Table S1. Pearson correlation coefficients between test weight and maize grain quantity (yield) and composition (protein, starch, and oil concentration) response variables across seven years of study.

Year	Yield	Protein	Starch	Oil
2011	0.485	-0.500	-0.093	0.313
2012	0.241	-0.124	-0.155	0.113
2013	0.094	0.049	-0.135	0.277
2014	0.352	0.279	-0.158	0.240
2015	-0.052	0.218	-0.111	0.139
2016	-0.029	0.264	-0.043	0.045
2017	0.224	-0.059	0.260	0.018

Table S2. Vector loadings of PCAs. PCA_1 can be broadly defined as a contrast between maize grain protein and oil versus starch concentrations. PCA_2 can be broadly defined as a greater protein versus oil concentrations.

Conc.	PCA_1	PCA_2
Protein	0.475015	0.740434
Starch	-0.707369	-0.000144
Oil	0.523440	-0.672129

Table S3. Absolute differences in observed versus predicted maize yield $|e_i|$.

Quartile	$\|e_i\|$	
	Metric tons / hectare	bushels / acre
1st Quartile	0.39	6.23
Median	0.89	14.12
3rd Quartile	1.58	25.11

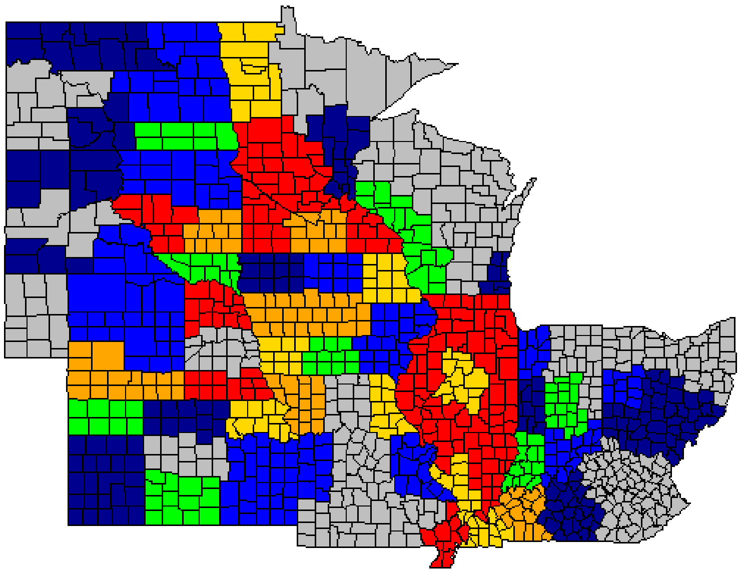

Figure S1. Sampling information by ASD. The total number of maize grain samples taken from each ASD over the course of the seven years of this study is shown in the figure above. Dark blue represents less than 10 samples, blue represents between 10 and 19 samples, green represents between 20 and 29 samples, gold represents between 30 and 39 samples, orange represents between 40 and 49 samples, and red represents 50 or more samples.

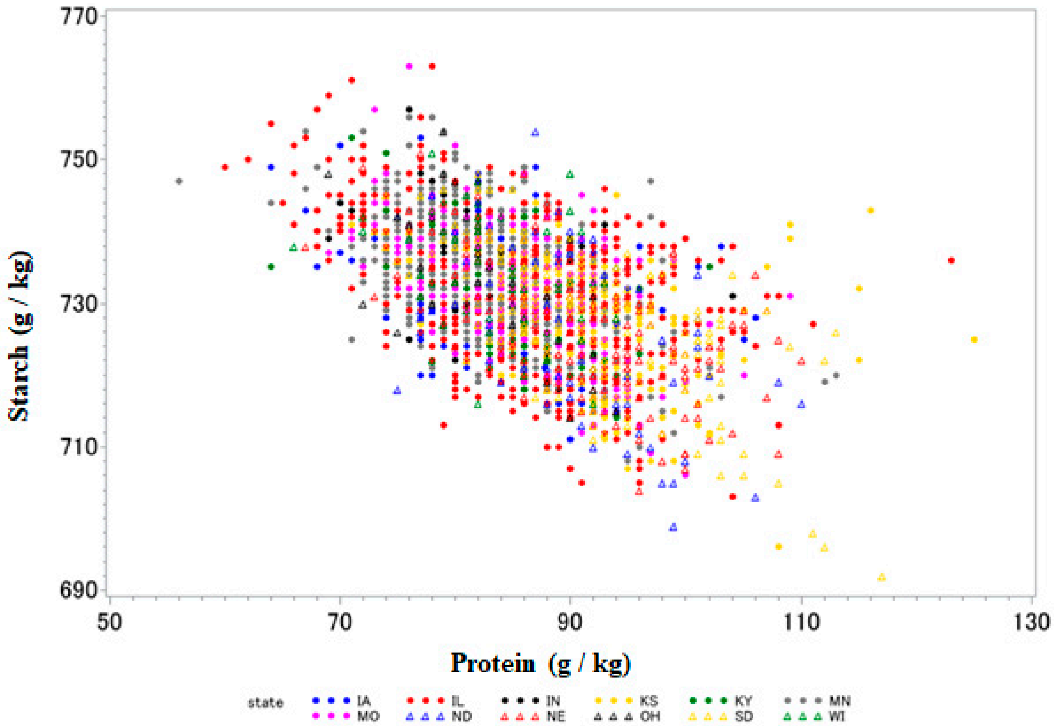

Figure S2. Plot of starch-to-protein ratio by state. Two states, North Dakota (blue triangle, ND) and South Dakota (yellow triangle, SD) consistently had higher maize grain protein concentrations than the other ten states, irrespective of year.

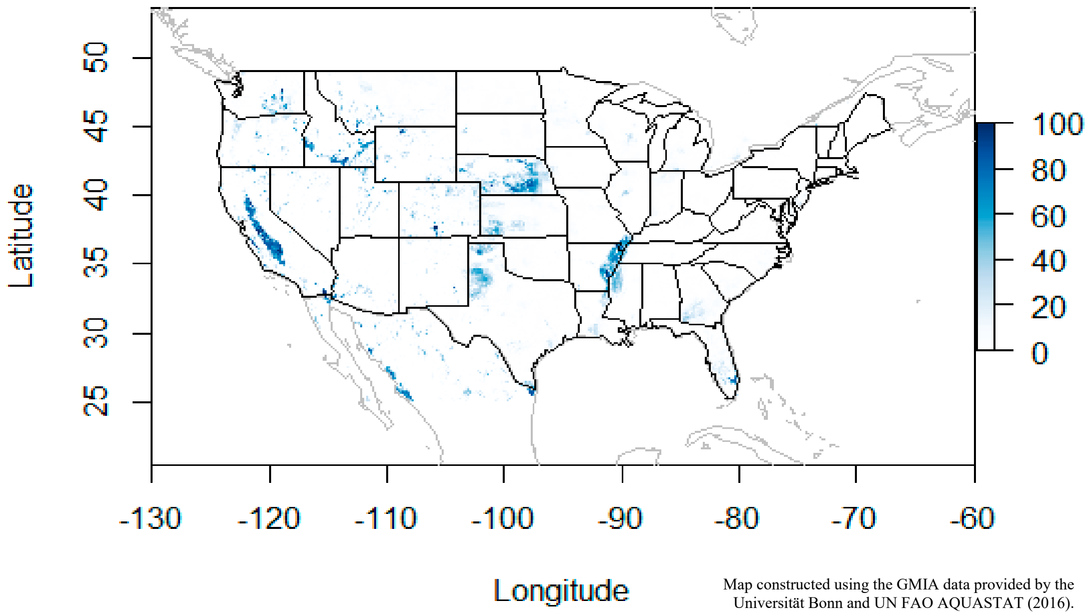

Figure S3. Map of irrigation prevalence in the United States. The percent irrigated land is shown in the figure above. Three ASDs, namely NE 30, NE 50, and MO 90, were noted to fall into clusters that were higher yielding than the clusters to which their neighboring ASDs belonged. This map indicates that a probable cause for this observation is the greater presence of irrigation in these areas.

Supplemental Code: R Code for latitude and longitude coordinates.

```
#Install necessary packages. This step need only be completed once.
install.packages(c("ggplot2", "devtools", "dplyr", "stringr","maps","mapdata","sp"))

#Load necessary packages after installation
library(maps)
library(mapdata)
library(sp)
library(ggplot2)
library(devtools)
library(dplyr)
library(stringr)

#Create a file with county information
counties=map_data("county")

#Name states of interest.
states=c("iowa","illinois","indiana","kansas","kentucky","minnesota","missouri","nebraska","north
dakota","ohio","south dakota","wisconsin")

#Subset the original counties dataset to include only the counties from the states of interest
counties2=counties[counties$region %in% states,]

#Create a function to calculate the spatial polygon centroid latitude and longitude coordinate of each
county.
#This code was modifed based off of a forum found here. The function is NOT our original code.
#URL: https://stackoverflow.com/questions/9441778/improve-centering-county-names-ggplot-maps
getLabelPoint = function(county) {Polygon(county[c('long', 'lat')])@labpt}

#Create a new variable called "name" is of the form "county,state"
counties2$name=paste(counties2$subregion,",",counties2$region,sep="")

#Use the getLabelPoint function
centroids=by(counties2,counties2$name,getLabelPoint)

#Row bind the output into a dataframe, rename the columns as "long" and "lat", and rename the rows as the
"county,state" names
centroids2=do.call("rbind.data.frame",centroids)
names(centroids2)=c("long","lat")
rownames(centroids2)=names(centroids)
```