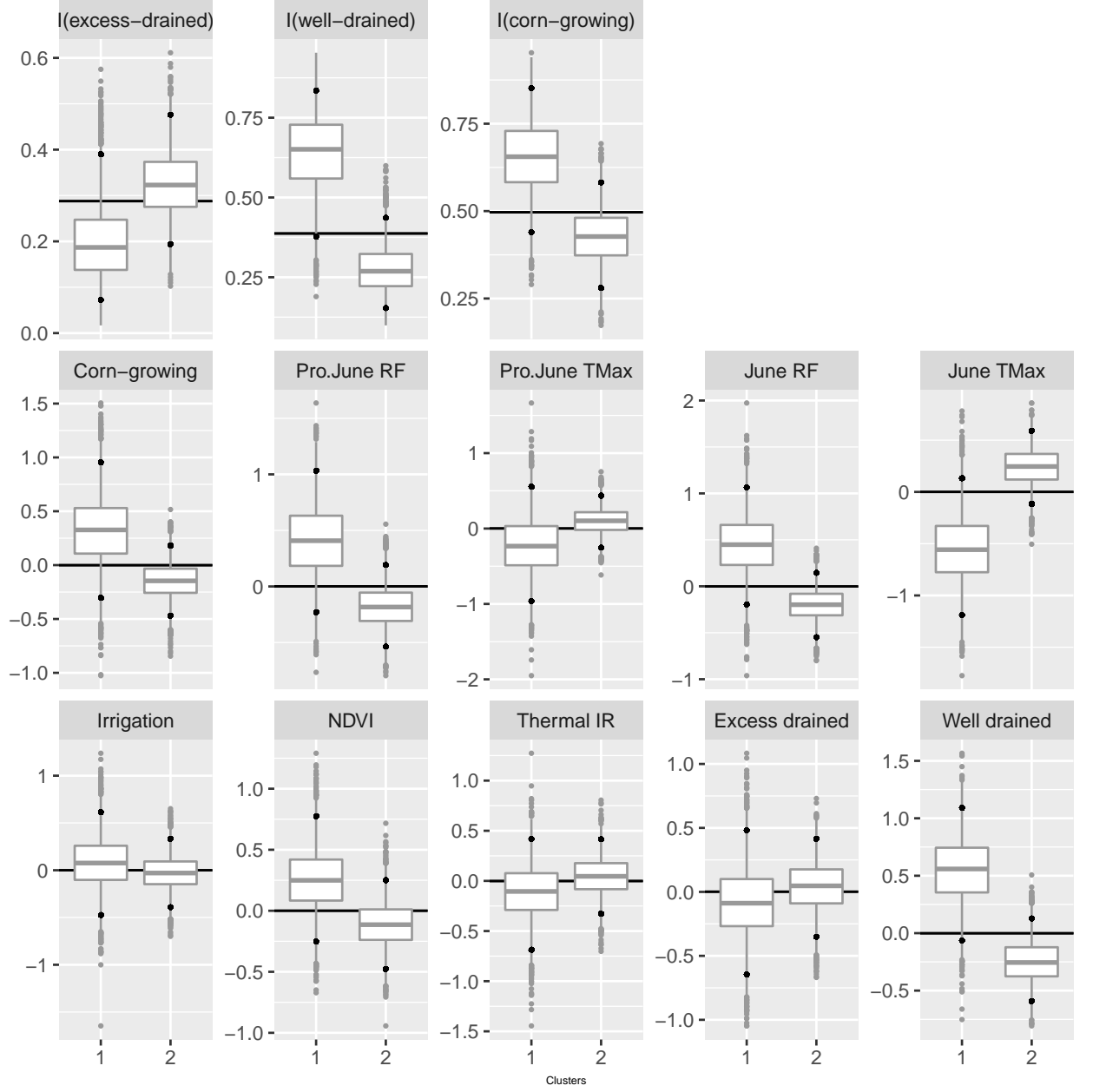


Figure 8: Posterior Distribution of Risk Factor Profile for Each of the Two Clusters of  $Y_i^{100}$  for the High Risk Years

# Defining and characterizing Aflatoxin contamination risk areas for corn in Georgia, USA: Adjusting for collinearity and spatial correlation

Yoo, EunHye

2018-07-02

Attribution-NonCommercial-NoDerivatives 4.0 International

---

Yoo E, Kerry R, Ingram B, et al., (2018) Defining and characterizing Aflatoxin contamination risk areas for corn in Georgia, USA: Adjusting for collinearity and spatial correlation. *Spatial Statistics*, Volume 28, December 2018, pp. 84-104

<http://dx.doi.org/10.1016/j.spasta.2018.06.003>

*Downloaded from CERES Research Repository, Cranfield University*