Analyzing Landscape Effects on Corn and Soybean Yield and Yield Risk From a Large Yield Monitor Dataset

D. Brenton Myers[1], Newell R. Kitchen[1], Kenneth A. Sudduth[1], Paul W. Tracy[2] and Fred J. Young[3]
[1]USDA-ARS Cropping Systems & Water Quality Research Unit, Columbia, MO, [2]MFA Incorporated, Columbia, MO, [3]USDA-NRCS, Columbia, MO

Introduction

Spatiotemporal crop yield variability is due to a variety of factors. Some, such as genetics, weeds and pests, drainage, irrigation, and nutrient supply have management opportunities. However factors such as soil physical properties and landscape position cannot be managed. Even manageable factors can have difficult interactions with climate or landscape. The interactions between plant, soil, and landscapes are significant unknowns that producers face, and are major causes of yield risk.

Until the advent of precision agriculture, most field or plot experiments designed to understand these spatiotemporal interactions have taken a plot, or single field measurement approach. Even after yield monitors have become common, many studies have relied on yield data from one or a few fields only. The collection of yield-monitor data from producers over large regions into data warehouses offers a new avenue to explore complex interactions.

Strategy and Objectives

Our strategy is to use the multi-temporal and spatial replication of crop yield monitor data to empirically model productivity and the risk to productivity due to soil and landscape factors. The general approach (fig. 1) is to collect yield, soil, and landscape data (continuous and full coverage), merge these, then model yield and yield variance with data mining techniques. Finally, using the full coverage soil landscape data layers the model is applied throughout the study area. Our objective is to produce regional coverage maps of mean yield and yield risk for Northeast Missouri.

Modeling Approach

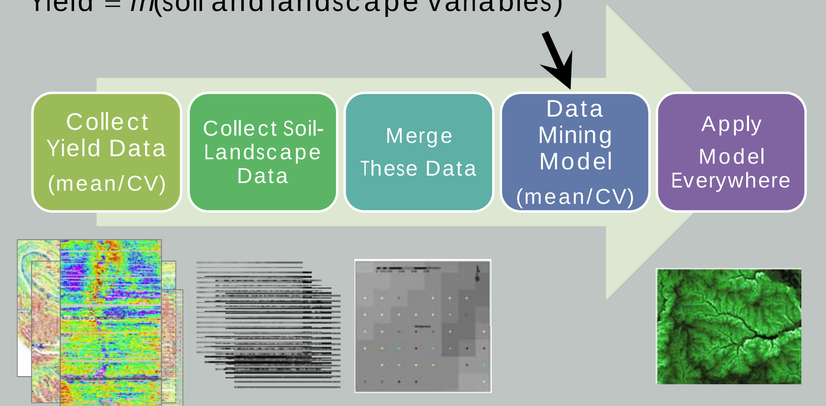


Figure 1. Data mining is used to model yield and yield variability (mean and coefficient of variation) as a function of soil landscape variables.

Study Area

The Northeast Missouri Central Claypan Area (fig. 2) has a flat to gently rolling topography and a variable depth to a restrictive argillic horizon, or claypan. Depth to the claypan interacts with landscape position and controls many factors such as plant available water capacity, cation exchange capacity, runoff and infiltration, and root exploration. These landscapes are commonly used for corn and soybean production and express spatial and temporal variability in yield correlated to landscape morphology (Kitchen et al. 2005).

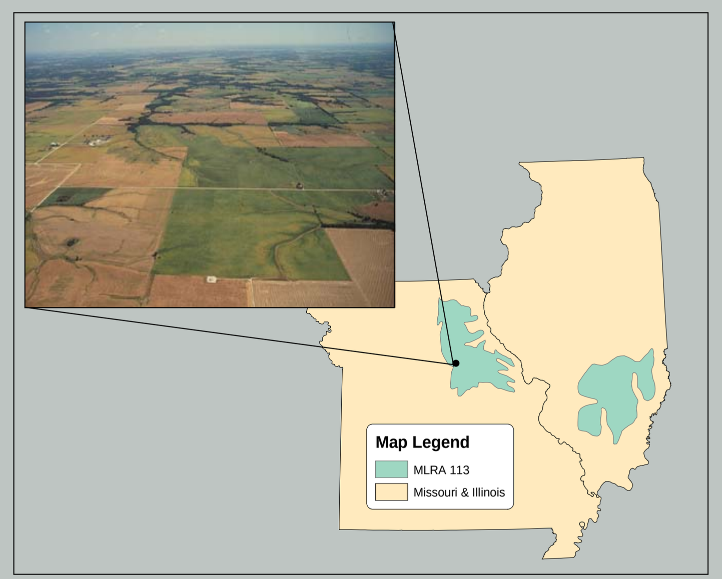


Figure 2. The Central Claypan Areas are the setting for this research.

Soil-Landscape Variables

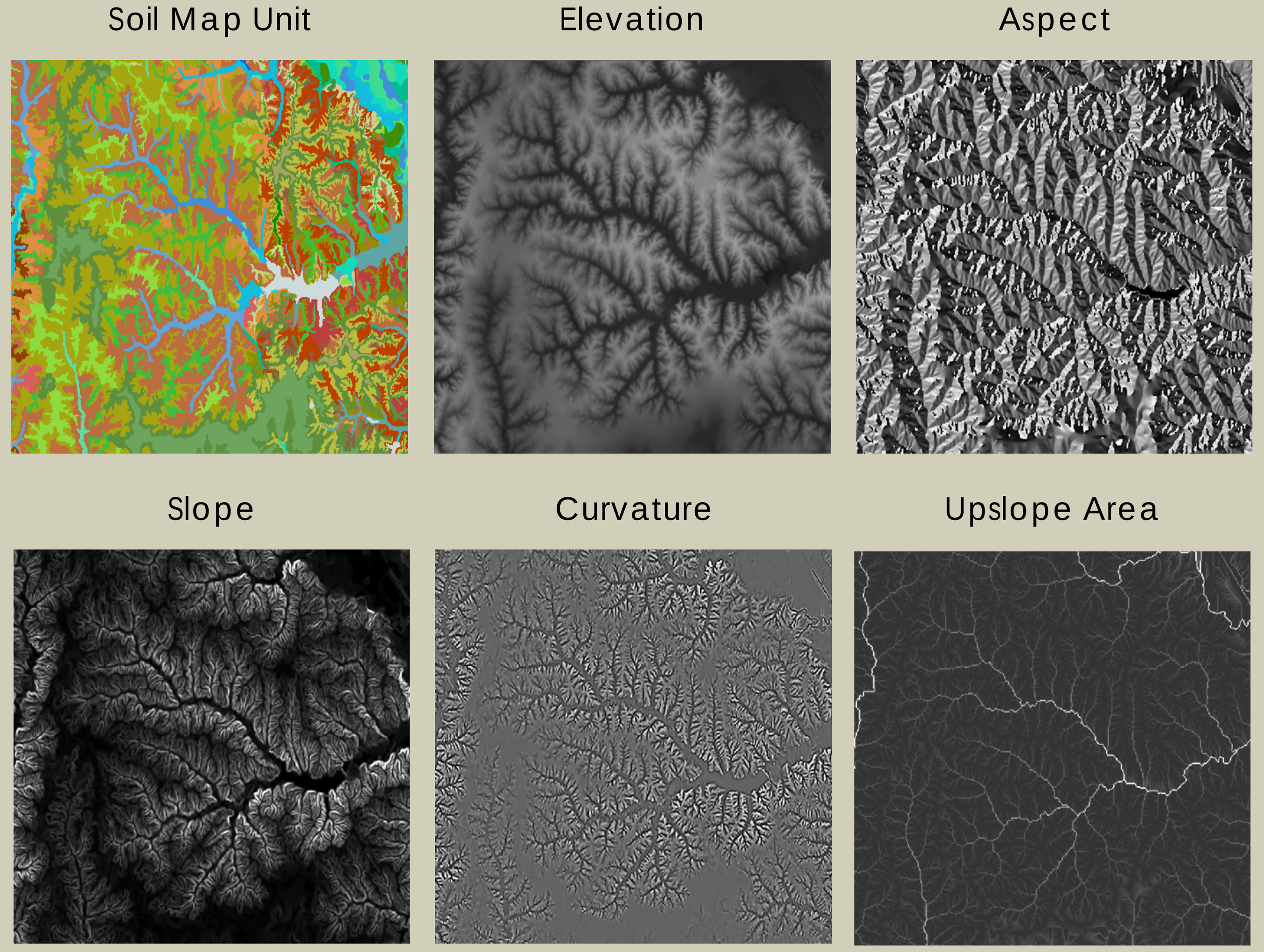


Figure 3. A representative set of soil landscape variables used to model yield and yield coefficient of variation (CV).

Methods and Data

A large database of corn and soybean yield monitor data was collected from cooperating producers (see 'Yield Data' sidebar, far right column). This preliminary analysis uses a smaller initial portion of this dataset (table 2). Yield maps are first cleaned using the Yield Editor software (Sudduth and Drummond, 2007; see 'Yield Map Cleaning' in the far right column). Next they are converted to raster images with a 30 meter pixel resolution by simple averaging of yield points. Mean yield and coefficient of variation (CV) across years were calculated at each pixel location separately for corn and soybean.

Soil-landscape variables were used to develop models of yield and yield CV. These variables included SSURGO map-unit and digital elevation model (DEM) derivatives such as: elevation, local elevation, slope, curvature (average, profile and planform), aspect, and upslope area. These soil-landscape variables were then combined with the yield statistics and split into training and testing datasets (50%/50%). Random forest models (Breiman, 2001) were fit to the training data and fit statistics were calculated from the test data. Variable importance was calculated for each model to examine which predictors supplied the most information to explain yield and yield CV.

Results

Random forest model results developed from the preliminary dataset were successful at estimating mean yield and yield CV for the Central Claypan Area (fig. 4 a-d, table 1). Elevation (actual and local), slope, aspect, and soil map-unit were the most important variables to model both corn and soybean yield and yield CV (fig. 4 e-h). Additional yield data will be used to update these models as it is available.

Random Forest Models

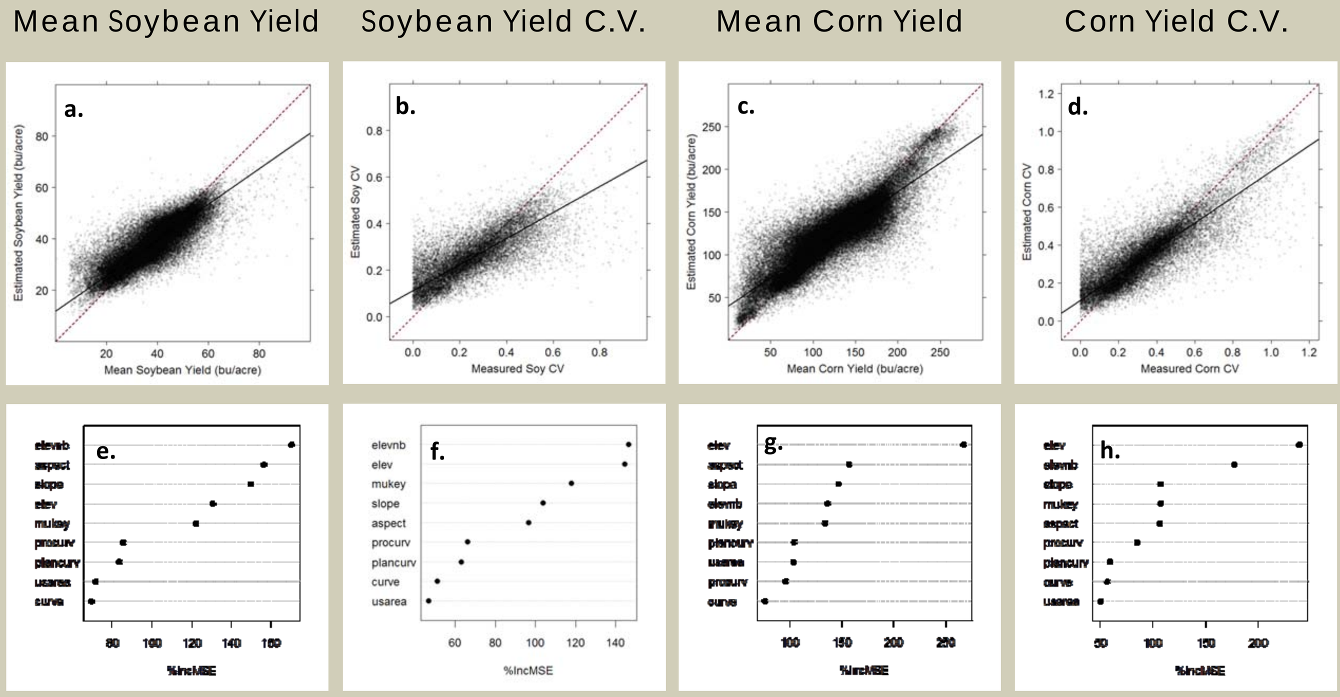


Figure 4. Random forest models were developed for the estimation of yield and yield coefficient of variation (CV) from soil-landscape properties. Test dataset results show the fit of these models (a-d) and their variable importance statistics (e-h) for corn and soybean. Percent increase in mean squared error (%IncMSE) is the improvement in MSE contributed by a variable as compared to its random permutation. Variable codes are elevation (elev), local elevation (elevnb), soil map-unit (mukey), slope, aspect, curvature, profile curvature (procurve), planform curvature (plancurv), and upslope area (usarea).

Table 1. Training and test statistics for random forest models fit to mean yield and coefficient of variation of yield for both corn and soybean. Grain yields are measured in bu a^{1}.

		Train		Test	
Crop	Variable	R^2	RMSE	R^2	RMSE
Soybean	Mean	0.66	7.0	0.67	7.0
	C.V.	0.55	0.10	0.58	.11
Corn	Mean	0.75	24.4	0.70	27.2
	C.V.	0.71	0.10	0.66	0.13

Conclusions

These results translate into a capability to produce regional maps of yield and yield CV and provide producers a risk assessment and management tool. Further, an objective assessment of profit sustainability can be used to evaluate suitable alternative grain or biofuel cropping systems and/or conservation practices for marginal sites.

Acknowledgements

Project funding provided by a USDA-Missouri Conservation Innovation Grant and by USDA-ARS. We would like to thank the large crew of yield cleaners: Lindsay, Matt, Alec , Wes, Eric, Jake. We would also like to thank Rick Green, and Chris Pund who were instrumental to help collect yield data.

References

Breiman, L. 2001. Random forests. Machine learning. 45(1): 5–32.
Drummond, S., and K. Sudduth. Yield Editor Version 1.02 Beta. Available at http://www.ars.usda.gov/services/software/download.htm?softwareid=20&modecode=36-20-15-00 (verified 26 October 2010).
Kitchen, N., K. Sudduth, D. Myers, S. Drummond, and S. Hong. 2005. Delineating productivity zones on claypan soil fields using apparent soil electrical conductivity. Computers and Electronics in Agriculture. 46(1-3): 285-308.
Sudduth, K.A., and S.T. Drummond. 2007. Yield Editor. Agronomy Journal. 99(6): 1471.

Yield Data

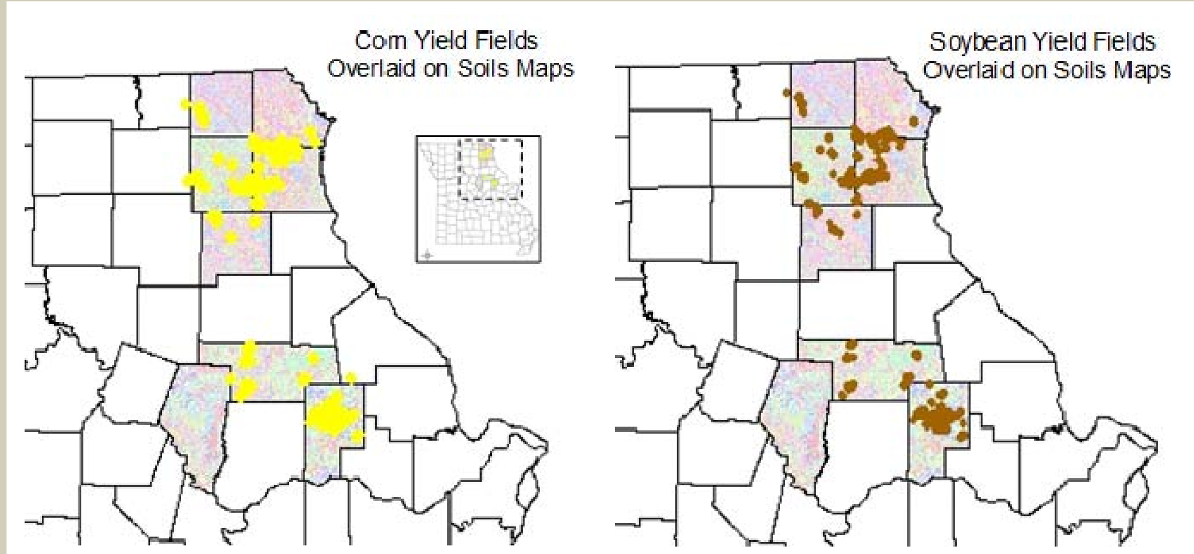
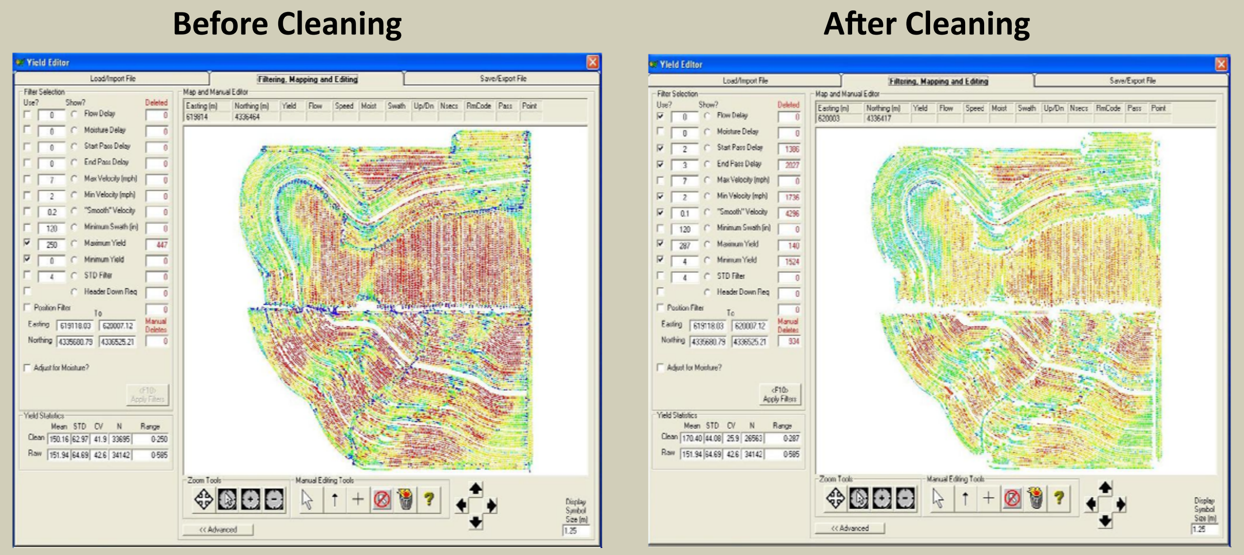


Figure 5. Approximate distribution of corn and soybean yield maps. Locations are generalized to protect the identity of participating cooperators.

Table 2. Summary of progress for yield data acquisition, cleaning, and analysis. The model results reported in this poster were produced from the preliminary dataset. Yield data cleaning and acquisition is ongoing.

	Crop	Preliminary Dataset	Current Progress	Target Dataset
Producer Cooperators	Corn and Soy	10	20	25
Field-Years Cleaned	Corn	192	544	900
	Soybean	239	707	1,100
Acre-Years Cleaned	Corn	10,944	31,008	50,000
	Soybean	13,862	41,713	65,000

Yield Map Cleaning



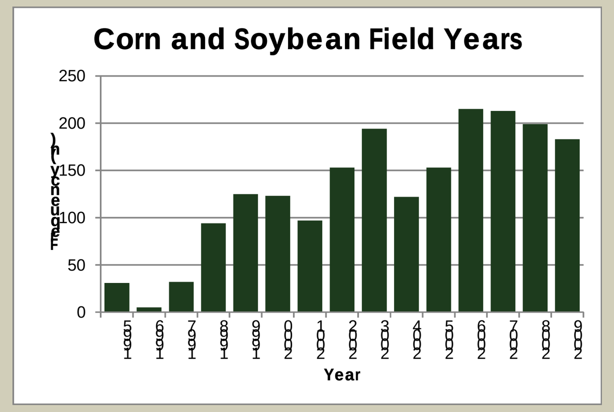


Figure 6. Yield data are collected from producers in raw format. Raw yield data contains many errors due to crop row configuration (point rows and end rows), grain flow delays (start, end, thresher), header switch, swath width, and position errors. Yield Editor (Sudduth and Drummond, 2007) is a software program developed to streamline correction of these issues. The software is available at: http://www.ars.usda.gov/Services/Services.htm?modecode=36-22-15-00 . The panels above show a yield map before (upper left) and after (upper right) the cleaning process. The histogram to the right shows the per-year frequency of field years of yield data currently in our database (cleaned and uncleaned).