water

Supporting Information for

Increased Dependence on Irrigated Crop Production Across the CONUS (1945-2015)

Samuel J. Smidt [1], Anthony D. Kendall [2] and David W. Hyndman [2]

[1] Soil and Water Sciences Department, University of Florida, Gainesville, FL, 32611, USA, ssmidt@ufl.edu
[2] Department of Earth and Environmental Sciences, Michigan State University, East Lansing, MI, 48824, USA, kendal30@msu.edu, hyndman@msu.edu

Contents of this file

S1. Model Limitations

A limitation to the applicability of this model is that crop yields are highly variable, thus finding several statistically significant predictor variables available at every year and for every state can be difficult. For example, a commodity like sorghum may have strong correlation for estimating irrigated corn, but its data are spatially and temporally discontinuous, rendering sorghum unusable as a predictor variable. This study originally included sorghum in the analysis, but a statistically significant correlation could not be determined using the available predictor variables, rendering sorghum unusable as both a predictor variable and analyzed commodity. This method may be successfully applied to other commodities, but new predictor variables may be necessary for statistically correlated yields across the CONUS.

This model is based on the linear trends observed in observed crop yields since 1945. In theory, this model approach will continue to increase linearly into the future. However, commodity yields likely have an upper yield limit based on seed type, technology, and climate conditions. While this method may be used to predict into the near future, extended yield predictions will need to consider the physical limitations to crop growth.

S2. Model Validation

In addition to validation with even year data, the model was validated by comparing estimated values to: 1) the 2013 Farm and Ranch Irrigation Survey dryland and irrigated yield data, and 2) the observed total annual production data from the NASS Survey [13]. The 2013 Farm and Ranch Irrigation Survey, part of the 2012 Agricultural Census [28], is the only comprehensive dataset for CONUS dryland and irrigated yields. The difference between our estimated 2013 yield values and 2013 Farm and Ranch Irrigation Survey data averaged across crops were -1.8% for dryland and 12.6% for irrigated crops (Table S1), where positive differences represent overestimation by our model while negative represents underestimation. Given that yield is a widely variable statistic based on many natural and artificial drivers, the low percent differences between estimated and observed production established confidence in the model estimates. Dryland yields had closer agreement than irrigated yields because irrigation has a wider range of effects on yields across mixed state, where the upper limits to dryland yields are limited by precipitation. Corn and soybeans were the commodities in closest agreement, likely due to consistent farming strategies and soil parameters across regions, whereas cotton and hay had the most disagreement, likely due to more variable growing conditions and productivity of soil types.

To evaluate the estimated irrigated and dryland acreages, we compared them relative to observed production data. The 2013 Farm and Ranch Irrigation Survey contains irrigated and dryland areas, but there is much disagreement between the 2013 NASS Survey and the 2013 Farm and Ranch Irrigation Survey (e.g., the best agreement between these surveys was only 58% for wheat acreages). As a result, observed production was calculated as the 2013 NASS Survey yield values multiplied by the area for each state and commodity. Estimated production was calculated by multiplying estimated dryland and irrigated yields by the estimated dryland and irrigated areas, respectively, for each commodity and state. Estimated production closely follows observed production (Figure S1), including specific peak patterns. Given that estimated production relies on highly variable yields and estimated areas derived from these, the strong correlation between estimated and observed production indicates a robust model. The largest differences between observed and estimated production are for hay and wheat post-1990, which have the greatest variability in upper limits of irrigated yield across mixed states.

S3. Additional Model Description

A list of all predictor variables considered, variables used for each commodity and condition, and variables listed in order of statistical significance (p-value) are displayed in SI Table 2. As demonstrated by 6 out of the 10 models, time proved a predominant predictor variable. We conducted a two-step regression to evaluate the influence of time in predicting commodity yields and to ensure other variables were meaningful predictors. The first step ran each linear model using time as the only predictor variable using

$$Y_E = b_0 + b_1 T \quad \text{(S1)}$$

where Y_E is estimated yield, b_0 is the y-axis intercept, b_1 is the derived coefficient for the corresponding predictor variable, T, which is time. We then subtracted the reported yield values from the time-only estimated values to extract the residuals of the model using

$$R = Y_E - Y_O \quad \text{(S2)}$$

where R is the residuals, and Y_O is the observed yield. The second regression step ran all other variables against the residuals of the first step using

$$Y_E = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n \quad \text{(S3)}$$

where b_n is the derived coefficient for the corresponding predictor variable, x_n. The results for corn are displayed in Figure S3. We found time is less predominant in dryland systems, which are influenced more heavily by other yield-limiting factors (e.g., drought). However, time alone is a relatively strong predictor in irrigated areas, as irrigation reduces the yield-limiting impacts of outside climatic forces. We also conducted an uncertainty analysis to confirm that our variable selection accurately captured the primary drivers to crop production prediction. Using a 95% confidence interval, we plotted the residual interval data for each commodity and condition to identify the outliers in our models (Figure S3). We found that, proportionally, few outliers exist relative to the data within the confidence interval, confirming our variables were effective in predicting crop production for all five commodities and conditions. We also identified random distribution of the data, which confirms our linear model approach as an appropriate and robust method for this type of analysis.

S4. Observed Census Data vs. Model Estimates

Compared to the Census data (1997-2012 in 5-yr intervals), our approach generally provides accurate estimates of the reported values including the peaks (e.g., Figure 4A, 2007) and the dips (e.g., Figure 4E, 2002). We overestimate total irrigated cotton area relative to the Census data. However, it is unlikely that only ~30% of all cotton area in mixed states is irrigated as reported in the Census, since the 2013 Farm and Ranch Survey reports that 58% of cotton area is irrigated across the same region; we estimate that 71% of cotton area is irrigated in 2013. Furthermore, our model *underestimates* high-yield cotton (Figure 3), confirming that our *overestimate* of irrigated cotton area is related to uncaptured or under-reported irrigated land in the Census data (i.e., since our model underestimates high-yield cotton, we would expect to also underestimate total area). Our irrigated soybean area is also underestimated from 1997-2007. This is confirmed by our recent underestimate of soybean production (Figure S1). Both the Census data and our model capture the recent increase in irrigated soybean area, though the economic implications of irrigated soybeans may be more heightened than described here, as evidenced by our early underestimation of irrigated soybeans compared to the reported the Census data.

S5. Irrigation Enhancement Revenue

Given that revenue is directly correlated to market values, the wide variability seen across commodities can be partially attributed to fluctuations in annual market prices. For example, corn prices peaked in the early 1970's, which correlates with the first major peak in revenue enhancement (Figure 7). There was a subsequent lag between the early 1970's when commodity prices were highest and irrigation revenue, as farmers continued to irrigate at high production rates throughout the 70's. As commodity prices rapidly declined through the 80's and 90's, so did the widespread intensity of irrigation applications. However, other peaks are found outside of the early 70's (i.e., in times outside of high market prices) indicating that other factors also influence revenue. For example, other peaks may be attributed to technology changes that led to increased irrigated yields relative to dryland yields, such as new seed cultivars, or a reduction in the overall variability of crop production, which ultimately can lead to a more consistent irrigated yield relative to dryland yield. It is also likely that after the spike in commodity prices during the 70's, farmers with irrigation systems were ultimately able to revamp applications once production costs declined due to technological advancements. Thus, peaks in irrigation enhancement revenue are due to: 1) increases in crop price, where any irrigation enhancement is magnified by dollar amounts (e.g., 2 kg x $2/kg = $4 vs. 2 kg x $5/kg = $10), or 2) an increase in yield margin relative to dryland yields (e.g., 2kg x $2/kg = $4 vs. 5kg x $2/kg = $10), which inherently can include changes in climate conditions like droughts.

S6. Data Availability

All original coding in MATLAB is stored on GitHub, and all data outputs are stored on the CUAHSI HydroServer.

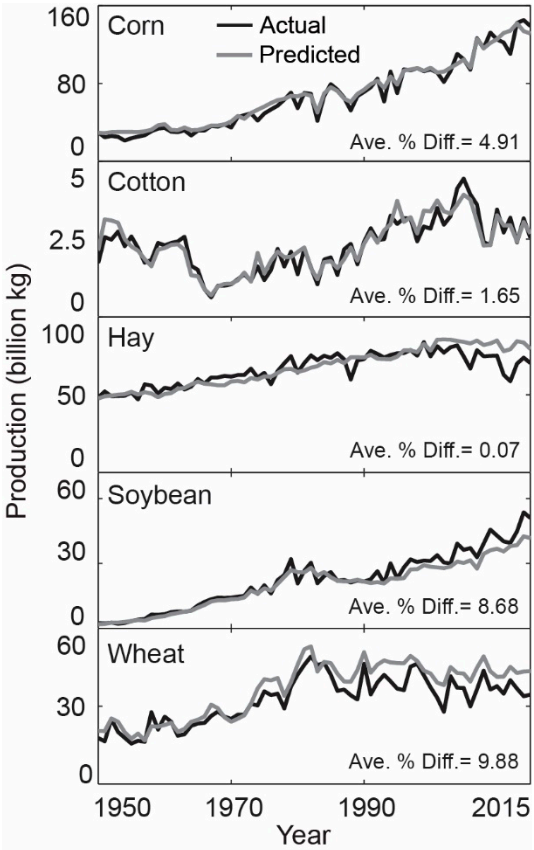

Figure S1. Observed total annual production compared to estimated (dryland + irrigated) production for each commodity across all mixed states.

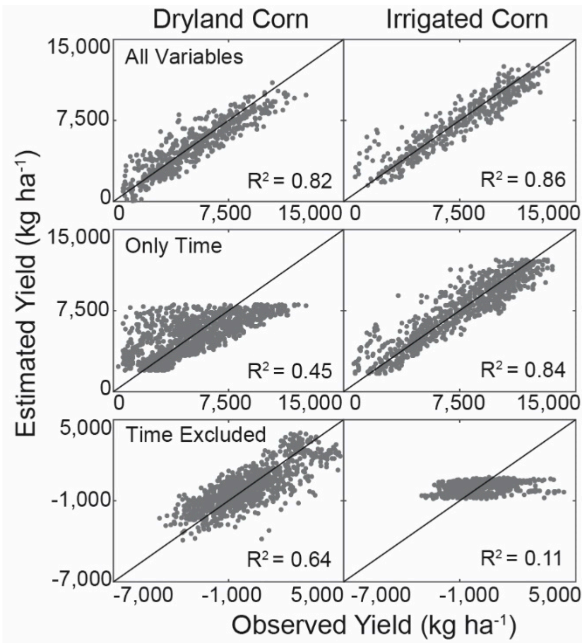

Figure S2. Two-step regression isolating time as a predictor variable relative to all variables used in the estimation of corn production.

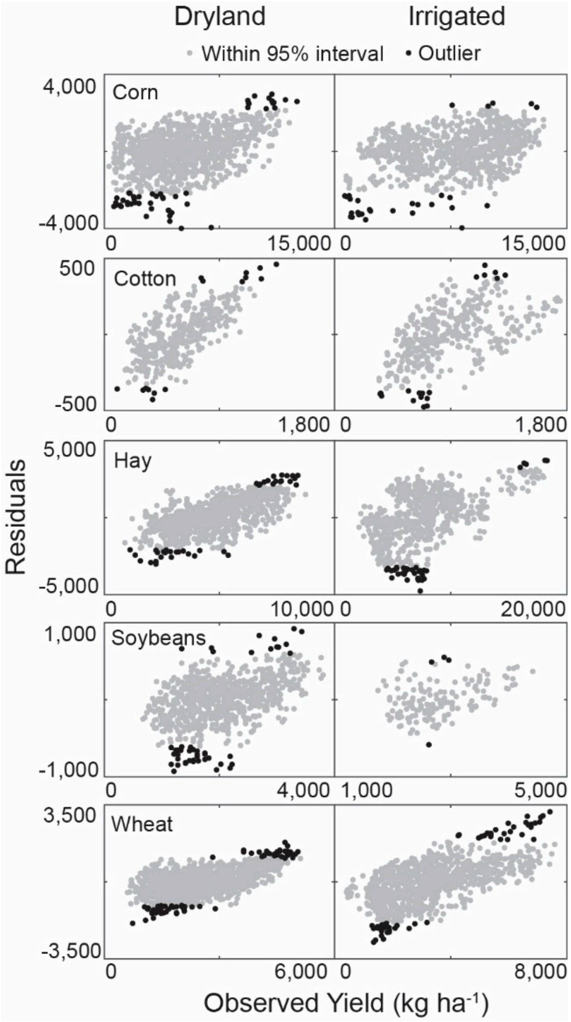

Figure S3. Residual intervals for each commodity and condition characterized by a 95% confidence threshold.

Table S1. Percent difference between 2013 estimated and observed dryland and irrigated yields for each commodity across all mixed states. Note that positive values are overestimations and negative values are underestimations.

Commodity	Dryland (% diff.)	Irrigated (% diff.)
Corn	-6.6	1.8
Cotton	8.3	21.2
Hay	-2.6	24.0
Soybean	-2.6	7.4
Wheat	-5.3	8.8

Table 2. Statistically significant (<0.05 p-value) predictor variables used to derive coefficients for estimating dryland and irrigated yields and acreages for each mixed state. Note the variable bank includes all the predictor variables tested. .

Crop	Type	Variable Used (p-value), listed in order of greatest significance						
Corn	Dryland	Time (<0.001)	Precip. (<0.001)	Temp. (<0.001)	Hay (<0.001)	Lat. (<0.001)		
	Irrigated	Time (<0.001)	Temp. (<0.001)	Rech. (<0.001)	Wheat (0.003)			
Cotton	Dryland	Time (<0.001)	Rech. (<0.001)	Corn (<0.001)				
	Irrigated	Long. (<0.001)	Time (<0.001)	Wheat (<0.001)				
Hay	Dryland	Lat. (<0.001)	Corn (<0.001)	Precip. (<0.001)	Temp. (<0.001)	Wheat (0.006)		
	Irrigated	Temp. (<0.001)	Time (<0.001)	Wheat (<0.001)				
Soybean	Dryland	Corn (<0.001)	Lat. (<0.001)	Precip. (<0.001)	Hay (<0.001)	Time (<0.001)		
	Irrigated	Time (<0.001)	Lat. (<0.001)	Temp. (<0.001)	Sorgh. (<0.001)			
Wheat	Dryland	Rech. (<0.001)	Time (<0.001)	Hay (<0.001)	Lat. (<0.001)	Precip. (<0.001)	Corn (<0.001)	Temp. (0.01)
	Irrigated	Time (<0.001)	Hay (<0.001)	Lat. (<0.001)	Precip. (<0.001)	Long. (<0.001)		

Variable Bank: Temperature, Precipitation, Latitude, Longitude, Recharge, Corn Yields, Cotton Yields, Hay Yields, Sorghum Yields, Soybean Yields, Wheat Yields