2013

Development and evaluation of ordinary least squares regression models for predicting irrigated and rainfed maize and soybean yields

Vivek Sharma
University of Nebraska-Lincoln

Daran R. Rudnick
University of Nebraska-Lincoln, daran.rudnick@unl.edu

Suat Irmak
University of Nebraska-Lincoln, suat.irmak@unl.edu

Development and Evaluation of Ordinary Least Squares Regression Models for Predicting Irrigated and Rainfed Maize and Soybean Yields

V. Sharma, D. R. Rudnick, S. Irmak

Abstract. *Understanding the relationships between climatic variables and soil physical and chemical properties with crop yields on large scales is critical for evaluating crop productivity to make better assessments of local and regional food security, policy, land and water resource allocation, and management decisions. In this study, ordinary least squares (OLS) regression models were developed to predict irrigated and rainfed maize and soybean yields at the county level as a function of explanatory variables [precipitation (P), actual crop evapotranspiration (ET_a), organic matter content (OMC), cation exchange capacity (CEC), clay content (CC), and available soil water capacity (ASW)] of the dominant soil type in each of the 93 counties in Nebraska. Models were developed for the statewide average dataset (state models) as well as for the four major climatic zones (zonal models). Spline interpolation was used to spatially interpolate all independent variables across all 93 counties. The results of the OLS state models showed a very good performance for predicting rainfed maize and soybean yields. For rainfed maize, about 73% of the variation in yield (RMSD = 867 kg ha^{-1}) was explained by ET_a alone, and 83% of yield variability (RMSD = 690 kg ha^{-1}) was explained by the model Yield = f(ET_a, P, ASW, CEC, CC). For rainfed soybean, about 69% of the variability (RMSD = 238 kg ha^{-1}) was explained by ET_a alone, and a maximum of 85% (RMSD = 164 kg ha^{-1}) of the variability was explained by the model Yield = f(ET_a, P, ASW, CEC, CC). No additional variation in yield was explained by adding OMC to the rainfed maize and soybean yield models. Less correlation was found between the predicted and observed yields for irrigated maize and soybean than for the rainfed yields for both crops. For irrigated maize and soybean, a maximum of 45% (RMSD = 533 kg ha^{-1}) and 36% (RMSD = 218 kg ha^{-1}) of the variability in yield was explained by the models Yield = f(ET_a, P, ASW) and Yield = f(ET_a, P, ASW, CEC, CC, OMC), respectively. For the rainfed crops, ET_a played a major role in predicting yield, whereas P and ASW played a major role in predicting irrigated yields. ET_a and P accounted for 96%, 73%, and 67% of the total explained variation in rainfed soybean yield for zones 2 (drier), 3, and 4 (wetter), respectively, whereas soil physical and chemical properties accounted for 4%, 27%, and 33%, respectively. Unlike rainfed conditions, irrigated maize and soybean yield predictions were improved by applying the zonal models rather than the state models.*

Keywords. *Evapotranspiration, Inverse distance weighting, Irrigation, Kriging, Maize, Ordinary least square, Rainfed, Soybean, Spline.*

The quantitative characterization of spatio-temporal variability in crop yield is an important component for various applications, including site-specific soil, water, and nutrient management for improving uniformity of crop production and for precise application of inputs in precision farming. An important initial step for evaluating yield variability on a field, basin, state, or regional scale is to understand the relationships between crop yields and various climatic and soil variables. Furthermore, quantifying the effect of these variables that drive crop yield can aid in decision making and enable policy makers to make better assessments or projections of crop productivity. Considerable attention has been given to assess the effect of climatic variables on crop yield (Adams et al., 1998; Bryant et al., 2000; McCarthy et al., 2001; Reidsma et al., 2009), which impacts the future agricultural productivity. Change in climatic conditions from year to year is one of the major determinants of crop yield fluctuations. Lobell et al. (2007) analyzed the relationships between crop yield and three climatic variables (minimum temperature, maximum temperature, and precipitation). Sharma et al. (2011) used geographical weighted regression (GWR) to evaluate the non-stationarity relationships between annual, seasonal, and monthly precipitation on maize and soybean yields. In addition to climatic variables, crop yield is also affected by physical and chemical properties of the soil media, including available soil water holding

Submitted for review in October 2012 as manuscript number SW 9973; approved for publication by the Soil & Water Division of ASABE in June 2013.

The mention of trade names or commercial products is for the information of the reader and does not constitute an endorsement or recommendation for use by the University of Nebraska-Lincoln or the authors.

The authors are **Vivek Sharma, ASABE Member,** Graduate Research Assistant, **Daran R. Rudnick, ASABE Member,** Graduate Research Assistant, and **Suat Irmak, ASABE Member,** Professor, Department of Biological Systems Engineering, University of Nebraska-Lincoln. **Corresponding author:** Suat Irmak, 239 L.W. Chase Hall, University of Nebraska-Lincoln, Lincoln, NE 68583-0726, phone: 402-472-4865, e-mail: sirmak2@unl.edu.

capacity, texture, bulk density, clay content, soil layer thickness (Stone et al., 1985; Miller et al., 1988; Wright et al., 1990; Kreznor et al., 1989), pH (Moore et al., 1993), subsoil acidity (Wright et al., 1990), cation exchange capacity (CEC), salinity (Okogun et al., 2004), and fertility (Kreznor et al., 1989). Kravchenko and Bullock (2002a and 2002b) reported that soil properties on a field scale explained up to 71% of crop yield variability, and organic matter content (OMC) was found to be the most yield-influencing factor. Letey (1985) explained how crop production is indirectly affected by pore size distribution and directly affected by soil matric potential and its relationship to plant-available soil water. Lal (1997) found maize grain yield in western Nigeria to be significantly correlated with soil organic carbon, exchangeable Ca^{2+}, and CEC.

Several studies have estimated the effects of climatic and soil physical and chemical properties on crop productivity using either simulation models or regression-based techniques. Several researchers have demonstrated the strength of coupling crop models with GIS for agricultural decision support and resource planning at various spatial scales (Dent and Thornton, 1988; Curry et al., 1990; Kaspar et al., 2003; Sarangi et al., 2005). Hansen and Jones (2000) demonstrated several approaches to scale-up field-scale crop model predictions to larger scales. Crop simulation models, such as CERES-Maize (Jones and Kiniry, 1986) and DSSAT (Jones et al., 2003), have been used to predict crop yields by incorporating varying weather, soil physical and chemical properties, plant genetic background, management, and other agronomic practices that function at uniform or non-uniform areas on a field scale (Hansen and Jones, 2000; Irmak et al., 2001, 2002, 2005, 2006). Other researchers (e.g., Lal et al., 1993; Thornton et al., 1995; Rosenthal et al., 1998) applied various crop models in regional estimation of crop yield and variability. These models provide powerful and useful information on predicting spatial variability of crop yields, but they require a considerable number of input parameters due to variability in soil, topographical conditions, weather, and management practices at a regional scale.

Various interpolation techniques are available to predict and interpolate point-based information or variables to large scales within predetermined boundaries. Many of the interpolation techniques are referred to as deterministic and geostatistical interpolation methods. Deterministic interpolation methods such as inverse distance weighting (IDW) (Wilmott and Matsuura, 1995; Dodson and Marks, 1997) and spline (Hulme et al., 1995) estimate the value at a point from values recorded at neighboring points (Kurtzman and Kadmon, 1999). Geostatistical interpolation methods, such as kriging (Webster, 1985; Holdaway, 1996; McBratney and Pringle, 1997; Hudson and Wackernagel, 1994; Hammond and Yarie, 1996), are based on statistical models that include autocorrelation. These techniques are similar to interpolation techniques used with minimum spatial variance (Curran et al., 1997; Curran and Atkinson, 1998). For example, Goovaerts (2000) showed significant improvement in predicting continuous surfaces of mean monthly and mean annual rainfall when elevation was incorporated into the analysis. A similar observation was made by Hevesi et al. (1992) after comparing multivariate geo-statistics results for rainfall interpolation (which included elevation as a covariate) with six other interpolation techniques. Li et al. (2006) found that variables such as latitude, longitude, elevation, and distance from the sea were important predictors of seasonal temperature in the Zhejiang Province of China. Vicente-Serrano and Cuadrat (2003) compared diverse interpolation methods in Spain. Ninyerola et al. (2000) used multiple regressions with latitude, solar radiation, and cloudiness factor as independent variables for climatological modeling of temperature. Collins and Bolstad (1996) compared eight interpolation techniques for maximum and minimum air temperature estimation across two regions (eastern and western North America) at three temporal scales (ten-year mean, seasonal mean, and daily); their result showed that several variable characteristics can influence the choice of a spatial interpolation technique. Lal et al. (1993), McKinion et al. (2010), Irmak et al. (2010), and Sharma et al. (2011) have used various interpolation techniques in combination with GIS to spatially interpolate weather, soil physical and chemical properties (including soil moisture, nutrients, pH, and soil carbon), terrain (slope and elevation), crop characteristics, and other parameters to predict the impact of these parameters on crop yields and yield variability and to analyze regional crop productivity. Sharma and Irmak (2012a, 2012b) used a spline technique to spatially interpolate and analyze long-term monthly (May, June, July, August, and September), seasonal (May through September), and annual reference (potential) evapotranspiration (ET_{ref}), precipitation, actual crop evapotranspiration, and seasonal net irrigation requirements for maize and soybean in all 93 Nebraska counties. Several studies have used IDW interpolation techniques to predict and map climatic variables (Willmott and Robeson, 1995; Blennow and Persson, 1998). IDW and kriging techniques have been compared in several studies. In some cases, kriging performed better than IDW (Tabios and Salas, 1985; Hosseini et al., 1994; Dalthorp et al., 1999; Kravchenko and Bullock, 1999), and in other studies IDW outperformed kriging (Nalder and Wein, 1998; Weber and Englund, 1992).

While aforementioned studies compare various interpolation methods for estimating and analyzing different variables, studies that couple several yield-driving factors to understand spatio-temporal attributes of irrigated and rainfed crop yields on large scales are limited. The objectives of this study were to: (1) evaluate three interpolation techniques (kriging, spline, and IDW) and their validity for predicting climatic variables, and (2) develop and evaluate ordinary least square regression models to analyze the relationship between observed rainfed and irrigated maize and soybean yields based on the spatial variation of yield-driving factors [actual crop evapotranspiration (ET_a), precipitation (P), available soil water capacity (ASW), organic matter content (OMC), cation exchange capacity (CEC), and clay content (CC)] in all 93 counties in Nebraska.

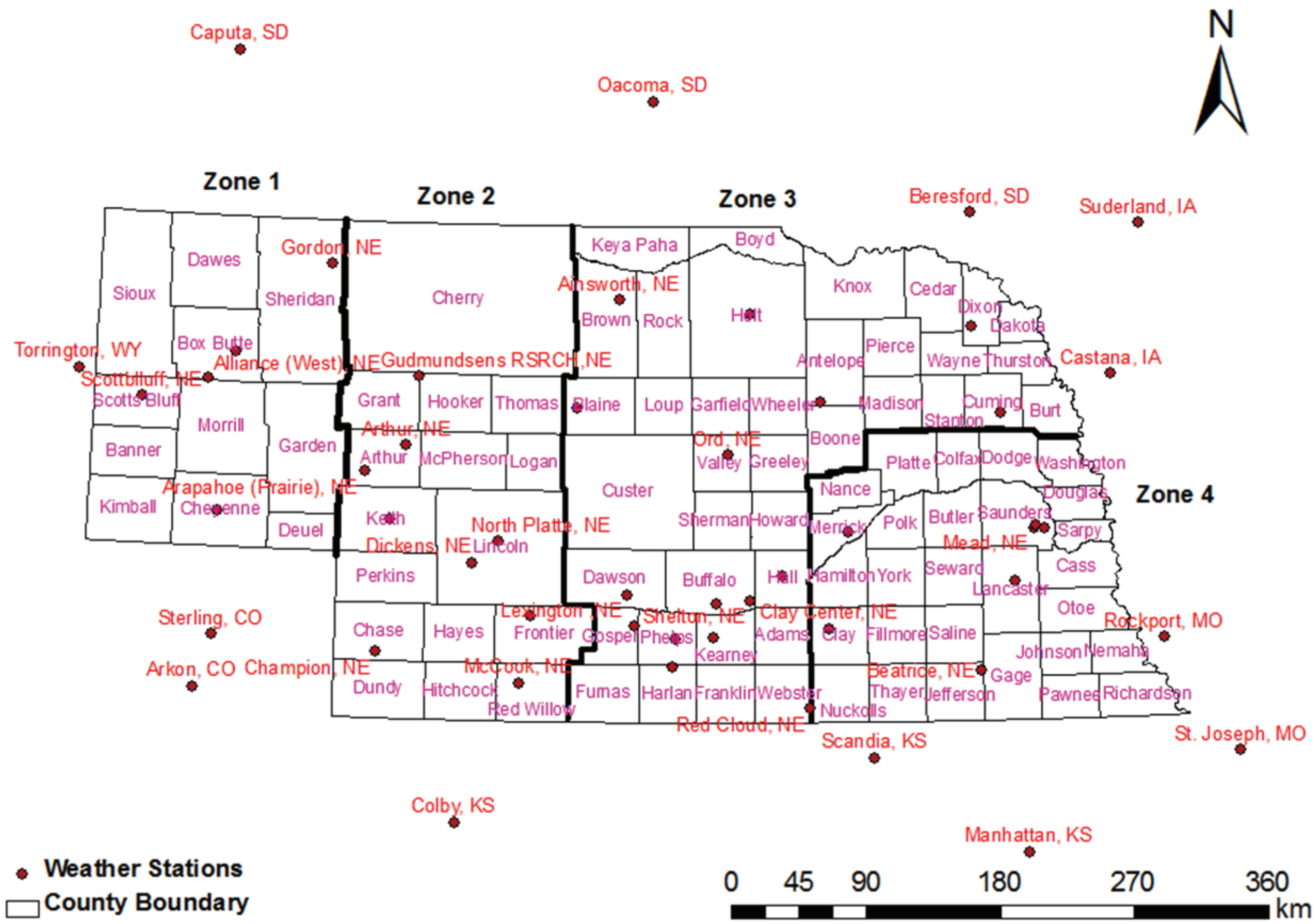

Figure 1. Map of Nebraska showing the zonal boundaries and locations of weather stations used for the analyses.

Materials and Methods

Study Area

The study was conducted for the entire state of Nebraska (fig. 1). Nebraska has 93 counties located between latitude 40° to 43° N and longitude 95° 19′ to 104° 3′ W. The state comprises Universal Transverse Mercator (UTM) Zones 13, 14, and 15. In this study, for the GIS analysis, UTM Zone 14 was used because more than 80% of state area is under this zone (Sharma and Irmak, 2012a, 2012b). Because of its latitude and interior continental location, which is impacted by Rocky Mountains cold air masses and Gulf of Mexico warm air streams, Nebraska has wide climatic seasonal variation, with warm summers (Strahler and Strahler, 1984) and cold and windy winters. The continental climate of Nebraska is mainly divided into two parts: the eastern and central parts have a humid/subhumid continental climate, and the western third has a semiarid/arid climate. The study area was divided into four different zones based on regional differences in climate, soil, and topographical characteristics (fig. 1). Detailed descriptions of these zones are presented by Sharma and Irmak (2012a, 2012b).

Observed and Predictive Variables

Yield data for irrigated and rainfed maize and soybean were obtained from 1996 to 2009 from the USDA National Agricultural Statistic Service (NASS; www.nass.usda.gov). For some of the counties in Nebraska, maize and soybean yield data were missing or incomplete from 1996 to 2009. Therefore, in our analysis, 91, 78, 85, and 69 counties were included for irrigated maize, rainfed maize, irrigated soybean, and rainfed soybean, respectively. Detailed descriptions of the irrigated and rainfed maize and soybean yield data with the spatial variation across Nebraska are presented by Sharma and Irmak (2012a, 2012b). The Census of Agriculture released in 2009 reported that Nebraska had approximately 3.6 million ha of irrigated land as of 2007 (USDA-NASS, 2009). From the east to the west side of the state, crop production becomes more reliant on irrigation due to the decrease in precipitation and increase in ET demands, as well as the lower water holding capacity of the soils.

Daily historical weather data from 1996 to 2009 were obtained from 50 Automated Weather Data Network (AWDN) stations located throughout Nebraska and in surrounding states (High Plains Regional Climate Center; http://hprcc1.unl.edu/cgi-hpcc/home.cgi). Daily climate data, including maximum and minimum air temperatures, relative humidity, incoming shortwave radiation, wind speed, and precipitation, were imported into ArcGIS (ver. 10, ESRI, Redlands, Cal.) for the exploratory spatial analysis. To reduce boundary effects during interpolation, stations outside of Nebraska (two in Colorado, three in Kansas, three in South Dakota, two in Missouri, and two in Iowa) were included in the analysis. Point coverage of ground-based meteorological stations was created in ArcGIS 10. The location (longitude and latitude) of the weather stations and the climate data were imported into a geo-database and explored using the ArcGIS Geospatial Analyst tool before interpolation. Daily climate data from the automated weather stations were input into the Penman-Monteith equation (Monteith, 1965), with a fixed canopy resistance (Irmak et al., 2012), to calculate daily alfalfa-reference (potential) evapotranspiration (ET_{ref}).

Crop coefficients (K_c) along with ET_{ref} were used to calculate actual crop evapotranspiration (ET_a). The typical emergence date was assumed to be the beginning (May 1) and physiological maturity was assumed to be the end of the growing season (Sept. 30) for the whole state, although the growing season becomes shorter from the eastern to western part of the state. Thus, no adjustments were made to account for the differences in growing season from the eastern to the western edge of the state. The stages of growth and development used in this study were approximated according to the phenological development of maize and soybean obtained from extensive field experiments conducted by S. Irmak (unpublished data) at the University of Nebraska-Lincoln, Institute of Agriculture and Natural Resources (UNL-IANR) South Central Agricultural Laboratory (SCAL) near Clay Center, Nebraska. The K_c values and ET_a estimation procedures used were the same as those outlined by Sharma and Irmak (2012a). Physical and chemical properties of the dominant soil type for each county were obtained from the USDA-NRCS Soil Survey Geographic Database (SSURGO; http://soildatamart.nrcs.usda.gov/State.aspx). The soil physical and chemical properties obtained to predict crop yield were: available soil water capacity (ASW) (calculated in this study), cation exchange capacity (CEC), organic matter content (OMC), and clay content (CC). ASW was computed for the 1.2 m soil profile, while the remaining predictors were taken from the top 0.30 m. It was assumed that the NRCS-reported values held constant for the period of 1996 to 2009.

Interpolation and Regression Techniques

The predicted values of precipitation and ET_{ref} based on 14 years of historical data were computed using spline, kriging, and inverse distance weighting interpolation methods. To compare the performance of these interpolation techniques, the predicted values were compared with weather station data and were evaluated based on the root mean square difference (RMSD) and coefficient of determination (R^2) using number of observations ($N = 50$). For all techniques, interpolations with a maximum of ten and minimum of seven neighboring weather stations were tested.

Spline interpolation is a deterministic method that fits a mathematical function through input data to create a smooth surface. It can generate accurate surfaces from only a few sampled points (Anderson, 2002). In this interpolation, each station was omitted, in turn, from the estimation of the fitted surface, and the mean square error (MSE) was calculated. This process was repeated for a range of values of a smoothing parameter, and then the value that minimized MSE was used to obtain the optimum smoothing. In our analysis, a regularized spline was selected because it creates a smoother surface closely constrained with the sample data range. On the other hand, kriging interpolation is based on a statistical model that includes autocorrelation (i.e., the statistical relationship among the measured points). This is because geostatistical techniques (kriging) have the ability to provide some measure of accuracy of the predictive surfaces (Merino et al., 2001). In kriging, the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. The kriging tool fits a mathematical function to a specified number of points, or all points within a specified radius, to determine the output value for each location (Sharma and Irmak, 2012a, 2012b). Kriging weights the surrounding measured values to derive a prediction for an unmeasured location. Detailed descriptions of the spline and kriging functions used in this study are presented by Sharma and Irmak (2012a) and Irmak et al. (2010).

The third interpolation technique used in the study was inverse distance weighting (IDW), which is based on the assumption that the climatic and soil property variables at an unsampled location are a distance-weighted average of the variables at the sampling points. The interpolated surfaces are more heavily influenced (weighted) by nearby points and less influenced by distant points. Detailed mathematical descriptions of the IDW method are provided by Watson and Philips (1985), Hosseini et al. (1994), Nalder and Wein (1998), and Kollias et al. (1999).

After interpolation, zonal statistics were used to calculate precipitation and ET_{ref} values for each county. Zonal statistics (Spatial Analyst tool of ArcGIS 10) calculate statistics on the value of a raster (1000 m × 1000 m cell size) within the zone of another dataset and summarize the results as mean, maximum, minimum, and range values. Each county mean value from zonal statistics was calculated from the precipitation and ET_{ref} rasters using all of the Nebraska counties defined by name (string attribute field) of the county feature class. Some studies used zonal statistic for computing average elevation, aspect, slope (topographic attributes), and normalized difference vegetation index (NDVI) (Bakhsh and Kanwar, 2004; Sharma et al., 2011), while others used zonal analysis to calculate crop yields for different grids (Kulkarni et al., 2008).

The ordinary least square (OLS) regression used in this study is a generalized liner modeling technique that may be used to model a response or dependent variable. It provides a global model of the variable for the prediction. Classical OLS regression theory relies on the assumption that the explanatory variables are measured with minimal or no error. The technique allows single or multiple explanatory variables to be used in the model. Mathematically, the simple linear model fitted by OLS is expressed as:

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \ldots \beta_n x_n + \varepsilon \quad (1)$$

Table 1. Coefficient of determination (R^2) and root mean square difference (RMSD) between observed and interpolated annual (Jan. 1 to Dec. 31), seasonal (May 1 to Sept. 30), and monthly precipitation and reference (potential) evapotranspiration (ET_{ref}) for kriging, spline, and inverse distance weighting (IDW) interpolation computed from cross-validation of weather stations.

	Precipitation						Alfalfa-Reference (Potential) Evapotranspiration (ET_{ref})					
	Kriging		Spline		IDW		Kriging		Spline		IDW	
Period	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD	R^2	RMSD
Annual	0.89	41.4	0.90	43.3	0.87	47.6	0.70	103.0	0.73	96.4	0.73	95.6
Seasonal	0.87	27.2	0.86	29.1	0.84	30.8	0.67	64.9	0.71	60.6	0.72	59.2
January	0.49	3.5	0.46	3.6	0.47	3.6	0.75	5.7	0.77	5.4	0.73	5.8
February	0.79	3.0	0.81	2.9	0.80	3.0	0.81	4.9	0.81	4.8	0.76	5.4
March	0.78	4.8	0.79	4.8	0.77	4.9	0.80	6.6	0.80	6.8	0.77	7.1
April	0.81	6.3	0.79	6.8	0.74	7.3	0.51	8.1	0.54	7.7	0.54	7.6
May	0.83	8.6	0.84	8.2	0.82	8.8	0.31	11.4	0.32	11.1	0.31	11.1
June	0.74	8.9	0.71	9.7	0.67	10.2	0.70	11.3	0.71	11.0	0.70	11.2
July	0.68	9.3	0.68	9.4	0.64	10.1	0.68	17.0	0.70	16.2	0.69	16.5
August	0.76	8.7	0.75	9.0	0.70	9.8	0.72	17.1	0.74	16.6	0.73	16.7
September	0.72	7.4	0.76	6.8	0.74	7.1	0.45	18.5	0.50	17.4	0.55	16.6
October	0.68	7.9	0.70	7.5	0.68	7.9	0.50	11.5	0.52	11.1	0.57	10.5
November	0.93	2.7	0.93	2.7	0.93	2.7	0.39	13.2	0.36	13.5	0.35	13.6
December	0.84	2.4	0.86	2.3	0.84	2.6	0.42	12.9	0.41	13.0	0.38	13.3

where y is the dependent variable; β_o is the intercept; β_1, β_2, $\beta_3 \ldots \beta_n$ are the coefficients (slope) of the independent variable x (x_1, x_2, $x_3 \ldots x_n$); and ε is the deviation of the point from the regression line (error term).

Results and Discussion

Three interpolation techniques (kriging, spline, and IDW) were used to predict growing season (May 1 to Sept. 30) precipitation and ET_{ref} from 1996 to 2009. All interpolation techniques were performed in ArcGIS 10. The performance indicators (RMSD and R^2) for each interpolation technique and the time frame of the interpolation are presented in table 1. The highest reported R^2 value for precipitation for all three techniques was 0.93 in November, with RMSD values of 2.68, 2.74, and 2.69 mm for kriging, spline, and IDW, respectively. The lowest R^2 value for all techniques was observed in January, with values of 0.49 (RMSD = 3.5 mm), 0.46 (RMSD = 3.6 mm), and 0.47 (RMSD = 3.6 mm) for kriging, spline, and IDW, respectively. Slightly higher error in January could be attributed to higher spatial variability in monthly precipitation. The spatial variability was not reflected in the standard deviation of the data (data not shown) but was reflected in local trends, indicating that the variability was more in the neighboring station. On the other hand, less variability was observed in November in the local trends. Slightly higher RMSD was observed for ET_{ref} (table 1) than precipitation. The RMSD ranged from 4.8 mm for spline in February to 18.5 mm for kriging in September (table 1). Overall high RMSD were observed in warmer months, and the opposite occurred in colder months. The highest R^2 value for kriging, spline, and IDW was 0.81 (February), 0.81 (February), and 0.77 (March), respectively. The lowest R^2 for all techniques was observed in May as 0.31 for kriging and IDW and 0.32 for spline method. Annual precipitation and ET_{ref} showed significant error as compared with seasonal precipitation and ET_{ref}, indicating that RMSD values were in proportion to the original values in the dataset. For instance, high annual precipitation and ET_{ref} values produced higher RMSD, as compared with low seasonal precipitation and ET_{ref}.

Similar results were obtained from all interpolation techniques when estimating precipitation and ET_{ref}. The spline and kriging methods had the closest agreement in most cases, but overall the spline method resulted in slightly higher R^2 values between the observed and interpolated data. Based on the statistics reported in table 1, all three methods were judged to be suitable for estimating the spatial distribution of precipitation and ET_{ref} across Nebraska. However, because the spline method resulted in slightly higher R^2 values, this method was used to interpolate the climatic parameters across all counties.

Distribution of Climatic Variables and Soil Properties

To evaluate the potential impacts of climatic and soil physical and chemical properties on maize and soybean yields, the spatial distributions of these yield-driving factors across Nebraska are presented in figure 2 and figures 3a to 3d. Spatial distributions of the long-term average seasonal (May 1 to Sept. 30) precipitation and ET_{ref} for the entire state are presented in figures 2a and 2b. Precipitation amounts in Nebraska gradually increase from the northwest to the southeast corner of the state. The difference in precipitation amounts along the gradient is 260 mm. The minimum precipitation of 227 mm and maximum of 486 mm were reported in Scottsbluff and Richardson Counties, respectively. Reference evapotranspiration follows the opposite trend to precipitation across the state. There is a gradual decrease in ET_{ref} from western to the eastern edge of the state. Unlike precipitation, ET_{ref} shows less variation from north to south on the eastern edge of the state. The difference in ET_{ref} between the western and eastern edges of the state is 280 mm, with a maximum of 1086 mm in Cheyenne County and a minimum of 807 mm in Douglas County.

Figures 3a to 3d present the spatial distribution of organic matter content (OMC, percent weight) (fig. 3a), available soil water capacity (ASW) (fig. 3b), percent clay content (CC) (fig. 3c), and cation exchange capacity (CEC) (fig. 3d) across Nebraska. Available soil water capacity is the difference between field capacity and permanent wilting point summed for each soil layer (0.30 m increments) in the

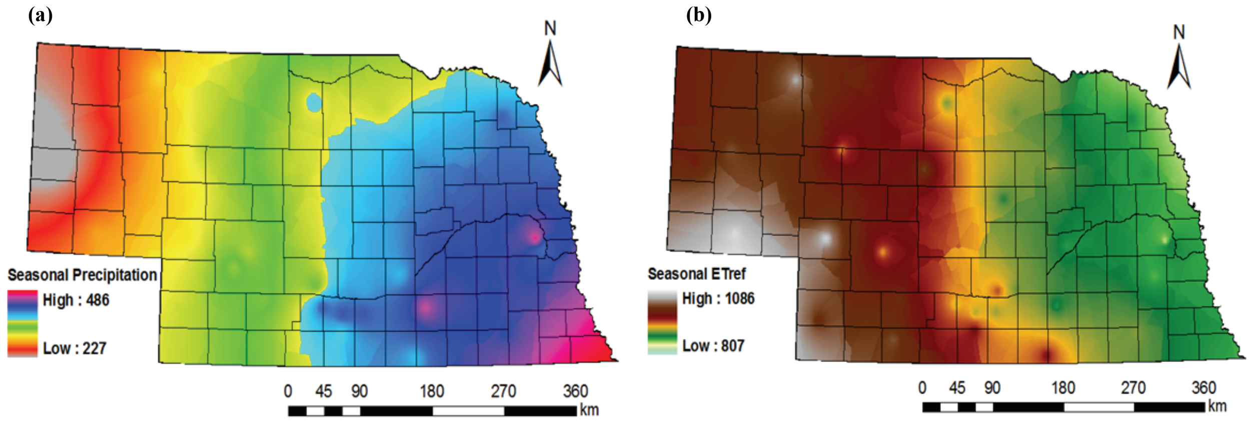

Figure 2. Spatial variation of long-term (1996-2009) average seasonal (May 1 to Sept. 30) (a) precipitation (mm) and (b) reference (potential) evapotranspiration (ET_{ref}, mm) with spline interpolation across Nebraska (data source: Sharma et al., 2011).

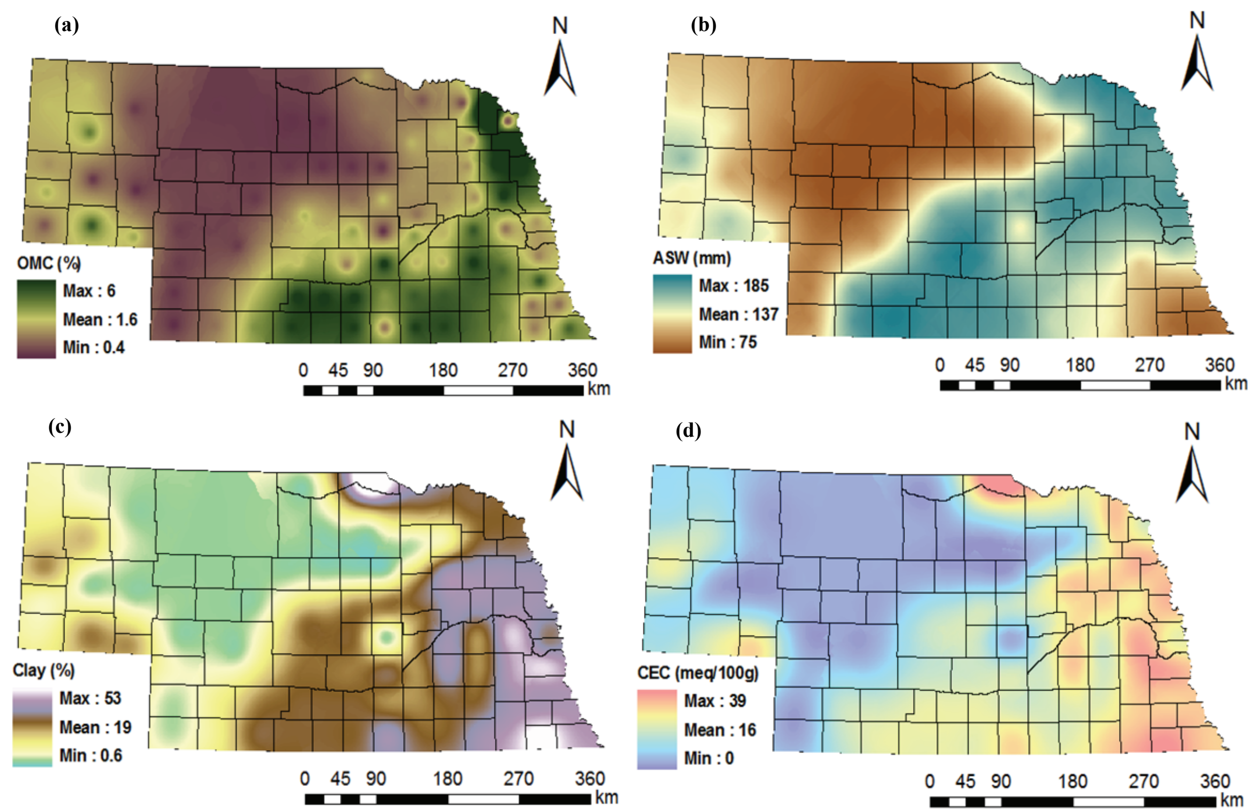

Figure 3. Spatial variation of: (a) percent topsoil organic matter content (OMC), (b) available soil water capacity (ASW) in the 0-1.20 m soil profile (mm), (c) percent topsoil (0-0.30 m) clay content (CC), and (d) topsoil cation exchange capacity (CEC) according to major soil series in each county across Nebraska.

top 1.20 m soil profile for both irrigated and rainfed maize and soybean. The remaining soil properties were reported for the top 0-0.30 m soil layer. Similar gradients across Nebraska were observed for all soil properties included in the study. Low levels of OMC, ASW, CC, and CEC were reported in the northern portion of west-central part of the state, which is known as the Sandhills. Figure 3a shows the distribution of OMC. High levels of OMC were found in the south central and northeastern parts of the state. OMC ranges from nearly zero in the Sandhills region to approximately 6% in Dixon County, with a statewide average of 1.57%. Statewide variation of organic matter is primarily influenced by climate, land use and soil management practices, and vegetation type and coverage density. On a smaller scale, variation of organic matter is affected by topography, crop type, crop and soil management practices, and precipitation amounts. One of the reasons for the high OMC in the south central and northeastern part of the state is due to the adoption of no-till practices in these areas and the deep silt loam soils. Based on USDA-NRCS (2009)

Table 2. Coefficient of determination (R^2) and ordinary least square (OLS) model coefficients for rainfed maize and soybean yields.

Model	R^2	β_0	β_1	β_2	β_3	β_4	β_5	β_6
Rainfed maize yield								
Yield = f(ET$_a$)	0.73	20030.8	-22.49	-	-	-	-	-
Yield = f(ET$_a$, P)	0.75	12050.66	-15.27	8.98	-	-	-	-
Yield = f(ET$_a$, P, ASW)	0.79	9735.72	-13.71	9.47	7.96	-	-	-
Yield = f(ET$_a$, P, ASW, CEC)	0.82	10208.23	-13.13	6.8	3.73	43.95	-	-
Yield = f(ET$_a$, P, ASW, CEC, CC)	0.83	9526.89	-12.95	8.4	3.71	123.22	-68.7	-
Yield = f(ET$_a$, P, ASW, CEC, CC, OMC)	0.83	9546.73	-12.96	8.35	3.66	121.1	-67.33	11.61
Rainfed soybean yield								
Yield = f(ET$_a$)	0.69	7321.04	-9.28	-	-	-	-	-
Yield = f(ET$_a$, P)	0.73	3941.68	-6.19	4.29	-	-	-	-
Yield = f(ET$_a$, P, ASW)	0.81	2639.83	-5.16	5.06	3	-	-	-
Yield = f(ET$_a$, P, ASW, CEC)	0.83	3356.83	-5.39	3.41	2.14	9.91	-	-
Yield = f(ET$_a$, P, ASW, CEC, CC)	0.85	2683.51	-4.68	4.06	2.3	47.96	-31.31	-
Yield = f(ET$_a$, P, ASW, CEC, CC, OMC)	0.85	2592.31	-4.59	4.17	2.36	51	-33.52	-10.17

statistics, about 77% of maize is planted as no-till in the eastern part of the state, and about 70% of the maize land area in Banner county (western Nebraska) is planted as no-till, resulting in higher OMC in these areas.

Available soil water capacity in the top 1.20 m soil profile has a minimum value of 74 mm of water in Dundy County, a maximum of 185 mm of water in multiple counties, with a statewide average of 137 mm (fig. 3). The state has about eight major soil types with 138 soil series. Out of 138 soil series, 17 constitute about 49% of the land area (USDA-NRCS web soil survey), but only the major soil type for each county was selected to map ASW in our analysis (fig. 3b). Thus, for a given county, if the soil type has more than 185 mm of ASW in the top 1.20 m soil profile, it is not considered in our analysis. The average percent clay content (CC) in Nebraska is 19%, with a minimum value of 3% (multiple counties) and maximum value of 53% (Boyd County) (fig. 2c). Cation exchange capacity (CEC) reflects the amount of nutrients (Ca^{2+}, Mg^{2+}, and K^{1+}) a soil can store and make available to crops. CEC has a minimum value of 2.5 meq per 100 g in Dundy County, a maximum of 37.5 meq per 100 g in Boyd County, with a statewide average of 16.2 meq per 100 g (fig. 2d). The highest values were observed in the eastern part of the state, where clay content and organic matter content are highest. Furthermore, tillage methods heavily impact CEC (Prasad and Power, 1991) and, in general, CEC is more favorable under no-till or reduced-till practices (Lal, 1989), which is heavily practiced in the eastern part of Nebraska.

OBSERVED VS. PREDICTED YIELD USING OLS (STATEWIDE MODELS)

The performance of the OLS models, as measured by R^2 and RMSD, indicated that a large amount of yield variation is explained by the explanatory variables. Figures 5, 7, 9, and 11 present the statewide distribution of observed vs. predicted yield with R^2 and RMSD values for irrigated and rainfed maize and soybean. In some counties, maize and/or soybean are not grown, and these counties with no yield data are represented with white color. The results of the OLS models showed a very good performance for predicting rainfed maize and soybean yields. The amount of total variation in yield that was explained by different variables ranged from 73% to 83% for rainfed maize and from 69% to 85% for rainfed soybean (table 2). Less variation in irrigated yield was observed as compared with the rainfed crop yields with R^2 ranging from 0.19 to 0.46 and 0.06 to 0.36 for irrigated maize and soybean yields, respectively (table 3). The performance of OLS models was interpreted as an indicator of the overall importance of the selected climatic and soil variables to the observed spatial pattern of yield stability.

The statewide average OLS coefficients (β_o, β_1, β_2, β_3, β_4, β_5, and β_6 from eq. 1) for rainfed and irrigated maize and soybean are presented in tables 2 and 3, respectively. The OLS approach produced six coefficients for all cases (i.e., irrigated and rainfed maize and soybean) with the addition of each explanatory variable. A hierarchical approach was used to produce each coefficient. Each coefficient represents the strength and type of relationship that explanatory variable has with the dependent variable. For the rainfed maize models, strong negative relationships were observed between yield and ET$_a$; however, the negative relationship decreases with the addition of other explanatory variables (table 2). On the other hand, precipitation, ASW, and CEC are main yield-driving factors, showing strong positive relationships with rainfed maize yields. High CEC coefficients of 43.95, 123.22, and 121.10 were observed for Yield = f(ET$_a$, P, ASW, CEC), Yield = f(ET$_a$, P, ASW, CEC, CC), and Yield = f(ET$_a$, P, ASW, CEC, CC, OMC) models, representing a strong positive relationship with rainfed maize yields. A similar positive effect on yield with CEC was reported by Casanova et al. (1999). For all crops, ET$_a$, P, ASW, CEC, CC, and OMC were of moderate to high importance in predicting yield. Tables 2 and 3 also present model coefficients and R^2 values of various predictive models developed for rainfed and irrigated maize and soybean yields. For rainfed maize, about 73% of the variation in yield, with an RMSD value of 867 kg ha^{-1}, was explained by ET$_a$ alone. By adding variables to the explanatory model, the overall performance, measured by R^2, increased. The maximum variability of 83%, with an RMSD value of 690 kg ha^{-1}, was explained by the model Yield = f(ET$_a$, P, ASW, CEC, CC) (table 2). No additional variation in yield was explained by adding OMC to the rainfed maize yield model.

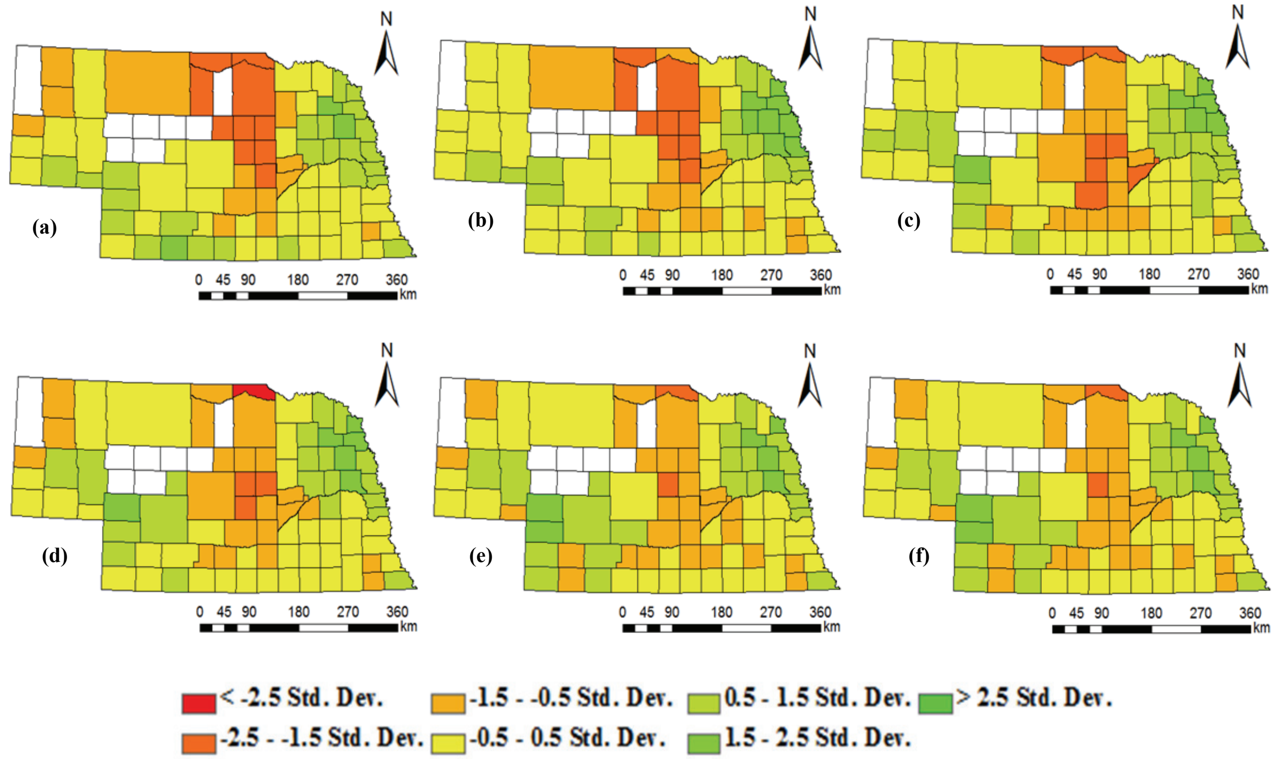

Figure 4. Standardized residual (standard deviation) maps for various predictive models: (a) Y = f(ET$_a$), (b) Y = f(ET$_a$, P), (c) Y = f(ET$_a$, P, ASW), (d) Y = f(ET$_a$, P, ASW, CEC), (e) Y = f(ET$_a$, P, ASW, CEC, CC), and (f) Y = f(ET$_a$, P, ASW, CEC, CC, OMC), where Y = rainfed maize yield (kg ha^{-1}), ET$_a$ = actual evapotranspiration (mm), P = precipitation (mm), ASW = available soil water capacity in the 0-1.20 m soil profile (mm), CEC = cation exchange capacity (meq per 100 g), CC = clay content (%), and OMC = organic matter content (%).

Consistent spatial patterns were expressed in the predicted rainfed maize models as explanatory variables were added (fig. 5). A curvilinear relationship was observed between predicted and observed yield when only a few variables were included. This resulted in underprediction at lower and higher observed yields and overprediction at yields between approximately 4,000 and 7,000 kg ha^{-1}. As the number of explanatory variables increased, the predicted and observed yield relationship approached unity (fig. 5). However, the statistical analysis showed that the p-value for the intercept and slope of the regression line was significantly different from unity, i.e., $p < 0.05$ (data not shown). The results of the standardized residual maps show that predicted yields were within ±2 standard deviations (SD) of the observed yields. Less than 3% of the counties fell over the ±2.0 SD range. Figure 4 shows the standardized residual maps as explanatory variables were included in the prediction of rainfed maize yield. In all rainfed maize yield models, the northeast corner of the state, with Wayne, Cuming, Pierce, Thurston, Dixon, and Dakota counties, was underpredicted, whereas the north central portion of the state, including Holt, Loup, Garfield, Wheeler, and Keya Paha counties, was overpredicted. For the first model, in which ET$_a$ is the only variable, crop yield of the north central portion of the state was overpredicted with -2.0 SD, and the northeast corner was underpredicted with +2.0 SD. The model explaining the maximum amount of yield variability, Yield = f(ET$_a$, P, ASW, CEC, CC), reduced the SD in these two areas closer to ±1.0 SD. The distinct spatial patterns observed in the models show that other yield-influencing parameters exist that are not included in the models (i.e., nutrient availability, soil and water salinity, management practices such as tillage method, crop rotation, etc.) and may be required to strengthen the predictive models.

Similar results were observed for rainfed soybean yields. About 69% of the variability (RMSD = 238 kg ha^{-1}) was explained by ET$_a$ alone, and 73%, 81%, and 83% of the variability was explained by the models Yield = f(ET$_a$, P), Yield = f(ET$_a$, P, ASW), and Yield = f(ET$_a$, P, ASW, CEC), respectively (fig. 7). The maximum variation of 85% (RMSD of 164 kg ha^{-1}) was explained by Yield = f(ET$_a$, P, ASW, CEC, CC). Similar to rainfed maize, no additional variation in soybean yield was explained by adding OMC to the model. The intercept and slope of the regression line between explanatory variables and yield were significantly different ($p < 0.05$) from unity for rainfed soybean (data not shown). Compared to rainfed maize, small to moderate differences in model coefficients were observed for rainfed soybean yield. All the variables have positive relationship with rainfed soybean yield (table 2), except ET$_a$. The OLS SD maps indicate that the predicted yield or residuals were within the ±2 SD range. The crop yields of the north central portion of the state were overpredicted in all models; however, the SD decreased from approximately ±2 to ±1 SD as the explanatory variables were increased (fig. 6). Less variability was explained when adding additional soil physical and chemical properties as compared with climatic variables.

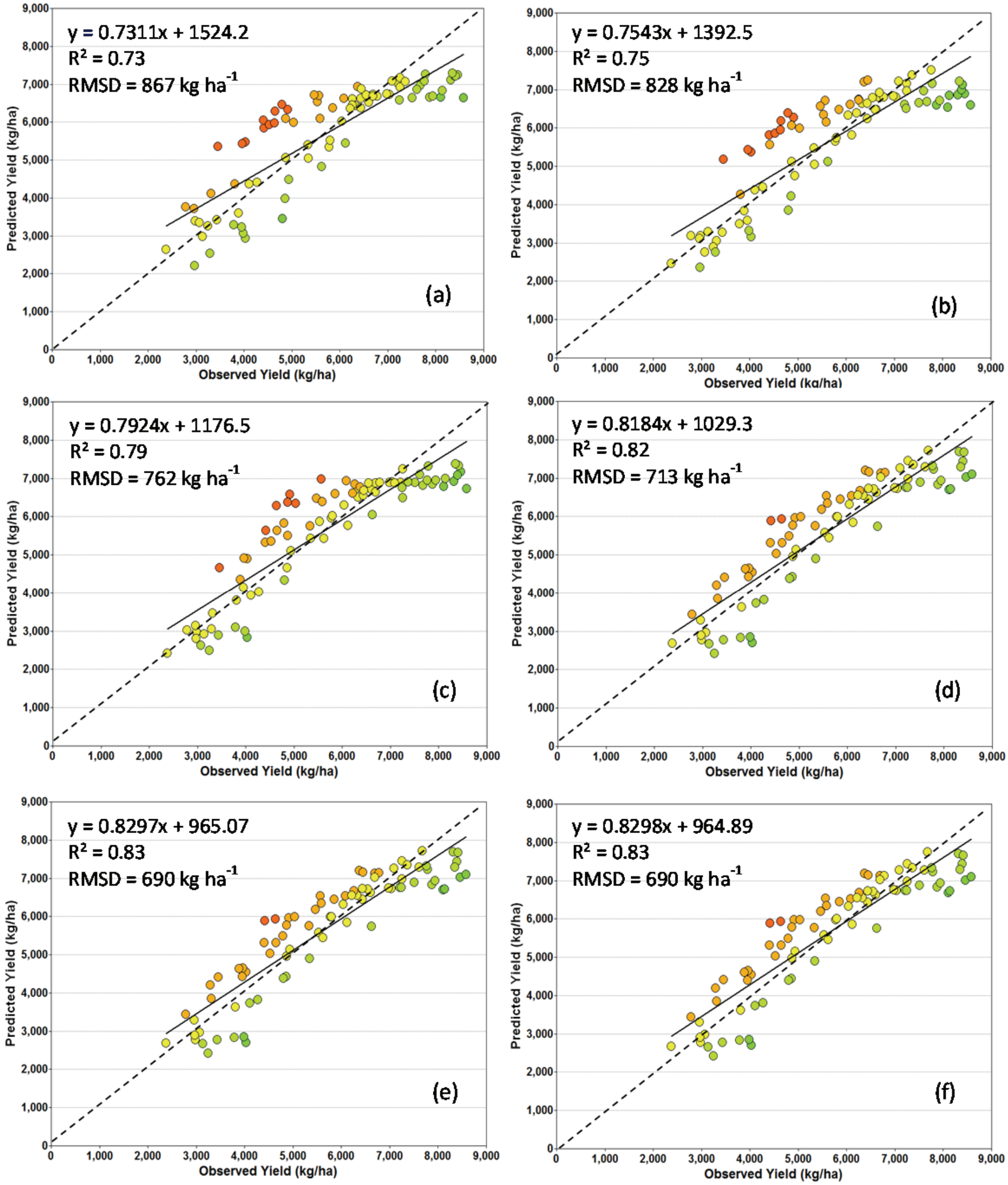

Figure 5. Relationship between observed vs. predicted rainfed maize yield models by using different predictive variables: (a) Y = f(ET$_a$), (b) Y = f(ET$_a$, P), (c) Y = f(ET$_a$, P, ASW), (d) Y = f(ET$_a$, P, ASW, CEC), (e) Y = f(ET$_a$, P, ASW, CEC, CC), and (f) Y = f(ET$_a$, P, ASW, CEC, CC, OMC), where Y = rainfed maize yield (kg ha^{-1}), ET$_a$ = actual evapotranspiration (mm), P = precipitation (mm), ASW = available soil water capacity (mm) in the 0-1.20 m soil profile, CEC = cation exchange capacity (meq per 100 g), CC = clay content (%), and OMC = organic matter content (%). The colors of the data points are associated with the county colors on the rainfed maize standardized residuals map (fig. 4).

As mentioned earlier, similar spatial patterns exist across Nebraska for the soil physical and chemical properties included in the analysis. The soil properties are affected similarly by geographical conditions (climate, topography, etc.) and influenced by each other. For instance, CEC is known to be higher in areas with high clay content and OMC; low levels of clay content tend to produce less dense vegetative cover, resulting in lower OMC; and ASW is proportional to soil texture and is lower in sandy soils. By incrementally adding soil properties to the models, the predictions are strengthened; however, the predictive surfaces are less likely to change from one model to the next.

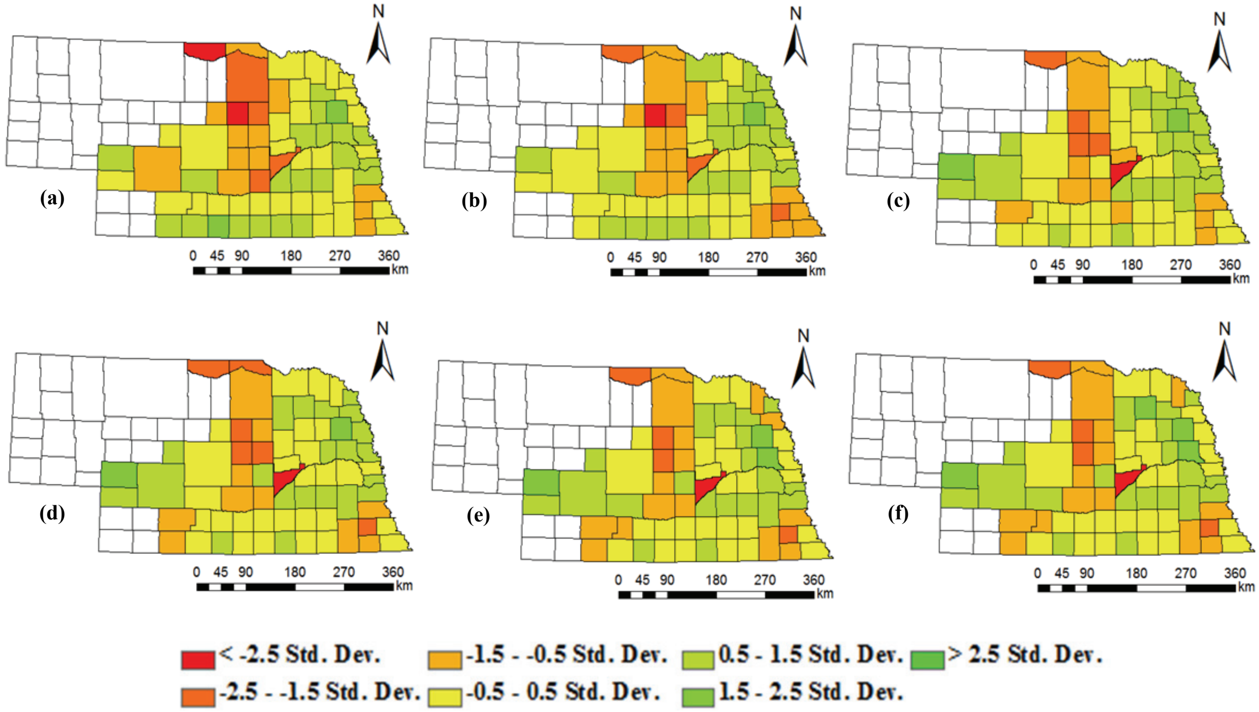

Figure 6. Standardized residual (standard deviation, SD) maps for various models: (a) Y = f(ET$_a$), (b) Y = f(ET$_a$, P), (c) Y = f(ET$_a$, P, ASW), (d) Y = f(ET$_a$, P, ASW, CEC), (e) Y = f(ET$_a$, P, ASW, CEC, CC), and (f) Y = f(ET$_a$, P, ASW, CEC, CC, OMC), where Y = rainfed soybean yield (kg ha^{-1}), ET$_a$ = actual evapotranspiration (mm), P = precipitation (mm), ASW = available soil water capacity in the 0-1.20 m soil profile (mm), CEC = cation exchange capacity (meq per 100 g), CC = clay content (%), and OMC = organic matter content (%).

Less correlation was found between the predicted and observed yield for irrigated maize as compared with the rainfed maize yields (figs. 5 and 9). For irrigated maize, an R^2 of 0.19 and an RMSD of 645 kg ha^{-1} was found between predicted and observed yield for the model Yield = f(ET$_a$). A maximum of 45% of the variability in yield (RMSD = 533 kg ha^{-1}) was explained by the model Yield = f(ET$_a$, P, ASW) (fig. 9). No additional variation in yield was explained by adding CEC, CC, and OMC, indicating that the irrigated maize yields are mostly dependent on precipitation, ET$_a$, and ASW (table 4). Most of the OLS coefficients had a positive relationship with irrigated maize yield. Contrary to rainfed crops, irrigated maize and soybean yields had a positive relationship with ET$_a$ for all models (tables 2 and 3). Except for CEC, all explanatory variables in the irrigated maize yield model Y = f(ET$_a$, P, ASW, CEC, CC, OMC) had a positive impact on yield. Less apparent spatial yield patterns existed for irrigated maize as compared with rainfed maize. The standardized residual (standard deviation, SD) maps indicate that the residuals (observed minus predicted yield) are within 1.5 SD (fig. 8). Yield was underpredicted for the counties in the central part of the state (fig. 8). Irrigation influences crop yield predictions more in areas with less precipitation (e.g., central and western Nebraska) as compared with areas with higher precipitation amounts (e.g., the eastern part of the state), where yield shows the overprediction trend (fig. 8). In general, irrigation allows crops to resume their normal or potential growth rates under the imposed atmospheric and geographical constraints; therefore, the inclusion of irrigation amount and method would further improve the performance of the irrigated maize models. Furthermore, irrigation minimizes the range of observed yields across the state, which most likely reduced the ability of the explanatory variables to explain yield variability for irrigated maize and soybean. Irrigated maize yields ranged from 8,000 to 12,000 kg ha^{-1} (fig. 9); whereas rainfed maize yields ranged from 2,500 to 8,500 kg ha^{-1} (fig. 5).

Irrigated soybean yield predictions had a lower R^2 of 0.06 and RMSD of 263 kg ha^{-1} for the model Yield = f(ET$_a$). Table 3 presents the coefficients and R^2 values for different irrigated soybean yield prediction models. The maximum R^2 was 0.36 (RMSD = 218 kg ha^{-1}) for the model that included ET$_a$, P, ASW, CEC, CC, and OMC (fig. 11). While irrigation water increases crop water productivity, especially in areas with high ET demands and low precipitation, the inability to account for average county-level irrigation amounts (due to unavailability of data in terms of total irrigation amount applied for maize or soybean crops), the predictive models are susceptible to over- or underestimation of yield in heavily or modestly irrigated areas. For instance, south central and southwest Nebraska are intensely irrigated to meet crop water demands, and the predictive models, on average, underestimated the observed irrigated soybean yields. In addition, the eastern portion of the state, which is less reliant on irrigation, resulted in the irrigated soybean models overpredicting yields. The crop yield response to irrigation is not only a function of seasonal total irrigation applied, but also a strong function of irrigation timing. Irrigation amount and timing can also impact crop yield differently depending on the crop growth stage when the irrigations are applied. Because none of the models

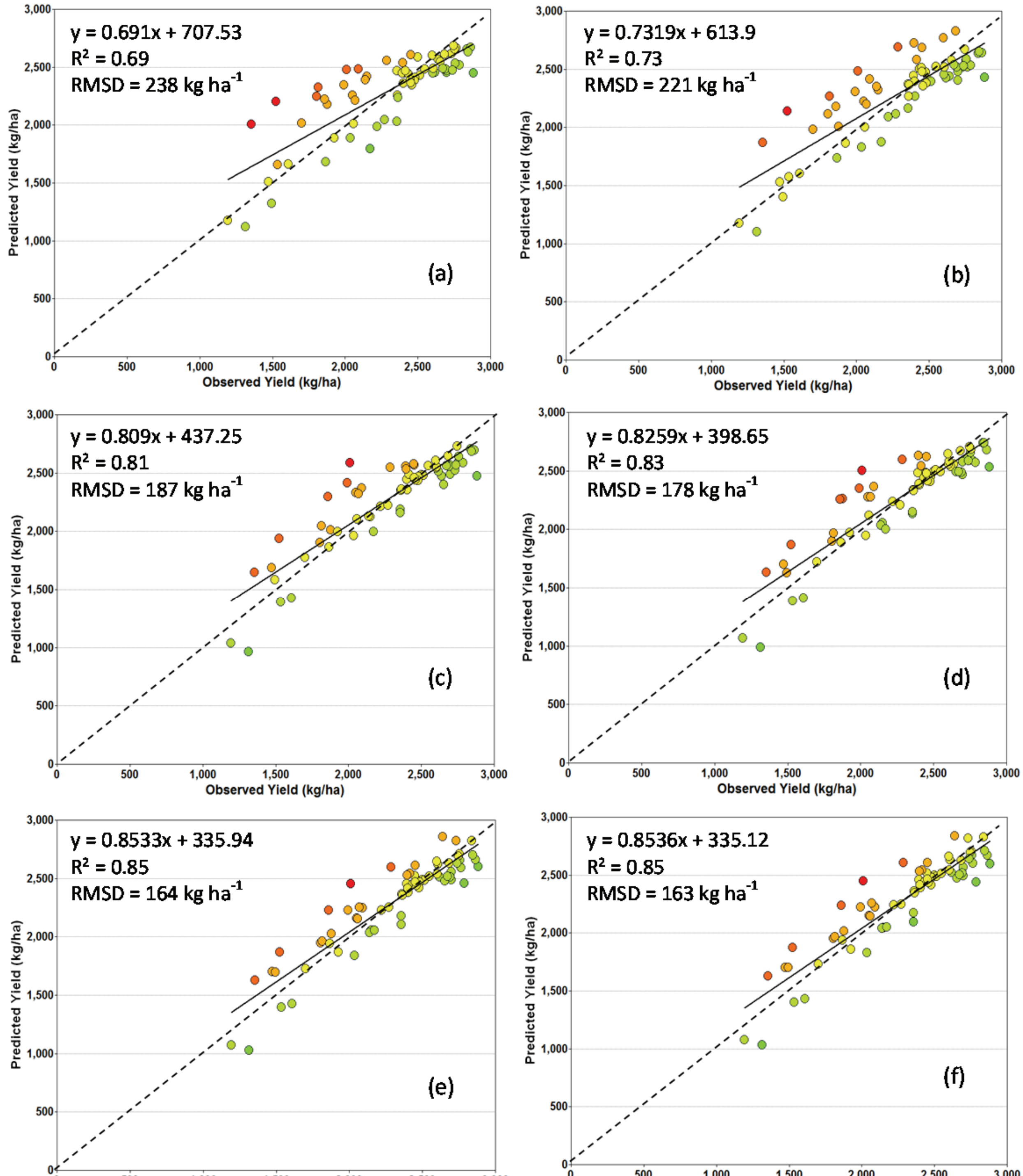

Figure 7. Relationship between observed vs. predicted rainfed soybean yield models by using different predictive variables: (a) Y = *a*(ET_a), (b) Y = *f*(ET_a, P), (c) Y = *f*(ET_a, P, ASW), (d) Y = *f*(ET_a, P, ASW, CEC), (e) Y = *f*(ET_a, P, ASW, CEC, CC), and (f) Y = *f*(ET_a, P, ASW, CEC, CC, OMC), where Y = rainfed soybean yield (kg ha^{-1}), ET_a = actual evapotranspiration (mm), P = precipitation (mm), ASW = available soil water capacity (mm) in the 0-1.20 m soil profile, CEC = cation exchange capacity (meq per 100 g), CC = clay content (%), and OMC = organic matter content (%). The colors of the data points are associated with the county colors on the rainfed soybean standardized residuals map (fig. 6).

accounted for these yield-impacting factors, they were not able to predict irrigated maize and soybean yields as accurately as the rainfed yields for the same crops.

Table 4 shows the incremental R^2 values that resulted from adding explanatory variables to each model. For the rainfed crops, ET_a played the major role in predicting yield, whereas precipitation and available soil water capacity played the major role in predicting irrigated yields. For rainfed crops, about 2% to 4% of the increment was caused by precipitation, and about 4% to 8% of the increment was caused by ASW. The lower correlation between observed and model-predicted yields for the irrigated crops as com-

Table 3. Coefficient of determination (R^2) and ordinary least square (OLS) model coefficients for irrigated maize and soybean yields.

Model	R^2	β_0	β_1	β_2	β_3	β_4	β_5	β_6
Irrigated maize yield								
Yield = f(ET$_a$)	0.19	13677.6	-4.93	-	-	-	-	-
Yield = f(ET$_a$, P)	0.38	4269.27	3.67	10.46	-	-	-	-
Yield = f(ET$_a$, P, ASW)	0.45	2995.8	4.6	10.59	4.54	-	-	-
Yield = f(ET$_a$, P, ASW, CEC)	0.45	2960.1	4.46	10.97	5.3	-7.28	-	-
Yield = f(ET$_a$, P, ASW, CEC, CC)	0.45	3104.54	4.44	10.6	5.31	-27.82	17.74	-
Yield = f(ET$_a$, P, ASW, CEC, CC, OMC)	0.46	3163.44	4.43	10.4	5.08	-40.02	25.43	66.86
Irrigated soybean yield								
Yield = f(ET$_a$)	0.06	4253.84	-1.4	-	-	-	-	-
Yield = f(ET$_a$, P)	0.23	-193.24	2.78	5.49	-	-	-	-
Yield = f(ET$_a$, P, ASW)	0.24	-323.42	2.88	5.5	0.51	-	-	-
Yield = f(ET$_a$, P, ASW, CEC)	0.33	-1477.8	3.41	7.89	1.92	-15.05	-	-
Yield = f(ET$_a$, P, ASW, CEC, CC)	0.34	-1349.6	3.3	7.69	1.97	-31.59	13.62	-
Yield = f(ET$_a$, P, ASW, CEC, CC, OMC)	0.36	-1168.02	3.15	7.6	1.82	-40.02	19.23	39.44

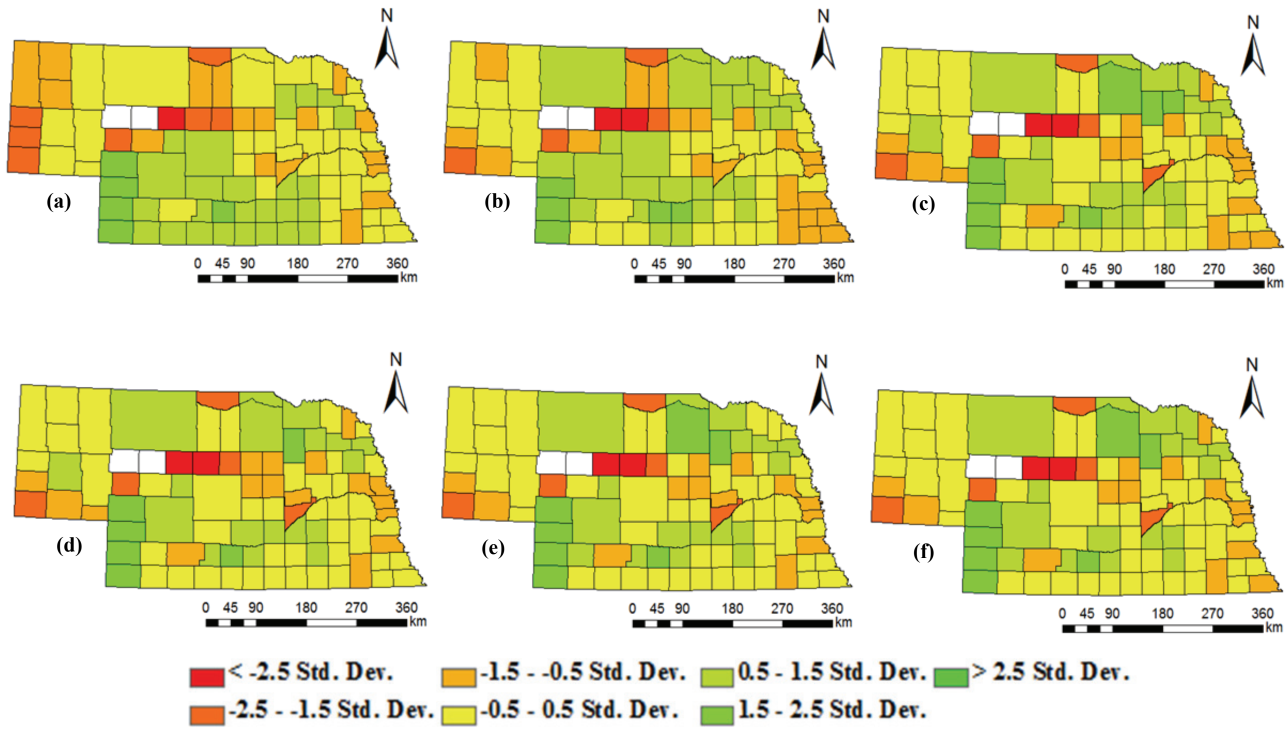

Figure 8. Standardized residual (standard deviation) maps for various models: (a) Y = f(ET$_a$), (b) Y = f(ET$_a$, P), (c) Y = f(ET$_a$, P, ASW), (d) Y = f(ET$_a$, P, ASW, CEC), (e) Y = f(ET$_a$, P, ASW, CEC, CC), and (f) Y = f(ET$_a$, P, ASW, CEC, CC, OMC), where Y = irrigated maize yield (kg ha^{-1}), ET$_a$ = actual evapotranspiration (mm), P = precipitation (mm), ASW = available soil water capacity (mm) in the 0-1.20 m soil profile, CEC = cation exchange capacity (meq per 100 g), CC = clay content (%), and OMC = organic matter content (%).

pared with the rainfed crops was most likely due to the inability of the models to account for the county-level within-season irrigation amounts, as discussed previously.

Zonal Models

Substantial variation was observed in terms of the impact of variables on crop yield across the state (i.e., the R^2 between predicted vs. observed yield varied considerably from the eastern to the western part of the state). To further evaluate the overall importance of the yield-driving factors in predicting maize and soybean yields for rainfed and irrigated conditions, crop yield models were developed for different zones. Similar to the state analysis, the importance of the explanatory variables for each crop within each zone was determined by the R^2 values, as reported in figure 12. Hereafter, models developed for the entire state will be referred to as "state models" for easy comparison to zonal models.

The zonal rainfed maize models explained less yield variability than the state models. The R^2 values when ET$_a$ was the only explanatory variable were 0.06, 0.12, 0.48, and 0.40 for zones 1, 2, 3, and 4, respectively, as compared with the state model R^2 of 0.73. Maximum variability of 36%, 78%, 80%, and 61%, respectively, was explained by the model Yield = f(ET$_a$, P, ASW, CEC, CC, OMC) for zones 1, 2, 3, and 4, respectively, as compared with 83% for the state model Yield = f(ET$_a$, P, ASW, CEC, CC) (fig. 12). Unlike the state model, the addition of OMC helped to explain rainfed maize yield variability in all zones. The inclusion of precipitation was the main descriptor for the western (drier) zones (1 and 2) where, on average, seasonal precipitation amounts are less than poten-

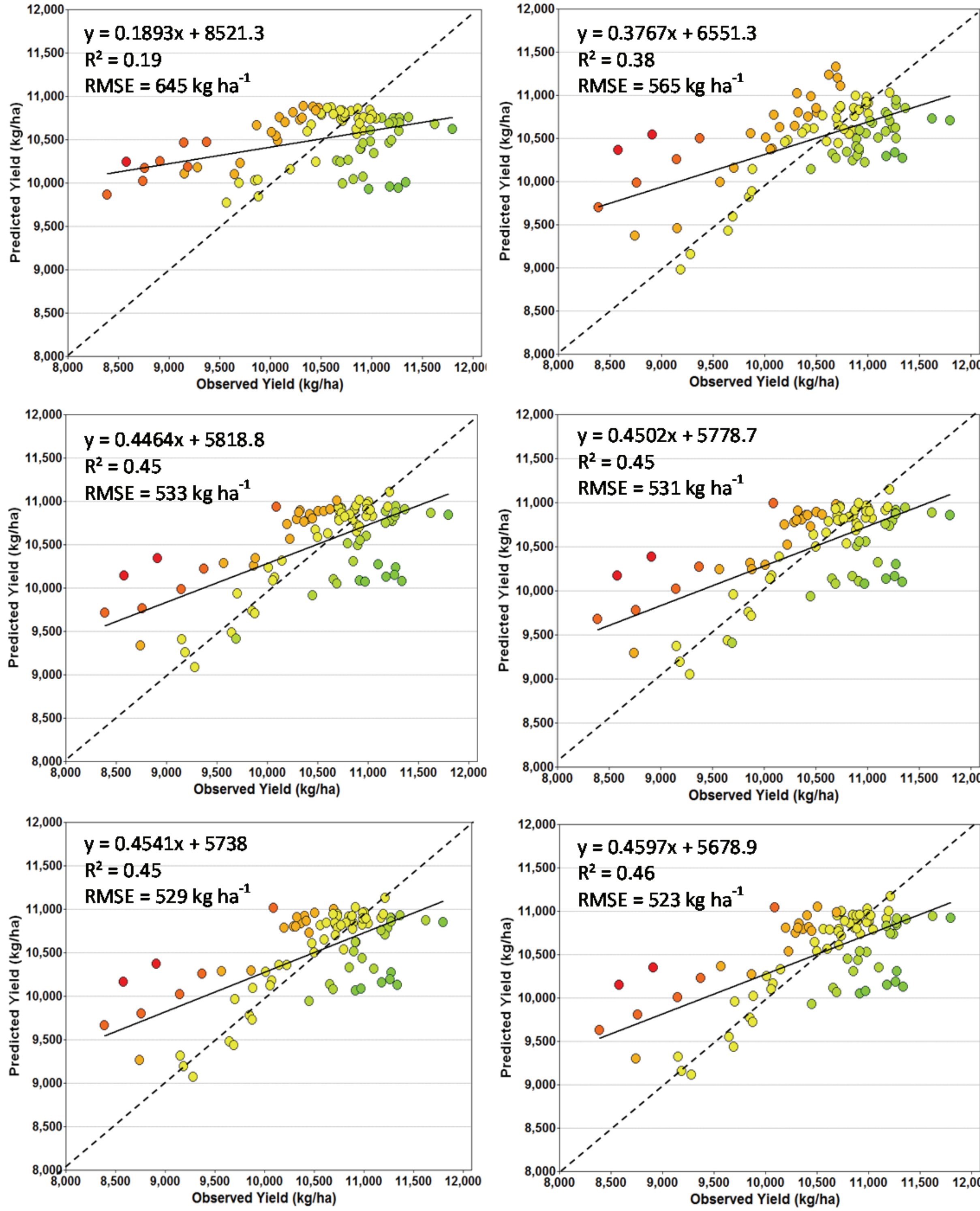

Figure 9. Relationship between observed vs. predicted irrigated maize yield models by using different predictive variables: (a) Y = *f*(ET$_a$), (b) Y = *f*(ET$_a$, P), (c) Y = *f*(ET$_a$, P, ASW), (d) Y = *f*(ET$_a$, P, ASW, CEC), (e) Y = *f*(ET$_a$, P, ASW, CEC, CC), and (f) Y = *f*(ET$_a$, P, ASW, CEC, CC, OMC), where Y = irrigated maize yield (kg ha^{-1}), ET$_a$ = actual evapotranspiration (mm), P = precipitation (mm), ASW = available soil water capacity (mm) in the 0-1.20 m soil profile, CEC = cation exchange capacity (meq per 100 g), CC = clay content (%), OMC = organic matter content (%). The colors of the data points are associated with the county colors on the irrigated maize standardized residuals map (fig. 8).

tial ET demands, with an increase in R^2 from 0.06 to 0.21 for zone 1 and from 0.12 to 0.50 for zone 2. The eastern zones (3 and 4) that receive greater precipitation amounts were influenced the most by seasonal ET$_a$. The change in R^2 values for the zonal rainfed maize models are shown in figure 12a.

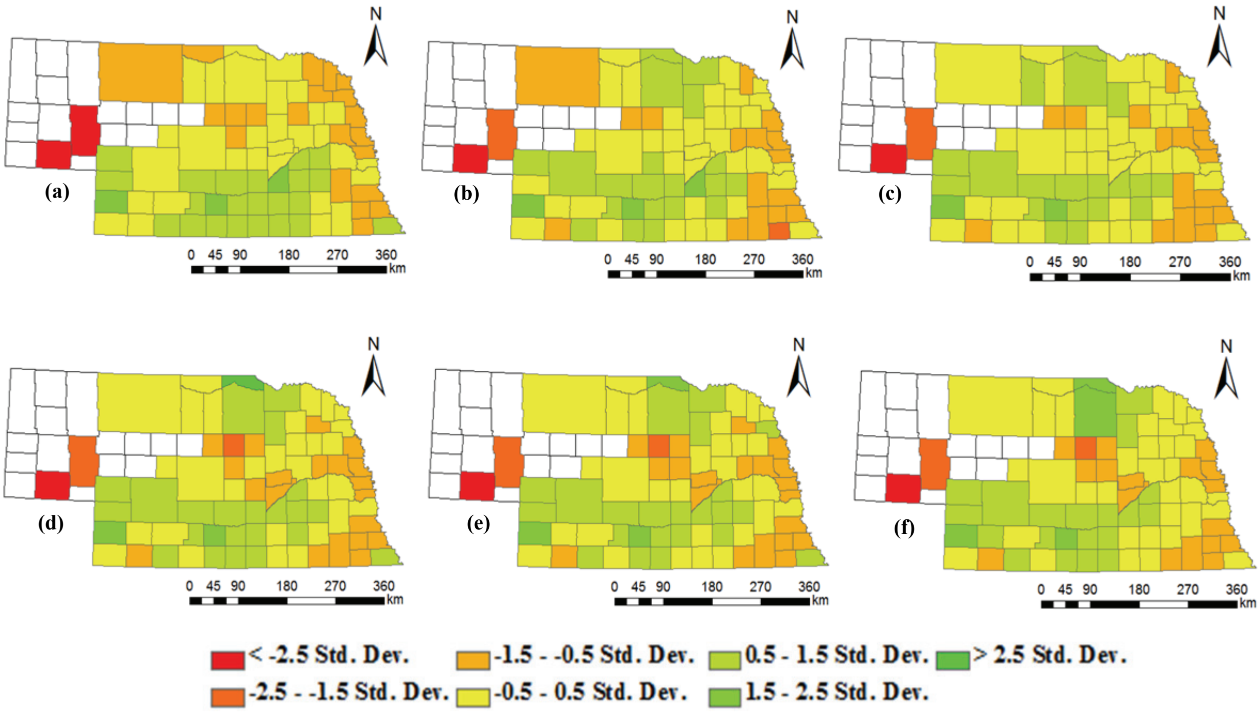

Figure 10. Standardized residual (standard deviation) maps for various models: (a) Y = f(ET$_a$), (b) Y = f(ET$_a$, P), (c) Y = f(ET$_a$, P, ASW), (d) Y = f(ET$_a$, P, ASW, CEC), (e) Y = f(ET$_a$, P, ASW, CEC, CC), and (f) Y = f(ET$_a$, P, ASW, CEC, CC, OMC), where Y = irrigated soybean yield (kg ha^{-1}), ET$_a$ = actual evapotranspiration (mm), P = precipitation (mm), ASW = available soil water capacity (mm) in the 0-1.20 m soil profile, CEC = cation exchange capacity (meq per 100 g), CC = clay content (%), and OMC = organic matter content (%).

Similar to the rainfed maize models, equal or less rainfed soybean yield variability was explained by the zonal models as compared with the state models. Actual evapotranspiration was the greatest predictor of rainfed soybean yield variability for all zones, with R^2 values between observed and predicted yield of 0.75, 0.44, and 0.26 for zones 2, 3, and 4, respectively (fig. 12b). Furthermore, the climatic variables (ET$_a$ and precipitation) accounted for 96%, 73%, and 67% of the total explained variation in yield, whereas 4%, 27%, and 33% of yield variability was explained by soil physical and chemical properties for zones 2, 3, and 4, respectively. The addition of OMC as an explanatory variable for zones 2 and 3 provided no further explanation of rainfed soybean yield variability; however, the addition of OMC in zone 4 (the wettest part of the state) improved the model prediction and resulted in a final R^2 value of 0.48.

Unlike rainfed conditions, irrigated maize and soybean yield predictions were improved in certain zones by applying the zonal OLS models rather than the state models. For irrigated maize, the maximum variability explained by the model Yield = f(ET$_a$, P, ASW, CEC, CC, OMC) was 80%, 86%, 44%, and 40% for zones 1, 2, 3, and 4, respectively, as compared with 46% for the state model (fig. 9 vs. fig. 12). The greater zonal R^2 values are due to similar management practices existing within a zone, including irrigation methods and amounts. The use of zones having common management practices results in better yield prediction from the explanatory variables used in the study. For the state models, management practices vary greatly across the state, leading to less accurate yield prediction. All zonal models were able to strengthen their predictive powers by adding additional soil properties above the Yield = f(ET$_a$, P, ASW) model (fig. 12c). The R^2 values for zones 1 and 2 were considerably greater than the best performing state model, whereas zones 3 and 4 individually had nearly the same R^2 values. As mentioned earlier, the inability to account for irrigation amounts, irrigation method, maize hybrid or soybean variety characteristics, within-season irrigation management practices, and other management practices might have limited the performance of the models, especially for the state models. The maximum amounts of yield variability explained by the irrigated soybean zonal models were 47%, 49%, and 49% for zones 2, 3, and 4, respectively, which in all cases outperformed the state model that explained only 36% of the yield variability. The maximum variability for zones 2 and 3 was explained by the model Yield = f(ET$_a$, P, ASW, CEC, CC, OMC), whereas the maximum variability for zone 4 was explained by the model Yield = f(ET$_a$, P, ASW, CEC, CC). The state model, as well as the model for zone 3, was mostly influenced by precipitation, with an increase in R^2 from 0.06 to 0.23 and from 0.16 to 0.34, respectively. The change in R^2 values for the development of zone-specific irrigated soybean models are presented in figure 12d. Zones 2 and 4 were governed mostly by ET$_a$, and organic matter content played a more important role in the western (wettest) portion of the state than in other locations.

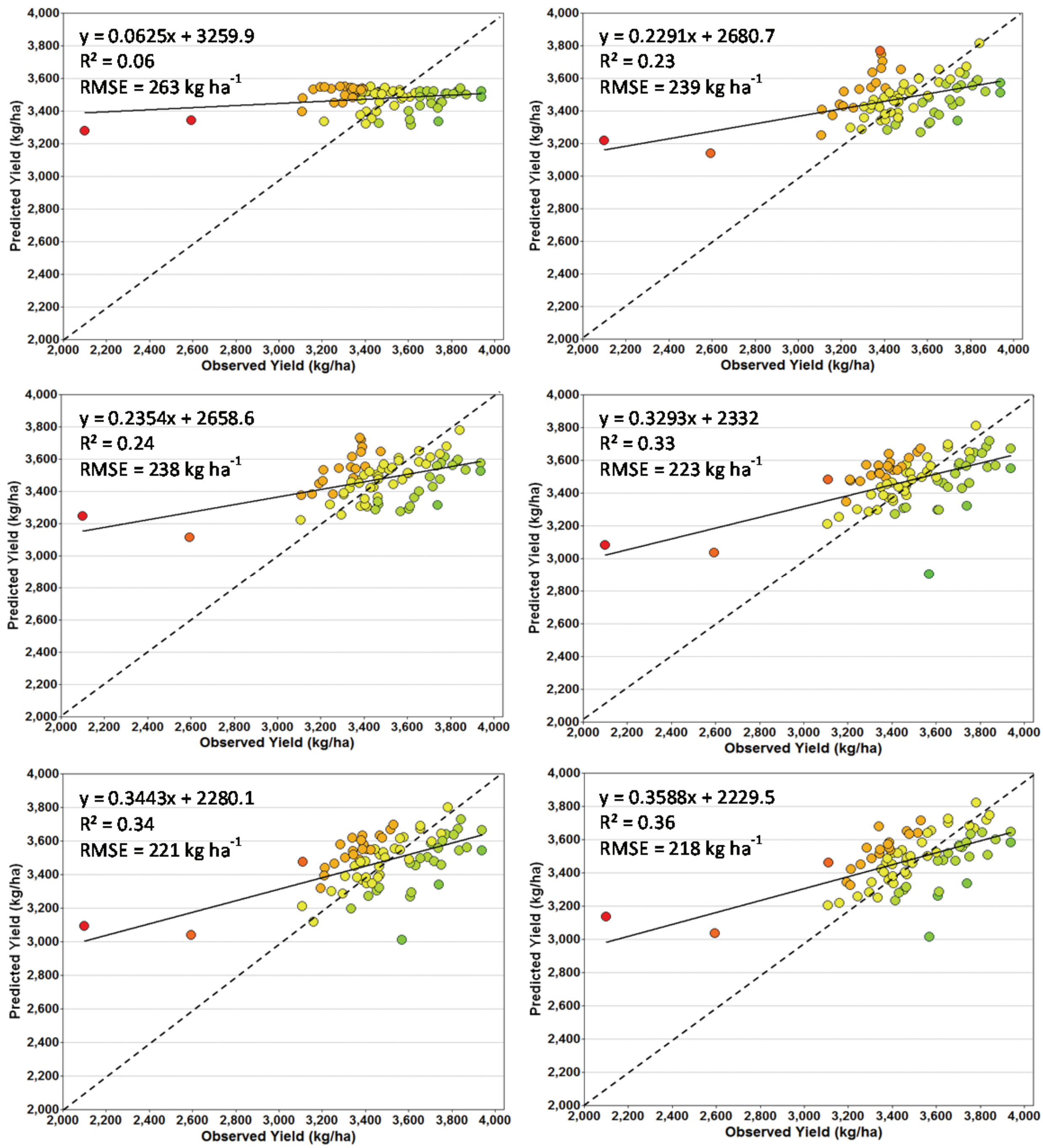

Figure 11. Relationship between observed vs. predicted irrigated soybean yield models by using different predictive variables: (a) Y = f(ET$_a$), (b) Y = f(ET$_a$, P), (c) Y = f(ET$_a$, P, ASW), (d) Y = f(ET$_a$, P, ASW, CEC), (e) Y = f(ET$_a$, P, ASW, CEC, CC), and (f) Y = f(ET$_a$, P, ASW, CEC, CC, OMC), where Y = irrigated soybean yield (kg ha^{-1}), ET$_a$ = actual evapotranspiration (mm), P = precipitation (mm), ASW = available soil water capacity (mm) in the 0-1.20 m soil profile, CEC = cation exchange capacity (meq per 100 g), CC = clay content (%), and OMC = organic matter content (%). The colors of the data points are associated with the county colors on the irrigated soybean standardized residuals map (fig. 10).

Table 4. Incremental R^2 (%) values by adding explanatory variables to the yield predicting model using ordinary least square (OLS) regression.

		Rainfed Maize		Rainfed Soybean		Irrigated Maize		Irrigated Soybean	
Model	No. of Explanatory Variables	R^2 (%)	Increm. R^2 (%)	R^2 (%)	Increm. R^2 (%)	R^2 (%)	Increm. R^2 (%)	R^2 (%)	Increm. R^2 (%)
Yield = f(ET$_a$)	1	73	-	69	-	19	-	0.1	-
Yield = f(ET$_a$, P)	2	75	2	73	4	38	19	23	23
Yield = f(ET$_a$, P, ASW)	3	79	4	81	8	45	7	24	1
Yield = f(ET$_a$, P, ASW, CEC)	4	82	3	83	2	45	0	33	9
Yield = f(ET$_a$, P, ASW, CEC, CC)	5	83	1	85	2	45	0	34	1
Yield = f(ET$_a$, P, ASW, CEC, CC, OMC)	6	83	0	85	0	45	0	36	2

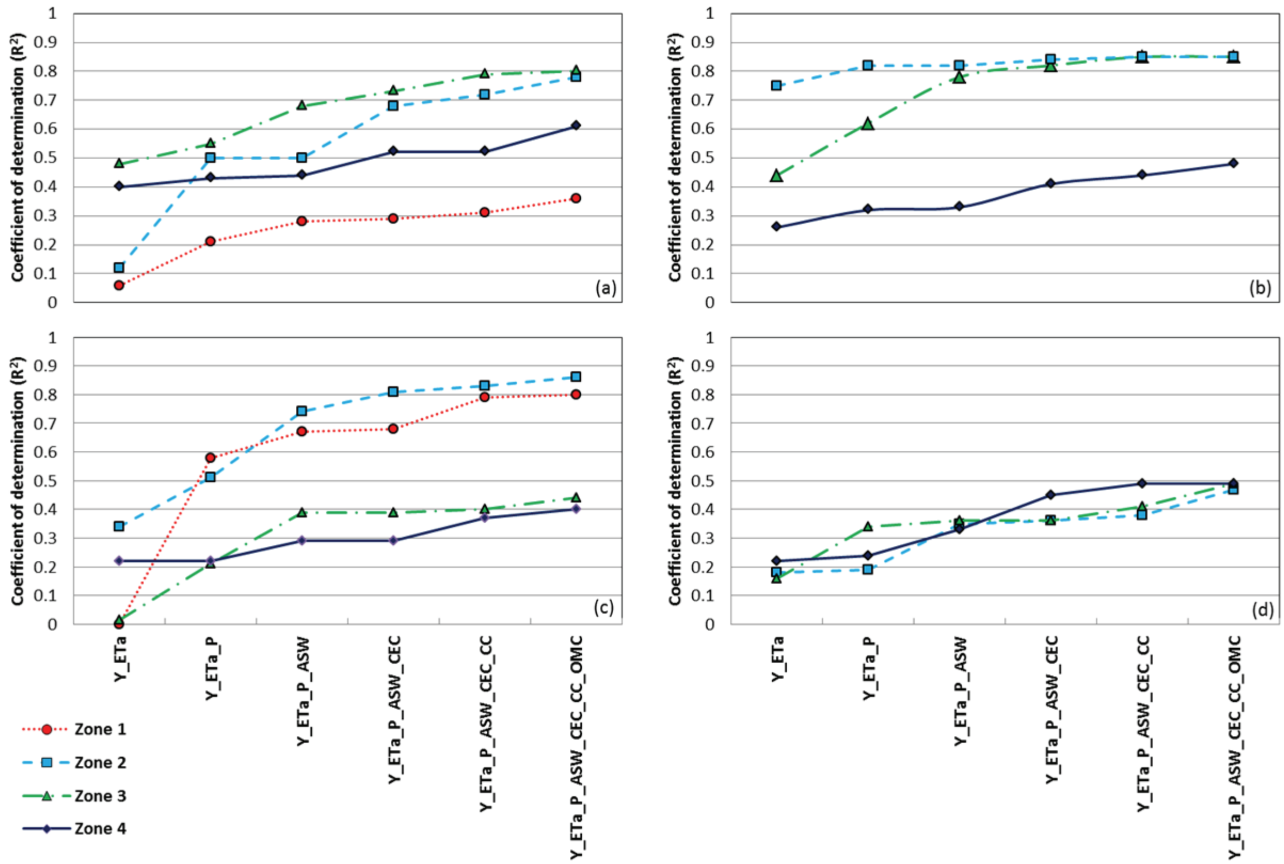

Figure 12. Coefficient of determination (R^2) between observed and predicted yield for (a) rainfed maize, (b) rainfed soybean, (c) irrigated maize, and (d) irrigated soybean for four zones across Nebraska. Zone 1 is the western (driest) and zone 4 is the eastern (wettest) part of the state.

Summary and Conclusions

The use of OLS regression techniques to understand the impacts and relationships between climatic variables and soil physical and chemical properties on irrigated and rainfed yields on a large scale (county, state, regional) is relatively new in the field of agricultural sciences. Accessibility to data from various weather stations and agricultural agencies coupled with information on soil physical and chemical properties can be used to develop OLS regression models to predict yield variability on a regional scale. In addition, these models account for the large-scale heterogeneity beyond the field level and, in combination with spatial analyses, allow identification of yield stability regions. To the best of our knowledge, this study is the first to couple various soil physical and chemical properties and climatic variables as well as soil water availability and actual crop evapotranspiration to predict irrigated and rainfed maize and soybean yields on a large scale. Models were developed for the entire state of Nebraska as well as for each of four zones. Kriging, spline, and inverse distance weighting (IDW) interpolation techniques were used to spatially estimate seasonal (May 1 to Sept. 30) precipitation and reference (potential) evapotranspiration (ET_{ref}). In general, all three interpolation methods performed similarly, with the spline method performing slightly better than the other two methods. The best state models that explained the maximum yield variability for rainfed and irrigated maize and soybean were: Yield = f(ET_a, P, ASW, CEC, CC), with an R^2 of 0.83 and RMSD of 690 kg ha^{-1}; Yield = f(ET_a, P, ASW, CEC, CC), with an R^2 of 0.85 and RMSD of 164 kg ha^{-1}; Yield = f(ET_a, P, ASW), with an R^2 of 0.45 and RMSD of 533 kg ha^{-1}; and Yield = f(ET_a, P, ASW, CEC, CC, OMC), with an R^2 of 0.36 and RMSD of 218 kg ha^{-1}. No additional yield variability was explained by adding OMC to the predictive state models for all crops, except irrigated soybean. Differences in the impact of explanatory variables on predicting crop yield were observed for different zones across the state. The zonal models provided insight into which explanatory variables were most important in predicting yield for a given crop within a climatic zone. Precipitation had the greatest influence on explaining rainfed maize yield variability for the western (drier) zones (1 and 2), whereas actual evapotranspiration (ET_a) explained the majority of the yield variability in the eastern (wetter) zones (3 and 4). Unlike the state model for rainfed maize yields, an increase in R^2 was observed by adding OMC to the models for each zone.

The zonal models for rainfed and irrigated maize and soybean provided important and useful insight into which parameters are most influential in predicting yield in different climatic zones. The division of Nebraska into climatic zones with similar characteristics can help reduce the error from other yield-influencing factors that are not available (e.g., within-season irrigation management practices, soil and crop management practices, nutrient management, etc.). These parameters on a county level are unavailable from any source. Therefore, we recommend that zonal models be used for predicting irrigated maize and soybean yields in climatic conditions similar to those that existed in this study so that the variability of the unaccounted factors decreases, resulting in a higher predictive power. However, for rainfed predictions, we suggest the use of state models since the climatic and soil physical and chemical properties are the primary influencing factors, and dividing the state into zones may mask certain properties, hindering the predictive power.

References

Adams, R. M., B. H. Hurd, S. Lenhart, and N. Leary. 1998. Effects of global climate change on agriculture: An interpretative review. *Clim. Res.* 11(1): 19-30.

Anderson, J. L. 2002. A local least squares framework for ensemble filtering. *Monthly Weather Rev*. 131(4): 634-642.

Bakhsh, A., and R. S. Kanwar. 2004. Using discriminant analysis and GIS to delineate subsurface drainage patterns. *Trans. ASAE* 47(3): 689-699.

Blennow K., and P. Persson. 1998. Modelling local-scale frost variations using mobile temperature measurements with a GIS. *Agric. Forest Meteorol.* 89(1): 59-71.

Bryant, C. R., B. Smit, M. Brklacich, T. R. Johnston, J. Smithers, Q. Chiotti, and B. Singh. 2000. Adaptation in Canadian agriculture to climatic variability and change. In *Societal Adaptation to Climate Variability and Change*, 181-201. S. M. Kane and G. W. Yohe, eds. Dordrecht, The Netherlands: Kluwer Academic.

Casanova, D., J. Goudriaan, J. Bouma and G. F. Epema. 1999. Yield gap analysis in relation to soil properties in direct-seeded flooded rice. *Geoderma* 91(3-4): 191-216.

Collins, F. C., and P. V. Bolstad. 1996. A comparison of spatial interpolation techniques in temperature estimation. In *Proc. 3rd Intl. Conf. and Workshop on Integrating GIS and Environ. Modeling*. Santa Barbara, Cal.: National Center for Geographic Information and Analysis.

Curran, P. J., and P. M. Atkinson. 1998. Geostatistics and remote sensing. *Prog. Physical Geography* 22(1): 61-78.

Curran, P. J., G. M. Foody, and P. R. van Gardingen. 1997. Scaling-up. In *Scaling-up from Cell to Landscape*, 1-5. P. R. van Gardingen, G. M. Foody, and P. J. Curran, eds. Cambridge, U.K.: Cambridge University Press.

Curry, R. B., R. M. Peart, J. W. Jones, K. J. Boote, and L. H. Allen Jr. 1990. Simulation as a tool for analyzing crop response to climate change. *Trans. ASAE* 33(3): 981-990.

Dalthorp, D., J. Nyrop, and M. Villani. 1999. Estimation of local mean population densities of Japanese beetle grubs (Scarabaeidae: Coleoptera). *Environ. Entomol.* 28(2): 255-265.

Dent, J. B., and P. K. Thornton. 1988. The role of biological simulation models in farming systems research. *Agric. Admin. and Extension* 29(2): 111-122.

Dodson, R., and D. Marks. 1997. Daily air temperature interpolated at high spatial resolution over a large mountainous region. *Clim. Res.* 8(1): 1-20.

Goovaerts, P. 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol.* 228(1-2): 113-129.

Hammond, T., and J. Yarie. 1996. Spatial prediction of climatic state factor regions in Alaska. *EcoScience* 3(4): 490-501.

Hansen, J. W., and J. W. Jones. 2000. Scaling-up crop models for climate variability applications. *Agric. Systems* 65(1): 43-72.

Hevesi, J. A., J. D. Istok, and A. L. Flint. 1992. Precipitation estimation in mountainous terrain using multivariate geostatistics: Part I. Structural analysis. *J. Appl. Meteorol.* 31(7): 661-676.

Holdaway, M. R. 1996. Spatial modelling and interpolation of monthly temperature using kriging. *Clim. Res.* 6(3): 215-225.

Hosseini, E., J. Gallichand, and D. Marcotte. 1994. Theoretical and experimental performance of spatial interpolation methods for soil salinity analysis. *Trans. ASAE* 37(6): 1799-1807.

Hudson, G., and H. Wackernagel. 1994. Mapping temperature using kriging with an external drift: Theory and an example from Scotland. *Intl. J. Climatol.* 14(1): 77-91.

Hulme, M., D. Conway, P. D. Jones, T. Jiang, E. M. Barrow, and C. Turney. 1995. Construction of a 1961-1990 European climatology for climate change modeling and impact application. *Intl. J. Climatol*. 15(12): 1333-1363.

Irmak, A., J. W. Jones, W. D. Batchelor, and J. O. Paz. 2001. Estimating spatially variable soil properties for crop model use in precision farming. *Trans. ASAE* 42(6): 1867-1877.

Irmak, A., J. W. Jones, W. D. Batchelor, and J. O. Paz. 2002. Linking multi-variables for diagnosing causes of spatial yield variability in soybean. *Trans. ASAE* 45(3): 839-849.

Irmak, A., J. W. Jones, and S. S. Jagtap. 2005. Evaluation of the CROPGRO-Soybean model for assessing climate impacts on regional soybean yields. *Trans. ASAE* 48(6): 2343-2353.

Irmak, A., J. W. Jones, W. D. Batchelor, S. Irmak, J. O. Paz, and K. J. Boote. 2006. Analysis of spatial yield variability using a combined crop model empirical approach. *Trans. ASABE* 49(3): 811-818.

Irmak, A., P. Ranade, D. Marx, S. Irmak, K. G. Hubbard, G. Meyer, and D. L. Martin. 2010. Spatial interpolation of climate variables in Nebraska. *Trans. ASABE* 53(6): 1759-1771.

Irmak, S., I. Kabenge, K. E. Skaggs, and D. Mutiibwa. 2012. Trend and magnitude of changes in climatic variables and reference evapotranspiration over 116-year period in the Platte River valley, central Nebraska, USA. *J. Hydrol.* 420-421: 228-244.

Jones, C. A., and J. R. Kiniry, eds. 1986. *CERES-Maize: A Simulation Model of Maize Growth and Development*. College Station, Tex.: Texas A&M University Press.

Jones, J. W., G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie. 2003. The DSSAT cropping system model. *European J. Agron.* 18(3-4): 235-265.

Kaspar, T .C., T. S. Colvin, D. B. Jaynes, D. L. Karlen, D. E. James, D. W. Meek, D. Pulido, and H. Butler. 2003. Relationship between six years of corn yields and terrain attributes. *Precision Agric*. 4(1): 87-101.

Kollias, V. J., D. P. Kalivas, and N. J. Yassoglou. 1999. Mapping the soil resources of a recent alluvial plain in Greece using fuzzy sets in a GIS environment. *J. Soil Sci.* 50(2): 261-273.

Kravchenko, A. N., and D. G. Bullock. 1999. A comparative study of interpolation methods for mapping soil properties. *Agron. J.* 91(3): 393-400.

Kravchenko, A. N., and D. G. Bu1lock. 2002a. Spatial variability of soybean quality data as a function of field topography: I. Spatial data analysis. *Crop Sci.* 42(3): 804-815.

Kravchenko, A. N., and D. G. Bullock. 2002b. Spatial variability of soybean quality data as a function of field topography: II. A proposed technique for calculating the size of the area for differential soybean harvest. *Crop Sci.* 42(3): 816-821.

Kreznor, W., K. Olson, W. Banwart, and D. Johnson. 1989. Soil, landscape, and erosion relationship in a northwest Illinois watershed. *SSSA J.* 53(6): 1763-1771.

Kulkarni, S. S., S. G. Bajwa, J. C. Rupe, and T. L. Kirkpatrick. 2008. Spatial correlation of crop response to soybean cyst nematode (*Heterodera glycines*). *Trans. ASABE* 51(4): 1451-1459.

Kurtzman, D., and R. Kadmon. 1999. Mapping of temperature variables in Israel: A comparison of different interpolation methods. *Clim. Res.* 13(1): 33-43.

Lal, H., G. Hoogenboom, J. P. Calixte, J. W. Jones, and F. H. Beinroth. 1993. Using crop simulation models and GIS for regional productivity analysis. *Trans. ASAE* 36(1): 175-184.

Lal, R. 1989. Conservation tillage for sustainable agriculture: Tropics vs. temperate environments. *Adv. in Agron.* 42: 85-197.

Lal, R. 1997. Long-term tillage and maize monoculture effects on a tropical Alfisol in western Nigeria: 2. Soil chemical properties. *Soil and Tillage Res.* 42(3): 161-174.

Letey, J. 1985. Relationship between soil physical properties and crop production. *Adv. in Soil Sci.* 1: 276-294.

Li, J., J. Huang, and X. Wang. 2006. A GIS-based approach for estimating spatial distribution of seasonal temperature in Zhejiang Province, China. *J. Zhejiang University Sci. A* 7(4): 647-656.

Lobell, D. B., K. N. Cahill, and C. B. Field. 2007. Historical effects of temperature and precipitation on California crop yields. *Clim. Change* 81(2): 187-203.

McBratney, A. B., and M. J. Pringle. 1997. Spatial variability in soil: Implications for precision agriculture. In *Proc. 1st European Conf. on Precision Agric*., vol. 1: 3-31. J. V. Stafford, ed. Oxford, U.K.: BIOS Scientific.

McCarthy, J. J., O. F. Canziani, N. A. Leary, D. J. Dokken, and K. S. White. 2001. *Climate Change 2001: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, U.K.: Cambridge University Press.

McKinion, J. M., J. L. Willers, and J. N. Jenkins. 2010. Spatial analysis to evaluate multi-crop yield stability for a field. *Computers and Electronics in Agric.* 70(1): 187-198.

Merino, G. G., D. Jones, D. E. Stooksbury, and K. G. Hubbard. 2001. Determination of semivariogram models to krige hourly and daily solar irradiance in western Nebraska. *J. Applied Meteorol.* 40(6): 1085-1094.

Miller, M., M. Singer, and D. Nielsen. 1988. Spatial variability of wheat yield and soil properties on complex hills. *SSSA J.* 52(4): 1133-1141.

Monteith, J. L. 1965. Evaporation and the environment. In *Proc. 19th Symp. Society for Experimental Biology: The State and Movement of Water in Living Organism*, 205-234. Cambridge, U.K.: Cambridge University Press.

Moore, I., P. Gessler, G. Nielsen, and G. Peterson. 1993. Soil attribute prediction using terrain analysis. *SSSA J.* 57(2): 443-452.

Nalder, I. A., and R. W. Wein. 1998. Spatial interpolation of climatic normals: Test of a new method in the Canadian boreal forest. *Agric. Forest Meteorol.* 92(4): 211-225.

Ninyerola, M., X. Pons, and J. M. Roure. 2000. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *Intl. J. Clim.* 20(14): 1823-1841.

Okogun, J. A., B. T. Otuyemi, and N. Sanginga. 2004. Soybean yield determination and response of rhizobial inoculation in an on-farm trial in the northern Guinea savanna of Nigeria. *West African J. Appl. Ecol.* 6: 30-39.

Prasad, R., and J. F. Power. 1991. Crop residue management: Literature review. *Adv. Soil Sci.* 15: 205-251.

Reidsma, P., F. Ewert, A. O. Lansink, and R. Leemans. 2009. Vulnerability and adaptation of European farmers: A multi-level analysis of yield and income responses to climate variability. *Regional Environ. Change* 9(1): 25-40.

Rosenthal, W., G. Hammer, and D. Butler. 1998. Predicting regional grain sorghum production in Australia using spatial data and crop simulation modelling. *Agric. Forest Meteorol.* 91(3): 263-274.

Sarangi, A., C. A. Cox, and C. A. Madramootoo. 2005. Geostatistical methods for prediction of spatial variability of rainfall in a mountainous region. *Trans. ASAE* 48(3): 943-954.

Sharma, V., and S. Irmak. 2012a. Mapping spatially interpolated precipitation, reference evapotranspiration, actual crop evapotranspiration, and net irrigation requirement in Nebraska: Part I. Precipitation and reference evapotranspiration. *Trans. ASABE* 55(3): 907-921.

Sharma, V., and S. Irmak. 2012b. Mapping spatially interpolated precipitation, reference evapotranspiration, actual crop evapotranspiration, and net irrigation requirement in Nebraska: Part II. Actual evapotranspiration and net irrigation requirement. *Trans. ASABE* 55(3): 923-936.

Sharma, V., A. Irmak, I. Kabenge, and S. Irmak. 2011. Application of GIS and geographically weighted regression to evaluate the spatial non-stationarity relationship between precipitation vs. irrigated and dryland maize and soybean. *Trans. ASABE* 54(3): 953-972.

Strahler, A. N., and A. H. Strahler. 1984. *Elements of Physical Geography*. New York, N.Y.: John Wiley and Sons.

Stone, J. R., J. W. Gilliam, D. K. Cassel, R. B. Daniels, L. A. Nelson, and H. J. Kleiss. 1985. Effect of erosion and landscape position on the productivity of piedmont soils. *SSSA J.* 49(4): 987-991.

Tabios, G. Q., and J. D. Salas. 1985. A comparative analysis of techniques for spatial interpolation of precipitation. *Water Resour. Bull.* 21(3): 365-380.

Thornton, P. K., A. R. Saka, U. Singh, J. D. T. Kumwenda, J. E. Brink, and J. B. Dent. 1995. Application of a maize crop simulation model in the central region of Malawi. *Exp. Agric.* 31(2): 213-226.

USDA-NASS. 2009. Census of agriculture: Farm and ranch irrigation survey. Washington, D.C.: USDA National Agricultural Statistics Service.

USDA-NRCS. 2009. Map of maize and soybean no-till acreages in Nebraska. Lincoln, Neb.: USDA Natural Resources Conservation Service. Available at: www.ne.nrcs.usda.gov/technical/notill.html.

Vicente-Serrano, S. M., and J. M. Cuadrat. 2003. Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): Application to annual precipitation and temperature. *Clim. Res.* 24(2): 161-180.

Watson, D. F., and G. M. Philip. 1985. A refinement of inverse distance weighted interpolation. *Geo-Processing* 2: 315-327.

Weber, D., and E. Englund. 1992. Evaluation and comparison of spatial interpolators. *Math. Geol.* 24(4): 381-391.

Webster, R. 1985. Quantitative spatial analysis of soil in the field. *Advances in Soil Sci.* 3: 1-70.

Willmott, C. J., and K. Matsuura. 1995. Smart interpolation of annually averaged air temperature in the United States. *J. Appl. Meteorol.* 34(12): 2577-2586.

Willmott, C. J., and S. M. Robeson. 1995. Climatologically aided interpolation of terrestrial temperature. *Intl. J. Clim.* 15(2): 221-229.

Wright, R., D. Boyer, W. Winant, and H. Perry. 1990. The influence of soil factors on yield differences among landscape positions in an Appalachian cornfield. *Soil Sci.* 149(6): 375-382.