Estimating the Spatially Varying Responses of Corn Yields to Weather Variations using Geographically Weighted Panel Regression

Ruohong Cai, Danlin Yu, and Michael Oppenheimer

Researchers have extensively studied crop yield response to weather variations, while only a limited number of studies have attempted to identify spatial heterogeneity in this relationship. We explore spatial heterogeneity in corn yield response to weather by combining geographically weighted regression and panel regression. We find that temperature tends to have negative effects on U.S. corn yields in warmer regions and positive effects in cooler regions, with spatial heterogeneity at a fine scale. The spatial pattern of precipitation effects is more complicated. A further analysis shows that precipitation effects are sensitive to the existence of irrigation systems.

Key words: climate change, corn yields, geographically weighted panel regression, spatial heterogeneity

Introduction

Weather variations have significant impacts on crop yields, but only a limited number of studies have explored the possible contributions to this relationship of geographical variation arising from different regional characteristics such as soil type, irrigation systems, and so on. Failure to recognize heterogeneous responses of crop yield to weather across the regions could lead to misguided policies aimed at local climate adaptation. Furthermore, while temperature is projected to increase in most U.S. regions, the expected magnitude of warming varies considerably across the regions. For instance, Midwest and Corn Belt states are projected to have a larger temperature increase (one to three degrees Fahrenheit more) by the end of this century than other parts of the country (Coulson et al., 2010; U.S. Global Change Research Program, 2009). Therefore, the issue of heterogeneous climate impacts could grow in importance over time.

The relationship between weather and crop yields presented in most existing literature is essentially a global estimate, as the relationship applies invariantly over space (Lobell and Burke, 2010). Such estimates may be informative for climate mitigation and adaptation planning at large spatial scales but may be misleading for localized programs, particularly those aimed at farmer adaptation. A study of regional differences in how crop yields respond to weather is expected to provide policymakers with more useful information about local climate impacts.

Some evidence of spatial varying relationship between weather and crop yields is found in previous literature. Based on the results of crop response models, Tobey, Reilly, and Kane (1992) demonstrated that climate change effects on crop yields may differ for different geographic zones. Tao et al. (2006) found that temperature was negatively correlated with crop yields at all stations except one in northeastern China. Schlenker and Roberts (2009) found that the relationship between temperature and corn yields varies for different geographic regions in the United States. Using a crop simulation model, Butterworth et al. (2010) found that climate change increases the productivity of oilseed rape in the United Kingdom, with greater benefits in Scotland than in England. The above studies often estimate separate models using data from predefined geographic regions. Although these studies admit the existence of inherent spatial heterogeneity in the relationships, their definitions of spatial heterogeneity are fairly arbitrary. This study uses a new method—Geographically Weighted Panel Regression (GWPR, Yu, 2010)—to further explore the hypothesis that weather variations have different impacts on crop yields in different regions.[1] Using county-level data from the continental United States, the GWPR approach attempts to address the inherent spatial heterogeneity of these relationships. Instead of using data from various predefined regions, the approach lets the data tell the story.

Process-based crop simulation models have been used to simulate how weather affects crop growth (Jones et al., 2003). These models usually require large amounts of information, making it difficult to apply them to the analysis of data on a large spatial scale. Alternatively, researchers use statistical models (Dixon et al., 1994; Lobell and Burke, 2010; Elbakidze, Lu, and Eigenbrode, 2011)—mostly based on time-series data (Lobell et al., 2008) and panel data (Schlenker and Roberts, 2009)—to quantify the relationship between climate and crop yields. Compared to time-series or cross-sectional data, panel data provide more information and enable control for time-invariant unobserved heterogeneity. Panel data analysis can also help improve estimation efficiency (Wooldridge, 2002). Using a fixed effects model and U.S. county-level panel data, Deschênes and Greenstone (2007) found that growing degree days have negative effects on corn and soybean yields. McCarl, Villavicencio, and Wu (2008) detected positive effects of temperature on the U.S. state-level soybean yields by estimating a fixed effects model with time trend. Using county-level data from 1950 to 2005, Schlenker and Roberts (2009) estimated a fixed effects model with state-specific quadratic time trend and showed nonlinear temperature effects on corn, soybeans, and cotton yields.

Both cross-sectional and panel data models assume that the relationship between variables is spatially invariant, ignoring possible spatial heterogeneity. Even if we are only interested in the global mean of the relationship, the conventional regression estimation of spatial data might still produce misleading results since regression residuals from spatial data are usually spatially autocorrelated, violating the statistical assumption of independently distributed errors. Spatial regression models such as the spatial lag model and the spatial error model have been developed to account for spatial covariance, but they still assume spatially constant coefficients (Anselin, 2001; Mueller and Loomis, 2008; Elhorst, 2010).

Geographically weighted regression (GWR) is an exploratory local spatial approach that uses each data point and its neighboring observations to construct a local model and then estimates local regression coefficients, which are allowed to vary across the space for different local models (Brunsdon, Fotheringham, and Charlton, 1996; Fotheringham, Brunsdon, and Charlton, 2002). Thus, spatial heterogeneity could be explored explicitly. Cho, Bowker, and Park (2006) used GWR to generate local estimates for the effects of proximity to water bodies and parks on housing price. Also using the GWR model, Partridge and Rickman (2007) found that local job growth in particular reduced poverty in persistent poverty counties. In areas relevant to our study, Olgun and Erdogan's (2009) GWR model demonstrated that the relationship between various climatic factors and wheat potential exhibited considerable spatial variability. Sharma (2011) studied the relationship between crop yields and precipitation for ninety-three counties in Nebraska and found that the GWR model

---

[1] We use the term "weather" rather than "climate" because we use relatively short-term averages. See the Data section for more details.

provides a better understanding of the spatially varying effects than the ordinary least squares (OLS) model.

For the conventional GWR model, local models are based on cross-sectional data. Studies that use GWR to analyze local panel data remain rare. As the first attempt, Yu (2010) combined GWR and a panel model to study the regional development of the Greater Beijing area. We follow Yu (2010) in implementing a panel data version of GWR (GWPR) to explore the spatially varying relationship between weather and crop yields for counties in the continental United States. We expect that GWPR will be able to take advantage of panel data and generate more thorough results compared to the cross-sectional GWR model.

Methodology

To introduce the GWPR model, we start with a fixed effects panel model with a time trend:

$$(1) \qquad Y_{(u,v)t} = \alpha_{(u,v)} + \beta X_{(u,v)t} + \gamma t + \varepsilon_{(u,v)t},$$

where subscript (u,v) denotes the latitude-longitude geographic coordinates and subscript t indicates year; $Y_{(u,v)t}$ denotes corn yields for location (u,v) at year t; $X_{(u,v)t}$ denotes weather conditions for location (u,v) at year t; β is a coefficient that stays constant across the space; $\alpha_{(u,v)}$ denotes time-invariant fixed effects such as local soil type; t is a linear time trend that removes the effects of technological improvements over time; and $\varepsilon_{(u,v)t}$ is the error term.

The GWPR alternative would be:

$$(2) \qquad Y_{(u,v)t} = \alpha_{(u,v)} + \beta_{(u,v)} X_{(u,v)} + \gamma_{(u,v)} t + \varepsilon_{(u,v)t}.$$

Equation (2) differs significantly from equation (1) in that coefficient β is not assumed to stay constant across the space. Estimation of equation (2) follows a typical local kernel regression approach as detailed in Fotheringham, Brunsdon, and Charlton (2002) and Yu (2010). In general, certain local samples around a regression point will be weighted based on geographic proximity to that point, and weighted least squares is used to produce the local coefficients. This study uses the linear term growing season temperature and precipitation as independent variables. Schlenker and Roberts (2009) described a nonlinear relationship between crop yield and weather; their nonlinear curve ranges over 40°C and we need a range of at least 10°C in order to observe the nonlinearity. However, our GWPR estimation consists of local regression models in which each local regression only includes a small number of nearby counties that are unlikely to have temperature variations over 10°C.[2] Therefore, a nonlinear relationship between crop yield and weather is hard to detect using our GWPR local estimation. To keep the discussion simple, we investigate only linear aspects of these relationships.

We do not estimate a random effects model here, since county fixed effects $\alpha_{(u,v)}$ may be correlated with weather conditions $X_{(u,v)t}$, violating the assumption of a random effects model that fixed effects need to be orthogonal to the other covariates of the model. For instance, local irrigation systems, captured by $\alpha_{(u,v)}$, may be correlated with precipitation in $X_{(u,v)t}$. Statistical tests based on global data also prefer a two-way fixed effects model (a model with both county and time fixed effects). Instead of using time fixed effects, we specify a local model with time trends for several reasons. First, the above tests that suggest time fixed effects are based on the global panel data, while a GWPR model uses a local panel model, which has many fewer observations. Second, we suspect that time fixed effects may largely absorb weather-induced corn yield variations for local models. However, while avoiding time fixed effects keeps some useful yield variations, it also keeps variations unrelated to weather—such as policy changes or technological improvements—that could have been removed by using time fixed effects. Therefore, we implement a linear time trend to help remove these effects.

---

[2] For a specific year, it is unusual for a small group (e.g., 20–30) of nearby counties to exhibit a 10°C difference in average growing season temperature. It is unusual for a specific county to show a difference of 10°C in average growing season temperature over a five-year period.

One of the critical components in both the GWR and GWPR models is the spatial weighting scheme, which determines how many neighboring observations are included in the local model and, more importantly, how these observations are weighted. Generally, the spatial weighting scheme is often modeled with a distance-decaying kernel function—observations nearer to the regression point are assigned more weight to represent their larger impacts on the regression point (Fotheringham, Brunsdon, and Charlton, 2002). In practice, two types of kernel functions are used: the fixed bandwidth and adaptive bandwidth kernel functions. For every local model, the fixed kernel function includes all of the observations that fall within a fixed distance from the regression point and then weights them. The apparent advantage is that such an approach is fast to execute since there is only one weighting function to be calculated; however, since most study regions do not have regular shapes, the fixed kernel will include fewer observations where data points are sparse and more observations where data points are dense. This could potentially mask subtle spatial heterogeneity that we might want to explore further.

Alternatively, the adaptive kernel function, as its name suggests, will use a fixed "amount" of observations instead of fixed "distance." In so doing, every local model will have exactly the same number of observations, though they will be weighted differently in different local models and hence computationally more expensive. General practices often point out that the adaptive kernel function is preferred because of its ability to detect subtle spatial heterogeneity (Yu, 2006). This study uses an adaptive kernel. The kernel function's single parameter, the bandwidth, can be optimized using the corrected Akaike Information Criterion (AICc) or cross validation criterion (CV). In addition, we assume that bandwidths are time-invariant (which is reasonable since we have a relatively short panel and the spatial structure that presumably generates the bandwidth shall remain invariant for short panels, though time-variant bandwidths are certainly possible). After the bandwidth optimization, observations are spatially weighted based on a geographical weighting function with the bi-square scheme (Fotheringham, Brunsdon, and Charlton, 2002). Spatial weights are also assumed to be time-invariant.

We use the "spgwr" package in R for the GWPR estimation (Bivand and Yu, 2013). Since this package is designed for cross-sectional data, we use the demeaned corn yield and weather data, which is equivalent to estimating a fixed effects model.

For the purpose of comparison and completeness of our analyses, we apply the following six models to our data: the global OLS model, the spatial regression model, the GWR model for cross-sectional data, the global panel model, the spatial panel model, and the GWPR model for panel data.[3]

Data

We use corn—a major crop in the United States—in our study so that we can conduct a large area analysis. Corn yield data for 958 U.S. counties from 2002 to 2006 were collected from the U.S. Department of Agriculture's National Agricultural Statistics Service (2011). The main reason for using a short panel of five-year data is that it allows us to include a wide range of counties across the United States while simplifying the analysis by maintaining a balanced panel. Using a longer balanced panel will largely reduce the number of counties included. For instance, a balanced panel data for 2001–2010 would include only 453 counties as compared to 958 counties for 2002–2006, as many counties have missing yield data during 2008–2010. Using the period 2002–2006 also helps avoid the possible impacts from abrupt policy changes; for instance, many provisions of the 2002 farm bill expired in 2007 (Johnson, 2008). We also exclude all of the western coastal counties that are far away from the rest of counties.

---

[3] Due to the length of the paper, we omit the description of the spatial regression model and the spatial panel model. We apply the spatial lag model, which has a spatially lagged dependent variable as independent variable, and the spatial error model, which has a spatially autocorrelated error term.

Table 1. Summary Statistics of Corn Yields, Growing Season Monthly Total Precipitation, and Growing Season Monthly Mean Temperature

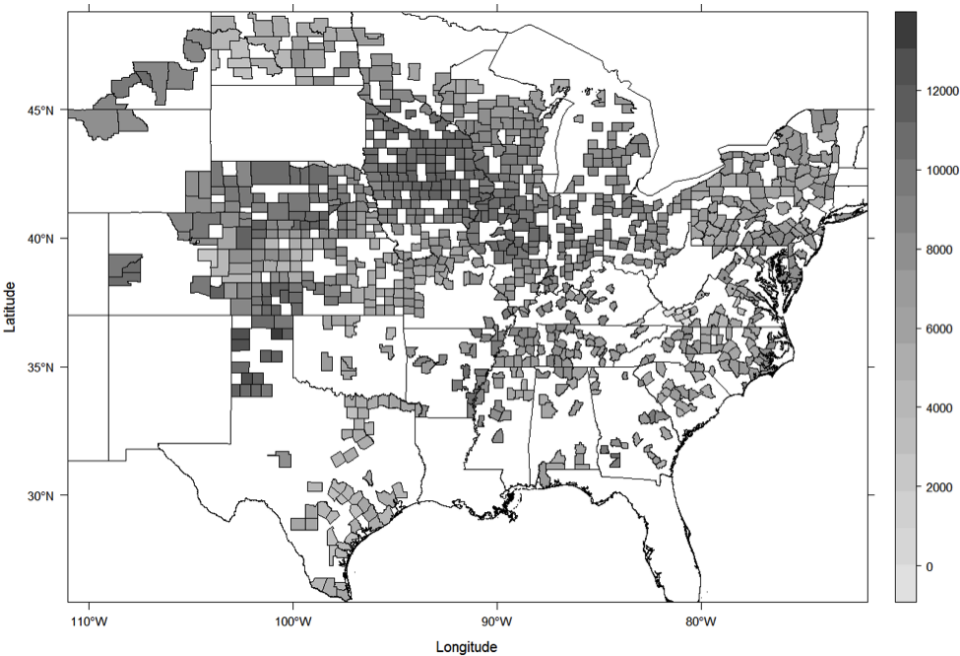|	Mean	Median	SD	Minimum	Maximum
Corn yields (Kg Ha^{-1})	7,883	8,039	2,370	0	14,386
Growing season monthly total precipitation (mm)	91.44	93.98	30.48	12.70	231.14
Growing season monthly mean temperature (°C)	18.56	18.44	−14.83	11.89	28.67

Figure 1. Spatial Distribution of 2002–2006 Average U.S. County Corn Yields (kg/hectare)

We use the growing season monthly mean temperature and total precipitation, collected from National Climatic Data Center (NCDC), for weather stations. We assume that the growing season for corn is from April to October for all counties. We collect monthly mean temperature for each of these seven months and then calculate the growing season mean temperature by taking the average of these seven observations. These station-level weather data are converted to county-level data by taking the average of all the stations in a county.

Results and Discussion

Spatial variations in temperature, precipitation, and corn yields of the study area are presented in figures 1–3. In general, as expected, temperatures are higher in the south than in the north, while precipitation decreases from south to north and from east to west. Counties with high corn yields are mostly in the Corn Belt states. Table 1 presents the summary statistics for corn yields, temperature, and precipitation.

Models for Cross-Sectional Data

We start our analysis with the five-year average data of weather and crop yields over the period of 2002–2006 and apply the models: the global OLS model, the spatial regression model, and
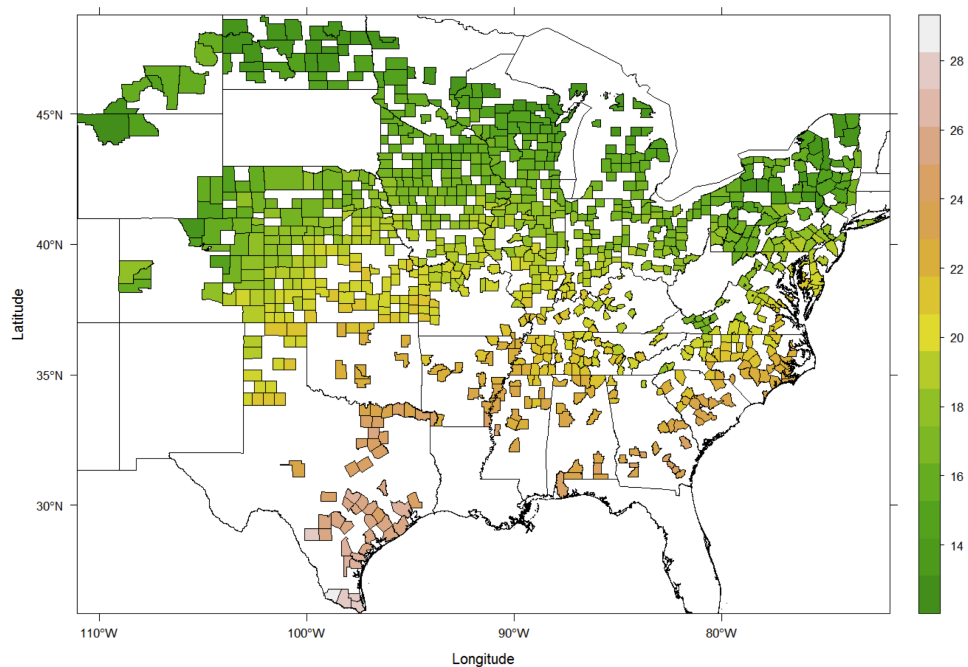
Figure 2. Spatial Distribution of 2002–2006 Average U.S. County Monthly Mean Temperature (°C)
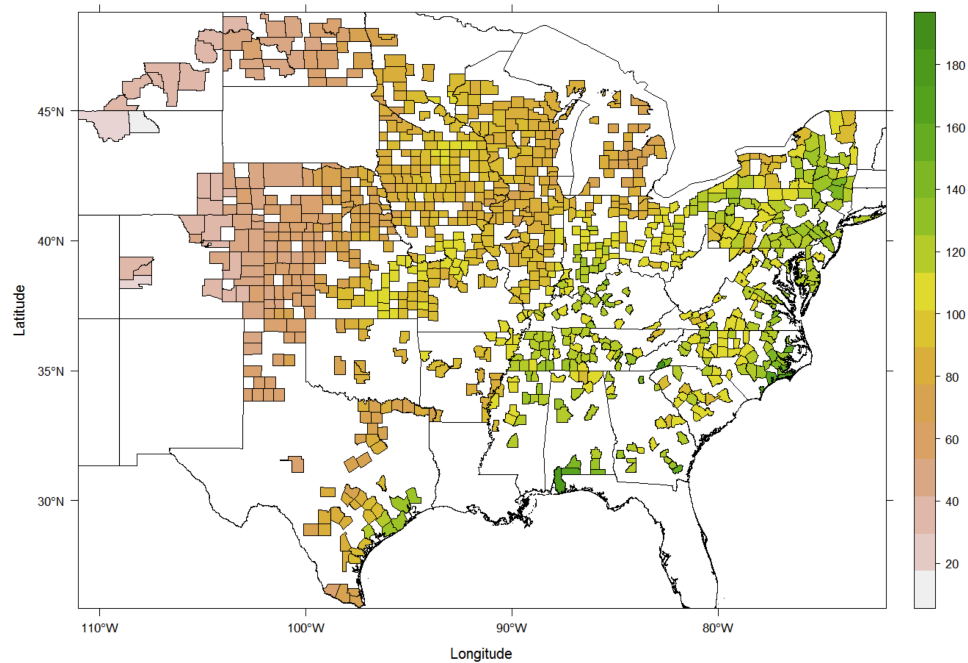

Figure 3. Spatial Distribution of 2002–2006 Average U.S. County Monthly Total Precipitation (mm)

the GWR model. The results from the global OLS regression model—which includes all of the counties—show that corn yields are negatively associated with both temperature and precipitation (table 2). Applying OLS to spatial data may induce inefficient estimation due to possible spatial autocorrelation in the error term. To test the potential spatial autocorrelation, we apply Moran's I of the residuals of the global OLS regression model and find that Moran's I is 0.791, indicating a strong positive spatial autocorrelation between neighboring counties.[4]

Next, using the averaged data, we apply two spatial regression models: the spatial lag and error models (Anselin, 2001). The spatial regression models addressed the residual spatial autocorrelation well—Moran's I for the lag model's residuals reduced to −0.009 and the error model to −0.006 (table 2). However, the spatial regression models still do not explore spatial heterogeneity in coefficient estimates. Using the GWR model, coefficients for temperature (precipitation) have an interquartile range of 658.04 (62.35), which is larger than two times the standard errors of the global OLS model—which is 42.65 (5.19)—indicating certain spatial variability in the GWR coefficients (Fotheringham, Brunsdon, and Charlton, 2002).[5] Still, this model also suggests that many GWR coefficients are not pseudo-significant at the 5% significance level.[6] This may be due to the fact that weather-induced yield variations over time have been completely removed by using the five-year average data. The AICc of the GWR model is 16,079, lower than 17,171 for the OLS model, 16,573 for the spatial lag model, and 16,582 for the spatial error model, indicating that GWR fits the data best, even considering the added complexity.

Models for Panel Data

The aforementioned analyses were all based on cross-sectional data. We next estimate the global panel model, the spatial panel model, and the GWPR model using panel data. For the global panel model, statistical tests prefer a fixed effects model with both county and time fixed effects.[7] In addition, instead of using time fixed effects, we also estimate a global panel model with a linear time trend, since it is common for researchers to include a time trend as an explanatory variable to remove the effects of technological improvements on crop yields (McCarl, Villavicencio, and Wu, 2008). Table 3 shows that the global fixed effects models with time fixed effects or time trend have similar results in that corn yields are negatively associated with growing season temperature while positively associated with growing season precipitation. The global panel model does not account for potential spatial autocorrelation in the residuals. In the spatial panel models, spatial coefficients are significant (λ and ρ in table 3), indicating a strong spatial autocorrelation in the panel model's residuals.

To explore spatial heterogeneity in panel data, we estimate a GWPR model. Although both fixed effects models with time fixed effects and time trend have been estimated for the global panel model, we only estimate a fixed effects model with time trend for the GWPR local model. As the GWPR local models only include a small group of counties, time fixed effects for the local model may absorb too much of the useful variations of corn yields induced by weather.[8] When we estimate a local fixed effects model with time fixed effects, we observe that only about 19% (16%) of counties have pseudo-significant temperature (precipitation) coefficients, indicating that using time fixed effects in GWPR local models may be problematic.

---

[4] Moran's I is a measure of spatial autocorrelation. It varies from −1 to +1, indicating perfect dispersion and perfect correlation, respectively (Moran, 1950). A Moran's I close to 0 indicates small spatial autocorrelation.

[5] Interquartile range is the difference between the first and third quartile.

[6] It should be noted that local t statistics (pseudo-t statistics) should be viewed with caution since nearby observations are used repeatedly for its calculation.

[7] These tests include a Hausman test to compare random effects model and fixed effects model, an LM test to compare random effects model and OLS model, and an F test to determine whether a time fixed effect is needed. The results are not shown here.

[8] The global fixed effects model with time fixed effects or time trends generates similar results, since the average yield variations over time for all the counties should be smaller than that of a small group of local counties and therefore time fixed effects would not absorb too much useful variations at the global level.

Table 2. Estimation Results for the OLS, Spatial Models, and GWR Based on Cross-Sectional Data of 958 U.S. Counties with 2002–2006 Average Observations

Dependent Variable Corn Yields	OLS	Spatial Error Model	Spatial Lag Model	GWR				
				Min	1st Qu.	Median	3rd Qu.	Max
Growing season precipitation	−12.30***	−9.25***	−4.96***	−301.10	−28.19	4.03	34.16	326.10
	(2.65)	(3.30)	(1.72)					
Growing season temperature	−155.82***	−114.12***	−60.50***	−3,705.72	−246.60	101.20	411.44	3,539.85
	(21.76)	(28.90)	(14.22)					
λ		0.571						
ρ			0.573					
N	958	958	958			958		
AICc	17,171	16,582	16,573			16,079		
Moran's I	0.791	−0.006	−0.009			0.089		
Adjusted R-squared	0.09					0.88		

Notes: Single, double, and triple asterisks (*, **, ***) represent significance at the 10%, 5%, and 1% level. Standard errors are shown in parentheses. For the GWR coefficients, its minimum, first quartile, median, third quartile, and maximum are presented. The value of the coefficient can be interpreted as changes of corn yield (kg/hectare) per unit of temperature change (°C), or per unit of precipitation change (mm) for a specific county.

Table 3. Estimation Results for the Fixed Effects Models, Spatial Panel Models, and the GWPR Model Based on Panel Data of 958 U.S. Counties, 2002–2006

Dependent Variable:	Fixed Effects Model		Spatial Panel Fixed Effects Error Model		Spatial Panel Fixed Effects Lag Model		GWPR				
Corn Yields	(1)	(2)	(1)	(2)	(1)	(2)	Min	1st Qu.	Median	3rd Qu.	Max
Growing season precipitation	13.48***	14.33***	8.76***	8.68***	7.45***	7.12***	−56.86	1.35	12.64	25.07	66.82
	(1.09)	(1.14)	(1.06)	(1.08)	(0.82)	(0.81)					
Growing season temperature	−528.84***	−608.74***	−488.21***	−546.53***	−291.19***	−275.69***	−3,438	−1,357	−692	75	1,766
	(45.03)	(41.06)	(50.01)	(48.31)	(33.86)	(29.35)					
λ					0.549***	0.598***					
					(0.013)	(0.012)					
ρ			0.562***	0.612***							
			(0.013)	(0.012)							
Time fixed effects	Yes	No	Yes	No	Yes	No			No		
Time trend	No	Yes	No	Yes	No	Yes			Yes		
N	4,790	4,790	4,790	4,790	4,790	4,790			4,790		
Number of Counties	958	958	958	958	958	958			958		
Number of Years	5	5	5	5	5	5			5		
Adjusted R-squared	0.08	0.19							0.51		

Notes: Single, double, and triple asterisks (*, **, ***) represent significance at the 10%, 5%, and 1% level. Standard errors are shown in parentheses. For the GWPR coefficients, its minimum, first quartile, median, third quartile, and maximum are presented. The value of the coefficient can be interpreted as changes of corn yield (kg/hectare) per unit of temperature change (°C), or per unit of precipitation change (mm) for a specific county.

The bandwidth for the GWPR local model is optimized by a CV criterion with adaptive bandwidths, which results in 160 observations for each local model over five years (that is thirty-two counties for each year). Compared to a GWR model based on the average cross section or a GWPR model with time fixed effects, there are now more pseudo-significant coefficients in the GWPR model with time trend, with about 65% (53%) for temperature (precipitation) (figures 5 and 7). This may be due to the fact that useful yield variations over time are now largely kept by avoiding time fixed effects in the local model. The results that there are more significant temperature coefficients than precipitation coefficients is consistent with Lobell and Burke's 2010 finding that cross-sectional and panel models are better at predicting yield responses to temperature change than precipitation change.

The median of the GWPR coefficient estimates for both precipitation and temperature are similar to the estimates from two global fixed effects models (table 3). We further find that the GWPR coefficient estimates have large spatial variability, with an interquartile of 1,432 (23.72) for temperature (precipitation) larger than two times the standard errors of the global fixed effects model with time trend, which is 82.12 (2.28), indicating certain spatial variability in GWPR coefficients. Corn yields are mostly positively associated with temperature in the north and northeast United States, while the negative relationships are presented in other regions (figures 4, 5, and). This may be due to the fact that the average temperature is near or higher than the optimal corn growth temperature in the south Corn Belt and other southern states; therefore, temperature variations tend to have negative effects. In contrast, the average temperature is generally lower than the optimal corn temperature in the northern states; therefore, temperature variations tend to have positive effects (Schlenker and Roberts, 2009). About 77% of counties have positive precipitation coefficients compared to 28% of counties with positive temperature coefficients, indicating that an increase in precipitation is more favorable than an increase in temperature for corn production in most U.S. counties (figure 8). Positive precipitation effects also indicate that the average precipitation is below the optimal precipitation for most U.S. corn production counties. This observation is consistent with Good (2011), where the most favorable precipitation in July in the heart of the Corn Belt should be about 25% above average. The spatial pattern of precipitation coefficients is irregular compared to that of temperature coefficients (figures 6 and 7). This may be explained by two reasons: first, there is more measurement error in spatial distribution of precipitation (Lobell, 2013), reducing the power of using precipitation to explain yield variations; second, the yield response to precipitation may be influenced by local irrigation systems. Comparing figures 6, 7, and 9, we find that some heavily irrigated land tends to have insignificant or negatively significant precipitation effects, such as the Nebraska and Mississippi river areas. Sharma (2011) also indicated that crop yield could be less related to precipitation in regions with better irrigation systems.

To further investigate the precipitation effects, a subset of data with counties reporting both irrigated and nonirrigated corn yields is also analyzed using the GWPR model (figures 10–12). The data for non-irrigated yields have larger variations in both temperature and precipitation coefficients, indicating that nonirrigated corn yields are more sensitive to weather variations. We further find that, most of the temperature coefficients are negative for non-irrigated corn yields, while precipitation coefficients have both negative and positive values, confirming previous results finding that precipitation coefficients have an irregular spatial pattern. Another observation is that non-irrigated corn has a similar number of significant temperature coefficients as that of irrigated corn but far more significant precipitation coefficients irrigated corn. Compared to the temperature effects, the precipitation effects may be more sensitive to the existence of local irrigation systems.
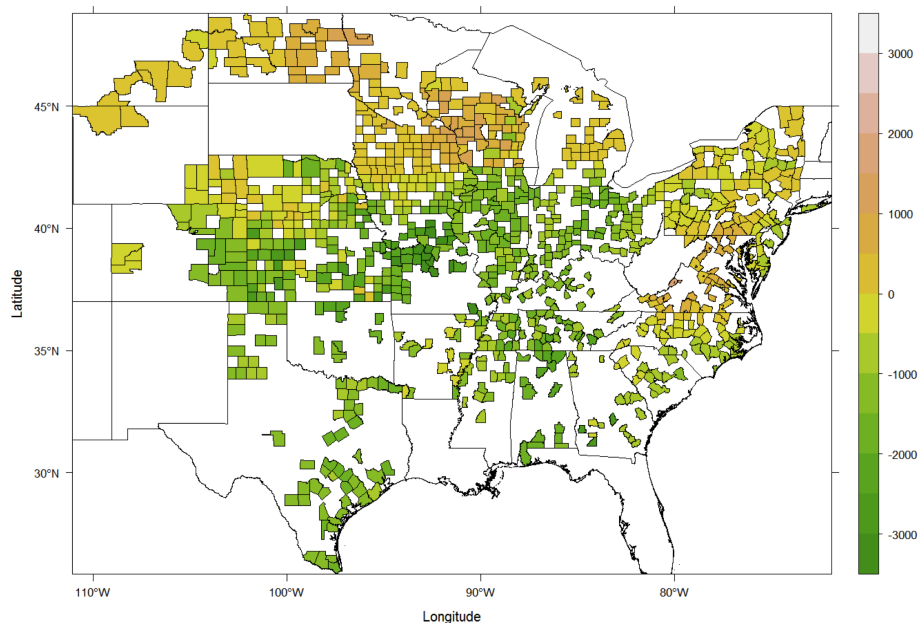
Figure 4. Spatial Distribution of GWPR Coefficients of Growing Season Temperature

Notes: This is from a GWPR analysis based a panel data of 958 U.S. counties during 2002–2006. The coefficients represent the changes of corn yield (kg/hectare) per unit of temperature change (°C).
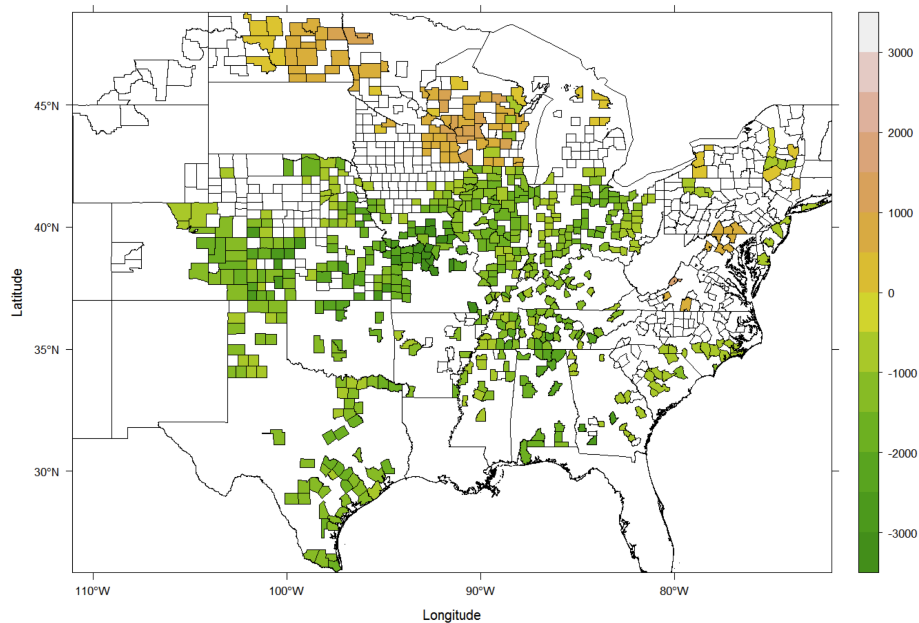

Figure 5. Spatial Distribution of GWPR Coefficients of Growing Season Temperature

Notes: Only counties that are pseudo-significant at the 5% significance level are shown. This is from a GWPR analysis based a panel data of 958 counties during 2002–2006. The coefficients represent the changes of corn yield (kg/hectare) per unit of temperature change (°C).
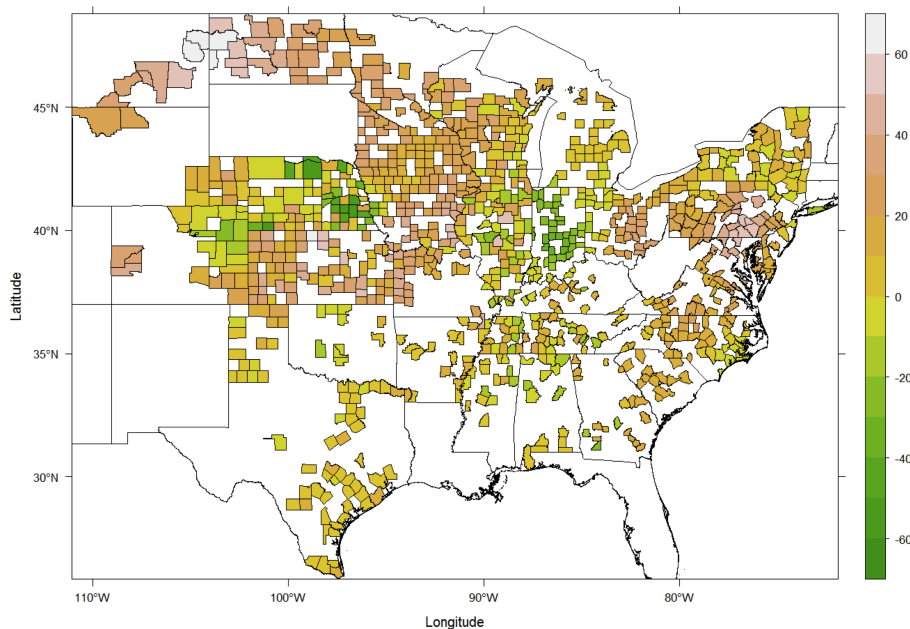
Figure 6. Spatial Distribution of GWPR Coefficients of Growing Season Precipitation

Notes: This is from a GWPR analysis based a panel data of 958 counties during 2002–2006. The coefficients represent the changes of corn yield (kg/hectare) per unit of precipitation change (mm).
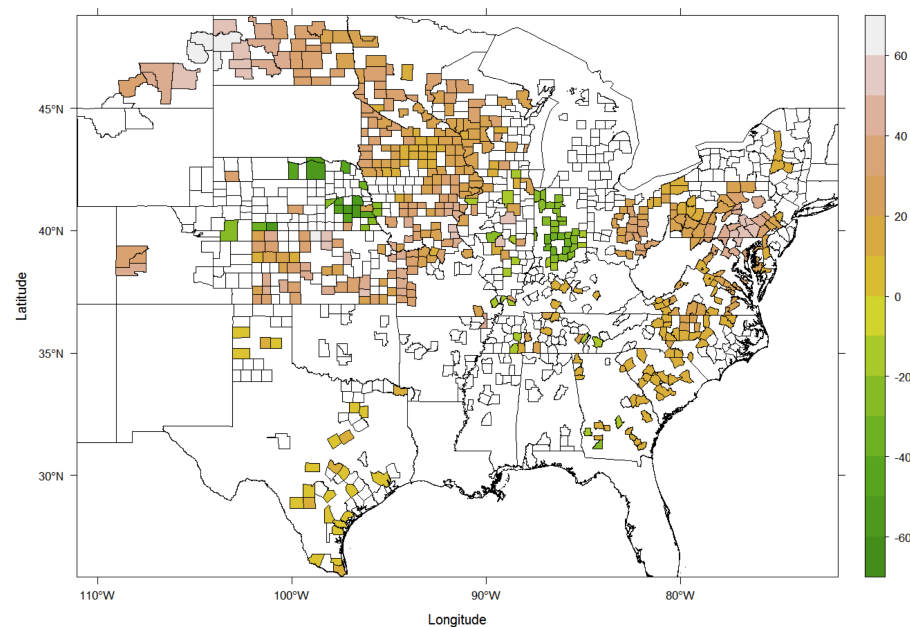

Figure 7. Spatial Distribution of GWPR Coefficients of Growing Season Precipitation

Notes: Only counties that are pseudo-significant at the 5% significance level are shown. This is from a GWPR analysis based a panel data of 958 counties during 2002–2006. The coefficients represent the changes of corn yield (kg/hectare) per unit of precipitation change (mm).

Model Validation

To further investigate the performance of the GWPR model, we first compare its out-of-sample prediction to that of a fixed effects model, which provides a global constant coefficient estimate. Specifically, we randomly draw 90% of the entire sample without replacement as the training data set and use the remaining 10% of the sample as the testing data. We estimate the GWPR model and a fixed effects model based on the training data set and then calculate the out-of-sample root mean squared error (RMSE) based on the testing data set. After repeating the simulation for 1,000 times, we find that the GWPR model is able to reduce the out-of-sample RMSE on average 10.7% from that of the fixed effects model, indicating that GWPR helps improve the out-of-sample prediction accuracy (figure 13a). We further compare the RMSE of the GWPR model with that of an alternative model in which the coefficients vary by region.[9] Then we estimate a fixed effects model for each region to obtain different coefficients for different regions. We find that the RMSE of GWPR is on average 9.6% lower than that of the region-specific fixed effects model (figure 13b), showing that GWPR still has the best performance, even though the region-specific fixed effects model helps improve the RMSE as compared to the global fixed effects model.[10]

We also investigate whether the overall spatial pattern of the weather-crop yield linkage changes when using an alternative model design: each state estimated separately with a fixed effects model that generates a state-specific coefficient estimate. Comparing figure 14 to figures 4 and 6, we find that both temperature and precipitation coefficients follow the same spatial pattern whether generating state-specific coefficients with a fixed effects model or generating county-specific coefficients with the GWPR model.

Lastly, we test the robustness of the GWPR coefficients for alternative adaptive bandwidths. In our main result, each local model includes 160 observations (from thirty-two counties each year for five years); now we alter each local model to include 10, 20, 50, 100, 150, and 200 nearby counties. Figure 15 shows that the overall spatial pattern of temperature coefficients does not change with alternative bandwidths, while larger bandwidth (a local model with more counties included) tends to produce less heterogeneous patterns.[11]

Conclusions

Based on a series of spatial and traditional nonspatial analyses, this study combined GWR and panel models to explore the spatial variability of the relationship between weather and corn yields for the continental U.S. counties during the period of 2002–2006. The GWPR approach confirms that temperature tends to have negative effects on corn yields in warmer regions and positive effects in cooler regions, as expected, but we are able to explore these spatially varying relationships at a finer spatial scale and in much more detail. Compared to temperature effects, the spatial pattern of precipitation effects is relatively irregular in terms of geographic distribution. Further analysis, which distinguished between irrigated and non-irrigated corn yields, indicates that irrigation may reduce the effects of precipitation variation, as expected. Large measurement error in precipitation may also contribute to this irregular spatial pattern.

This work contributes to the literature in two aspects: first, it is one of the first attempts to combine the GWR approach with panel data, which generates more thorough results compared to the cross-sectional GWR model. Second, for U.S. counties, we find that climatic factors have (pseudo-) significantly spatially variant impacts on corn yields. This result underscores the importance of

---

[9] We roughly divide our sample of 958 counties into three regions (figure 13c).

[10] We may also vary coefficients by state, which may further reduce RMSE from the model with three regions. However, since several states have a small number of counties in our balanced panel dataset, it is hard to implement a cross-validation with 90% of the counties as the training dataset and 10% of the counties as the testing dataset.

[11] The results for precipitation coefficients show similar pattern. We do not report them here for conciseness, but are available upon request.
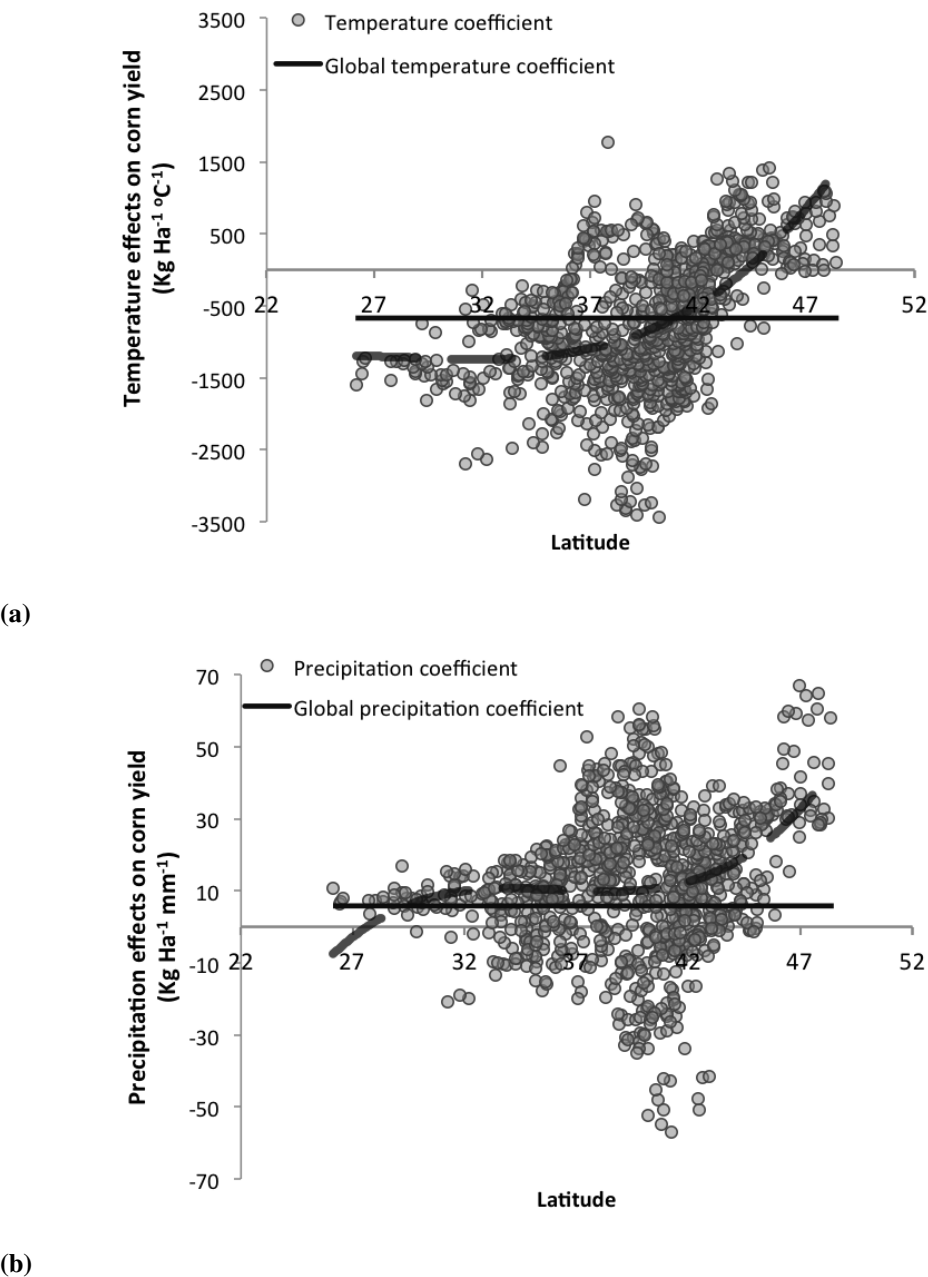
(a)

(b)

Figure 8. Distribution of GWPR Coefficients over Latitude

Notes: Each data point represents the GWPR coefficient estimate for a specific county. (a) temperature. (b) precipitation.

developing local models to guide adaptation efforts as discussed below. These results are very promising in that they are consistent with previous findings yet produce a distinctive spatial pattern of the effects of weather variations on corn yields, which could be used for guiding detailed and precise agricultural planning and decision making. For instance, the possibility of exploring localized crop responses to weather and generating a precise forecast of crop yields may help policymakers determine the severity of climate change impacts in specific regions; thus, spatially explicit adaptation programs could be developed and funding for adaptation to climate change more
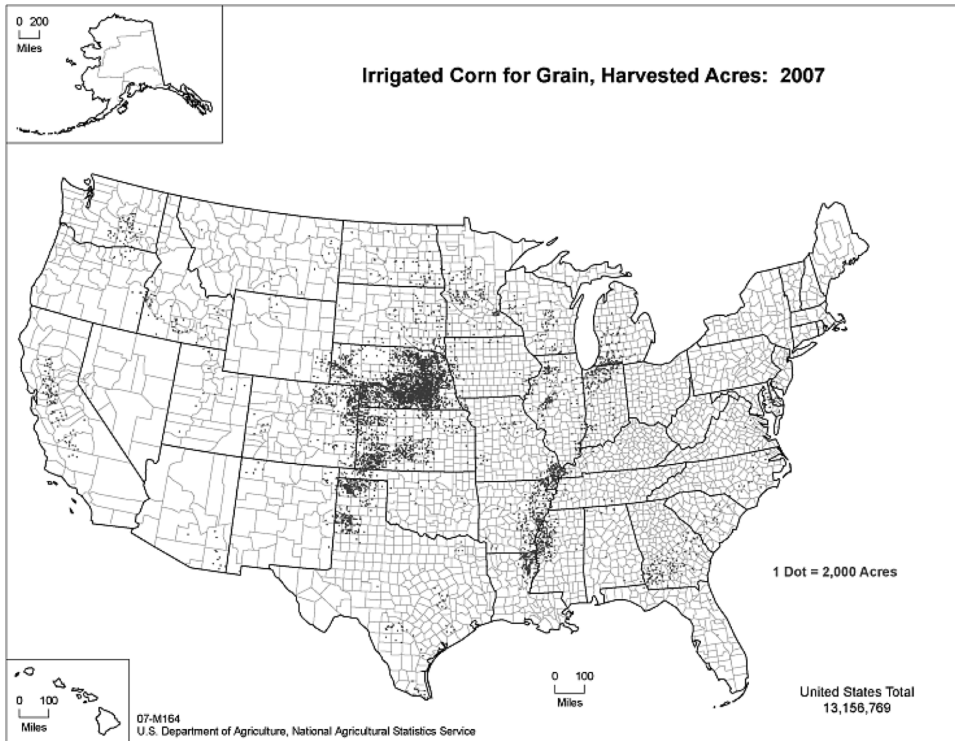
Figure 9. Harvested Acres for Irrigated Corn for Grain in 2007

Notes: Each dot represents 2,000 acres (U.S. Department of Agriculture, National Agricultural Statistics Service, 2011).

efficiently allocated among different regions. The results also provide agricultural producers with spatially explicit guidance for planting practices. Furthermore, identifying spatial heterogeneity in the effects of weather and climate may alert society about emergent social outcomes of climate change, such as possible human migration between regions with different climate impacts. The above policy implications would not be captured by relying on globally constant estimates, which are likely to be distorted at the regional level when yield responses are spatially non-stationary.

In addition to policy implications, identifying localized weather effects may also contribute to the development of agent-based modeling. An agent-based model attempts to simulate complex real-world systems by generating macro-phenomena from micro-specifications. It is usually initialized with heterogeneous agents and certain behavior rules that determine how the agents interact with one other and their common environment. However, most agent-based models use the same rule for all of the agents, which is unrealistic. For instance, assuming that we develop an agent-based model to study the aggregated production responses of agricultural producers to weather in the United States, it would be preferable to allow the responses of crop yields to weather to vary by state or county, while most previous studies of crop yield and weather only provide global estimates. By providing spatially varying crop responses to weather variations, a better (at least more realistic) agent-based model may be developed. Depending on data availability, such an agent-based model could be further improved by including more agent-specific parameters, such as risk attitude, which help determine how agricultural producers respond to the same variations of weather-induced crop yield change. Meanwhile, it should be noted that applying such results to help improve an agent-based model is not imminent, since more reliable GWPR coefficients are required for such application—an important goal for future research.

Although using GWPR advances data analysis, this method remains subject to the major limitations of GWR. For instance, the estimated coefficients are likely to vary with different
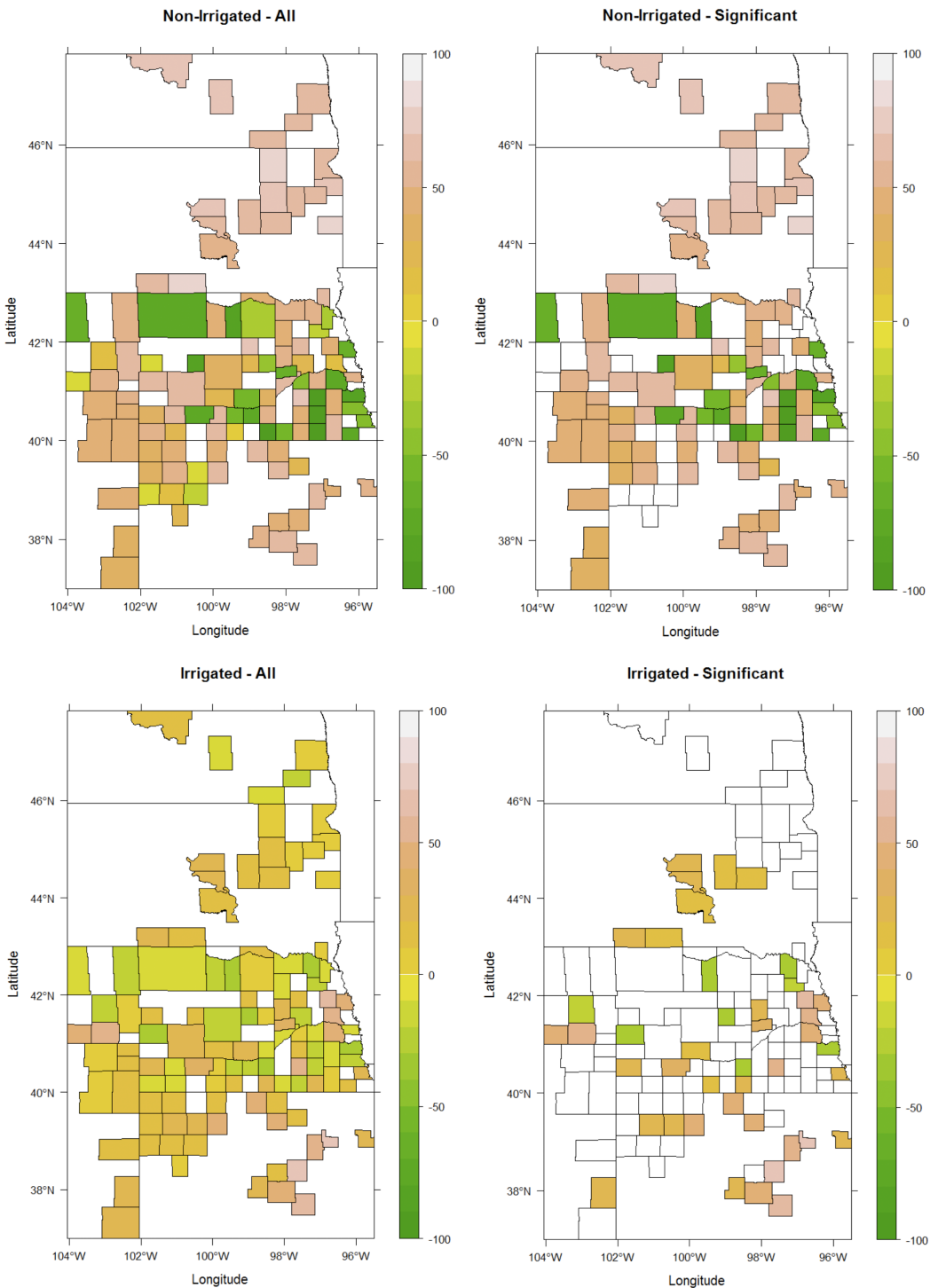
Figure 10. Spatial Distribution of GWPR Coefficients (and Pseudo-Significant Coefficients) of Growing Season Precipitation for Irrigated and Nonirrigated Corn Yields
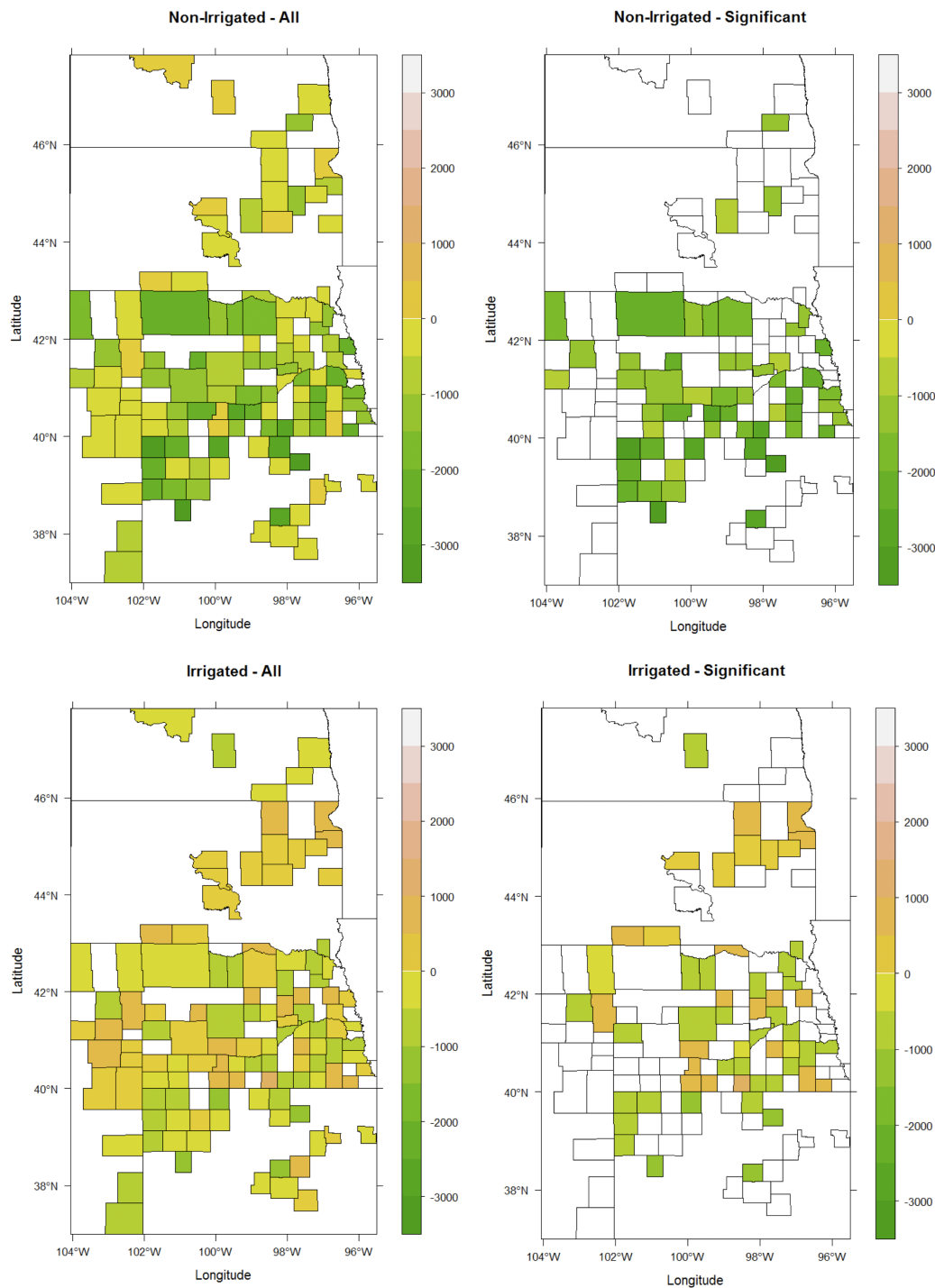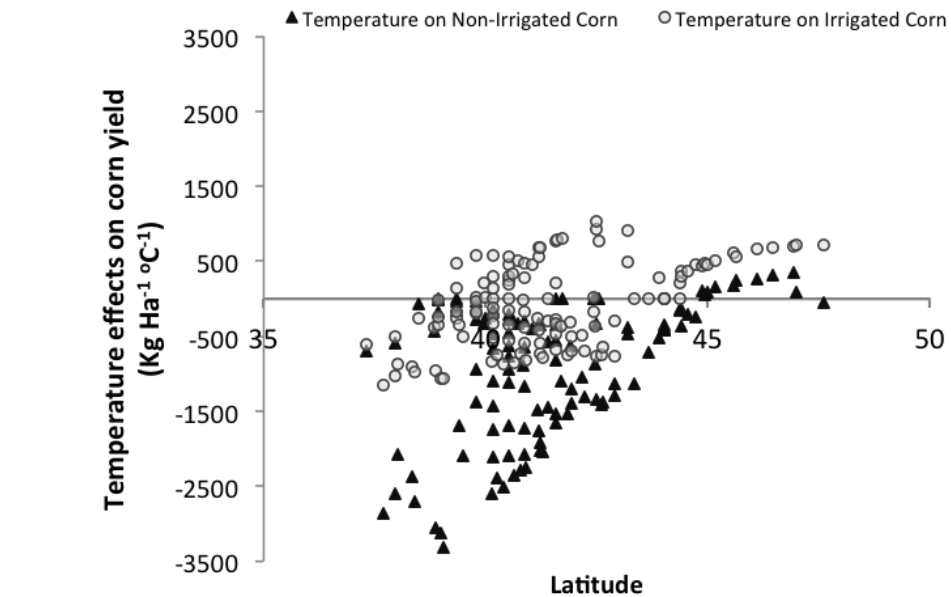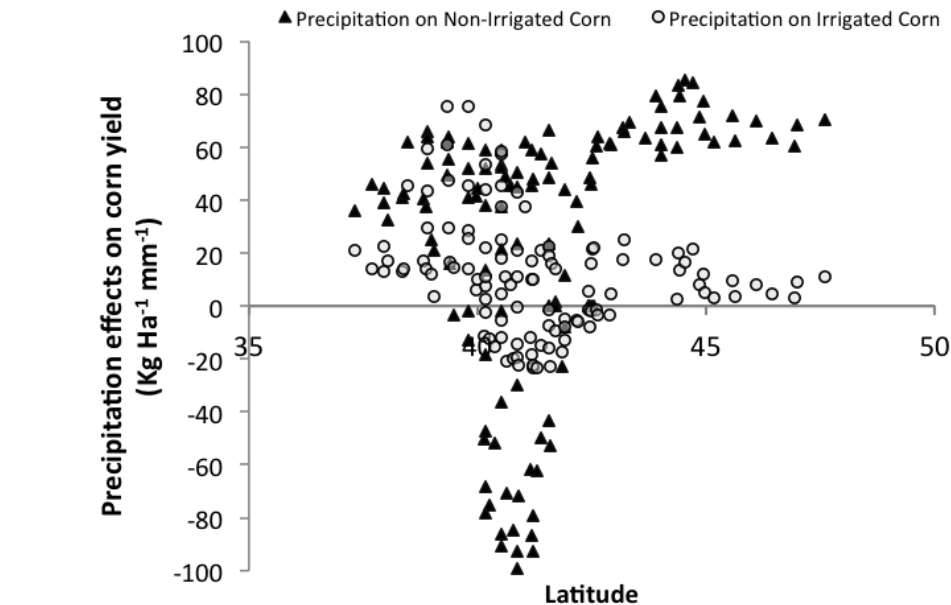
Figure 11. Spatial Distribution of GWPR Coefficients (and Pseudo-Significant Coefficients) of Growing Season Temperature for Irrigated and Nonirrigated Corn Yields

(a)

(b)

Figure 12. Distribution of GWPR Coefficients over Latitude, Separately Estimated based on Irrigated or Nonirrigated Corn Yields

Notes: Each data point represents the GWPR coefficient estimate for a specific county. (a) temperature. (b) precipitation.

bandwidths, while the optimized bandwidth is obtained by exploring the data, not with a solid underlying theory. A second concern is the statistical power of GWR, as the same samples are repeatedly used to calibrate nearby coefficients, which could quickly consume degrees of freedom.

(a)   (b)
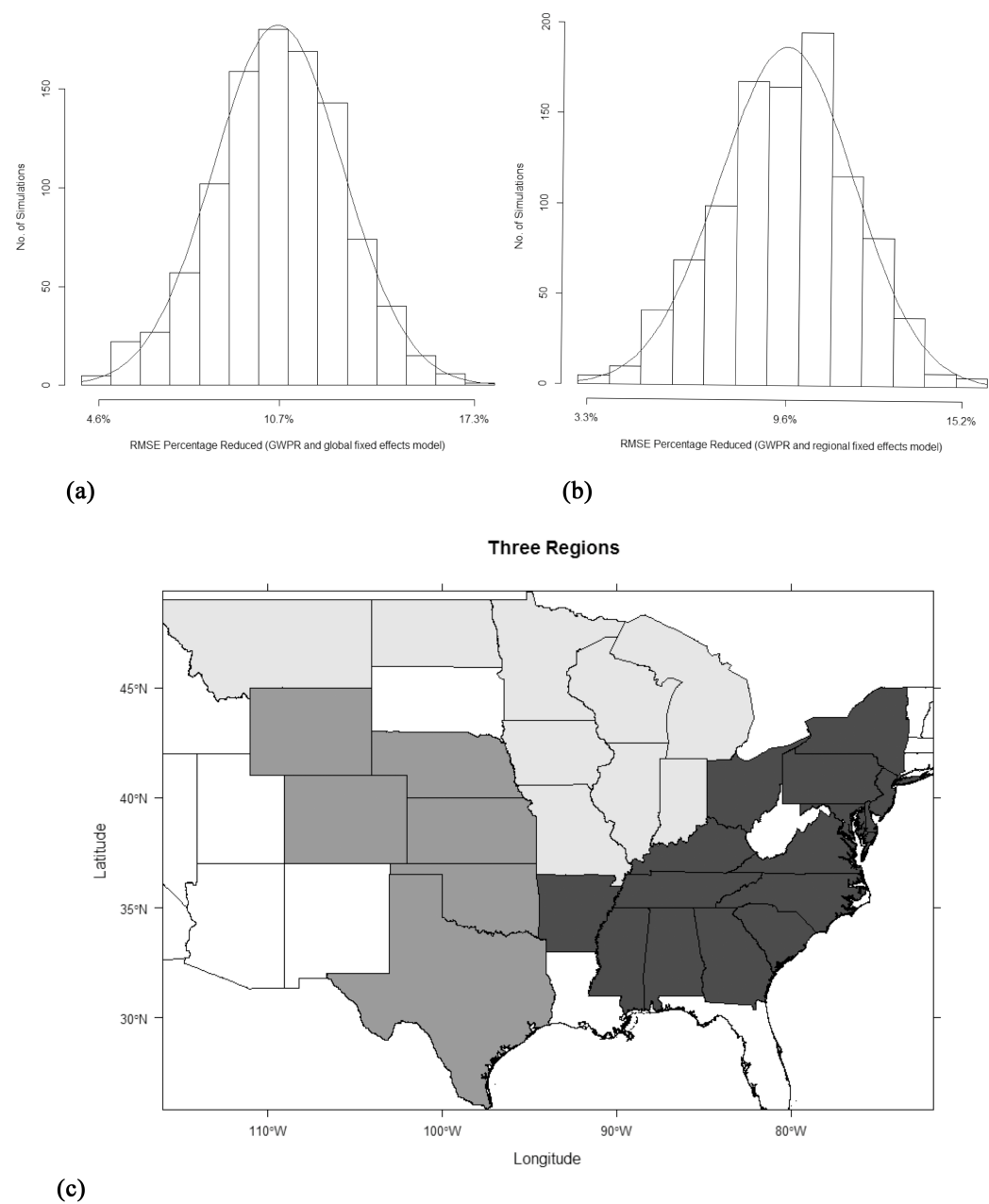

(c)

Figure 13. The distribution of how much the GWPR model reduces the out-of-sample RMSE compared to (a) a global fixed effects model and (b) a regional fixed effects model. A global fixed effects model generates a constant coefficient for the whole sample, while a region-specific fixed effects model generates different coefficients for each region by estimating a fixed effects model for each region separately. We have divided our 958 counties into three regions (c). We randomly draw 90% of the entire sample without replacement as the training data set, and the remaining 10% of the sample as the testing data. The three percentages on the X-axis are the minimum, mean, and maximum of the RMSE percentage reduction.
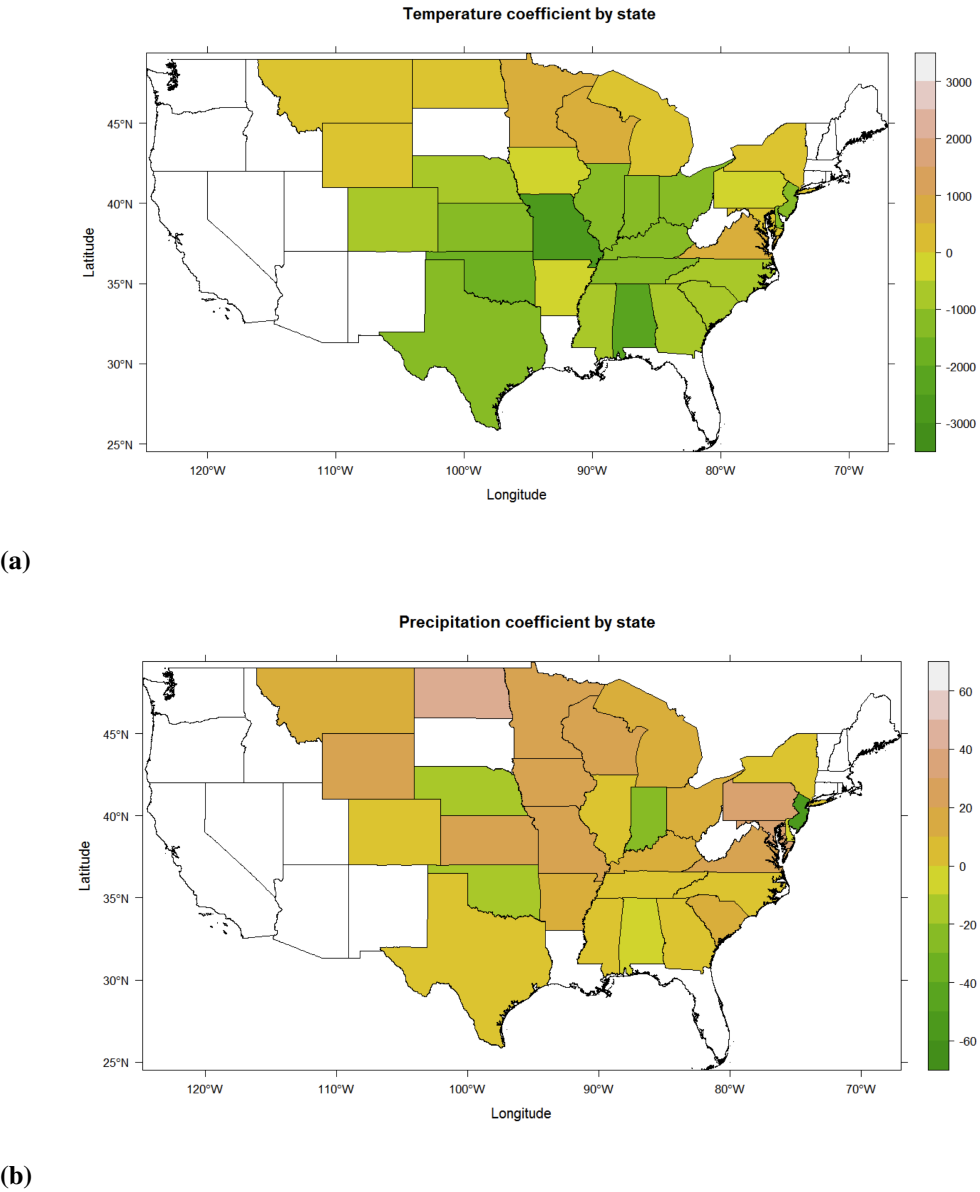
(a)

(b)

Figure 14. Temperature and Precipitation Coefficients Estimated Separately by State Using a Fixed Effects Model

Notes: The same model as the second fixed effects model in table 3.

Therefore, inferences from GWPR should not be treated with the same confidence as traditional statistical analysis. However, the value of the approach and the attempt to combine the exploratory GWR and rich panel data are not diminished by its exploratory nature. The most important contribution this approach provides is that further exploitation of the data can be carried out with more confirmatory analyses. One such analysis is to investigate specific areas with few or no pseudo-significant coefficients (of both temperature and precipitation) and try to identify other factors (other than climatic ones) that might have a strong influence over corn yields.

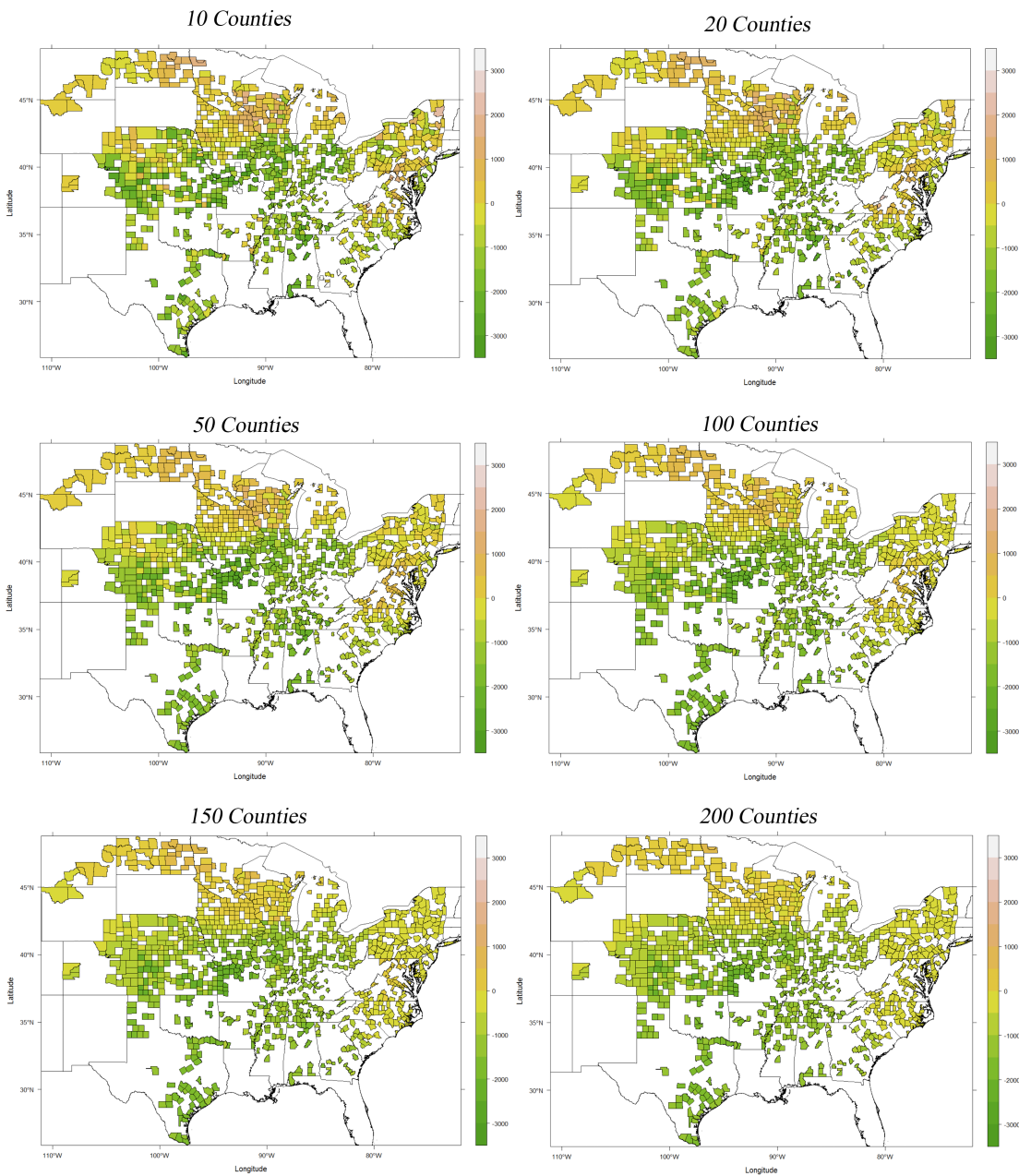*[Received September 2013; final revision received July 2014.]*

Figure 15. Spatial Distribution of GWPR Coefficients of Growing Season Temperature with Alternative Bandwidths

Notes: Each local model includes 10, 20, 50, 100, 150, or 200 counties. The baseline local model has thirty-two counties. It should be noted that the local model with ten counties has fifty observations since we have five time periods.

References

Anselin, L. "Spatial Econometrics." In B. H. Baltagi, ed., *A Companion to Theoretical Econometrics,* Oxford: Blackwell, 2001.

Bivand, R., and D. Yu. "SPGWR - An R Package for Geographically Weighted Regression." 2013.

Brunsdon, C., A. S. Fotheringham, and M. E. Charlton. "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity." *Geographical Analysis* 28(1996):281–298.

Butterworth, M. H., M. A. Semenov, A. Barnes, D. Moran, J. S. West, and B. D. L. Fitt. "North-South Divide: Contrasting Impacts of Climate Change on Crop Yields in Scotland and England." *Journal of The Royal Society Interface* 7(2010):123–130.

Cho, S. H., J. M. Bowker, and W. M. Park. "Measuring the Contribution of Water and Green Space Amenities to Housing Values: An Application and Comparison of Spatially Weighted Hedonic Models." *Journal of Agricultural and Resource Economics* 31(2006):485–507.

Coulson, D. P., L. A. Joyce, D. T. Price, D. W. McKenney, R. M. Siltanen, P. Papadopol, and K. Lawrence. *Climate Scenarios for the Conterminous United States at the County Spatial Scale Using SRES Scenarios A1B and A2 and Prism Climatology.* Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, 2010.

Deschênes, O., and M. Greenstone. "The Economic Impacts of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather." *American Economic Review* 97(2007):354–385.

Dixon, B. L., S. E. Hollinger, P. Garcia, and V. Tirupattur. "Estimating Corn Yield Response Models To Predict Impacts Of Climate Change." *Journal of Agricultural and Resource Economics* 19(1994):58–68.

Elbakidze, L., L. Lu, and S. Eigenbrode. "Evaluating Vector-Virus-Yield Interactions for Peas and Lentils under Climatic Variability: A Limited Dependent Variable Analysis." *Journal of Agricultural and Resource Economics* 36(2011):504–520.

Elhorst, J. P. "Spatial Panel Data Models." In M. M. Fischer and A. Getis, eds., *Handbook of Applied Spatial Analysis,* Berlin: Springer, 2010, 377–407.

Fotheringham, A. S., C. Brunsdon, and M. Charlton. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships.* Chichester, West Sussex: Wiley, 2002.

Good, D. "How Much Risk to the Corn Crop?" News release, University of Illinois Extension, Urbana, IL, 2011. Available online at http://web.extension.illinois.edu/state/ newsdetail.cfm?NewsID=25952.

Johnson, R. "Farm Bill Legislative Action in the 110th Congress." CRS Report for Congress RL33934, Washington, DC, 2008. Available online at http://fpc.state.gov/documents/ organization/ 105160.pdf.

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. "The DSSAT Cropping System Model." *European Journal of Agronomy* 18(2003):235–265.

Lobell, D. B. "Errors in Climate Datasets and Their Effects on Statistical Crop Models." *Agricultural and Forest Meteorology* 170(2013):58–66.

Lobell, D. B., and M. B. Burke. "On the Use of Statistical Models to Predict Crop Yield Responses to Climate Change." *Agricultural and Forest Meteorology* 150(2010):1443–1452.

Lobell, D. B., M. B. Burke, C. Tebaldi, M. D. Mastrandrea, W. P. Falcon, and R. L. Naylor. "Prioritizing Climate Change Adaptation Needs for Food Security in 2030." *Science* 319(2008):607–610.

McCarl, B. A., X. Villavicencio, and X. Wu. "Climate Change and Future Analysis: Is Stationarity Dying?" *American Journal of Agricultural Economics* 90(2008):1241–1247.

Moran, P. A. P. "Notes on Continuous Stochastic Phenomena." *Biometrika* 37(1950):17–23.

Mueller, J. M., and J. B. Loomis. "Spatial Dependence in Hedonic Property Models: Do Different Corrections for Spatial Dependence Result in Economically Significant Differences in Estimated Implicit Prices?" *Journal of Agricultural and Resource Economics* 33(2008):212–231.

Olgun, M., and S. Erdogan. "Modeling Crop Yield Potential of Eastern Anatolia by Using Geographically Weighted Regression." *Archives of Agronomy and Soil Science* 55(2009):255–263.

Partridge, M. D., and D. S. Rickman. "Persistent Pockets of Extreme American Poverty and Job Growth: Is There a Place-Based Policy Role?" *Journal of Agricultural and Resource Economics* 32(2007):201–224.

Schlenker, W., and M. J. Roberts. "Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change." *Proceedings of the National Academy of Sciences of the United States of America* 106(2009):15,594–15,598.

Sharma, V. "Quantification of Net Irrigation Requirement and Application of GIS and Geographically Weighted Regression to Evaluate the Spatial Non-Stationarity between Precipitations vs. Irrigated and Rainfed Maize and Soybean Yields across Nebraska." Biological Systems Engineering–Dissertations, Theses, and Student Research Paper 18, University of Nebraska-Lincoln, Lincoln, NE, 2011.

Tao, F., M. Yokozawa, Y. Xu, Y. Hayashi, and Z. Zhang. "Climate Changes and Trends in Phenology and Yields of Field Crops in China, 1981-2000." *Agricultural and Forest Meteorology* 138(2006):82–92.

Tobey, J., J. Reilly, and S. Kane. "Economic Implications of Global Climate Change for World Agriculture." *Journal of Agricultural and Resource Economics* 17(1992):195–204.

U.S. Department of Agriculture, National Agricultural Statistics Service. "Quick Stats." 2011. Available online at http://quickstats.nass.usda.gov/.

U.S. Global Change Research Program. *Global Climate Change Impacts in the United States: A State of Knowledge Report.* Cambridge: Cambridge University Press, 2009.

Wooldridge, J. M. *Econometric Analysis of Cross-Section and Panel Data.* Cambridge, MA: MIT Press, 2002.

Yu, D. "Spatially Varying Development Mechanisms in the Greater Beijing Area: A Geographically Weighted Regression Investigation." *Annals of Regional Science* 40(2006):173–190.

———. "Exploring Spatiotemporally Varying Regressed Relationships: The Geographically Weighted Panel Regression Analysis." Proceedings of the Joint International Conference on Theory, Data Handling, and Modelling in GeoSpatial Information Science, 134–139, Hong Kong, China, 2010.