Elsevier Editorial System(tm) for Applied Geography
Manuscript Draft

Manuscript Number:

Title: Geographic Determinants of Rural Land Covers and the Agricultural Margin in the Central United States

Article Type: Article

Keywords: Corn Belt; GIS; Land Cover Modeling; Multivariable Fractional Polynomial; Raster Regression

Corresponding Author: Dr. Timothy Stoebner,

Corresponding Author's Institution: Southern Illinois University

First Author: Timothy Stoebner

Order of Authors: Timothy Stoebner; Christopher L Lant, PhD

Abstract: Abstract. Geographic research on the Corn Belt and other regional landscapes of the central U.S. has not to date identified quantitatively the climatic, edaphic, topographic, and economic characteristics that determine rural land cover, and that therefore govern land cover change. Using the USDA/NASS Cropland Data Layer, this study identifies these characteristics using Multivariable Fractional Polynomials within a logistic regression framework. It maps the suitability distribution for corn, soybeans, spring and winter wheat, cotton, grassland, and forest land covers that dominate the central U.S., at a 56m resolution across 16 central U.S. states. The non-linear logistic regression models are successful in identifying determinants of land cover with relative operating characteristic (ROC) scores that range from 0.769 for soybeans to 0.888 for forest, with a combined corn/soybean model achieving an ROC of 0.871. For corn and soybean models, when prior land cover of a pixel is added, predictability and ROC scores increase substantially (0.07-0.10), indicating a strong temporal dependency in land cover dynamics due to crop rotation. This process also aids in the delineation of fields from pixels. Adding neighboring land covers further improves predictability and ROC scores only slightly (0.014-0.019), however, indicating a weak spatial dependency mechanism. By including annual crop prices within the logit models, economically marginal cropland that comes into crop production only when prices are high is identified in a spatially-explicit manner. This capacity improves further analyses of economic and environmental impacts of policies that effect crop prices.

Geographic Determinants of Rural Land Covers and the Agricultural Margin in the Central United States

Dr. Timothy J. Stoebner, Environmental Resources and Policy Ph.D. program, Southern Illinois University. tstoebner@siu.edu, 402-661-0847, 405 W Grand, Carbondale, IL 62901

Christopher L. Lant, Professor and Head, Department of Environment and Society, Utah State University, Logan, Utah clant@siu.edu

Geographic Determinants of Rural Land Covers and the Agricultural Margin in the Central United States

Abstract. Geographic research on the Corn Belt and other regional landscapes of the central U.S. has not to date identified quantitatively the climatic, edaphic, topographic, and economic characteristics that determine rural land cover, and that therefore govern land cover change. Using the USDA/NASS Cropland Data Layer, this study identifies these characteristics using Multivariable Fractional Polynomials within a logistic regression framework. It maps the suitability distribution for corn, soybeans, spring and winter wheat, cotton, grassland, and forest land covers that dominate the central U.S., at a 56m resolution across 16 central U.S. states. The non-linear logistic regression models are successful in identifying determinants of land cover with relative operating characteristic (ROC) scores that range from 0.769 for soybeans to 0.888 for forest, with a combined corn/soybean model achieving an ROC of 0.871. For corn and soybean models, when prior land cover of a pixel is added, predictability and ROC scores increase substantially (0.07-0.10), indicating a strong temporal dependency in land cover dynamics due to crop rotation. This process also aids in the delineation of fields from pixels. Adding neighboring land covers further improves predictability and ROC scores only slightly (0.014-0.019), however, indicating a weak spatial dependency mechanism. By including annual crop prices within the logit models, economically marginal cropland that comes into crop production only when prices are high is identified in a spatially-explicit manner. This capacity improves further analyses of economic and environmental impacts of policies that effect crop prices.

Keywords. *Corn Belt, GIS, Land Cover Modeling, Multivariable Fractional Polynomial, Raster Regression*

Geographic Determinants of Rural Land Covers and the Agricultural Margin in the Central United States

1.0 INTRODUCTION

Geographers have long described the characteristic landscapes of the central U.S. (USDA, 2006), including rain-fed corn and soybean fields stretching from central Ohio to eastern Nebraska, irrigated crops overlying the Ogallala aquifer of the southern Plains , spring wheat fields of the northern and winter wheat fields of the southern Plains, grasslands in the western plains, cotton in the Lower Mississippi Valley, and forests of the Canadian Shield and Ozark and other plateaus. While the geographic extents of these regions are easily defined, the climatic, edaphic, topographic, and economic factors that define them, and therefore govern their geographic change, have not been defined in a systematic quantitative manner.

Because land use is a critical lynchpin between social driving forces and environmental outcomes (Foley, et al., 2011; Lambin, Geist, & Rindfuss, 2006), understanding the biophysical and economic determinants that influence the spatial distribution of individual land covers and facilitating their accurate prediction takes on more urgency in a time when those determinants are in flux. Land use conversions alter biogeochemical cycles for water, carbon, nitrogen, and phosphorus, often in environmentally detrimental ways, generating intensified management challenges for maintenance of valuable ecosystem services such as carbon storage, denitrification, soil binding, and wildlife habitat (Bjorklund, Limburg, & Rydberg, 1999; Foley, et al., 2011; Secchi, Tyndall, Schulte, & Asbjornsen, 2008; Tilman, Cassman, Matson, Naylor, & Polasky, 2002). Therefore, distinguishing stable ‘prime’ versus ‘marginal’ agricultural lands that are more often converted among uses is critical to assessing both the environmental effects of expanding and contracting crop production and the economic effects of agricultural activities on local rural economies (Hauser, 2007).

The Corn Belt dominates the central U.S. The productivity of corn production drives out less profitable rural land covers on lands that possess prime soil and climatic characteristics (Baker, 1927; Laingen, 2012). As shown in the solid lines in Figure 1, on lands that possess soil fertility and climatic characteristics favorable for corn, that crop provides greater profitability than its competitors (wheat, grassland, forest) and usually slightly higher than its rotational complement (soybeans). Similarly, wheat is economically superior to grassland and grassland to forest, producing a landscape that is largely organized by agricultural economics on a farm-by-farm basis, with corn rotated with soybeans on the most productive land, followed by wheat, grassland, and forest remnants or regrowth on lands with the lowest agricultural productivity.

Increased biofuel production, population, and meat consumption in developing countries have increased global demand for soybeans and even more so for corn (Rosegrant, 2008; Searchinger, et al., 2008). Together with higher production costs borne of bad weather and rising input prices (especially for petroleum), this demand has generated higher and more variable crop prices over the last decade (McPhail & Babcock, 2008; Mehaffey, Smith, & Van Remortel, 2011). As shown in the dashed lines in Figure 1, an increase in the price of corn raises its profitability relative to soybeans, replacing rotations with continuous corn, and wheat, pressing less favorable land into corn production. This land reallocation decreases the supply of soybeans and wheat, indirectly raising their prices to a lesser degree. Wheat then expands into lands that had been occupied by grassland. In this manner, the price of corn is a leverage point in the geographical pattern of land cover in the central United States.

This article explores, from both an explanatory and a predictive standpoint, the geographical determinants of seven individual land covers that dominate the central U.S.: corn, soybeans, spring and winter wheat, cotton, grassland, and forest. We focus on four research questions. First, how can climatic, edaphic, topographic and economic determinants be quantified for use in explaining and predicting individual land covers? Second, focusing on only corn and soybeans, once the areas of suitability are established, how much added predictability in locating these land covers is gained through knowledge of past and neighboring land uses? That is, what are the spatial and temporal dependencies in determining cropping patterns of the Corn Belt? Third, can the models be utilized to spatially project the suitability of each land cover based on a four-year (2008-2011) time frame? Fourth, what is the geographical expression of supply adjustments or, put in a more geographical context, what specific 'marginal' locations come into corn production when prices rise, and come out of production when prices fall?

2.0 METHODS

2.1 Land Change Modeling Techniques

A central issue in the use of past land covers for land change modeling is the nature of spatial and temporal dependencies. Urban land uses are rarely reversed, making nearly all conversions from rural to urban uses. Land is rarely converted into or out of forest more than once in a few decades, decreasing the likelihood of conversion among rural uses (Alig, Plantinga, Ahn, & Kline, 2003; Osteen, Gottlieb, & Vasavada, 2012; Sohl, et al., 2012). Crop rotation, on the other hand, produces a strong temporal dependency, especially in the Midwestern context where soybeans are stereotypically rotated with corn to capture the benefits of nitrogen fixation and to break pest cycles.

Many land uses, such as (sub)urbanization and tropical deforestation, exhibit the property of spatial contagion where the conversion of one parcel increases the probability that neighboring parcels will also be converted (Brown, Walker, Manson, & Seto, 2004). This type of diffusion is often captured by cellular automata models (Berger, 2001; Verburg, Schot, Dijst, & Veldkamp, 2004) that produce characteristic spatial patterns of contiguous sprawl borne of positive spatial dependencies. Land use researchers have also incorporated temporal dependencies or spatial diffusion into their analyses through the use of process-oriented agent-based models (Valbuena, Verburg, Bregt, & Ligtenberg, 2010; Verburg, 2006) that integrate social factors that influence land change through a variety of potential scenarios (National Research Council, 2013).

Several land change modeling approaches including machine learning, statistical (regression), and cellular automata incorporate climatic, topographic, edaphic, or economic determinants and often utilize satellite-based land cover information (National Research Council, 2013). Provided each of the predictor variables can be represented in a raster format and the modeling approach can be duplicated in a GIS, the results can be mapped on a cell-by-cell basis through raster regression to produce suitability distribution maps of the land covers being modeled. Among the statistical approaches is logistic regression, a common tool in modeling discrete categories (Geoghegan, et al., 2001; Serneels & Lambin, 2001; van Asselen & Verburg, 2012). Recent papers modeled agricultural land use intensity gradients using logistic regression to identify climatic, edaphic, topographic, and other determinants for several European countries at a moderate resolution of 1 km^2 (Temme & Verburg, 2011) and at the worldwide scale with a coarse resolution of five arc-minutes (van Asselen & Verburg, 2012).

Logistic regression, however, does not provide for non-linear effects in its native form. Generalized linear models (GLM) provide the ability to incorporate non-linear functions of independent variables, commonly found in models that include environmental variables (Guisan, Edwards Jr, & Hastie, 2002; Hastie & Tibshirani, 1987). Non-linear approaches also include specific techniques including multivariate adaptive regression splines (MARS), classification and regression tree (CART), generalized additive models (GAM), and multivariable fractional polynomials (MFP) (Guisan, et al., 2002; Hastie & Tibshirani, 1987; Hosmer & Lemeshow, 2000). While MARS and CART techniques offer non-linear approaches, they do not provide the necessary output for raster calculation techniques. The GAM technique does provide similar transformations and output as MFP, but is computationally intensive and therefore convergence cannot be obtained with a large dataset. On the other hand, MFP provides easily interpretable output and easily processes very large datasets. For these reasons, non-linear logistic regression with MFP is employed here to address the research questions posed.

2.2 Study Area and Data Sources

The study area is the sixteen states for which high spatial resolution raster data are available that accurately identify corn, soybeans, spring and winter wheat, cotton, grassland, and forest in locations where variables needed to define the determinants of these land covers can also be represented spatially at the same pixel resolution. The key data set for identifying these land covers is the USDA/NASS Cropland Data Layer (CDL). This 56m resolution raster dataset began limited coverage in 1999 and became available for the entire study area by 2007 (Han, Yang, Di, & Mueller, 2012). The several dozen classifications of the CDL, which include a variety of natural, agricultural, and built-up land covers, were aggregated into 14 major classifications for this study (Figure 2). According to USDA metadata, the CDL has an average 88 percent overall accuracy for the states within the study area and 94 percent accuracy for corn and soybean pixels (NASS, 2013). The study area is composed of 34 percent cropland, 27 percent grassland, 22 percent forest, and 17 percent urban, water, and wetlands. Almost 80 percent of the cropland is comprised of corn and soybeans, which form the distinctive Corn Belt region in the north central portion of the study area (Figure 2).

The U.S. Corn Belt has been the subject of a great deal of geographical research. Prior to the rapid introduction of nitrogen-fixing soybeans in the 1950s (Hart, 1986), corn shared the prime agricultural lands of the Midwest with oats, wheat, and hay when it was first mapped by Baker (1927). Since then, its geographical extent has shifted substantially. The most recent delineation of the Corn Belt (Figure 2) was produced from county-level Census of Agriculture data consisting of harvested acres, crop yield, and crop prices (Laingen, 2012). Laingen (2011) showed that, since 1950 when the U.S. Department of Agriculture first mapped it on a county basis, the Corn Belt has expanded to the northwest in western Minnesota, northeastern South Dakota and southeastern North Dakota, while contracting from the southwest in northeastern Kansas and northern Missouri (Laingen, 2012). These geographical changes may reflect a change in conditions suitable for corn itself, due to genetic modification, a change in the locations that provide those conditions due to climate change or soil degradation, or socioeconomic changes in technology and prices.

The Ogallala aquifer in the western portion of the study area (Figure 2) provides a unique capacity for the production of water-intensive crops, such as corn or soybeans, in areas of water deficits. In a preliminary analysis, datasets of irrigated crops in Kansas (KARS, 2006) and in Nebraska (CALMIT, 2010), were compared to a polygon of the Ogallala aquifer (USGS, 2005). Approximately 92 percent of all irrigated crops in those states fall within the outline of the aquifer. Center pivot irrigation (Lichtenberg, 1989), however, has a different set of climatic and edaphic determinants than rain-fed production. Because it is difficult to distinguish rain-fed from irrigation crops in the CDL, and in order to avoid entangling two different sets of land cover determinants, the area overlain by the Ogallala aquifer was excluded from the study. Although cotton is commonly irrigated, it is seldom found in the Ogallala region. Therefore, rain-fed and irrigated cotton are modeled jointly.

2.3 Land Cover Determinants

In developing the non-linear logits, over 200 candidate explanatory variables were tested as potential determinants of land cover, including climate, soil, topographic, economic, and land ownership characteristics (Table 1). Climate variables, which display both spatial and temporal variation across years and seasons, were tested in the models as ten-year averages of monthly, seasonal, and annual variables. Topography and soil characteristics were treated as spatially variant but temporal invariant. Temporally variant, but spatially invariant, crop prices were tested based on the specific year for the particular commodity and compiled as monthly, seasonal, or annual variables depending on the commodity to reflect the timing of farmers' decisions. Protected lands and urban development were considered to be temporally static.

2.3.1 Climate Data

Climatic variables obtained from the Global Summary of the Day-National Climatic Data Center (NCDC, 2012) included precipitation, temperature, solar radiation, and relative humidity from 2001 through 2010 for 214 weather stations throughout the study area. The station data were interpolated to a spatial grid of one-tenth of one degree using a minimum curvature interpolation model (Sandwell, 1987), then downscaled to 56m raster format using the Kriging interpolation tool in ArcGIS ("ArcGIS," 2012). Due to equal distance spacing (one-tenth degree) and little variation in values between points, interpolation errors are minimal.

Well-established agronomic research shows that suitable climatic conditions for corn and soybean production are similar, though not identical (Huang & Khanna, 2012; Sun & Van Kooten, 2013). Temperature is often incorporated into a Growing Degree Days (GDDs) value to assess conditions for crop production. GDDs are calculated from a base temperature, such as 10^{o}C, and a maximum threshold temperature, such as 30^{o}C. This modified formula calculates GDDs by subtracting the base temperature (10^{o}C), from the mean of the daily minimum temperature and the cutoff temperature (30^{o}C), then sums these daily values to the appropriate monthly, seasonal, or annual totals.

Utilizing a growing season of April through September and a temperature range of 8 to 32 degrees to calculate GDDs, Huang (2012) found that corn and soybean yields increase steadily with increases in temperature and peak at roughly 1783 GDDs for corn and 2165 GDDs for soybeans. Sun & Van Kooten (2013) found optimal corn yields are within a range of 1200 to 1300 GDDs for a May through September growing season. While these are values where crop production peaks, studies also show a sharp decline in yield as temperatures decrease to 10 ^{o}C or increase beyond 30^{o}C, creating a non-linear parabolic function between GDDs and yield (Bonhomme, 2000; Huang & Khanna, 2012; Schlenker & Roberts, 2009; Sun & Van Kooten, 2013).

Intolerant to drought, corn and soybean yields climb as precipitation increases. This typical relationship was reported in Van Assleen & Verburg (2012) where precipitation was a positive coefficient for several types of cropland for their global model (van Asselen & Verburg, 2012). The Food and Agriculture Organization suggests corn needs roughly 500 to 800 mm of precipitation during the growing season, while requirements for soybeans (450 to 700 mm) and wheat (450 to 650 mm) are slightly lower. They also report that when a water deficit (evapotranspiration exceeds precipitation) occurs, yields decrease significantly, with corn yields diminishing at a faster rate than soybeans (FAO, 2013). A combination of climatic conditions in the form of annual precipitation minus annual potential evapotranspiration (PET), defined as water surplus, was found to be the key factor in determining the ecological boundary between grassland and forest in Minnesota (Danz, Reich, Frelich, & Niemi, 2011). Danz et al. (2011) similarly found that increased moisture availability and surface roughness increased the potential for forest over grassland.

2.3.2 Soil Characteristics

Soil quality indicators such as organic matter, texture, bulk density, and rooting depth (Gregorich, Monreal, Carter, Angers, & Ellert, 1994) are common predictors for agricultural land use and cover models (Mertens & Lambin, 2000; Müller & Zeller, 2002). The Soils Survey Geographic database (SSURGO) soils data set provides high spatial resolution data on physical and chemical characteristics as well as percent slope (Soil Survey Staff, 2011). The data were preprocessed in a polygon format that was developed at a map scale of 24,000 and then converted to the same 56m raster format as the Cropland Data Layer.

Available water capacity (AWC) is a measure of the soil's ability to retain moisture during periods of low precipitation and is highly influenced by organic matter content and soil texture. The likelihood of intensive agriculture is increased when soils containing higher levels of organic matter and finer textures are present (van Asselen & Verburg, 2012). The NRCS (2008) reports that low soil AWC can result in reduced crop production and is often utilized to predict droughts and project irrigation needs (NRCS, 2008).

All crops studied are more productive in pH neutral conditions of 6.0 to 7.5, although corn has less tolerance of acidic conditions than the other crops (NRCS, 2005). While forests are tolerant of a wide range of soil conditions, deforestation for crop production has limited forest land cover to less productive soils that are excessively or poorly drained and acidic (Alig, et al., 2003). Results from van Asselen & Verburg (2012) and Danz et al. (2011) indicate forests tend to persist in sandy soils with lower pH values (Danz, et al., 2011; van Asselen & Verburg, 2012).

As row crops, corn, and even more so soybeans and cotton, are susceptible to high rates of soil erosion (Claassen, Carriazo, Cooper, Hellerstein, & Ueda, 2011; Tilman, et al., 2002). The loss of productivity due to soil erosion largely limits their production to lands with less than 6 percent slope (Pierce, Dowdy, Larson, & Graham, 1984), whereas groundcover crops such as wheat and grasslands limit erosion on lands with steeper slopes (Larson, Pierce, & Dowdy, 1983; Veldkamp, Polman, Reinhard, & Slingerland, 2011).

2.3.3 Protected Lands and Urban Development

In the U.S., crop production is primarily conducted on private land. Although some government managed land is accessible for agricultural use (Sleeter, et al., 2012; Sohl, et al., 2012), many public lands have land use restrictions where agricultural use is extremely limited or prohibited (USGS, 2013). Polygons of protected public lands were obtained from the National Gap Analysis Program Protected Areas Data Portal and included in the models as a land ownership variable.

Following the von Thunen "Isolated State" theory, the distance to urban land can also affect agricultural land use as a spatial influence (Lambin, et al., 2006). Distance to developed lands was calculating using the CDL database and tested in the models.

2.3.4 Price Data

A daily crop price dataset from the Marketing Loan Program of the U.S. Department of Agriculture Farm Services agency (FSA) was acquired through a Freedom of Information Act request (FSA, 2011). The dataset values are the Posted County Prices that the FSA uses as the benchmark for determining the value of the commodity at the time the farmer decides to sell the commodity and repay the loan. The daily Posted County Price is calculated at the county level and is based on the nearest terminal markets (Monke, 2006). Two monthly means, the price at the time of planting and the price at the time of harvest, were tested in the models. For corn, soybeans, spring wheat, and cotton, March was used for time of planting and September for time of harvest. For winter wheat, time of planting month was September and time of harvest was June.

2.4 Sampling Process

To reduce the effect of autocorrelation, a random set of 300,000 points was created to sample the 811 million pixel study area using the random point generator in ArcGIS with a 1,000 meter minimum distance. After extracting all the necessary data from the raster datasets, the sample was reduced to roughly 200,000 (population sample) by removing unwanted data types from the analysis including (1) developed lands, which are beyond the scope of this study, (2) irrigated lands, by eliminating all samples located over the Ogallala aquifer, and (3) no-data samples, which occurred mainly through missing soils information. From this population sample, a random sample of 8000 pixels was selected to calibrate each of the land cover models; however, the ratios of occurrence to non-occurrence were very low for most of the models.

Better accuracy results from logistic regression analyses are obtained when the ratio of occurrence versus non-occurrence observations are as close to 1:1 as possible (Franklin & Miller, 2010). To increase the low ratios of the samples, the population sample was divided into occurrence and non-occurrence subsets for each of the land covers. To reflect crop rotation, the occurrence subsets include all pixels where the land cover was present in at least one of the four years of 2008 - 2011. Each of the 8000-pixel land cover sample sets were compiled by selecting one-half of the samples from the occurrence subsets and the other half from the non-occurrence subsets. Due to the lack of occurrences of cotton in the population sample set, the cotton land cover sample set was limited to just 3000 pixels, taking 1500 pixels from each of the cotton occurrence and non-occurrence subsets.

The land cover sample sets were then expanded temporally by compiling what we will refer to as ‘pixel-years,’ where there are four observations for each pixel location representing the years 2008-2011. Soil, climate, and protected lands vary spatially, but not temporally among pixel-years. Each observation was given a different value for crop price, previous year land cover, and neighborhood land covers based on pixel-year. This process produced land cover sample sets of 32,000 pixel-years, with the exception of the 12,000-pixel-years cotton sample set.

2.5 Multivariable Fractional Polynomials

An assumption of logistic regression is that the relationship between the logit and independent variables is linear (Brown, Goovaerts, Burnicki, & Li, 2002; Hosmer & Lemeshow, 2000), yet we know that land cover occurrences have a non-linear relationship with biophysical variables such as temperature, precipitation, and soil characteristics (Huang & Khanna, 2012; Sun & Van Kooten, 2013). Multivariable Fractional Polynomials (MFP) is a non-parametric fitting model originally developed for medical research (Sauerbrei & Royston, 1999) that uses exponential powers to transform variables that produce a non-parametric fit. This allows the modeler to use continuous data even when the distribution is not normal or the variable relationship is not linear (Hosmer & Lemeshow, 2000; Meier-Hirmer, Ortseifen, & Sauerbrei, 2003). Transformations can be either first-degree (one transformation) or second-degree (two transformations) depending on the response of the variable within the model. For example, the typical regression equation reads as:

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + e \tag{1}$$

In the case of MFP, these variables could be transformed by applying an exponential power to specific variables to create non-linear functions such as:

$$Y = b_0 + b_1x_1 + b_2x_2^{p1} + b_3x_3^{p1} + b_4x_3^{p2} + e \tag{2}$$

Where x_2 is transformed by a single exponential power (p1) and x_3 is transformed using two exponential powers (p1 and p2). A first-degree transformation (e.g., squared or cubed) produces a curvilinear function, while a second-degree transformation uses two transformations for a single variable producing a parabolic relationship. The best-fitting transformation is determined by the lowest difference in deviance for each variable starting with the most significant variable and continuing until all variables have been tested. This cycle is repeated until all significant variables and transformation have not changed from the previous cycle. To our knowledge, this article provides the first use of MFP in land use/cover prediction. The closest example found was a study to determine the effect of weather on maize (corn) yields in northern China (Sun & Van Kooten, 2013).

2.6 Measures of Model Performance

Two measures were used to evaluate the fit of the models: percent correct and relative operating characteristic (ROC). The advantage of these measures is that they are not model-dependent. The percent correct value is dependent on the presence/absence cutoff threshold for the probability values produced by the model (Franklin & Miller, 2010). Also referred to as receiver operating characteristic (van Asselen & Verburg, 2012) the ROC curve is a plot of the true-positive rate versus the false-positive rate for every potential threshold within the model. The metric used is the sum of the area under the curve (AUC), thus providing a single metric that is threshold-independent. It provides a measure of quality based on a 0.5 (random) to 1.0 (perfect fit) range. An ROC score above 0.7 is considered a good model while any level above 0.9 is considered a very high quality model (Franklin & Miller, 2010). An advantage to using ROC over other metrics such as Akaike's Information Criterion (AIC) is that results can be compared across different models.

2.7 Temporal and Spatial Dependency

Land covers can exhibit either positive dependency, such as the stability of forest cover on a pixel, or negative dependency, such as the tendency for one crop in a rotation to be followed by a different crop. Temporal dependency was incorporated into the logit models by using the previous year land cover as a binary predictive variable. Spatial dependency was tested by incorporating into the logit models the sum of pixels for each land cover in a nine-by-nine pixel neighborhood around each sample pixel. The neighborhood was based on the previous year land cover for each pixel-year because farm fields create very strong homogeneity among pixels. Therefore, using the current year neighborhood creates an almost perfect model because almost all sample pixels are surrounded by like pixels.

2.8 Spatial Extrapolation

The logistic regression output of each land cover model was used to generate suitability distribution maps across the 811 million pixels in the central U.S. study area using the raster calculator tool in ArcGIS ("ArcGIS," 2012). This raster regression methodology uses the model coefficients and the pixel-specific values of the predictor variables to produce high spatial resolution (56m) maps. The suitability maps were then compared to the actual distribution of each land cover for the years 2008-2011.

To make this comparison, two new maps were created for each land cover model. The probabilities of the suitability maps were converted to a binary format of two for above the threshold and zero for below the threshold. Several criteria that provide similar results, have been established to determine the threshold of a logistic regression model, where above the threshold is a positive prediction and below is a negative prediction (Freeman and Moisen 2008). For example, thresholds for the corn model were found to be 0.310 when sensitivity equals specificity, 0.232 when sensitivity and specificity are maximized, 0.238 when the ratio of observed prevalence is utilized, and 0.238 when the mean probability of the model is applied. For this project the mean of the probabilities was used to establish thresholds for all the land cover models.

Maps were also created in a binary format of two if the land cover was present in any of the four years and one for all other pixels. By adding the resulting maps in map algebra, accuracy maps are created with values representing true and false positives and true and false negatives. The same comparison was conducted for the corn and soybean temporal and spatial dependency models. The production maps were based on a single year of 2011, however, because 2010 was used as the source for the previous year land cover. Since the extra pixel-years were excluded, the accuracy of temporal and spatial maps is slightly reduced.

2.9 Price Sensitivity

The economic analysis was then applied to the corn suitability maps to identify prime vs. marginal lands by setting prices per bushel for corn from $2 to $10, which encompasses the full range of empirical prices in the 21st Century, and mapping the resulting changes in corn occurrence.

3.0 RESULTS

3.1 Model performance

Non-linear logit models for each of the seven land covers performed in the ROC range 0.749 - 0.874, with highest ROC values produced by the forest and cotton models and lowest by the soybean and grassland models (Table 2). Models achieved percent correct ranges from 65 - 79 with forest and winter wheat the highest and soybeans and corn the lowest. The true positive rates are lower, ranging from 37 percent for soybeans to 75 percent for forest (Figure 3A). Due to their complementarity in crop rotations, there were high levels of false positives where the logit models confuse corn and soybeans (Figure 3D). For this reason, an eighth model was developed where a pixel was coded with an occurrence if either corn or soybeans was present. This corn-soy model achieved a high ROC of 0.871 (Figure 3C) with 78 percent correct pixel prediction (Figure 3A).

3.2 Geographic Determinants

Table 2 lists the coefficients and Figure 4 captures the non-linear statistical relationships that define the logistic regression models. Significant variables include the climate variables of summer (June-August) GDDs and growing season (April-October) water surplus, while significant soil variables include Available Water Capacity and pH. Percent slope was significant in all the models. The spatial distribution of each significant variable (Figure 5) indicates the influence each one had on land cover suitability distributions. By comparing these distributions to the response curves of Figure 4, one can begin to see specific relationships between the biophysical characteristics of the study area and the distribution of each land cover.

All of the biophysical variables were transformed using MFP to improve the model response to each variable (Figure 4). These graphs utilize the "effects plot" statement from SAS, which graphs the function of each variable by holding all other variables at their mean. These response curves demonstrate how MFP can produce applicable non-linear results that are consistent with agronomic research (Huang & Khanna, 2012; Sun & Van Kooten, 2013) in depicting the curvilinear relationship of croplands to climatic conditions.

3.2.1Climatic Determinants

Summer growing degree days affect each land cover in a unique non-linear fashion (Figure 4). Corn and soybeans share a common peak at 1108 GDDs but, consistent with Sun and Van

Kooten (2013), corn has a much narrower range, with probabilities exceeding 0.05 with GDDs only in the 867-1405 range, compared to 783-1627 GDDs for soybeans. Spring wheat peaks at 889 GDDs compared to cotton at 1429, with a very narrow range. Winter wheat and grassland mirror one another, increasing in probability at a diminishing rate as GDDs climb. The inclusion of CDL data from Texas, available only since 2008, may have been able to define a thermal maximum for winter wheat. The curve for forest displays the GDD ranges under which it is economically dominated by crops, rather than its own physiological response, reaching a minimum probability at 1148 GDDs, very close to the corn and soybean maxima. Canadian Shield forests peak at 700 GDDs, while southern plateau forests peak at 2000 GDDs, the lowest and highest observed in the study area, respectively.

Consistent with FAO (2013), corn and soybeans have nearly identical responses to growing season water surplus, peaking at 285 mm and 281 mm, respectively, and exceeding 0.05 from 0-560 mm and 30-530 mm, respectively. Spring and winter wheat mirror one another, peaking at -19 mm and -13 mm, respectively, though this is partly accounted for by the economic dominance of corn and soybeans where more ideal water surpluses occur. Probability of forest increases with water surplus across the entire range observed, but grassland displays a bi-modal distribution, with a minimum at about 313 mm – near the corn and soybean peak. Negative water surpluses (water deficits) up to 600 mm are characteristic of the Great Plains grasslands, while water surpluses exceeding 600 mm characterize steeper and less fertile lands in the eastern and southern parts of the study area.

3.2.2 Edaphic and Topographic Determinants

Consistent with NRCS (2005), all crops studied, as well as grassland, display a preference for neutral to slightly alkaline soil, relegating forest to acidic soils. The preference of corn production over other crops under prime soil conditions is evident with an available water capacity peak of 0.33 versus soybeans' peak at 0.21. Spring and winter wheat are restricted to less fertile soils with AWC peaks of 0.17 and 0.10, respectively, while grassland probability declines with AWC, reflecting the economic dominance of crops on higher quality soils. AWC was an insignificant predictor for cotton and forest.

All crops studied display a similar negative exponential response to percent slope. Grasslands are most likely to occur on moderate slopes, peaking at 12 percent, while the probability of forest increases throughout the range of slopes observed. These results reflect the economic superiority of crops over pasture/rangeland and pasture/rangeland over forest on flatter lands capable of supporting them.

3.2.3 Land Ownership and Crop Price Determinants

Forests are more likely and crops less likely on publically protected lands (Table 2). The high variance in crop prices over the years studied (2008-2011) offers an opportunity to quantify their effect on land cover. Winter wheat, cotton, and corn probabilities increased linearly with prices for those respective crops. Corn price had a negative effect on spring wheat.

3.3 Temporal and Spatial Dependency

Corn and soybeans patterns were examined for temporal effects, such as those shown by Houet et al. (2010) and Sohl et al. (2010), by including knowledge of the status of pixels in the previous year. Previous year in corn has a slightly positive, and previous year in soybeans has a large positive effect on the probability of corn occurring in a pixel (Table 3). For soybeans, previous year in corn had the larger positive effect. For the combined corn-soybean model, either crop in the previous year has a large positive effect. Also indicating a strong temporal dependency borne of crop rotation, model performances improved substantially. For corn, ROC increased from 0.789 to 0.857 and percent of pixels correctly predicted increased from 69 to 76. For soybeans ROC increased even more from 0.749 to 0.848 and percent correct from 65 to 79. For combined corn-soybeans, ROC increased from 0.871 to a very high 0.944 and percent correct from 78 to 90. These results imply that, if climatic, edaphic and topographic determinants are known, next year's Cropland Data Layer is quite predictable from this year's CDL, with price scenarios adding additional predictability.

When the neighboring pixels in a 9x9 window, near the average field size of the study area, are added, however, the additional model performance improvement is very modest: ROC improves only 0.016, 0.019, and 0.014 in the corn, soybean and corn-soybean models, respectively. Correct prediction falls in the corn and the soybean models (Table 3). The only statistically significant neighboring land uses are grassland and forest, both with negative coefficients. This implies that, due to the need to maintain minimal field sizes, even if other factors are favorable, a lone pixel is unlikely to be used for corn or soybeans if the surrounding pixels are not in crop production.

Analysis with and without past and neighboring land uses shows that the process of agricultural land use determination in the U.S. Midwest has strong temporal dependencies via crop rotation and forest stability, but weak to non-existent contagious spatial diffusion, which stands in contrast to many land use and cover change studies (Berger, 2001; Parker, Manson, Janssen, Hoffmann, & Deadman, 2003), often employing agent-based modeling that utilizes spatial diffusion as a central feature.

3.4 Spatial Extrapolation

Calculating the nonlinear logistic regression output of the suitability models (Table 2) on a 56m pixel basis shows the probabilities of a specific cover occurring in each of the 811 million pixels in the study area and indicates the distribution of the suitability for each land cover (left-hand map in each panel of Figure 6). The accuracy assessment of the suitability maps versus actual 2008-2011 land cover distribution provides a statistical output to determine the viability of reproducing land cover distributions using logistic regression on a raster dataset (Figure 7).

Although the area overlying the Ogallala aquifer was not included in the modeling of the crop models this area was included in the extrapolation process to see how well the models could predict cropland overlying the aquifer. Two sets of statistical results are provided (Figure 7) and show that including the Ogallala area in the extrapolation had very little influence on the overall accuracy of the suitability maps.

The overall percent correct of the suitability distributions compared to actual land cover distributions is above 79 percent for all models except grassland (Figure 7A). The spatial distribution of true positive (green) and negative (gray), as well as false positive (red) and negative (orange) pixels is shown on the right-hand map in each panel of Figure 6. As with the modeling results (Figure 3A), these values are heavily driven by high true negative values. The relative consistency between the percent correct values of the models (Figure 3A) and the raster regression results (Figure 7A) is an indication that the extrapolation of the models across the entire study area was successful. The one exception is the cotton model, which exhibits very low true positive results. This is likely due to the very low occurrence rate within the study area as well as economic choices of farmers to produce a more profitable crop in areas that are suitable for cotton production.

The temporal and spatial dependency model extrapolations (Figure 7B) are even more consistent with CDL patterns than the suitability maps alone (Figure 3B). The ability of these models to predict land covers at the pixel level can be seen in the distribution maps (Figure 8) and in the highly detailed map of Cass County on the expanding Corn Belt fringe in southeast North Dakota (Figure 9A) that shows the fine detail of the suitability maps. Polygons in the SSURGO soil dataset are the limiting factor in spatial resolution (left-hand map in Figure 9A), although probabilities are relatively generalized across the county. The added value of previous year land cover data, however, vastly improves the model's ability to discretely define each pixel (center map Figure 9A), thereby creating the more realistic looking landscape depicted by the detailed map on the far right of Figure 9A.

3.5 Marginal Lands

To spatially identify marginal croplands, suitability maps were produced for corn at two-dollar intervals from $2 to $10/bushel (Figure 9B). Using a threshold of 24 percent, pixels were selected that transitioned into positive-predicted status as corn prices increase, reflecting the coefficient on corn price in the logit model. The maps show pixels that were positive predictions at $2/bushel in dark gray delineating prime cropland, and those that were negative predictions even at $10/bushel in light gray delineating lands unsuitable for corn production. Marginal lands have negative predictions at $2/bushel, but transition into positive predictions at $4 (light green), $6 (dark green), $8 (orange), and $10/bushel (red). Due to climatic factors, marginal lands occur more commonly along the Corn Belt fringe, but also occur in its interior where slopes are high or AWC or pH is low.

Figure 10 shows the relationship between the predicted area in corn and actual corn production for the 16-state study area and Iowa. Utilizing the GIS and remote sensing theory of soft classification of pixels, where a pixel is a percentage of two or more classifications rather than hard classified with a single value (Pontius & Cheuk, 2006; Silván-Cárdenas & Wang, 2008; Thenkabailc, et al., 2007), predicted area in corn is approximated by multiplying pixel area (3,136 m^2) by the probability of each pixel and summing the values for each respective area. This predicted area methodology over-estimates total corn production on average by 17.5 percent for the study area and by 10 percent for the state of Iowa. Figure 10 shows that area in corn can be quite accurately predicted, compared to a linear supply curve, for any price that occurred during the 2007-2011 time frame. The similarity in slopes of these lines shows the ability of the models to replicate the influence of increasing corn prices on overall corn production.

4.0 DISCUSSION AND CONCLUSIONS

In relation to the research questions posed in section 1.0, the nonlinear logit models succeed in predicting land cover suitability distributions and facilitate quite accurate high spatial resolution mapping. The suitability distributions provided by the extrapolations are consistent with actual land cover distributions (Figure 6) with all except the grassland model predicting at least 79 percent of pixels correctly (Figure 7A). For less common land covers like cotton, the geographic range of suitability is predicted well, even if the actual true positive rate is low (6 percent). Incorporating previous-year land cover information improved the performance of corn and soybean models, indicating the importance of temporal dependency. As confirmed by other studies using Markov Chains (Brown, Pijanowski, & Duh, 2000; Logofet & Lesnaya, 2000; Muller & Middleton, 1994; Ramasubramanian & Jain, 1999), this also results in improvements in predictability at the pixel level, as seen in highly detailed county-level maps (Figure 9A). The inclusion of crop prices in the models allows accurate, high spatial resolution mapping of changes in corn production along the supply curve (Figure 9B). Strong spatial contagion effects, as distinct from clustering of suitability factors like climate and soil fertility, are not evident.

Multivariable Fractional Polynomial transformations proved invaluable in matching non-linear land cover responses to determinants, from the exponential, rather than linear, decrease in crop suitability as a function of percent slope, to the parabolic relationships with climatic variables such as growing degree days and water surplus that define the suitability range similarly to the habitat of a species. In species distribution modeling, the concept of suitability is not only to determine the existence of a particular species at a particular location, but also to locate any suitable habitats available (Franklin & Miller, 2010). Here we incorporated this concept into a highly anthropogenic landscape, with the understanding that these landscapes are greatly influenced by social, political, and economic factors, especially prices and land protection status.

With the acceleration of biofuel production, expanded production of corn and other crops has been a focus of study (Hellerstein & Malcolm, 2011; McPhail & Babcock, 2008; Mehaffey, et al., 2011; Searchinger, et al., 2008; Sylvester, Brown, Deane, & Kornak, 2013; Wright & Wimberly, 2013). The ability to predict where these land use changes may occur facilitates environmental and economic impact assessment. Finally, because climatic factors are important determinants of rural land covers, the nonlinear logit and pixel extrapolation framework utilized here could be interfaced with spatially-specific climate change scenario data to build projections of the geographic adaptation of crop production and land cover to climate change.

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant No. 1009925. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation (NSF).

References

Alig, R. J., Plantinga, A. J., Ahn, S. E., & Kline, J. D. (2003). *Land use changes involving forestry in the United States: 1952 to 1997, with projections to 2050*. Citeseer.

. ArcGIS. In. (2012) (10.0 ed.). ESRI, Redlands, CA USA.

Baker, O. E. (1927). Agricultural Regions of North America. Part IV--The Corn Belt. *Economic Geography, 3* (4), 447-465.

Berger, T. (2001). Agent-based spatial models applied to agriculture: A simulation tool for technology diffusion, resource use changes and policy analysis. *Agricultural Economics, 25* (2-3), 245-260.

Bjorklund, J., Limburg, K., & Rydberg, T. (1999). Impact of production intensity on the ability of the agricultural landscape to generate ecosystem services: An example from Sweden. *Ecological Economics, 29*, 269-261.

Bonhomme, R. (2000). Bases and limits to using 'degree.day' units. *European Journal of Agronomy, 13* (1), 1-10.

Brown, D. G., Goovaerts, P., Burnicki, A., & Li, M.-Y. (2002). Stochastic simulation of land-cover change using geostatistics and generalized additive models. *Photogrammetric Engineering and Remote Sensing, 68* (10), 1051-1061.

Brown, D. G., Pijanowski, B. C., & Duh, J. D. (2000). Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. *Journal of Environmental Management, 59* (4), 17p.

Brown, D. G., Walker, R., Manson, S., & Seto, K. (2004). Modeling land use and land cover change. In G. Gutman, A. Janetos, C. Justice, E. Moran, J. Mustard, R. Rindfuss, D. Skole & B. L. I. Turner (Eds.), *Land Change Science* (pp. 395-409). Kluwer Academic Publishers.

CALMIT, Center for Advanced Land Management Information Technologies,. (2010). 2005 Land Use Mapping. In. University of Nebraska Lincoln.

Claassen, R. L., Carriazo, F., Cooper, J. C., Hellerstein, D., & Ueda, K. (2011). Grassland to cropland conversion in the Northern Plains: The role of crop insurance, commodity, and disaster programs. *Economic research report no. 120*.

Danz, N. P., Reich, P. B., Frelich, L. E., & Niemi, G. J. (2011). Vegetation controls vary across space and spatial scale in a historic grassland-forest biome boundary. *Ecography, 34* (3), 402-414.

FAO, Food and Agriculture Orginization,. (2013). Crop Water Information: Soybean. In W. D. a. M. Unit (Ed.), (Vol. 2013).

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., Mueller, N. D., O'Connell, C., Ray, D. K., & West, P. C. (2011). Solutions for a cultivated planet. *Nature, 478* (7369), 337-342.

Franklin, J., & Miller, J. A. (2010). *Mapping apecies distributions: Spatial inference and prediction*. Cambridge University Press.

FSA, Farm Service Agency,. (2011). Archived LDP & PCP Rates. In. USDA,, United States Department of Agriculture.

Geoghegan, J., Villar, S. C., Klepeis, P., Mendoza, P. M., Ogneva-Himmelberger, Y., Chowdhury, R. R., Turner Ii, B. L., & Vance, C. (2001). Modeling tropical deforestation in the southern Yucatán peninsular region: Comparing survey and satellite data. *Agriculture, Ecosystems & Environment, 85* (1–3), 25-46.

Gregorich, E. G., Monreal, C. M., Carter, M. R., Angers, D. A., & Ellert, B. H. (1994). Towards a minimum data set to assess soil organic matter quality in agricultural soils. *Canadian Journal of Soil Science, 74* (4), 367-385.

Guisan, A., Edwards Jr, T. C., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling, 157* (2–3), 89-100.

Han, W., Yang, Z., Di, L., & Mueller, R. (2012). CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. *Computers and Electronics in Agriculture, 84*, 111-123.

Hart, J. F. (1986). Change in the Corn Belt. *Geographical Review*, 51-72.

Hastie, T., & Tibshirani, R. (1987). Generalized Additive Models: Some Applications. *Journal of the American Statistical Association, 82* (398), 371-386.

Hauser, R. (2007). Corn-based ethanol in Illinois and the U.S.: A report from the Department of Agricultural and Consumer Economics, University of Illinois. In.

Hellerstein, D., & Malcolm, S. (2011). The Influence of rising commodity prices on the Conservation Reserve Program. In. Economic Research Service.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression* (Second ed.).

Huang, H., & Khanna, M. (2012). Determinants of US corn and soybean yields: Impact of climate change and crop prices. *Available at SSRN*.

KARS, Kansas Applied Remote Sensing,. (2006). 2005 Kansas Land Cover Patterns - Level IV | KARS Research. In (2005 ed.). Kansas University.

Laingen, C. (2012). Delineating the 2007 Corn Belt Region. *Papers of the Applied Geography Conferences* (35), 174-182.

Lambin, E. F., Geist, H., & Rindfuss, R. R. (2006). *Land-use and land-cover change: Local processes and global impacts*. Springer Berlin Heidelberg.

Larson, W. E., Pierce, F. J., & Dowdy, R. H. (1983). The threat of soil erosion to long-term crop production. *Science, 219* (4584), 458-465.

Lichtenberg, E. (1989). Land quality, irrigation development, and cropping patterns in the northern High-Plains. *American Journal of Agricultural Economics, 71* (1), 187-194.

Logofet, D. O., & Lesnaya, E. V. (2000). The mathematics of Markov models: what Markov chains can really predict in forest successions. *Ecological Modelling, 126* (2), 285-298.

McPhail, L. L., & Babcock, B. A. (2008). Ethanol, mandates, and drought: Insights from a stochastic equilibrium model of the U.S. corn market. In *CARD Working Paper No. 08-WP-464*. Iowa State University: Center for Agricultural and Rural Development.

Mehaffey, M., Smith, E., & Van Remortel, R. (2011). Midwest U.S. landscape change to 2020 driven by biofuel mandates. *Ecological Applications, 22* (1), 8-19.

Meier-Hirmer, C., Ortseifen, C., & Sauerbrei, W. (2003). Multivariable Fractional Polynomials in SAS. In. University of Freiburg, Germany, Freiburg, Germany.

Mertens, B., & Lambin, E. F. (2000). Land cover change trajectories in Southern Cameroon. *Annals of the Association of American Geographers, 90* (3), 467-494.

Monke, J. (2006). Farm commodity programs: Direct Payments, Counter-Cyclical Payments, and Marketing Loans. In.

Müller, D., & Zeller, M. (2002). Land use dynamics in the central highlands of Vietnam: A spatial model combining village survey data with satellite imagery interpretation. *Agricultural Economics, 27* (3), 333-354.

Muller, M. R., & Middleton, J. (1994). A Markov model of land-use change dynamics in the Niagara Region, Ontario, Canada. *Landscape Ecology, 9* (2), 151-157.

NASS, National Agricultural Statistical Service,. (2013). CropScape - NASS CDL Program. In. USDA,, United States Department of Agriculture.

National Research Council. (2013). *Advancing Land Change Modeling: Opportunities and Research Requirements*. The National Academies Press, Washington, DC.

NCDC, National Climatic Data Center. (2012). Global Summary of the Day. In.

NRCS, Natural Resources Conservation Service. (2005). Soil Quality Indicators: Soil pH. In *USDA*.

NRCS, Natural Resources Conservation Service. (2008). Soil Quality Indicators: Available Water Capacity. In *USDA*.

Osteen, C., Gottlieb, J., & Vasavada, U. (2012). Agricultural resources and environmental indicators 2012. In U. S. D. o. A. USDA (Ed.).

Parker, D. C., Manson, S. M., Janssen, M. A., Hoffmann, M. J., & Deadman, P. (2003). Multi-agent systems for the simulation of land-use and land-cover change: A review. *Annals of the Association of American Geographers, 93* (2), 314-337.

Pierce, F. J., Dowdy, R. H., Larson, W. E., & Graham, W. A. P. (1984). Soil productivity in the Corn Belt: An assessment of erosion's long-term effects. *Journal of Soil and Water Conservation, 39* (2), 131-136.

Pontius, R. G., & Cheuk, M. L. (2006). A generalized cross-tabulation matrix to compare soft-classified maps at multiple resolutions. *International Journal of Geographical Information Science, 20* (1), 1-30.

Ramasubramanian, V., & Jain, R. (1999). Use of growth indices in Markov Chain models for crop yield forecasting. *Biometrical journal, 41* (1), 99-109.

Rosegrant, M. W. (2008). *Biofuels and grain prices: impacts and policy responses*. International Food Policy Research Institute.

Sandwell, D. T. (1987). Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data. *Geophys. Res. Lett., 14* (2), 139-142.

Sauerbrei, W., & Royston, P. (1999). Building multivariable prognostic and diagnostic models: Transformation of the predictors by using fractional polynomials. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 162* (1), 71-94.

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences, 106* (37), 15594-15598.

Searchinger, T., Heimlich, R., Houghton, R. A., Dong, F., Elobeid, A., Fabiosa, J., Tokgoz, S., Hayes, D., & Yu, T.-H. (2008). Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science, 319* (5867), 1238-1240.

Secchi, S., Tyndall, J., Schulte, L. A., & Asbjornsen, H. (2008). High crop prices and conservation. *Journal of Soil and Water Conservation, 63* (3), 68A-73A.

Serneels, S., & Lambin, E. F. (2001). Proximate causes of land-use change in Narok District, Kenya: A spatial statistical model. *Agriculture, Ecosystems & Environment, 85* (1-3), 65-81.

Silván-Cárdenas, J., & Wang, L. (2008). Sub-pixel confusion–uncertainty matrix for assessing soft classifications. *Remote Sensing of Environment, 112* (3), 1081-1095.

Sleeter, B. M., Sohl, T. L., Bouchard, M. A., Reker, R. R., Soulard, C. E., Acevedo, W., Griffith, G. E., Sleeter, R. R., Auch, R. F., Sayler, K. L., Prisley, S., & Zhu, Z. (2012). Scenarios of land use and land cover change in the conterminous United States: Utilizing the special report on emission scenarios at ecoregional scales. *Global Environmental Change, 22* (4), 896-914.

Sohl, T. L., Sleeter, B. M., Sayler, K. L., Bouchard, M. A., Reker, R. R., Bennett, S. L., Sleeter, R. R., Kanengieter, R. L., & Zhu, Z. (2012). Spatially explicit land-use and land-cover scenarios for the Great Plains of the United States. *Agriculture, Ecosystems & Environment, 153*, 1-15.

Soil Survey Staff. (2011). Natural Resources Conservation Service, United States Department of Agriculcture, Soil Survey Geographic (SSURGO) Database for the Central U.S. In.

Sun, B. J., & Van Kooten, G. C. (2013). Weather effects on maize yields in northern China. *The Journal of Agricultural Science, FirstView*, 1-11.

Sylvester, K. M., Brown, D. G., Deane, G. D., & Kornak, R. N. (2013). Land transitions in the American plains: Multilevel modeling of drivers of grassland conversion (1956–2006). *Agriculture, Ecosystems & Environment, 168*, 7-15.

Temme, A., & Verburg, P. H. (2011). Mapping and modelling of changes in agricultural intensity in Europe. *Agriculture, Ecosystems & Environment, 140* (1), 46-56.

Thenkabailc, P. S., Biradar, C. M., Noojipady, P., Cai, X., Dheeravath, V., Li, Y., Velpuri, M., Gumma, M., & Pandey, S. (2007). Sub-pixel area calculation methods for estimating irrigated areas. *Sensors, 7* (11), 2519-2538.

Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., & Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature, 418* (6898), 671.

USDA, United States Department of Agriculture, Natural Resources Conservation Service. (2006). Land resource regions and major land resource areas of the United States. *Handbook 296*.

USGS, United States Geological Survey. (2005). Principal aquifers of the 48 conterminous United States, Hawaii, Puerto Rico, and the U.S. Virgin Islands. In. United States Geological Survey.

USGS, United States Geological Survey. (2013). National Gap Analysis Program (GAP) Protected Areas Data Portal. In.

Valbuena, D., Verburg, P. H., Bregt, A., & Ligtenberg, A. (2010). An agent-based approach to model land-use change at a regional scale. *Landscape Ecology, 25* (2), 185-199.

van Asselen, S., & Verburg, P. H. (2012). A land system representation for global assessments and land-use modeling. *Global Change Biology, 18* (10), 3125-3148.

Veldkamp, A. T., Polman, N. B. P., Reinhard, S., & Slingerland, M. A. (2011). From scaling to governance of the land system: Bridging ecological and economic perspectives. *Ecology and Society, 16* (1), 1.

Verburg, P. H. (2006). Simulating feedbacks in land use and land cover change models. *Landscape Ecology, 21* (8), 1171-1183.

Verburg, P. H., Schot, P. P., Dijst, M. J., & Veldkamp, A. T. (2004). Land use change modelling: Current practice and research priorities. *GeoJournal, 61* (4), 309-324.

Wright, C. K., & Wimberly, M. C. (2013). Recent land use change in the western Corn Belt threatens grasslands and wetlands. *Proceedings of the National Academy of Sciences, 110* (10), 4134-4139.

Table 1. Over 200 candidate explanatory variables were tested for each land cover model. Climate variables were tested as monthly, seasonal, and annual means (temperature and humidity) and sums (growing degree days, potential evapotranspiration, precipitation, and water surplus) based on a ten-year mean (2001-2010). This large dataset was reduced to just seven significant variables.

Model Variables	Tested	Significant	N/A

Category	**Variable**	**Monthly**	**Seasonal**	**Annual**	**Non-Temporal**
Climate	Temperature (Celsius)				
	Growing Degree Days		July-August		
	Potential Evapotranspiration (mm)				
	Humidity (percent)				
	Precipitation (mm)				
	Water Surplus (mm)		April – October		
Soils	pH				
	Cation Exchange Capacity (meq/100g)				
	Available Water Capacity (cm/cm)				
	Organic Matter Content (percent)				
Topography	Percent Slope				
Economics	Crop Prices (per bushel)	Crop Specific			
	Oil Prices (per barrel)				
	Fertilizer Prices (per ton)				
Protected Lands		Based on 2011 Data			
Distance to Urban (Km)		Based on 2011 Data			

Table 2. Land cover models: coefficients, ROC scores, and percent correct values. All variables included in the models are significant at <0.05. Those denoted as N/A were not included in the models.

Parameter	MFP Trans-forma-tions	Corn	Soy-bean	Winter Wheat	Spring Wheat	Cotton	Forest	Grass
Intercept	NA	-128.9	-177.5	-108.2	336.1	-83.917	8.0887	4.69
Summer Growing Degree Days	Log	N/A	N/A	N/A	-44.410	N/A	N/A	N/A
	^2	-7.4E-6	-2.3E-6	N/A	N/A	N/A	N/A	N/A
	^-2	-1.1E+7	-3.9E+6	-2.3E+7	-1.8E+7	N/A	1.5E+8	-2.1E+7
	^-2*Log	NA	NA	3.08E+6	NA	N/A	-2.3E+7	3.0E+6
	^3	N/A	N/A	N/A	N/A	5.6E-7	N/A	N/A
	^3*Log	N/A	N/A	N/A	N/A	-7.4E-8	N/A	N/A
Growing Season Water Surplus/ Deficit	Linear	N/A	N/A	-0.0148	N/A	N/A	3.9E-3	N/A
	Log	20.9283	26.6843	NA	NA	N/A	N/A	N/A
	^-2	NA	NA	-7.1E+6	-7.4E+6	N/A	N/A	N/A
	^3	-3.3E-9	-4.3E-9	NA	-5.4E-9	N/A	N/A	-6.6E-8
	^3*Log	N/A	N/A	N/A	N/A	N/A	N/A	8.7E-9
Soil pH	Linear	N/A	N/A	N/A	N/A	N/A	-0.903	N/A
	^0.5	N/A	N/A	96.2471	N/A	N/A	NA	N/A
	^0.5*Log	N/A	N/A	-23.7959	N/A	N/A	NA	N/A
	^-2	-446.8	-225.1	N/A	N/A	-447.4	NA	-164.9
	^-2*Log	317.5	126.3	N/A	N/A	345.1	NA	106.2
Percent Slope	Linear	-0.0754	-0.1017	-0.0917	N/A	N/A	0.0744	-0.0658
	Log	N/A	N/A	NA	N/A	N/A	N/A	0.8278
	^-0.5	N/A	N/A	NA	N/A	15.1935	N/A	N/A
	^-1	N/A	N/A	NA	N/A	-8.4956	N/A	N/A
	^2	N/A	N/A	NA	-2.6E-3	NA	N/A	N/A
Available Water Capacity	Linear	N/A	N/A	-7.2026	N/A	N/A	N/A	-6.3632
	^2	-29.701	-70.5589	N/A	N/A	N/A	N/A	N/A
	^2*Log	-35.636	-64.2056	N/A	N/A	N/A	N/A	N/A
	^-2	N/A	N/A	-0.0041	N/A	N/A	N/A	N/A
	^3	N/A	N/A	N/A	-830	N/A	N/A	N/A
	^3*Log	N/A	N/A	N/A	-573.5	N/A	N/A	N/A
Protected Lands	Binomial	-0.729	-1.055	-1.3651	N/A	-1.9352	1.5238	-0.547
Sept. Corn Price	Linear	0.0334	N/A	N/A	-0.0455	N/A	N/A	N/A
June Wheat Price	Linear	N/A	N/A	0.0571	N/A	N/A	N/A	N/A
Sept. Cotton Price	Linear	N/A	N/A	N/A	N/A	0.0168	N/A	N/A
ROC	NA	0.789	0.749	0.814	0.831	0.842	0.874	0.752
Percent Correct	NA	69%	65%	74%	70%	71%	79%	72%

Table 3. Spatial and temporal dependency models: coefficients, ROC scores, and percent correct values. All variables included in the models are significant at <0.05. Those denoted as N/A were not significant and not included in those models.

Models		Suitability Models			Temporal Dependency Models			Spatial Dependency Models		
Parameter	**MFP Transformation**	**Corn**	**Soy**	**Corn-Soy**	**Corn**	**Soy**	**Corn-Soy**	**Corn**	**Soy**	**Corn-Soy**
Intercept		-128.9	-177.5	-245.3	-79.96	-130.2	-143.8	-63.7	-134.7	-118.0
Summer Growing Degree Days	^2	-7.4E-6	-2.3E-6	-7.3E-6	-6.9E-6	7.0E-7	-4.6E-6	-6.5E-6	2.5E-7	-5.6E-6
	^-2	-1.1E+7	-3.9E+6	-1.1E+7	-1.0E+7	6.5E+5	-7.2E+6	-9.5E+6	N/A	-8.9E+6
Growing Season Water Surplus	^3	-3.3E-9	-4.3E-9	-5.9E-9	-2.1E-9	-3.1E-9	-3.4E-9	-1.7E-9	-3.2E-9	-2.7E-9
	Log	20.9238	26.6843	38.52	13.335	18.71	22.56	10.875	19.61	19.25
Soil pH	^-2	-446.8	-225.1	-225.7	-502.4	-45.79	-41.84	-420.9	-14.52	N/A
	^-2*Log	317.5	126.3	120	370.4	N/A	N/A	330.8	N/A	N/A
Percent Slope	Linear	-0.0754	-0.1017	-0.133	-0.055	-0.085	-0.084	-0.017	-0.045	-0.035
Available Water Capacity	^2	-29.701	-70.558	-102.3	15.51	-68.87	-78.50	7.792	-49.16	-51.49
	^2*Log	-35.636	-64.205	-100.9	N/A	-55.83	-73.69	N/A	-35.16	-45.64
Protected Lands	Binomial	-0.729	-1.055	-1.000	-0.703	-0.914	-0.666	-0.447	-0.667	N/A
Corn Price	Linear	0.0334	N/A	0.036	0.042	N/A	N/A	0.039	N/A	N/A
Previous Year Corn	Binomial	N/A	N/A	N/A	0.256	2.746	3.694	-0.233	2.049	2.870
Previous Year Soybeans	Binomial	N/A	N/A	N/A	2.295	0.574	3.182	1.783	-0.136	2.332
Forest Neighborhood	Linear	N/A	N/A	N/A	N/A	N/A	N/A	-0.038	-0.045	-0.058
Grass Neighborhood	Linear	N/A	N/A	N/A	N/A	N/A	N/A	-0.020	-0.018	-0.030
ROC		0.789	0.749	0.871	0.857	0.848	0.944	0.873	0.867	0.958
Percent Correct		69%	65%	78%	76%	79%	90%	75%	77%	90%

Figure captions.

Figure 1. Theoretical relationship of land covers to physical land characteristics. Corn-soybean rotations occur on the most suitable lands followed by wheat, grassland and forest. Increases in corn prices expand corn production onto less suitable lands driving out soybean rotations and wheat, thus increasing their prices and expanding production onto less suitable lands at the expense of grassland.

Figure 2. Land cover distributions within the 16-state study area, outline of the Ogallala aquifer (USGS, 2005), and the county-level delineation of the Corn Belt (Laingen, 2012).

Figure 3. Predictive accuracy of (A) eight land cover suitability models and (B) temporal and spatial dependency models for corn, soybeans and corn-soy. (C) ROC curves and (D) categorization of false positives for land cover suitability models.

Figure 4. Nonlinear responses derived using MFP of eight land covers studied to determinants: growing degree days, water surplus, soil pH, available water capacity, and slope.

Figure 5. Spatial distribution of determinants of land cover.

Figure 6. Left-hand panels: Maps derived through raster regression of suitability for corn, soybean, corn-soy, cotton, winter wheat, spring wheat, forest and grassland. Right-hand panels: Accuracy assessment maps of true and false positives, true and false negatives. Each of these maps contains 811 million 56m pixels. Details can be examined in supplementary materials at http://info.erp.siu.edu/tim/.

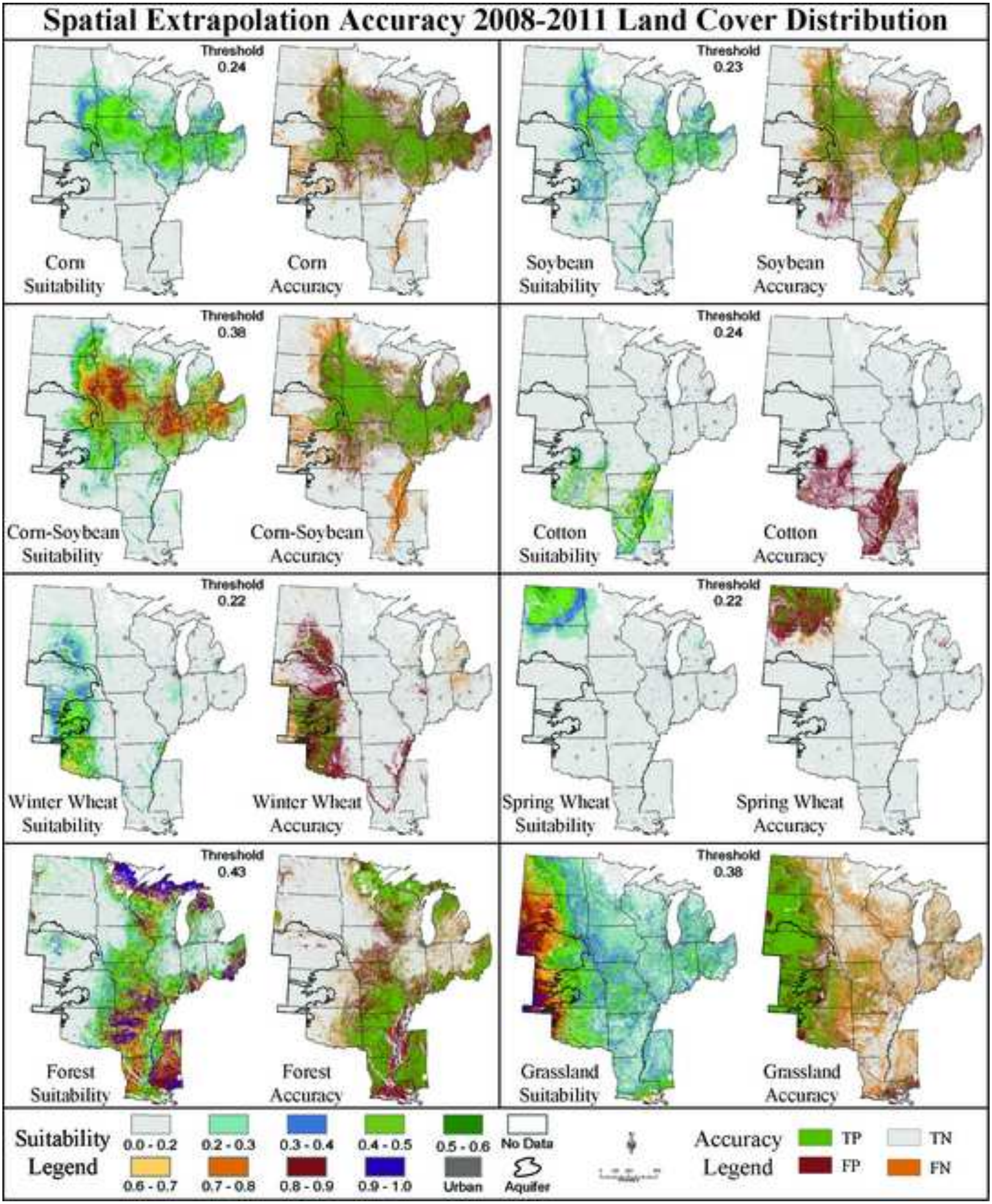
Figure 7. Predictive accuracy of (A) land suitability models with Ogallala aquifer, (B) temporal and spatial dependency models with Ogallala aquifer, (C) land suitability models without Ogallala aquifer, and (D) temporal and spatial dependency models without Ogallala aquifer.

Figure 8. Left-hand panels: Maps derived through raster regression of temporal and spatial dependency models corn, soybean, and corn-soy. Right-hand panels: Accuracy assessment maps of true and false positives, true and false negatives. Each of these maps contains 811 million 56m pixels. Details can be examined in supplementary materials at http://info.erp.siu.edu/tim/.

Figure 9. (A) Detailed examination of pixel predictions shows how the temporal models coalesce pixels into contiguous fields with the same land cover. (B) The effect of increasing prices on the spatial pattern of the probability of corn production in six counties on the margin of the Corn Belt.

Figure 10. (A) Empirical vs. modeled effects of prices on area planted to corn in the 16-state study area and Iowa.

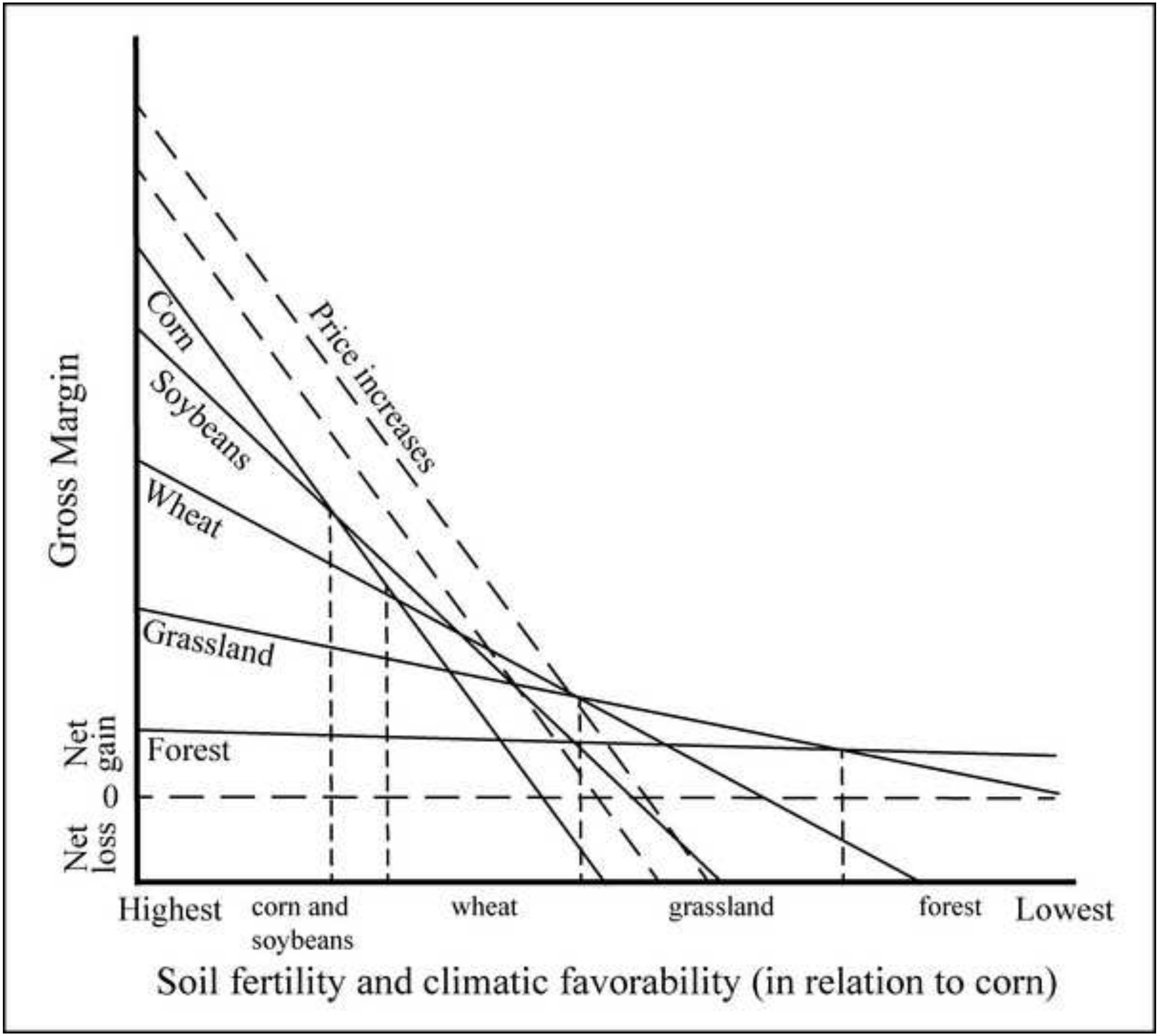
Graphical Abstract (for review)



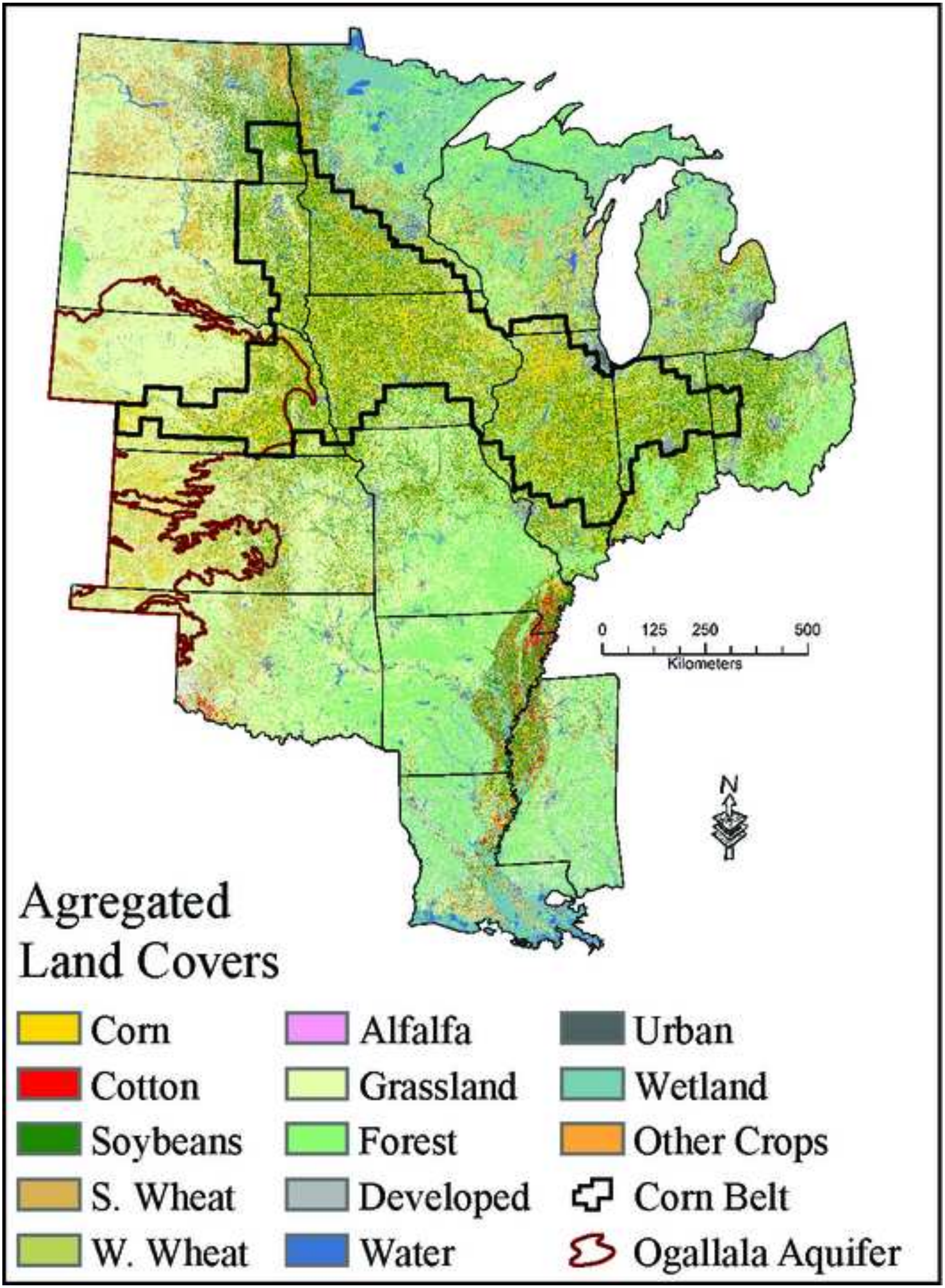
Highlights:

- Non-linear MFP logit models for seven rural land covers are specified
- High ROC levels range from 0.769 for soybeans to 0.888 for forest
- Suitability maps are generated for the central U.S. for 811 pixels at 56m scale
- Temporal dependence is high; spatial dependence is low
- Marginal lands for corn production are spatially identified

StoebnerFigure1
Click here to download high resolution image

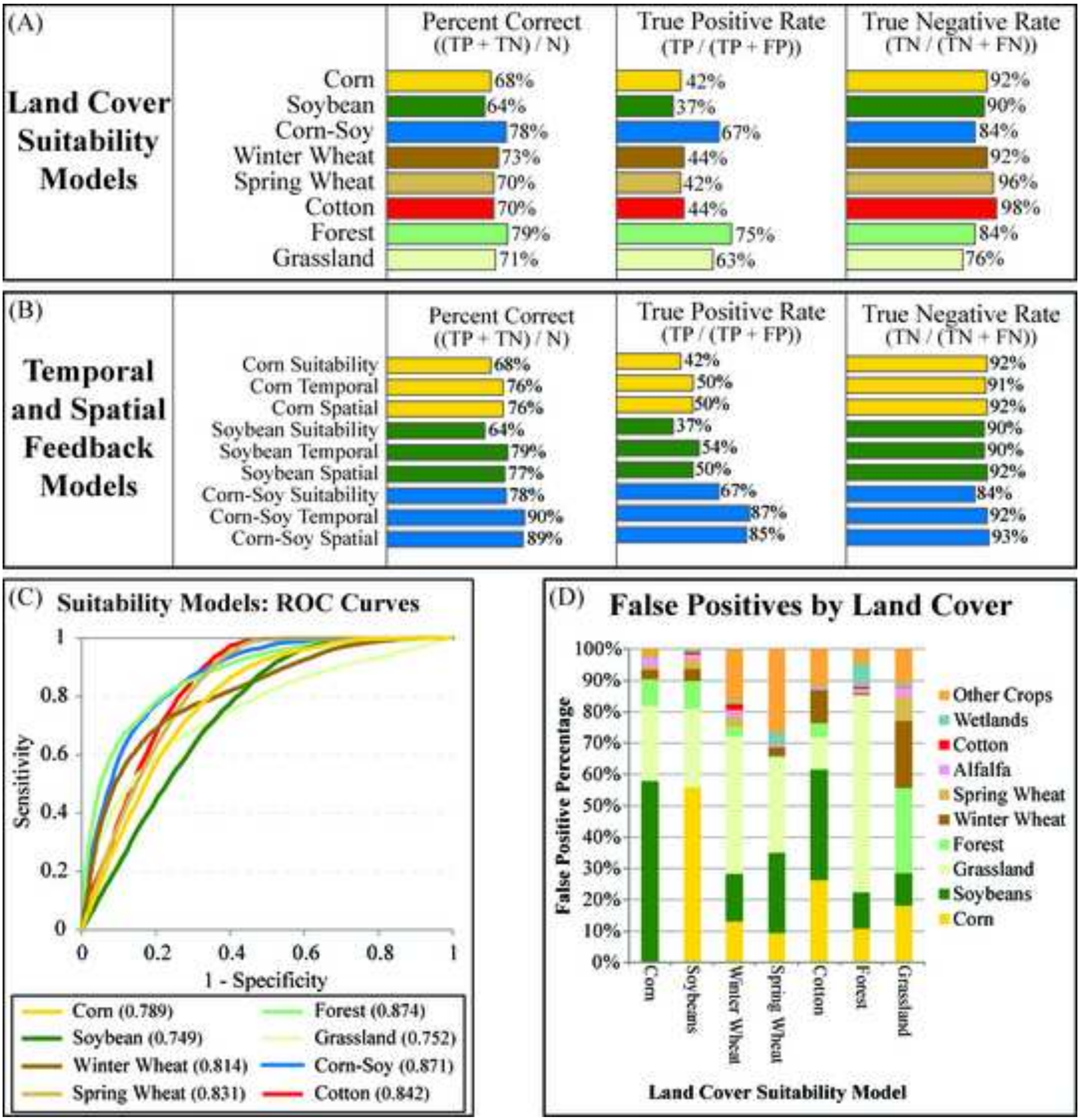


StoebnerFigure2
Click here to download high resolution image



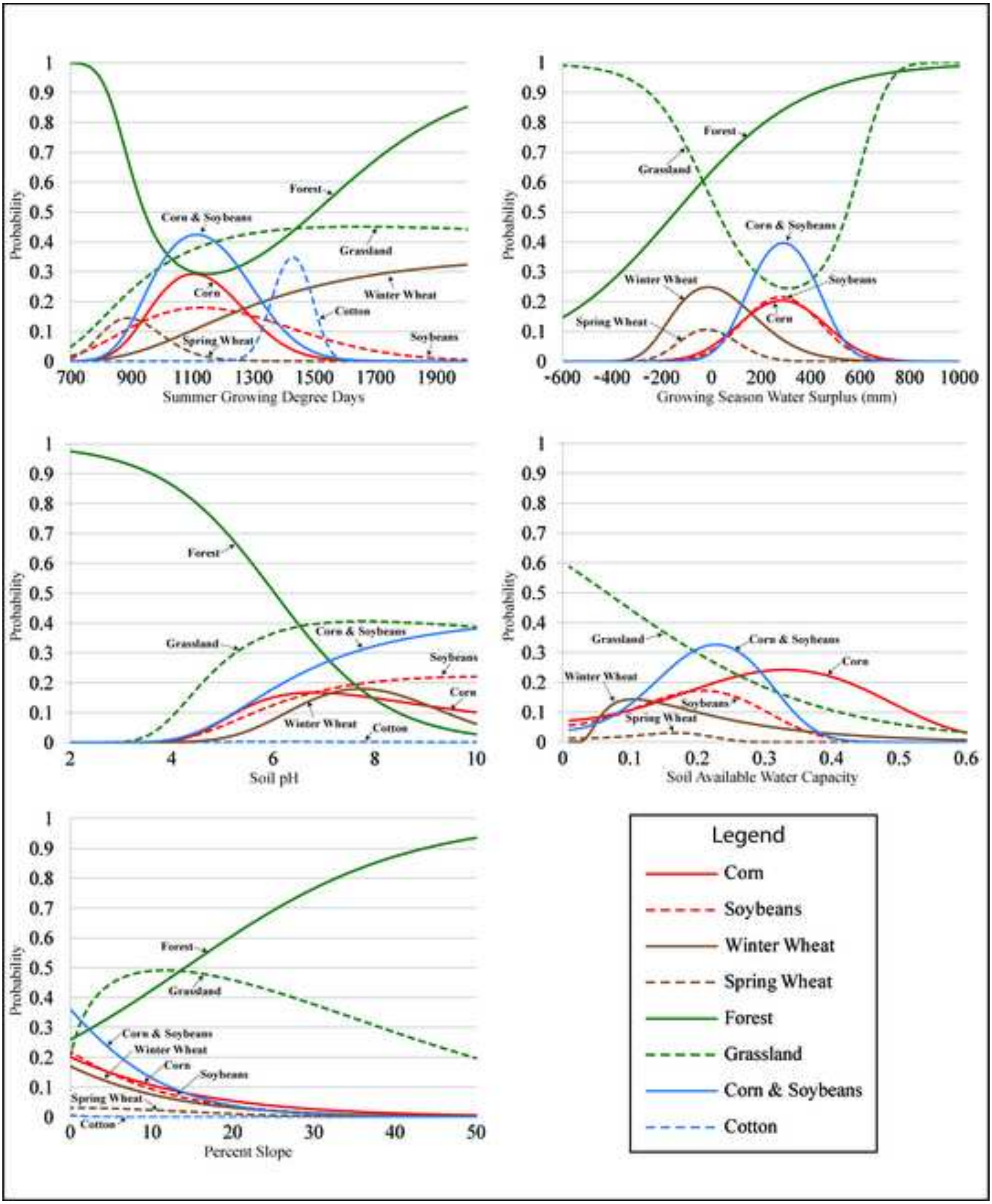
StoebnerFigure3

Click here to download high resolution image

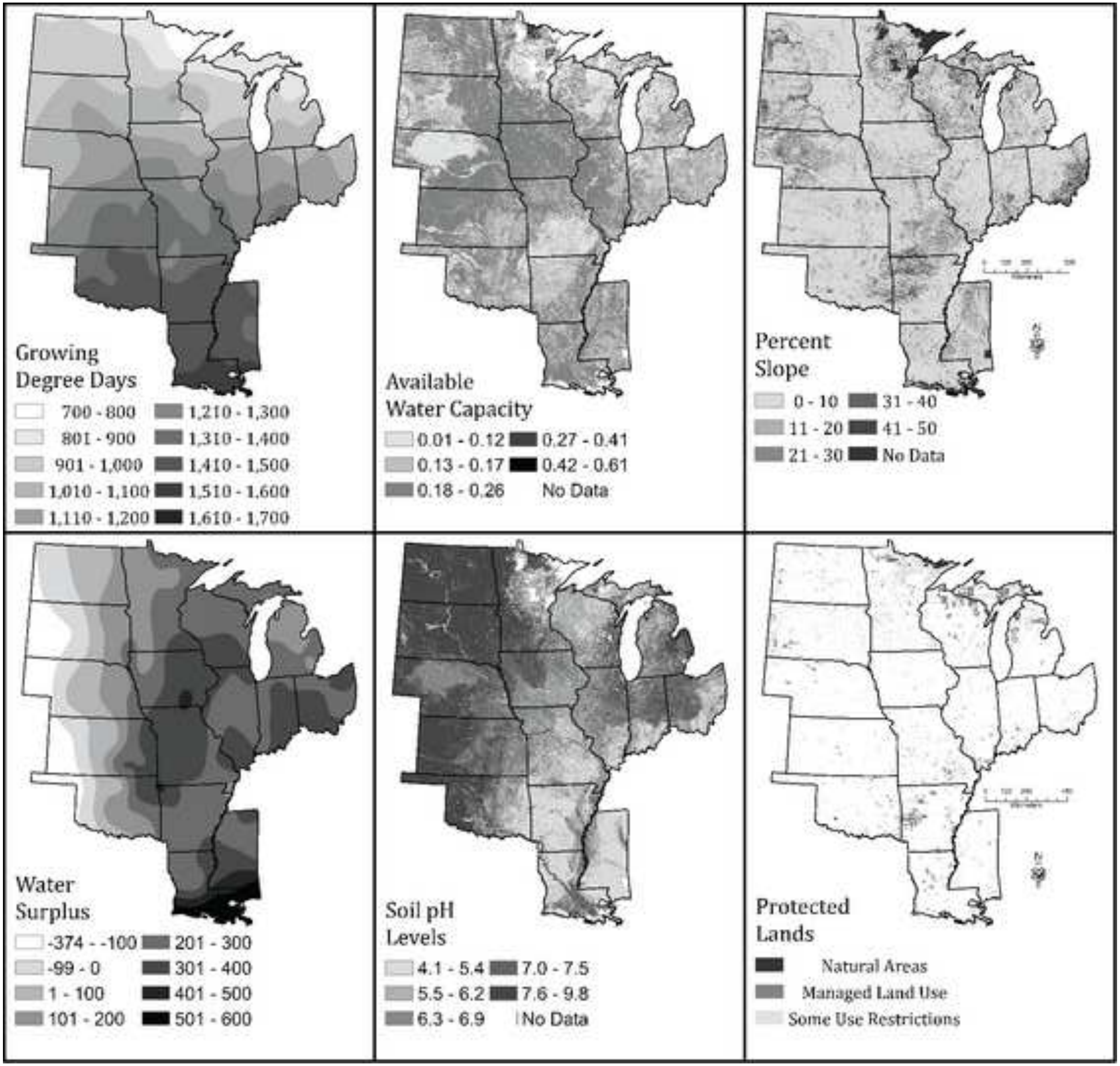


StoebnerFigure4

Click here to download high resolution image

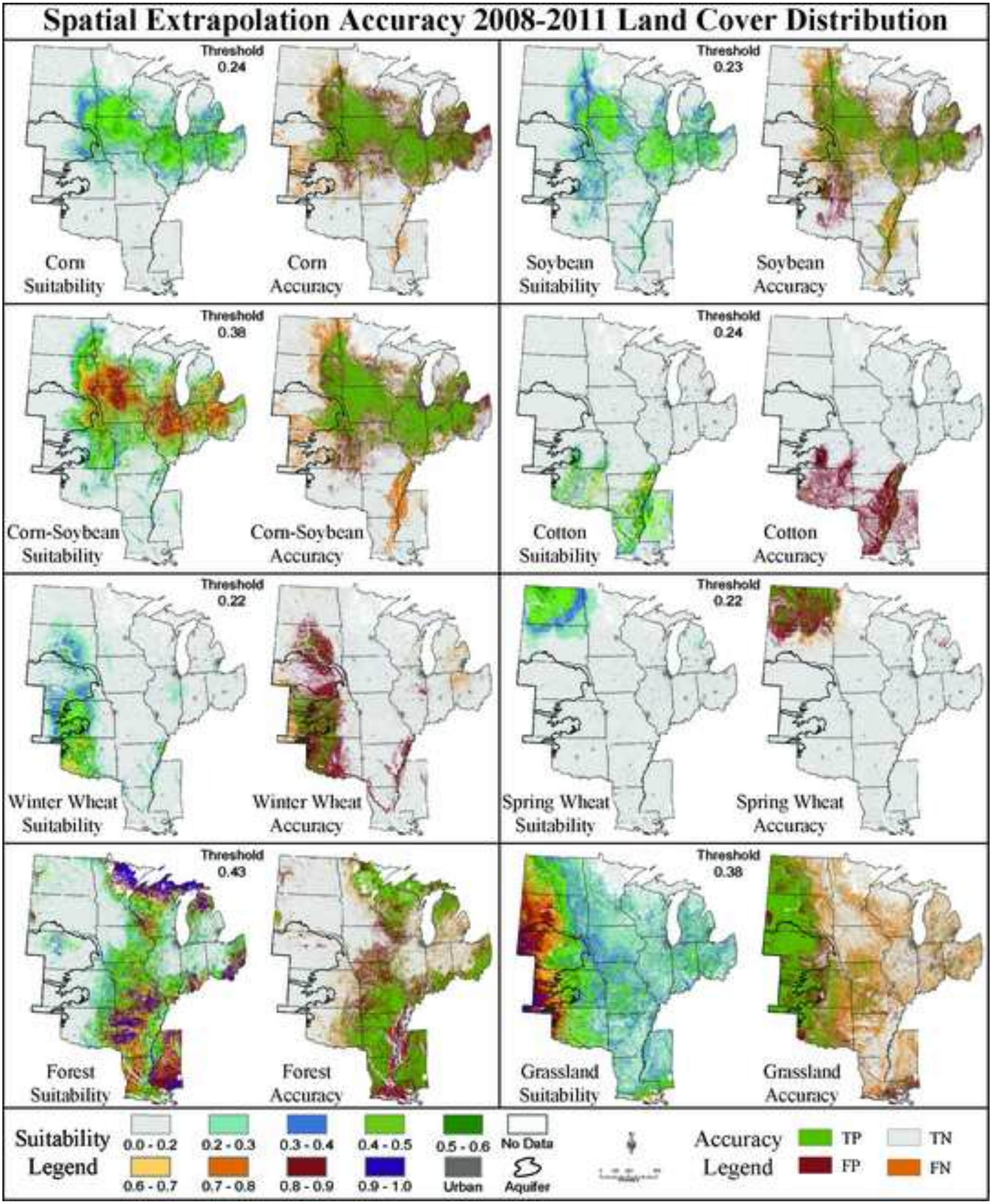


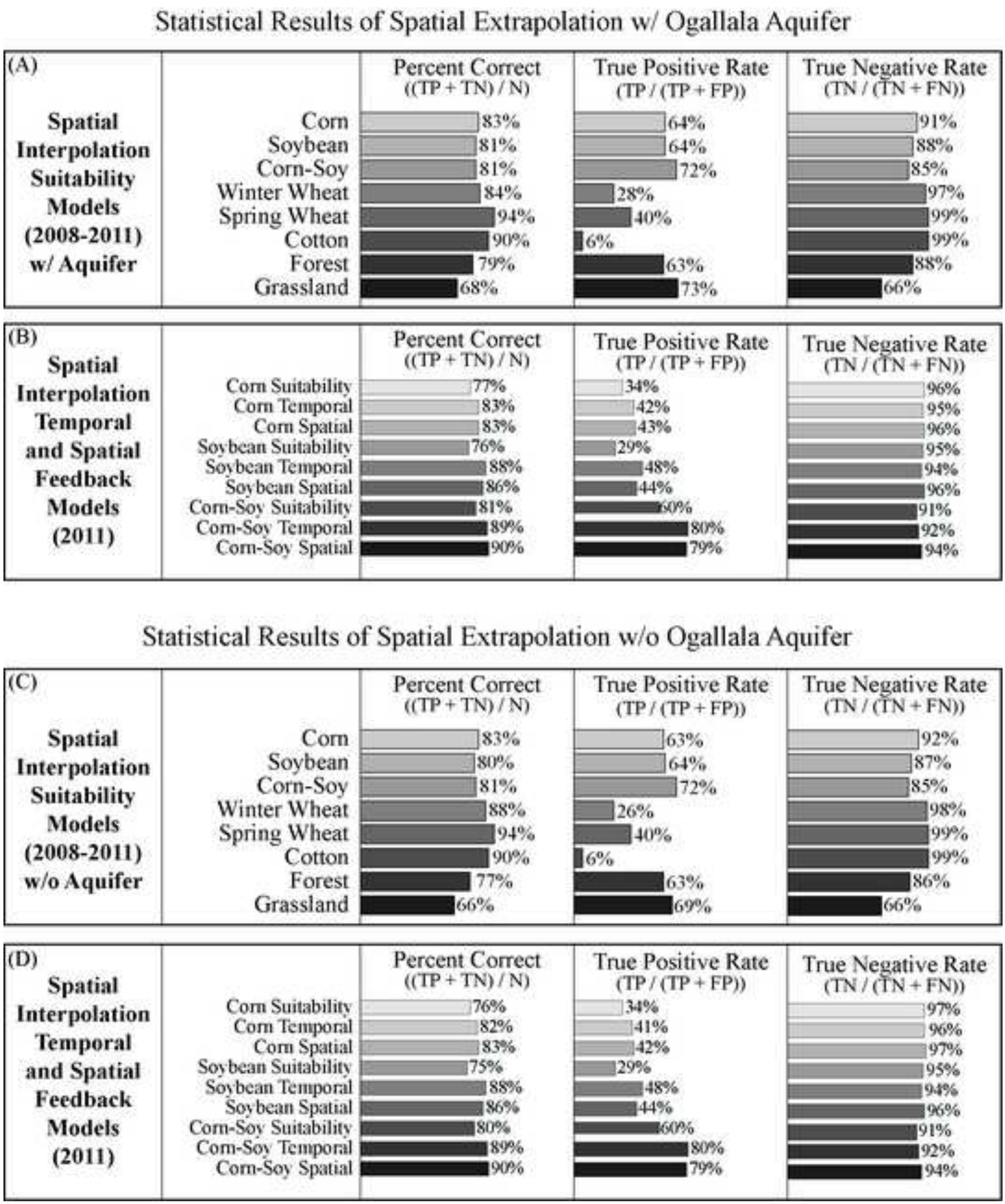
StoebnerFigure5
Click here to download high resolution image



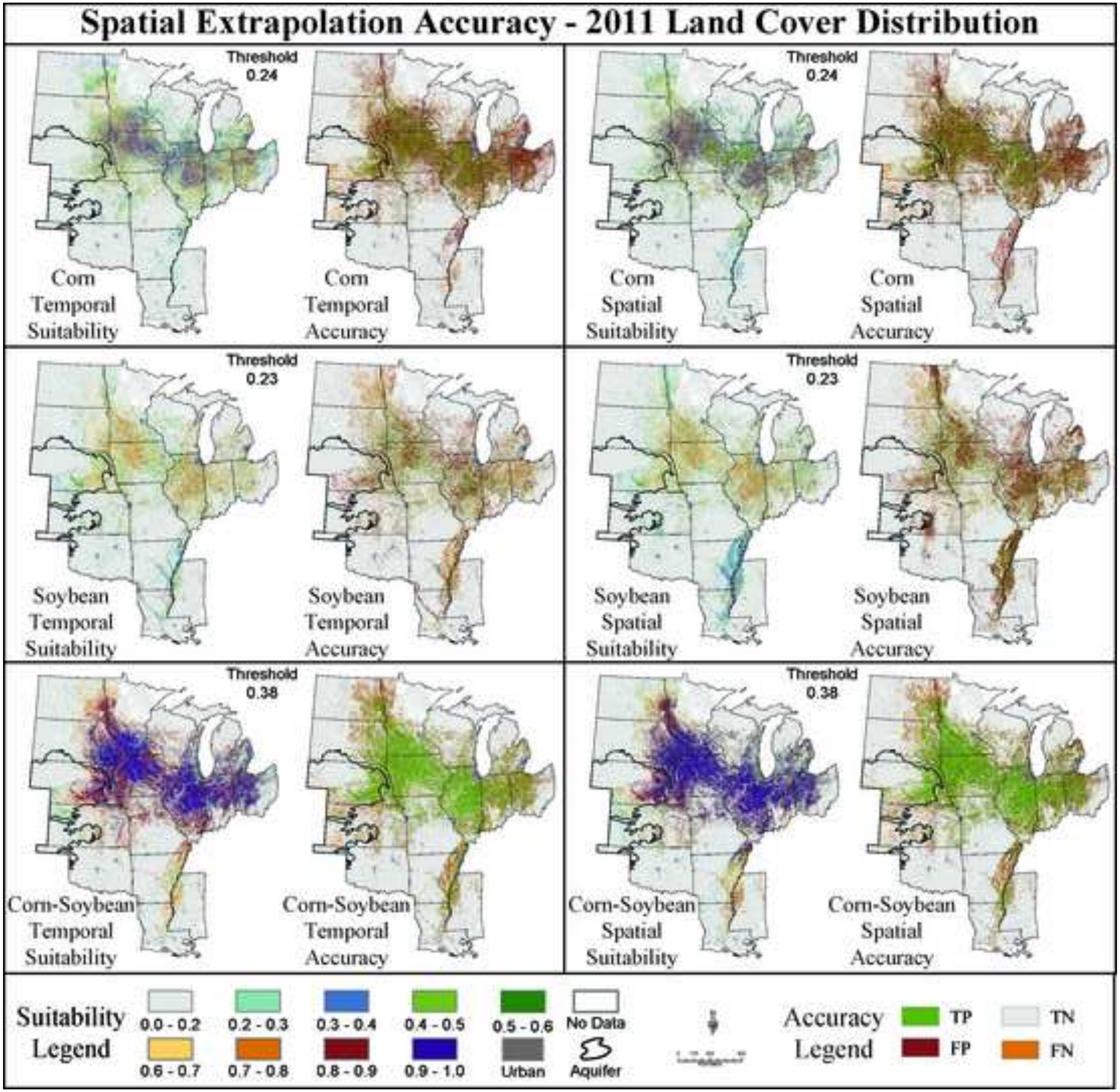
StoebnerFigure6

Click here to download high resolution image



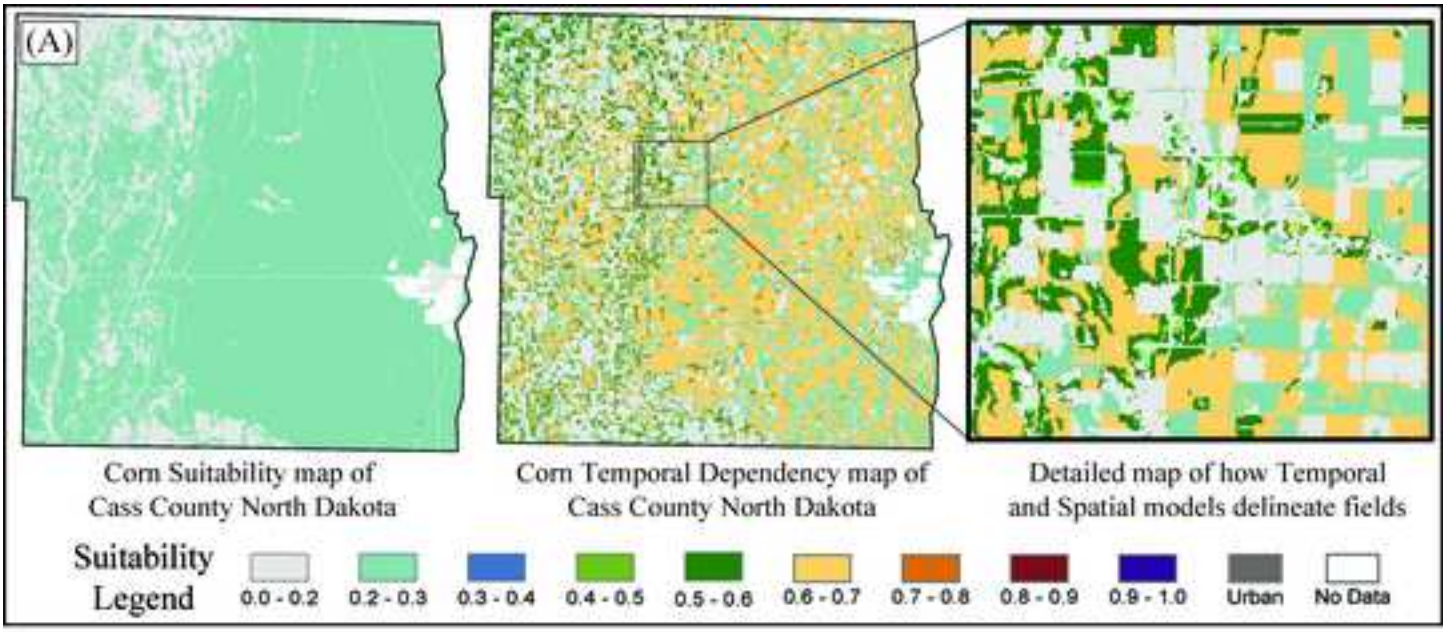

Statistical Results of Spatial Extrapolation w/ Ogallala Aquifer
(A)
Spatial Interpolation Suitability Models (2008-2011) w/ Aquifer
Percent Correct ((TP + TN) / N)
True Positive Rate (TP / (TP + FP))
True Negative Rate (TN / (TN + FN))
Corn 83% 64% 91%
Soybean 81% 64% 88%
Corn-Soy 81% 72% 85%
Winter Wheat 84% 28% 97%
Spring Wheat 94% 40% 99%
Cotton 90% 6% 99%
Forest 79% 63% 88%
Grassland 68% 73% 66%
(B)
Spatial Interpolation Temporal and Spatial Feedback Models (2011)
Percent Correct ((TP + TN) / N)
True Positive Rate (TP / (TP + FP))
True Negative Rate (TN / (TN + FN))
Corn Suitability 77% 34% 96%
Corn Temporal 83% 42% 95%
Corn Spatial 83% 43% 96%
Soybean Suitability 76% 29% 95%
Soybean Temporal 88% 48% 94%
Soybean Spatial 86% 44% 96%
Corn-Soy Suitability 81% 60% 91%
Corn-Soy Temporal 89% 80% 92%
Corn-Soy Spatial 90% 79% 94%
Statistical Results of Spatial Extrapolation w/o Ogallala Aquifer
(C)
Spatial Interpolation Suitability Models (2008-2011) w/o Aquifer
Percent Correct ((TP + TN) / N)
True Positive Rate (TP / (TP + FP))
True Negative Rate (TN / (TN + FN))
Corn 83% 63% 92%
Soybean 80% 64% 87%
Corn-Soy 81% 72% 85%
Winter Wheat 88% 26% 98%
Spring Wheat 94% 40% 99%
Cotton 90% 6% 99%
Forest 77% 63% 86%
Grassland 66% 69% 66%
(D)
Spatial Interpolation Temporal and Spatial Feedback Models (2011)
Percent Correct ((TP + TN) / N)
True Positive Rate (TP / (TP + FP))
True Negative Rate (TN / (TN + FN))
Corn Suitability 76% 34% 97%
Corn Temporal 82% 41% 96%
Corn Spatial 83% 42% 97%
Soybean Suitability 75% 29% 95%
Soybean Temporal 88% 48% 94%
Soybean Spatial 86% 44% 96%
Corn-Soy Suitability 80% 60% 91%
Corn-Soy Temporal 89% 80% 92%
Corn-Soy Spatial 90% 79% 94%

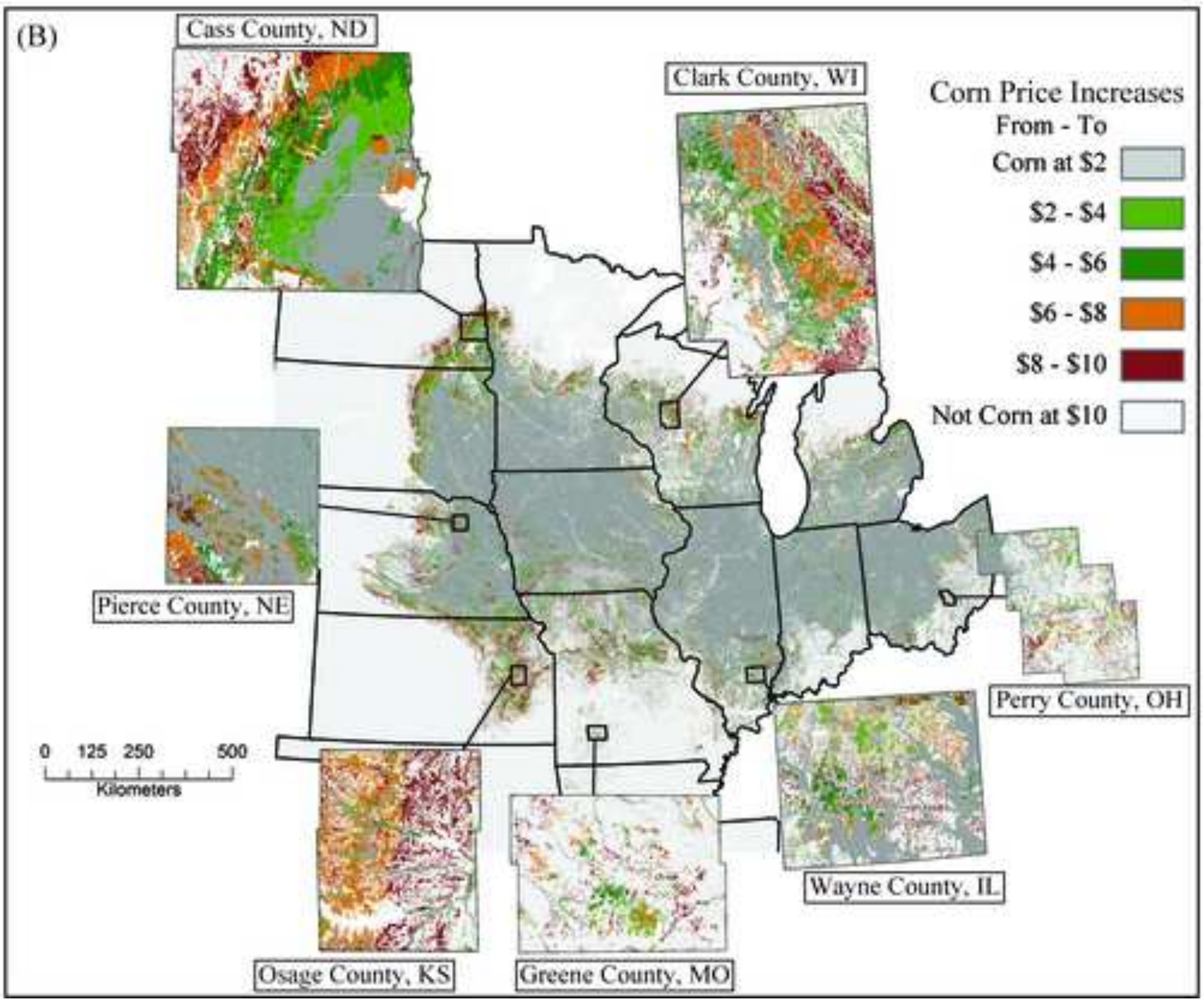
StoebnerFigure8
Click here to download high resolution image



StoebnerFigure9

Click here to download high resolution image





StoebnerFigure10
Click here to download high resolution image