Irregularly Sampled Data in Space and Time: Using Poisson Kriging to Reduce the Influence of Uncertain Observations in Assessing the Risk of Aflatoxin Contamination of Corn in Southern Georgia, USA

Ruth Kerry¹, Brenda Ortiz², Ben Ingram³, Brian Scully⁴, EunHye Yoo⁵

¹Brigham Young University, USA  
²Auburn University, USA  
³Universidad de Talca, Chile  
⁴USDA-ARS, USA  
⁵University at Buffalo, SUNY, USA

*Corresponding author: ruth.kerry@byu.edu

I. INTRODUCTION

Aflatoxin is a mycotoxin produced by fungi (A. flavus or A. parasiticus) which can contaminate several staple crops such as corn (Payne, 1992) and can cause liver cancer in humans and animals (Barrett, 2005). The Food and Drug administration office of the USA have set a limit of 20 ppb, total Aflatoxin, for interstate commerce of food and feed (FDA, 2015). Infection of corn with A. flavus or A. parasiticus is driven by high temperature and drought conditions (Guo et al. 2008) associated with particular climatic areas (Abbas et al. 2007), agro-ecological zones (Setamou et al. 1997) and soil types (Palumbo et al. 2010). In the Southeast of the U.S.A, corn planted as a summer crop is highly susceptible to Aflatoxin contamination. Rainfall variability and high temperatures in this region during summer, along with light textured soils can induce water stress and contamination. Also, lack of irrigation infrastructure in some areas further aggravates the situation (Brenneman et al. 1993). Salvacion et al. (2011) found that June maximum temperatures and precipitation were key predictors of Aflatoxin levels in southern Georgia (GA), USA.

There are several methods for accurate Aflatoxin measurement, but most are time-consuming and expensive (Papadoyanis, 1990) and conducted at harvest which does not allow implementation of in-season adaptation strategies to reduce the risk. Given the expense of Aflatoxin measurement, an important goal for agricultural extension services is to identify those years and counties most at risk of contamination to reduce unnecessary expense on testing in years and areas when there is little risk of contamination. Identifying such years and counties would also allow adaptation of management strategies in season to reduce contamination risk and identify areas where more resistant varieties (Chen et al. 2002) of corn should be planted or irrigation infrastructure improved.

This paper aims to use Poisson kriging to give a space-time summary of Aflatoxin contamination data collected irregularly, both spatially and temporally, over a 27 year period (1977-2004) in 53 counties in southern GA. This will then be used to evaluate a risk factors approach to identify counties with the greatest risk of Aflatoxin contamination based on weather, soil and crop variables. Due to the irregular sampling in space and time, and highly skewed data approaching a Poisson distribution typical geostatistical methods are unstable for analysing spatial and temporal patterns in this data.
II. METHODS

Data Collection

From 1977-2004, corn samples were collected at harvest using a grab sampling technique with an average of 3 replications in 53 counties in southern GA (Fig. 1a) for Aflatoxin assessment by the Plant Pathology Department, University of Georgia at Tifton, GA. Data was collected for 23-45 counties each year but exact locations within counties was not recorded. For all years combined there was a total of 705 observations and these data approached a Poisson distribution (Fig. 1b).

![Figure 1](image-url)

Figure 1. (a) Counties in Southern Georgia where Aflatoxin was measured, the proportion of years measurements were made and the location of weather stations (b) Frequency distribution of Aflatoxin values measured in southern Georgia 1977-2004.

Based on the findings of Salvacion et al. (2011), monthly maximum temperatures for June (June Tmax, °C) and June rainfall data (June RF, mm) were obtained for each year 1977-2004 from the Georgia Weather Network (http://georgiaweather.net). Fig. 1a shows the location of weather stations within GA. As all counties do not have a weather station, some have more than one, the weather stations are not located at the center of the county and have different installation dates, the weather variables were ordinary kriged (OK) to county centroids and a 1 km grid.

The area planted with corn per county was determined using The CropScape - Crop Land data layer produced by the National Agricultural Statistics Service (NASS, http://nassgeodata.gmu.edu/CropScape/). Unfortunately this information was only available from 2008-2009 onwards and so does not coincide with the period of Aflatoxin collection. So this and the soils data derived using it are quite uncertain and may not give an exact reflection of the proportion of each county under corn production because it assumes that the same rotations are used and that the proportion of agricultural land per county has not greatly changed over time.

A geo-corrected 1:250,000 map of soil associations (NRCS, 2006) was simplified and used to generate a map with 3 drainage classes: excessively, well and poorly drained soil. The percentage of
land areas with soil in each drainage class in the corn growing area (as identified using Cropscape above) was calculated for each county.

Geostatistical Analysis
Making sense of the Aflatoxin data with geostatistical methods was difficult. If a map of Aflatoxin contamination for each year for all counties in southern GA is to be produced then a variogram must be computed with an average of 37 data points and as few as 23. Variograms for individual years were unreliable and showed little spatial structure. This is typical of highly skewed (Kerry and Oliver, 2007a,b) and sparse data (Webster and Oliver, 1992). Poisson kriging (Monestiez et al. 2006) is ideal for data with a Poisson distribution and that have been irregularly observed in space or time. The proportion of years a county had Aflatoxin levels > 20ppb and > 100 ppb were Poisson kriged to county centroids and also to a 1 km grid. As proportions have a numerator and a denominator, the numerator was the number of years Aflatoxin levels were above one of the thresholds in a given county and the denominator was the number of years Aflatoxin was measured in that county. The influence of proportions for counties with fewer observations is down-weighted during variogram computation using the following weighted estimator

\[
\hat{\gamma}_N(h) = \frac{1}{2} \sum_{\alpha, \beta} \frac{d(v_{\alpha})d(v_{\beta})}{d(v_{\alpha}) + d(v_{\beta})} \left\{ \sum_{\alpha, \beta} \frac{d(v_{\alpha})d(v_{\beta})}{d(v_{\alpha}) + d(v_{\beta})} \left[ r(v_{\alpha}) - r(v_{\beta}) \right]^2 - m^* \right\}
\]

where \(N(h)\) is the number of pairs of counties \((v_\alpha, v_\beta)\) whose denominator weighted centroids are separated by the vector \(h\), and \(m^*\) is the denominator-weighted mean of the \(N\) area ratios. The usual squared differences, \([r(v_\alpha) - r(v_\beta)]^2\), are weighted by a function of their respective denominator sizes, \(d(v_{\alpha})d(v_{\beta})/[d(v_{\alpha})+d(v_{\beta})]\), which gives more importance to more reliable data pairs based on larger denominators. Poisson kriging is a form of kriging with non-systematic errors and is parametric, modelling the noise attached to each observation with a Poisson distribution. Observations with small denominators receive less weight in kriging, by adding an error variance term to the diagonal of the kriging system (Monestiez et al. 2006, Goovaerts, 2005). Poisson kriging was done in SpaceStat (BioMedware, 2013).

Risk Factors Approach
By applying OK, risk factor data (June TMax, June RF, % Corn and Soil Type) were generated for each county and the nodes of a 1 km grid. The OK data was used to determine if risk factors exceeded a certain threshold converting it into indicator data (0/1). Table 1 shows the thresholds chosen for each variable. The thresholds for June TMax and June RF were chosen with respect to 30-year normals in southern GA and values receiving a (1) show hotter or drier than normal years. The indicator thresholds for other risk factors were determined based on natural marked breaks in the frequency distribution or were based on values associated with the tails of a normal distribution.

Once the number of risk factors above the specified threshold for each county in each year was determined, the relationship between these data and the Poisson kriged Aflatoxin data was assessed. This suggested broad groupings of years and counties with different levels of Aflatoxin contamination risk. These broad groupings were used to define grouping variables for Mann-Whitney U and Kruskal-Wallis H comparison tests to determine if there were significant differences in Aflatoxin levels based on the thresholds identified by the risk factors approach.
Table 1. Threshold values used for Risk Factor Indicators

<table>
<thead>
<tr>
<th>Risk Factor</th>
<th>Threshold for Indicator (1/0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>June monthly maximum temperature (°C)*</td>
<td>&gt;33°C</td>
</tr>
<tr>
<td>June monthly Rainfall (mm)*</td>
<td>&lt;50 mm</td>
</tr>
<tr>
<td>Percent of county area growing corn (%)</td>
<td>&gt;1.75%</td>
</tr>
<tr>
<td>Percent of county with well-drained soils (classes 1-4) (%)</td>
<td>&gt;90%</td>
</tr>
<tr>
<td>Percent of county with excessively drained soils (classes 1-2.5)</td>
<td>&gt;2.5%</td>
</tr>
<tr>
<td>Percent of years with 2 weather risk factors</td>
<td>&gt;30%</td>
</tr>
</tbody>
</table>

*Thresholds for June Tmax and June RF were chosen with respect to 30-year normal Tmax and RF in the area to show hotter and drier years than normal.

III. RESULTS AND DISCUSSION

In trying to summarize the risk of Aflatoxin contamination in space and time and to verify if the risk factor approach proposed in this work is useful, an obvious starting point is to calculate summary statistics for each year and county. The means (not shown) were influenced by the maximum value and this was particularly pronounced in counties where a smaller proportion of observations had been made. This suggests that there can be great spatial and temporal variability in the Aflatoxin levels and that the small number problem is an issue with data for counties where few observations have been made giving very high or low Aflatoxin contamination levels.

Temporal patterns in Aflatoxin risk are mostly associated with weather variables, specifically June Tmax and RF (Salvacion et al. 2011) which change from year to year and can help identify the specific years at greatest risk of contamination. The temporal results are not shown here but will be shown in the associated presentation to illustrate this point.

Analysis of Risk Data - Spatial Patterns

Analysing spatial patterns in risk of Aflatoxin contamination to identify the counties at greatest risk is more complicated than temporal analysis as it requires analysis of factors that are relatively stable in time (% corn and soil type) as well as weather variables. Fig. 2 a, c, e show the patterns in percent corn, well- and excessively-drained soil and Fig. 2 b, d, f show the indicators produced from these data using the thresholds in Table 1. Based on % corn, there is greatest risk (1) in south western GA (Fig. 2b) for well-drained soil risk is greatest in the west (Fig. 2d) and for excessively-drained soil the north is the highest risk area (Fig. 2f). Fig. 3a shows a map where the two weather risk factors (June Tmax and June RF) have been combined and the proportion of years with two weather risk factors has been kriged to the 1 km grid. This gives a temporal summary of the areas most prone to drought in southern GA. When converted to an indicator based on the thresholds in Table 1, Fig. 3b shows the areas with a high (1, >30% of years) risk of drought. Fig. 3c shows the percentages of years with three or more risk factors for Aflatoxin contamination where June Tmax and RF are considered separate risk factors. The map has similarities to Figs. 2f and 3b suggesting that weather and excessively drained soils are the greatest risk factors for Aflatoxin contamination. Nevertheless, crucial importance of both weather factors is shown by the striking similarity in the patterns shown in Fig. 3a and b which show drought summary and Fig. 3e and f which show the Poisson kriged summary of Aflatoxin levels in all years.
Figure 2. Risk factors and their associated indicators plotted kriged to a 1 km grid (a-b) percent of area growing corn, (c-d) percent of area with well-drained soils, (e-f) percent of area with excessively drained soils.
Figure 3. (a) Percent of years with 2 weather risk factors and (b) associated indicator, (c) percent of years with ≥ 3 risk factors, (d) Relationship between percent of years with ≥ 3 risk factors and percent of years and >20 ppb Aflatoxin, (e) percent of years with Aflatoxin > 20 ppb and (f) percent of years with Aflatoxin > 100 ppb.
Fig. 3d shows the relationship between the proportion of years with ≥3 risk factors and proportion of years with Aflatoxin levels >20ppb for each county. The correlation coefficient for this relationship was 0.59 which is significant at \( p<0.001 \). Distinct groupings of counties are visible in the plot and have been circled. Such grouping were not very well defined when just weather factors were considered and the correlation coefficient was lower, suggesting that the other risk factors (% corn and soil drainage types) help to distinguish the spatial differences between counties. The higher risk counties, circled in red and green in Fig. 3d, are the northern most counties in southern GA as well as those in the central area of the southern half of the state.

Comparison tests were performed to determine whether there were significant differences in Aflatoxin levels defined in particular ways based on risk levels identified from the scatter-graph produced (Fig. 3d) with the risk factor approach. Counties at ‘high risk’ of Aflatoxin contamination (with > 3 risk factors) were compared with those not at high risk (< 3 risk factors) by Mann-Whitney U tests. When average Aflatoxin levels were used for comparison there was no significant difference \( (p=0.569) \) but when Poisson kriged % years with > 20ppb and 100 ppb Aflatoxin data were used for comparison there were significantly higher, \( p=0.002 \) and \( p=0.012 \), proportions of years with Aflatoxin levels exceeding thresholds for ‘high’ risk years. Kruskal Wallis H tests comparing all risk classes identified in Figure 3d also showed more significant differences \( (p<0.001) \) and the expected order of classes when Poisson kriged data were used rather than average Aflatoxin data. This shows that average Aflatoxin levels do not give a good summary of the counties most and least at risk of Aflatoxin contamination while the Poisson kriged data which down-weight the influence of proportions based on low numbers of observations, do give a good summary.

**IV. CONCLUSIONS**

This study showed that when data have been irregularly sampled in space and time and have a Poisson distribution, Poisson kriging is a reliable way to generate a temporal summary of spatial patterns. Simple averages were shown to be unreliable where fewer observations were made and standard geostatistical methods do not work well when data have a Poisson distribution or have few data for individual years. Comparison tests showed that counties and years identified as having the greatest risk levels using the risk factors approach did have significantly higher Aflatoxin levels. This was not the case however, for average Aflatoxin due to unreliable averages because of irregular sampling. Identifying the weather conditions and counties associated with the highest contamination risk will allow for in-season adaptation strategies such as irrigation to avoid drought as temperatures and rainfall in June are carefully monitored with respect to 30 year normals. Also, testing can be focused in the highest risk counties and very little expensive Aflatoxin testing will be needed in low risk years and the highest risk counties could plant more resistant varieties of corn.

Future work should investigate including new variables in the risk factors approach, fine tuning of the thresholds of existing variables and quantifying the relative uncertainty of predictions associated with each risk factor. Using OK risk factor data means that uncertainty would obviously be greatest where distance from sampling points is greatest and for data like % corn where data from a different time period to sampling had to be used. The data kriged to a 1 km grid shows the potential for identifying high risk areas at the sub-county level. This should be more reliable when sub-county data on in-season corn NDVI values are used as an indicator of in-season drought stress in the risk factors approach. An online interactive Aflatoxin risk assessment tool that uses the risk factors approach outlined here is currently being developed and will include NDVI data. There is the potential that such an online tool could be adapted to other crops, states and even farms so that Aflatoxin levels may be better managed.
References