Decision Support Tools for Developing Aflatoxin Risk Maps in Rainfed Peanut Fields

George Vellidis[1], Mark Abney[2], Thirimachos Burlai[1], Jake Fountain[3], Robert Kemerait[3], Sunaab Kukal[1], Lorena Lacerda[1], Sara Maktabi[1], Alicia Peduzzi[1], Cristiane Pilon[1], Morgan Sysskind[1]

[1]Institute of Integrative Precision Agriculture, University of Georgia, USA
[2]Department of Entomology, University of Georgia, USA
[3]Department of Plant Pathology, University of Georgia, USA

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.
Aspergillus flavus and Aspergillus parasiticus hereafter referred to jointly as A. flavus, are soil fungi that infect and contaminate preharvest and postharvest peanuts with the carcinogenic secondary metabolite aflatoxin. A. flavus can cause extensive economic losses to peanut growers and shellers by contaminating peanut kernels with aflatoxins. In the southeastern U.S., contamination from aflatoxin continues to be a major threat to the peanut industry and climate variability along with the increased focus on consumer health and low aflatoxin thresholds may exacerbate the problem in the future. Industry-wide loss to aflatoxin in Georgia in 2019 was estimated at 24%, and the U.S. has all but lost the European peanut export market due to stringent aflatoxin thresholds. This project leverages a recently established interdisciplinary team of researchers, Extension professionals, and graduate students from three University of Georgia (UGA) colleges and partners with the USDA-ARS National Peanut Research Laboratory to address the aflatoxin problem in peanut. Based on a study conducted by Vellidis et al. in 2006, our hypothesis is that aflatoxin is spatially distributed within peanut fields and may be correlated with easily measurable field parameters. The goal of this project is to develop decision support tools (DSTs) that predict areas in the field in which aflatoxin hotspots are likely to occur in the three weeks prior to harvest. This paper provides an overview of the work completed through the second year of this multiyear project.

Keywords.
Peanut, Aflatoxin, Random forest, Crop growth models, DSSAT

Introduction

Mycotoxins are toxic secondary metabolites produced by fungi that are harmful to human and animal health. These compounds may contaminate foods and feeds, either in the field or postharvest, creating a major food safety risk. Because of their ability to cause illness or fatalities at very low concentrations (ppb or ppm), mycotoxin levels on foods and feeds are strictly regulated by food safety agencies around the world. The Food and Agriculture Organization of the United Nations has estimated that 25% of the world's crops are affected by mycotoxins each year, with losses of approximately 1 billion tons of food products annually (Schmale & Munkvold 2009). Without considering human health impacts, loss estimates attributed to mycotoxins in the in the US and Canada range between \$0.5 and \$5 billion/year (Schmale & Munkvold 2009). Global and regional increases in temperature and the increased focus on consumer health and low mycotoxin tolerances will exacerbate the mycotoxin problem in the future.

The primary mycotoxin issue in the state of Georgia, USA, is aflatoxin contamination of peanut (*Arachis hypogea* L). Aflatoxin is a carcinogenic secondary metabolite produced by the soil fungi *Aspergillus flavus* and *Aspergillus parasiticus* hereafter referred to jointly as *A. flavus*. The fungi infect and contaminate preharvest and postharvest peanuts with aflatoxin. The peanut industry-wide loss to aflatoxin in Georgia was estimated at 24% in 2019, and one peanut shelling company, Premium Peanut LLC, reported their losses to aflatoxin at \$150 million that year. Aflatoxin tolerances are particularly low in the EU, and loss of the European market to countries such as Argentina and China, which produce peanuts with lower aflatoxin risk, is a real threat. In the U.S., peanuts are grown primarily in the southeastern states. Georgia produced 52% of the U.S. peanut crop or 0.66 million Mg on 275,200 ha in 2022 (Georgia Peanut Commission 2022). Consequently, the potential socio-economic implications of aflatoxin contamination to Georgia and the southeastern U.S. are grave.

Spatial Distribution of Aflatoxin Contamination

Despite the ubiquitous presence of *A. flavus* in soil, extensive invasion by these fungi and contamination of the peanut crop with aflatoxin in the field occurs primarily when the plant is subjected to drought stress and high soil temperatures (Hill et al. 1983; Sanders et al. 1985; Clevenger et al. 2016). Peanuts grown under drought stress may also be predisposed to subsequent aflatoxin contamination during harvest, handling, and storage.

Previous work also showed that soil type and crop rotation affect colonization of peanut kernels by *A. flavus* (Griffin and Garren 1974; Abbas et al. 2004) and that populations of the fungus in soil exhibit a moderate degree of spatial structure (Abbas et al. 2004). Vellidis et al. (2007) found that aflatoxin contamination was spatially aggregated within a rainfed peanut field. In that study, aflatoxin levels measured on peanut kernels sampled systematically throughout the field were used to create an interpolated map of the aflatoxin distribution in the field, which showed several areas with high concentrations or “hotspots” (Fig. 1a). Maps of normalized difference vegetation index (NDVI) (Fig. 1b) and apparent soil electrical conductivity (ECa) (Fig. 1c) of the field were also developed. NDVI is an indicator of plant biomass and was assessed with a tractor-mounted multispectral sensor. ECa is a surrogate for soil texture with low ECa values indicating sandy soils and high ECa values indicating soils with higher clay content. Geostatistical analysis indicated spatial correlation among aflatoxin concentration, NDVI, and ECa. Generally, higher levels of aflatoxin were observed in areas with lower ECa and lower NDVI. This indicates that aflatoxin is more prevalent in areas that may have experienced physiological water stress due to the lower water-holding capacity and the drought-prone characteristic of sandier soils. Thus, the study by Vellidis et al. (2007) documented that aflatoxin is spatially clustered within peanut fields and may be correlated with easily measurable field parameters.

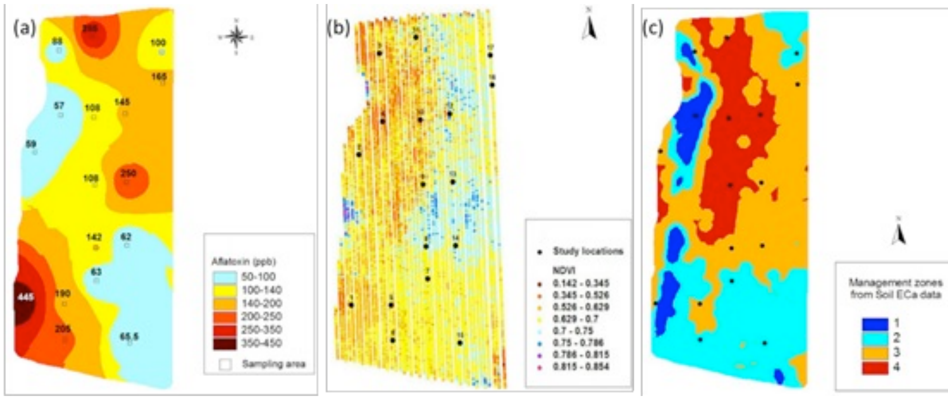

Fig 1. Aflatoxin concentration map from data collected in a 14-ha peanut field in Tift Co., GA in 2006 (a); NDVI map from a Crop Circle reflectance sensor (b); and soil ECa map created with a Veris 3100 EC mapper (c). Blues and orange/red indicate higher clay and sand content, respectively (Vellidis et al., 2007).

In the 15 years since the study by Vellidis et al. (2007), the tools available to collect spatially explicit data and to develop predictive models have improved dramatically. For example, unmanned aerial vehicles (UAVs) equipped with high-resolution sensors allow development of highly detailed maps of crop response in real time as well as accurate maps of the micro-relief of the terrain. Miniaturization of weather stations and wireless soil moisture sensing networks facilitate measurement of in-field microclimate and detailed mapping of soil moisture variability. Machine learning tools enable us to use large environmental data sets to develop predictive models whose performance improves as more data are added.

The goal of this project is to develop decision support tools (DSTs) that create aflatoxin risk maps in rainfed peanut fields. The maps may then be used by peanut growers to differentially harvest peanut fields to prevent cross-contamination from areas in the field in which aflatoxin may be present.

Materials and Methods

During the 2022 and 2023 growing seasons, three grower-managed rainfed peanut fields were selected for the study. The fields were in the Coastal Plain region of Georgia where most of the state's peanut production occurs.

Physical and Environmental Measurements

Soil apparent electrical conductivity (ECa) was collected in all fields using a Veris 3100 (Veris Technologies - Salina, KS) instrument with RTK guidance. Data were collected continuously in 9 m parallel swaths. In the configuration used, the Veris 3100 instrument collected integrated values of soil ECa for 0-0.3m and 0-0.9m. Soil ECa and elevation data were then used in the Management Zone Analyst (MZA) software (Fridgen et al. 2004) to create soil ECa and elevation-based management zones (MZs) in the field. MZA uses fuzzy means clustering to group like values. The MZs were used to assess the size of the sampling grid that was overlain on each field. Based on the observed spatial variability, a 0.5 ha (1.2 ac) sampling grid was selected for all three fields in 2022 while a 0.4 ha (1 ac) sampling grid was selected for all three fields in 2023. The size of the cells was optimized to be small enough to capture the spatial variability of the fields but also to account for the labor needed to collect data. The center point of each grid cell served as the sampling location for physical and biological measurements in the fields. A total of 66 sampling locations were established in 2022 and 77 in 2023. Fig. 2 shows the soil ECa map of the 10.5 ha 2023C field in southwestern Georgia overlain with the 0.4 ha grid and numbered sampling locations at the center of each grid.

Intact 90 cm soil cores were collected at the center point of each grid cell. Each core was divided into six 15 cm increments and each increment analyzed for texture), organic matter content (OM), pH, macronutrients (N, P, K, Ca, P) and micronutrients (Al).

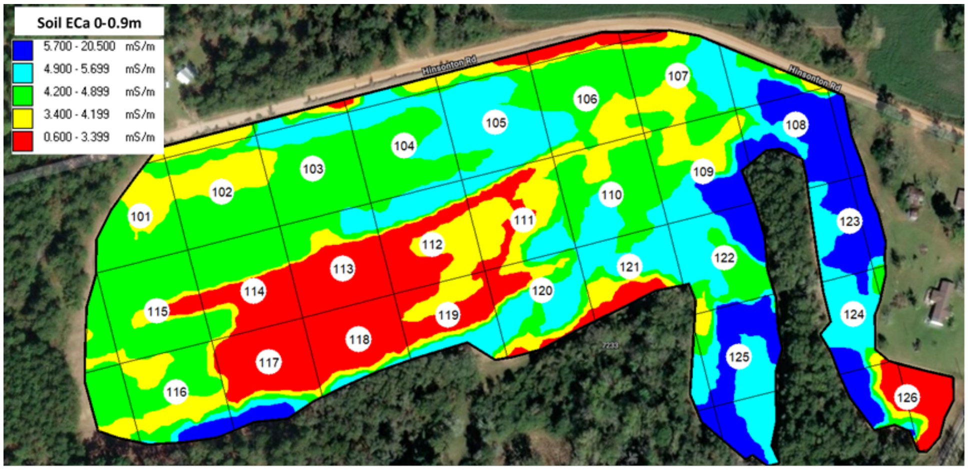

Fig 2. Soil ECa map of the 10.5 ha 2023C field in southwestern Georgia overlain with the 0.4 ha grid and numbered sampling locations at the center of each grid.

University of Georgia Smart Sensor Array (UGA SSA) soil moisture sensor nodes (Vellidis et al., 2008; 2013) were installed in 40 of the 66 of the grid cell center points in 2022 and 50 of the 77 in 2023. Locations were selected to represent a wide range of measured EC_a. The UGA SSA measures soil moisture in terms of soil water tension (the absolute value of soil matric potential). Each UGA SSA node measures soil water tension (SWT) hourly at three depths (10, 20, 40 cm) and soil temperature hourly at 5 and 10 cm. The UGA SSA nodes were used to measure SWT for the entire growing season (Maktabi et al., 2024).

Air temperature, precipitation, solar radiation, relative humidity, and wind speed among others were collected at 15 min intervals using an ATMOS 41W (METER Group, Inc., WA, USA) all-in-one compact weather station.

Physiological Measurements

Beginning with 60 days after planting (DAP), one group of plant physiological data were collected biweekly at all the sampling locations in 2022 and 2023. These data consisted of stomatal conductance, transpiration, leaf vapor pressure, minimum and maximum fluorescence in light and dark, quantum efficiency in dark and light, leaf light absorptance, and in leaf area index (LAI) (Sysskind, 2024). During each sampling event, peanut plants were carefully evaluated for their phenological stage (emergence, flowering, pegging, beginning pod, beginning seed, seed maturity) (Sysskind, 2024).

Beginning in 2023, a second group of plant physiological data were collected at approximately 60% of the grid sampling points. As with the soil moisture sensor nodes, sampling locations were selected to represent a wide range of measured EC_a. These data consisted of whole plant samples collected biweekly in all fields. Three whole peanut plants were collected and divided into leaves, stems, pods, and seeds. Leaf area was measured and then all plant components were oven dried to determine dry matter (Maktabi et al., 2024).

Beginning with approximately 60 DAP, whole plant samples were collected for aflatoxin analyses from each sampling location during the biweekly sampling events. Peanut pods were detached from the plants and sent to a commercial laboratory. Aflatoxin concentrations were measured using the Enzyme-Linked Immunosorbent Assay (ELISA) method (Hidayat and Wulandari, 2021).

Remotely Sensed Data

Planet Explorer (Planet Labs, San Francisco, CA) makes available multispectral imagery of the earth's surface at a spatial resolution of 3 m and temporal resolution of one day. This platform was used to download multispectral images of the study fields on or around the field sampling days. When cloud-free images were not available for the sampling day, images from ±2 days from the sampling date were used. These multispectral images reported reflectance data in four wavebands: Blue (455-515 nm), Green (500-590 nm), Red (590-670 nm) and Near Infrared (780-860 nm). The reflectance rasters were used to develop NDVI for each field for each sampling date. The NDVI values were in a 3 m raster that corresponded to the spatial resolution of the images. The rasters were used to extract NDVI values for each sampling area by creating a 6 m buffer around the grid cell center points and taking the arithmetic mean of all the raster NDVI values within the buffer (Kukal, 2024).

Yield Data

Yield data were collected beginning in 2023 by harvesting 30 m (100 ft) of one peanut bed (2 rows), 15 m (50 ft) on each side of the sampling point, with a 2-row bagging combine. Two or three bags were collected at each sampling point. The bags were weighed in the field and samples of peanut pods were extracted from one bag for aflatoxin analysis, burrower bug damage assessment, moisture content, and foreign material content (Kukal, 2024).

Decision Support Tools (DSTs)

Two types of mathematical simulation approaches were evaluated for their potential to create aflatoxin risk maps. These were application of a machine learning technique and application of an established peanut crop growth model.

Random Forest Regression with Recursive Feature Elimination

The machine learning technique was a Random Forest regression model approach that used physical and environmental measurements (soil texture, meteorological parameters, soil water tension (SWT), soil temperature) and NDVI. The only physiological measurement used was aflatoxin concentration. The model was developed using the “randomForest” package in RStudio. Data from all three 2023 fields were pooled and split into 75% training and 25% testing sets. The model included a loop to iterate over 25250 combinations of values for the number of trees, number of variables split, and maximum number of nodes in a tree.

A Random Forest regression model alone was not capable of subsetting important variables from the original predictor set that contribute to the skill of the model. Although the Random Forest modeling has the capability to recognize and delineate inter-variable collinearity by selecting a subset of the variables at each node-split in each tree, it was not able to not only identify the predictors that were actually adding to the variance explained in the aflatoxin concentration, but also eliminate predictors that were not adding any unique information to the model or, in cases, were a source of noise in the dataset. To do so, the Random Forest model was combined with a more-informed feature engineering approach called Recursive Feature Elimination (RFE). RFE was implemented using a “RecursiveFeatureElimination” function from a Python library called “feature_engine” (Kukal, 2024).

The models developed were evaluated based on the R^2 value obtained in each iteration and a three-fold cross validation was used to increase the robustness of the model. Using the mean R^2 value from the three-fold cross validation model as a baseline, certain variables were retained or dropped from the predictor set based on their contribution to the model R^2 values relative to the baseline. Once the reduced predictor set was identified, these predictors were used in the Random Forest model to predict aflatoxin concentration. The predictions using the engineered set of predictors were compared against independent testing data to document model performance (Kukal, 2024).

DSSAT-CROPGRO-Peanut

The Decision Support System for Agrotechnology Transfer (DSSAT) is a universally used DST that includes dynamic crop growth simulation models for over 42 crops (Boote, 2019). DSSAT has been well calibrated for a variety of crops and cultivars, allowing the users to simulate the growth and development of the crop of interest under different management practice scenarios and environmental conditions. DSSAT-CROPGRO-Peanut is a model within DSSAT that has been used to predict peanut development and yield under various environmental and management conditions (Hoogenboom et al., 2012). The model requires layer-wise soil data including physical and chemical properties and soil texture as well as daily weather data including maximum and minimum air temperature, solar radiation, relative humidity, and precipitation. DSSAT CROPGRO-Peanut-Aflatoxin developed by Boote (2018) has been used to predict aflatoxin production by associating it with model variables such as water stress occurrence during the water deficit sensitive periods such as pod filling (Maktabi, 2024).

CROPGRO-Peanut-Aflatoxin was applied to the MZs of the 2023 fields to test the model's ability to predict the spatial distribution of aflatoxin concentrations in rainfed peanut fields. The model was calibrated and evaluated using data from the 2023 fields (Maktabi et al., 2024).

Results and Discussion

Aflatoxin concentrations from the three 2023 fields were used for both modeling approaches. A total of 318 samples were collected and analyzed for aflatoxin concentrations from these fields. Overall, aflatoxin concentrations were low. Two hundred fifty-six (80.50%) of the samples were within the ELISA method's quantitation range while 62 (19.5%) were below the method's official detection limit (1.4 ppb) but still enumerated through the use of repetitively measured low standards. The mean concentration of the samples was 0.84 ppb, standard deviation was 0.75 ppb and the maximum measured concentration was 3.5 ppb.

Random Forest Regression with Recursive Feature Elimination

The model was trained and tested using the pooled data from all three 2023 fields. The RFE approach reduced the relatively large number of variables to fewer, more robust predictors. The final set of variables in the order of their importance based on the change in R^2 was %silt at 60-75 cm of depth, vapor pressure deficit, %clay at 60-75 cm of depth, %sand at 60-75 cm of depth, %silt at 30-45 cm of depth, soil temperature, solar radiation, and air temperature. Using these eight predictors, resulted in improved model performance (R^2 = 0.27 and RMSE = 0.65) compared to the original Random Forest model (R^2 = 0.16 and RMSE = 0.74) (Kukal, 2024). Fig 3. Shows a scatter plot of observed and predicted values of aflatoxin concentration from the developed Random Forest model following RFE.

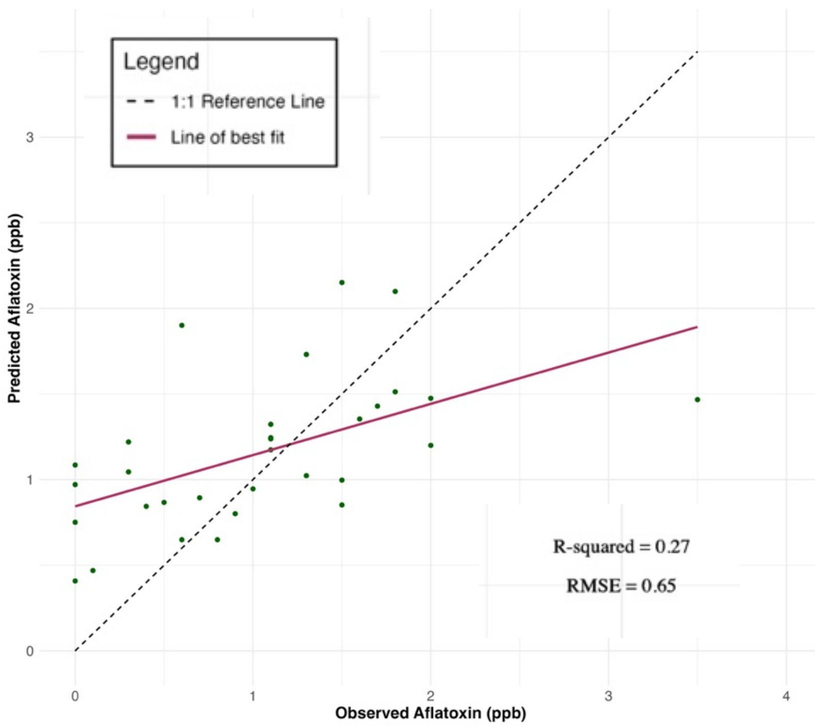

Fig 3. Scatter plot of observed and predicted values of aflatoxin concentration from the developed Random Forest model following RFE (Kukal, 2024).

The final eight predictors consisted of weather variables, soil texture parameters and soil temperature. If soil temperature could be eliminated, then the model could be run without in-season field measurements. To evaluate the performance of the model without soil temperature as a predictor, another model was developed. This model resulted in a $R^2 = 0.13$ and RMSE of 0.84 indicating that, as expected, soil temperature is a critical variable (Kukal, 2024).

The Random Forest model with RFE explained 27% of aflatoxin concentration variance using 8 variables some of which can be measured once for each field (soil texture) or easily measured during the growing season (meteorological data and soil temperature). This provides some promise that with further refinement with data from additional fields, this type of model may be able to predict aflatoxin occurrence at the field level with a reasonable degree of confidence (Kukal, 2024). At this point, the data requirements to train this type of model likely precludes it from being used to create within-field risk maps. It is better suited to larger-scale applications.

DSSAT-CROPGRO-Peanut

Fig. 4 shows the simulated versus observed values for aflatoxin content in the two MZs of Field 2023C. Because measured aflatoxin values were low, differences between MZs are not visually noticeable because of the scale of the y-axis. Nevertheless, the spatial distribution of aflatoxin across different zones occurred and was influenced by variations in soil texture and the resulting differences in water and heat stress levels. DSSAT simulations also showed different aflatoxin concentrations between MZs. For field 2023C, DSSAT predicted that the MZ consisting primarily of loamy sand resulted in higher season-end aflatoxin concentrations. However, the magnitude of the predicted aflatoxin concentrations was much higher than measured in the field indicating that the aflatoxin module requires additional calibration with more years of data (Maktabi et al., 2024). Simulated aflatoxin concentrations in MZs increasingly diverged as harvest approached indicating that DSSAT may be a useful tool in creating aflatoxin risk maps. The model will be further calibrated from data collected in three additional grower-managed rainfed fields in 2024.

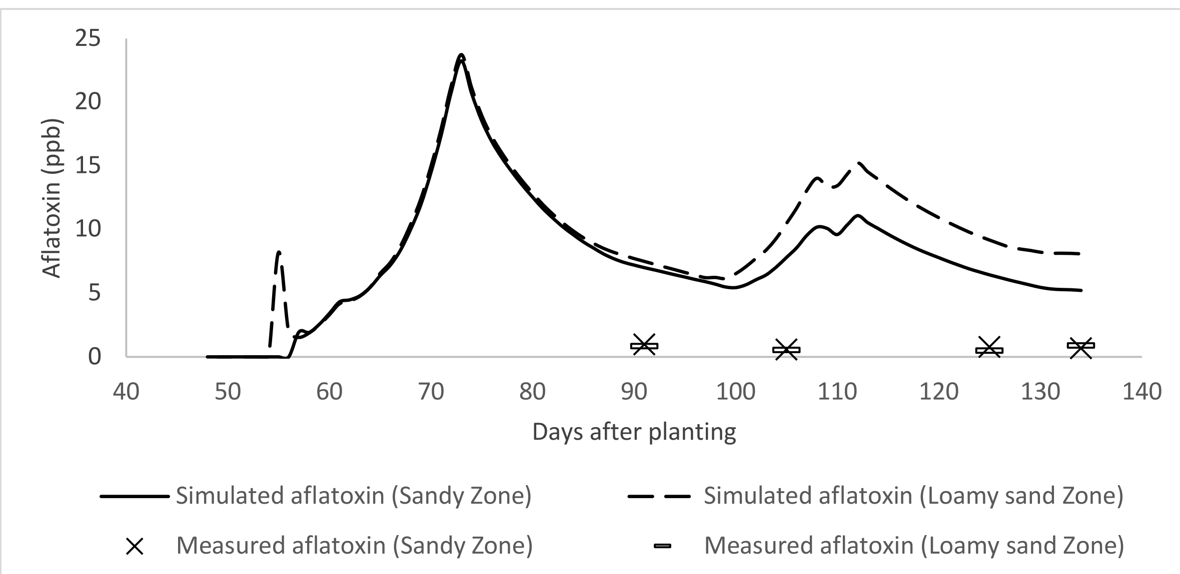

Fig 4. DSSAT-simulated and measured aflatoxin concentrations (ppb) in two zones of Field 2023C (Maktabi et al., 2024).

Acknowledgments

We would like to thank the peanut corn growers that allowed us to collect data on their farms, Premium Peanut for connecting us to the growers, and Mr. Matt Gruver, Mr. Rodney Hill, and the other UGA staff and students who contributed to field and lab work. Funding for this project was provided by the National Peanut Board, the Georgia Peanut Commission, and the USDA-ARS National Peanut Research Laboratory.

References

Abbas, H.K., Zablotowicz, R.M., and Locke, M.A. 2004. Spatial variability of Aspergillus flavus soil populations under different crops and corn grain colonization and aflatoxins. Can. J. Bot. 82:1768-1775.

Boote, K. J. (2018). Documentation of CROPGRO-Peanut-Aflatoxin Model.

Boote, K. (Ed.). (2019). *Advances in crop modelling for a sustainable agriculture* (1st ed.). Burleigh Dodds Science Publishing.

Boote, K. (Ed.). (2019). Advances in crop modelling for a sustainable agriculture (1st ed.). Burleigh Dodds Science Publishing. https://doi.org/10.1201/9780429266591

Clevenger, J., Marasigan, K., Liakos, V., Sobolev, V., Vellidis, G., Holbrook, C., and Ozias-Akins, P. 2016. RNA sequencing of contaminated seeds reveals the state of the seed permissive for pre-harvest aflatoxin contamination and points to a potential susceptibility factor. Toxins 8(11):317.

Georgia Peanut Commission. (2022). Georgia Peanut Commission statistics. https://gapeanuts.com/about/

Griffin, G.J., and Garren, G.H. 1974. Population levels of *Aspergillus flavus* and the *A. niger* group in Virginia peanut field soils. Phytopathology 64:322-325.

Hidayat, R., & Patricia Wulandari. (2021). Enzyme Linked Immunosorbent Assay (ELISA) Technique Guideline.

Hill, R.A., Blankenship, P.D., Cole, R.J., and Sanders, T.H. 1983. Effects of soil moisture and temperature on preharvest invasion of peanuts by the *Aspergillus flavus* group and subsequent aflatoxin development. Appl. Environ. Microbiol. 45(2):628-633.

Hoogenboom, G., Jones, J. W., Traore, P. C. S., & Boote, K. J. (2012). Experiments and Data for Model Evaluation and Application. In J. Kihara, D. Fatondji, J. W. Jones, G. Hoogenboom, R. Tabo, & A. Bationo (Eds.), Improving Soil Fertility Recommendations in Africa using the Decision Support System for Agrotechnology Transfer (DSSAT) (pp. 9–18). Springer Netherlands. https://doi.org/10.1007/978-94-007-2960-5_2

Kukal, S. 2024. A machine learning regression model for predicting the occurrence of aflatoxin in peanut fields using environmental and field-dependent variables. M.S. thesis, University of Georgia, 148p.

Maktabi, S. 2024. Using the DSSAT-CROPGRO-PEANUT crop growth model to develop spatially explicit aflatoxin risk maps in rainfed peanut fields. Dissertation research proposal, Crop and Soil Sciences Department, University of Georgia, 42p.

Maktabi, S., K. Boote, J. Fountain, G. Hoogenboom, S. Kukal, C. Pilon, M. Sysskind, G. Vellidis. 2024. Predicting the Spatial Distribution of Aflatoxin Hotspots in Peanut Fields using DSSAT CSM-CROPGRO-PEANUT-AFLATOXIN. Proceedings of the 16th International Conference on Precision Agriculture, 21-24 July, Manhattan, KS, USA, 9p.

Sanders, T.H., Cole, R.J., Blankenship, P.D., and Hill, R.A. 1985. Relation of environmental stress duration to *Aspergillus flavus* invasion and aflatoxin production in preharvest peanuts. Peanut Sci. 12:90-93.

Schmale III, D.G., and Munkvold, G.P. 2009. Mycotoxins in crops: A threat to human and domestic animal health. Plant Health Instructor, https://doi.org/10.1094/PHI-I-2009-0715-01

Sysskind, M. 2024. Spatial and temporal variability in plant physiology and association with aflatoxin development in peanut. M.S. thesis, University of Georgia, 86p.

Vellidis, G., Ortiz, B., Renga, M., Perry, C., Rucker, K., & Morari, F. (2007). Spatial distribution of aflatoxin in growing peanut. In Poster Proceedings of the Sixth European Conference on Precision Agriculture (6ECPA), Skiathos.

Vellidis, G., Tucker, M., Perry, C., Reckford, D, Butts, C., Henry, H., et al. 2013. A soil moisture sensor-based variable rate irrigation scheduling system. In: J.V. Stafford (Ed.), Precision Agriculture 2013. Wageningen Academic Publishers, Wageningen.

Vellidis, George, M. Tucker, C. Perry, C. Kevin, and C. Bednarz. 2008. “A Real-Time Wireless Smart Sensor Array for Scheduling Irrigation. Computers and Electronics in Agriculture 61 (1): 44–50. https://doi.org/10.1016/j.compag.2007.05.009.