Pre-harvest Forecasting of County Wheat Yield and Wheat Quality Conditional on Weather Information

Byoung-Hoon Lee
Graduate Research Assistant
Department of Agricultural Economics
Oklahoma State University
Stillwater, OK 74078
bh.lee@okstate.edu

Philip Kenkel
Professor and Fitzwater Chair for Cooperative Studies
Department of Agricultural Economics
Oklahoma State University
516 Agricultural Hall
Stillwater, OK 74078
phil.kenkel@okstate.edu

B. Wade Brorsen
Regents Professor and Jean & Patsy Neustadt Chair
Department of Agricultural Economics
Oklahoma State University
414 Agricultural Hall
Stillwater, OK 74078
wade.brorsen@okstate.edu

Abstract:

Wheat regression models that account for the effect of weather are developed to forecast wheat yield and quality. Spatial lag effects are included. Wheat yield, protein, and test weight level are strongly influenced by weather variables. The forecasting power of the yield and protein models was enhanced by adding the spatial lag effect. Out of sample forecasting tests confirm the models' usefulness in accounting for the variations in average wheat yield and qualities.

Key words: prediction, protein, spatial lag, test weight, weather, wheat yield

1. Introduction

Winter wheat production in the Southern Plains is a mostly dry land crop with substantial year-to-year variation in yields and quality due to rainfall, temperature and other weather events. If wheat yield and wheat quality response to weather conditions could be predicted early and accurately, the information could be widely used. The information could be particularly important to farmers optimizing late season agronomic and marketing decisions and to grain elevators and millers for purchasing decisions. Thus, there has been increasing interest in the use and development of robust crop weather response models.

Numerous models have been estimated to predict crop yield based on weather conditions. Two main prediction approaches are simulation models and multiple regression models. A number of comprehensive agricultural simulation models are now available to predict yield and variability of wheat. Jones and Kinir (1986) suggested a model to simulate the effects of genotype and weather conditions on crop yield, Duchon (1986), Claborn (1998), Bannayan, Crout, Hoogenboom (2003), and Tsvetsinskaya et al. (2003) predicted yields using weather forecasts and scenarios using the Crop Environment Resource Synthesis (CERES) simulation model. For the Great Plains, Eastering et al. (1998) and Wang et al. (2006) used the Erosion Productivity Impact Calculator (EPIC) model and Eastering et al. (1998) found spatial disaggregation of climate data enhance predictions. Using CERES-Wheat model, Weiss et al. (2003) investigated the responses of wheat yield and end-use quality using nitrogen management and planting dates data. The simulated results depended on spatial locations and climate changes, and also soil water stress and management of nitrogen strongly influenced yield distributions and kernel nitrogen content. Walker (1989) combined simulation and multiple regression to develop physiologically and regionally weighted drought indices from temperature and precipitation data.

The forecasts showed the indices well explain the variation of inter-regional and annual yield within a growing season.

A simulation model is designed to simulate crop yield using details about crop biology. However, as noted by Walker (1989), a simulation approach requires extensive information such as soil type, plant parameters, and weather data related with crop development stage, which are often not readily available. Tannura, Irwin, and Good (2008) argue that an important limitation of crop simulation models is that they are likely to ignore the influence of technology development over time. Bechter and Rutner (1978) and Just and Rausser (1981) found single-equation models forecast more accurately than large econometric models and we should expect a similar result for agronomic models.

Thus, many previous studies have preferred a regression approach rather than a large simulation model when the goal is forecasting. Studies using the multiple regression approach include Yang, Koo, and Wilson (1992), Dixon et al. (1994), Kandiannan et al. (2002), and Chen and Chang (2005) who used various production functions to capture the effect of climate variables on observed crop yield level and to predict crop yield. Irwin, Good, and Tannura (2008) and Tannura, Irwin, and Good (2008) modified Thompson's (1964) corn and soybean regression model and found crop yield strongly related to weather conditions such as temperature, rainfall, technology, and other weather variables. As Tannura, Irwin, and Good (2008) and other studies have proven, multiple regression models have high explanatory power and can represent relationships between weather conditions and crop yield. Thus, the multiple regression model approach is not only easier to use, it is also likely more accurate than the simulation model approach.

Several studies investigated the influences of weather conditions, genotype, and their interaction on wheat quality. The crop maturation period, such as milk development, heading, and ripening stages are the critical stages in determining wheat quality (FAO, 2002). Graybosch et al. (1995), Johansson and Svensson (1998), Smith and Gooding (1999) and Guttieri et al. (2000), and Johansson, Prieto, and Gissen (2008) developed quality models that showed the effect of weather and environment strongly influenced protein content and test weight of wheat. Smith and Gooding (1999) argued predicting grain quality before wheat harvest would be important information to grain buyers, and to farmers to help optimize agronomic activity, particularly, a late application of nitrogen fertilizer to increase protein content (Woolfolk et al., 2002). Britt et al. (2002) estimated six yield and quality of cotton response functions and profit functions as a function of weather information and input and output prices. Regnier, Holcomb, and Rayas-Durate (2007) investigated the variations in flour and dough functionality traits associated with environmental factors and found the interaction between crop years and production regions was a significant factor for flour and dough qualities since growing conditions and climate conditions differ among the regions and across years.

 Unlike previous yield regression models, most quality-related model studies did not measure prediction performance of their models and also used analysis of variance (ANOVA), Spearman rank correlation analysis or simple regression models without precise diagnostic tests for model misspecification.Therefore their methods may lead to biased and inconsistent estimates (McGuirk, Driscoll, Alwang, 1993).

The extensive previous studies have limitations. One is that the previous regression studies cited have solely estimated the impacts on yield and quality level, respectively, and did not deal with agronomic tradeoffs between yield and quality of wheat. Also few focused on

prediction and most studies did not consider out of sample forecasts but measured in sample fit. In-sample fit can be inaccurate because most models, including ours, are developed from pretesting over a large number of alternative specifications.

The other is that many of the above studies have either used data from a single location or have not used the extra information provided by spatial data. The increasing availability of spatial climate information makes it important to incorporate this new level of information to improve forecasts. Anselin (1988) explained that when using spatial data, the dependent variable at each location may be correlated with observations of the dependent variable at neighboring locations. This is defined as spatial contiguity (lag) effect. If this effect is ignored in a model specification, the estimates in the general model are likely to be biased. Therefore, in order to get more accurate forecasts, the crop response model using spatial data needs to include a spatial lag effect.

In addition, Oklahoma has two unique resources for examining the relationship between weather and wheat yields and quality. The Oklahoma Mesonet consists of 120 automated stations covering Oklahoma with one or more stations in each of Oklahoma's 77 counties. Plains Grains, Inc. (PGI) is a private, nonprofit wheat marketing organization based in Stillwater, Oklahoma. PGI evaluates wheat quality, including milling and baking quality from an extensive network of samples at the county level. These two unique data sets provide the opportunity to examine the ability to predict wheat yield and quality with weather data. These two data sets (meso-scale weather data and elevator scale quality data) are highly disaggregated. Thus, the disaggregated data sets could provide more precise wheat yield and quality predictions than was possible with the data sets used in past research.

The objective of the study is to develop wheat regression models to account for the impact of weather on wheat yield and quality and to predict (forecast) wheat yield and quality level accurately. In other words, the primary purpose of the study is to use weather information to predict wheat yield and wheat quality and to select variables and functional forms to estimate parameters and then measure how well the developed models forecast.

2. Conceptual framework

Previous studies have used knowledge about biological development stages of crops to help select the explanatory variables. Dixon et al. (1994) and Kafumann and Snell (1997) specified weather variables for their corn yield regression models that were based on biophysical stages of corn[1]. On the other hand, Yang, Koo, and Wilson (1992) and others used planting season and growing season precipitation and average temperature. Hansen (1991), Tannura, Irwin, and Good (2008) and others estimated the effect of calendar month precipitation and temperature variables on soybean and corn yield during crucial development periods to forecast potential crop yield. Even though biological stages of crops do not precisely correspond with calendar months, a number of previous regression response models have used weather variables defined on a monthly average calendar basis. Previous studies also assume every cross sectional location has the same development stages since it is very difficult to match the precise time point of crop development stages at every location. Another reason is the estimated results using monthly weather variables were similar with that of stage basis variables. For example, Dixon et al. (1994) compared weather variables based on biological stages with variables that based on fixed

---

[1] The corresponding weather variables were specified based on the six weeks before and three weeks after silking point rather than calendar months basis because corn is critically sensitive to precipitation in June and mid-July in Midwestern U.S.

calendar months and found the forecasting performance and R^2 of the two models only changed slightly.

Weather strongly affects four stages[2] of wheat development that determine wheat production level and qualities (FAO, 2002). Aitken (1974), Miralles and Slafer (1999), and Acevedo et al. (2002) argued mainly temperature and precipitation influence wheat development; the most crucial stages of wheat yield are from double ridge to anthesis (flowering) (GS2) and from anthesis to maturity (GS3) since kernel number and weight are being determined at that time (figure 1). Meanwhile, the influence of temperature and precipitation during grain filling are widely known to influence wheat quality characteristics. Graybosch et al. (1995), Johansson and Svensson (1998), Stone and Savin (1999), and Smith and Gooding (1999) found weather has deep impacts on grain quality; for instance, increased temperatures during grain filling tend to increase protein and reduce mean grain weight. Stone and Savin (1999) argued that 70-80 % of total protein is accumulated during the grain filling period.

Winter wheat of the southern Great Plains is typically planted in early September through the middle of November. In general winter wheat harvest begins toward the end of May in southern Oklahoma and continues until about the middle of July (IPM Center, 2005). According to crop weather summary in Oklahoma (DOA, 2000), wheat begins to double ridge and joint in February. Southwestern counties begin to head by the end of March. In April, anthesis is begun and some wheat in south Oklahoma begins the grain filling period, and finally wheat harvest begins approximately May 20th in the southern counties.

_____

[2] The stages can be categorized as germination to emergence (E), from germination to double ridge (GS1), from double ridge to anthesis (GS2), and grain filling period from anthesis to maturity (GS3) (FAO, 2002).

Using the above described general relation of weather variables and wheat by growth stages, the study selects calendar months during GS2 and GS3 and specifies appropriate calendar month weather variables for growing periods that correspond to these biological wheat development stages.

Eastering et al. (1998) used a fine spatial scale to reduce statistical bias from aggregation and confirmed the difference between the observed and estimated yield was greatly reduced when data scale was disaggregated to around 37mile × 50mile. Unfortunately, their method requires a very fine data scale and cannot be used with our data. On the other hand, Anselin (1988) assumed generally the dependent variable or residual at each location may be correlated with neighboring locations' dependent variables or residuals. For this spatially correlated data or residuals, the dependence is termed as spatial autocorrelation or spatial lag (contiguity) effect. This indicates that dependents or residuals are spatially autocorrelated and then violate the general assumption of statistically independent observations. If the spatial lag effect is not considered, estimates will be biased and inconsistent.

In addition, in order to estimate crop response to weather conditions previous studies have used regional models using regional cross-sectional data. However, the regional data such as observed yield, quality level, and weather variables are generally aggregated considerably beyond the county level. If point estimates (weather, yield, quality) are observed near the border of neighboring regions, there is an opportunity for spatial autocorrelation. For instance, grain produced in one county could be shipped to an adjoining county (this would only affect the quality data since the yield data are based on ARS yields which are in turn based on producer reports of harvested production). Some cropland will be closer to a weather station in a neighboring county than weather stations in its own county. Thus, weather measures in a

neighboring county should help predict yield. Thus, a spatial lag model is superior and ignoring this lag would cause parameter estimates to be biased and inconsistent.

Anselin et al. (2008) and Anselin and Bera (1998) express the neighbor relation with a spatial weights matrix, $\boldsymbol{W}$ and the elements w_{ij} of $\boldsymbol{W}$ reflect the potential spatial relations between observations that correspond to the spatial weights structure. The spatial weights matrix can be expressed as binary contiguity sharing a common border, distance contiguity including nearest neighbor locations, and inverse distance between two observations.

Anselin and Bera (1998) suggest two main alternative models of spatial autocorrelation: the spatial lag model, and the spatial error model. The main purpose of the former is to predict the spatial patterns such as cluster and random correlation, while the latter is to increase the efficiency of estimates (Bongiovanni and Lowenberg-DeBoer, 2001). A spatial lag model is used here since the explanatory variables in neighboring counties are expected to help predict our dependent variables. The general regression function can be expressed as:

$$(1) \; \boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y}$ is a vector of dependent variables, $\boldsymbol{X}$ is the matrix of independent variables, and $\boldsymbol{\varepsilon} \sim N(0, \sigma_\varepsilon^2 \boldsymbol{I})$ is a vector of stochastic error terms. The spatial lag model is

$$(2) \; \boldsymbol{y} = \rho \boldsymbol{W} \boldsymbol{y} + \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

where ρ is the spatial autoregressive coefficient, $\boldsymbol{W}$ is $N \times N$ spatial weight matrix (Greene, 2008). This is similar to including a lagged dependent variable in a time series model, except that endogeneity is created because the lagged effects go both directions. The weights matrix is standardized so that rows sum to 1 such as $w_{ij}s = w_{ij} / \sum w_{ij}$ where w_{ij} are elements of $\boldsymbol{W}$. If $\rho > 0$, the dependent variable at each location is positively correlated with other location's

dependent variables. Hence, the spatial lag model can be estimated with instrumental variables such as two stage least square (2SLS) and generalized method of moments (GMM) or with maximum likelihood (ML) (Lambert and Lowenberg-DeBoer, 2001), and 2SLS is used here (see appendix 1).

3. Data

The wheat yield data (from 1994-2009) are from 67 counties in Oklahoma and were obtained from 'Crop Production Report' of United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS). Oklahoma has 77 counties, but ten of them are not included due to having little wheat acreage. The cross-sectional time-series data is composed of 1,072 observations (16 years*67 counties).

The wheat quality data are obtained from Plains Grains, Inc (PGI)[3]. PGI tests 96 samples that were collected on a "grainshed" basis from grain elevators when at least 30% of the local harvest was completed. The term "grainshed" was developed by PGI and represents regions within each state in which the majority of the wheat is marketed through a terminal elevator, river elevator or train loading facility (figure 2). There are 8 grain sheds in Oklahoma. PGI collects representative wheat quality samples from country or terminal elevators. Generally elevators take samples from each truckload arriving at the elevator and the grain is sampled using a hand grain probe. Each elevator directly tests these samples about test weight and moisture content and then these samples typically accumulate in a barrel. Lastly, the elevators barrel is sampled by PGI's representative using a hand grain probe.

---

[3] Plains Grains Inc.(PGI) is located in Oklahoma that does a wheat quality survey and quality testing of hard red winter wheat to provide end-use quality information to the wheat buyer and producer and publish Wheat Quality Report PGI (2009).

The samples from county and terminal elevators are sent to USDA, ARS hard winter wheat Quality Lab in Manhatan, KS. Twenty-five quality parameters are analyzed in order to provide data that specifically describes the quality of wheat (PGI, 2009). These quality data were used for the quality models and correlation analysis. The available historical quality data set from 2004 to 2010 crop years. For more precise analysis, the study matched elevators' quality data with weather data from the closest Mesonet stations. This means one weather station per elevator was used to estimate wheat quality models, not a county average. Wheat quality characteristic data are protein content (% mb: moisture base) and test weight (lb/bu); the quality data were from 96 elevators based on 2010 (figure 3).

Weather data (from January 1, 1994 to May 31, 2010) were obtained from the Oklahoma Mesonet. Each of Oklahoma's 77 counties has one or more Mesonet stations. The selected daily data are daily rainfall (in), daily maximum (minimum) air temperature (°F), daily average air temperature (°F), total solar radiation (MJ m^{-2}d^{-1}), and growth degree days (GDD)[4]. For all Mesonet stations, the daily observations are aggregated to monthly averages. Generally there is one station per county. For counties with multiple stations an average of all stations in the county is used for yield models; however, quality models use only data from the closest weather station. Several weather stations were added during the study period so the closet weather station sometimes varied by year.

4. Empirical model specification

To specify accurately the underlying relationships between yield and quality variables and weather variables, the study first examined the relationships between weather variables and yield

---

[4] GDD=[(Tmax+Tmin)/2]-Tb, 32°F or 39.2°F as the base temperature (Tb) for physiological process in wheat(Cao and Moss, 1989), the GDD vary with growing stage and allow a rough estimation of when a given growth stage is going to occur at a particular site.

and quality level using the correlation coefficients and graphical displays using proc GAM in

SAS (SAS Institute Inc. 2004). GAM allows exploration of data and visualizing structure, and

is useful for investigating the relations between dependent and independent variables (see

appendix table 1, figures 4-15). Appendix table 1 shows all weather variables have a high

correlation with dependent variables: yield and quality level during the growing season.

Precipitation shows less correlation with yield than do average temperature while both variables

are associated with yield. Maximum temperature and minimum temperature have all low

correlation coefficients and negatively signed with protein and test weight, however, the two

variables in quality models were statistically significant. Even though solar radiation and GDD

have high correlation coefficients, the variables in the models were not statistically significant

and therefore those variables were excluded in the model specification. That disagrees with

Dixon et al. (1994) since the solar radiation variable in their model specification is essential.

Precipitation is quadratically related with yield; however, temperature has a linear relation with

yield. Thus, the yield response model used linear and quadratic terms of precipitation and a

linear term for temperature (see figures 4and 5). On the other hand, in the quality response

model there is no evidence that weather variables have a nonlinear relation with quality.

Therefore, the quality response model used a linear specification.

Meanwhile, the study considered several alternative functional forms such as parametric

methods: linear, Cobb-Douglas, translog, square root, spline, and semi-parametric method which

does not assume a specific functional form.

Cobb-Douglas and linear model estimates showed not only statistically significant individual

coefficients, but also relatively high pseudo R^2 (variance ratio) between in sample annual

predicted yield and annual actual yield during 1994-2009, therefore, we selected linear form and

Cobb-Douglas form for yield response model, meanwhile, the quality response model adopted a linear form. The models have the same individual fixed effect and random effect; the functional form can be written as[5]

$$(3) \ yield_{it} = \alpha_i + X_{it}\beta + v_t + \varepsilon_{it} \qquad \text{(Linear form)}$$

$$(4) \ \ln{(yield)}_{it} = \alpha_i + \sum_{j}^{n} \beta_j \ lnX_{jit} + v_t + \varepsilon_{it} \quad \text{(Cobb} - \text{Douglas form)}$$

and also can be expressed as spatial lag model form using spatial lag term:

$$(3.1) \ yield_{it} = \rho \boldsymbol{W_N} yield_{it} + \alpha_i + X_{it}\beta + v_t + \varepsilon_{it}$$

$$(4.1) \ \ln{(yield)}_{it} = \rho \boldsymbol{W_N}\ln{(yield)}_{it} + \alpha_i + \sum_{j}^{n} \beta_j \ lnX_{jit} + v_t + \varepsilon_{it}$$

where $yield_{it}$ is the wheat yield of county i and time t, α_i are individual fixed effects for counties, x_{jit} are the weather variables, and $\boldsymbol{W_N}$ is a $N \times N$ spatial weights for cross-sectional dimension, $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ is a stochastic error term, $v_t \backsim N(0, \sigma_v^2)$ is year random effect, and these error terms are assumed to be independent and identically distributed. The yield response model is composed of county fixed effect, year random effect, and three weather variables from February to April such as monthly average rainfall, squared average rainfall, and average temperature that correspond to before and after the anthesis period in Oklahoma because yield is mostly determined before the grain filling stage.

---

[5] Linear form equation (3) and Cobb-Douglas form equation (4) can be represented as matrices and vectors: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}$ and $\ln\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e}$, $\boldsymbol{e} \sim N(0, \sigma^2)$ and be also rewritten in expected mean form as $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X\beta}$ and $\mathrm{E}(\boldsymbol{y}) = \exp(\boldsymbol{X\beta} + \sigma^2/2)$ respectively. Therefore, when we compare predictions (expected values) between two functional forms accurately, these mean forms are carefully considered.

As discussed, wheat quality depends on the growth periods such as milk development, heading, and ripening stages. In Oklahoma the wheat growth stages during March to May or June in the northern region contribute to grain filling which relates strongly to wheat quality. Additionally, the quality model employed agronomic tradeoff relationship between yield and quality of wheat using the predicted yield level from yield response model and can be expressed as

$$(5)\ Quality_{it} = \alpha_i + \delta(yield)_{it} + \sum_{j}^{n} \gamma_j\, X_{jit} + v_t + \varepsilon_{it}$$

$$(5.1)\ Quality_{it} = \rho W_{Nt} Quality_{it} + \alpha_i + \delta(yield)_{it} + \sum_{j}^{n} \gamma_j\, X_{jit} + v_t + \varepsilon_{it}$$

where $Quality_{it}$ is composed of either protein content (12 % mb: moisture base) and test weight (lb/bu), and W_{Nt} is a $N \times N$ spatial weights for cross-sectional dimension and time t since the number of elevators vary by year, therefore weight structure also varies from year to year, for protein; weather variables are monthly average maximum and the monthly average rainfall from March to May, for test weight; weather variables used in this model included monthly average rainfall for March, April, and May and maximum and minimum temperatures for April and May were based on the heading and ripening period such as before and after anthesis season.

Estimation method and procedure

The study first tests spatial autocorrelation using proc VARIOGRAM in SAS (SAS Institute Inc. 2004). The most generally used test for spatial autocorrelation is Moran's I test[6]

---

[6] Moran's I statistic is $I = \left[\frac{N}{S}\right]\left[\frac{y_t' W y_t}{y_t' y_t}\right]$ where, y_t is a vector of dependent values for each time period t, W is a spatial weights matrix, N is observations, and S is the aggregation of all elements in W. In general, a Moran's I

(Griffith, 1987). Proc VARIOGRAM is used to calculate the Moran's I statistic, Z score, and p-value for testing the hypothesis of no spatial autocorrelation.

The study second adopts maximum likelihood estimation method (Greene, 2008, p. 400) and tests the heteroskedasticity and nonnormality of residuals using a likelihood ratio test, and Shapiro–Wilk test. If hereroskedasticity is formed in the wheat response models' error terms, multiplicative heteroskedasticity [7] will assumed (Greene, 2008, p. 170). If non-normality is formed, the GMM or alternative estimation ways which do not require specific distribution, or a transformation method can be used to modify.

If the dependent variable values are correlated with values of nearby locations based on the Moran's I statistic results, the models will include the weighted dependent variable of equation (2) and be estimated using instrumental variables (see appendix 1). Using proc IML in SAS (SAS Institute Inc. 2004) spatial weights matrix for first (W) and second order (W^2) are constructed based on inverse distance between two observations i and j where inverse distance matrices: $w_{ij} = 1/distance_{ij}$ up to cut off miles. At that time, *GeoDa* software (Luc Anselin, 2004) was used to measure Arc distances among observations for yield and cut off distance using the Oklahoma counties is 49.6 miles, otherwise 0.For quality observations, cut off distances vary over every year since the number of elevators differs by year, and therefore actual distances were used.

In addition, the developed models need to be evaluated for accuracy using out-of-sample forecasting test rather than only a fitness test using historical data. Since the models were selected by pretesting, in sample tests will overestimate their accuracy. To test the out-of-sample

statistic positive and large near one indicates positive autocorrelation while that is negative near one indicates negative autocorrelation (ESRI 2006).

[7] If residuals are heteroskedastic, residual term (ε_{it}) can be expressed as general multiplicative heteroskedasticity form: $\varepsilon_{it} \sim N(0, \sigma_{it}^2)$ or $\sigma_{it}^2 = \exp(\alpha' Z_{it})$ where α and Z_{it} are a vector of parameters and the matrix of independent variables.

forecasting power for the developed models, the yield and quality forecasts will be evaluated for

2010 out of sample. Also the forecasts will benchmark against previous actual six year average.

These tests are truly out-of -sample since the models were developed before the 2010 harvest.

RMSE, MAE, and Theil's U_1 coefficient[8] as measures of forecasting accuracy for all developed

models were used to evaluate the forecasting performance of the models. The first two forecast

error statistics (RMSE and MAE) depend on the scale of the dependent variable as relative

measures. The Theil coefficient is scale invariant and always lies between zero and one, that is,

zero means a perfect fit (Eviews 2000).

5. Empirical Results

The study first tested spatial autocorrelation for dependent variables. Table 1shows a

strong spatial lag effect for yield and protein data with a Moran's I statistic of 0.0078 and 0.0254

and p-values of 0.0001. For test weight data, however, the p-value is 0.2642, indicating the null

hypothesis $H_0: \rho = 0$ no spatial lag effect could not be rejected. Therefore, the study needed to

employ the yield response models in (3.1) and (4.1), and the protein response models in (5.1).

Table 1. Tests of No Spatial Autocorrelation for Wheat Yield, Protein, and Test Weight

	Moran's Index	Expected Index	SD	z-score	p-value
Yield	0.00784 ***	-0.00091	0.000512	17.09	<.0001
Protein	0.02540 ***	-0.00219	0.00106	26.03	<.0001
Weight	-0.00101 ***	-0.00219	0.00106	1.12	0.2642

Note: *** significant at 1%, Ho: no spatial autocorrelation.

The study second estimated equations (3) – (5.1) using SAS proc MIXED (SAS Institute

Inc. 2004) and then the residuals of the estimated models were tested for heteroskedasticity and

---

[8] $U_1 = \frac{\sqrt{\Sigma(\hat{y}_i - y_i)^2}}{\sqrt{\Sigma(\hat{y}_i)^2} + \sqrt{\Sigma(y_i)^2}}$ where $\hat{y}_i$ and y_i are the prediction value and the corresponding actual value of county i respectively (Eviews, 2000, p. 337).

nonnormality (appendix table2). The test results showed linear yield models' LR statistics are smaller than X^2 critical value at the 5% level ($X^2_{3,\ 0.05} = 7.82$) that is, the null hypothesis of homoskedasticity was not rejected for linear yield models; while, the Cobb-Douglas yield models' calculated LR statistics were 19.1 for general model and 16.1 for spatial model, and thus the null hypothesis was rejected at the 5% level. On the other hand, all quality models' LR statistics were greater than X^2 critical value at the 5% level ($X^2_{4,\ 0.05} = 9.47$). The null that residuals are homoskedastic was rejected, we assume multiplicative heteroskedasticity (see Greene, 2008 p. 523). Nonnormality tests showed we can reject the null of normality for all models except test weight, the only linear yield model that did not have heteroskedasticity. As appendix table 2 shows, normality of residuals is still present after correction for heteroskedasticity[9].

Comparing Yield Response Models and Spatial Yield Response Models

Table 2 shows the estimated yield response models and spatial yield response models. Log likelihood statistics were used to select the proper model and in this case (-2 log likelihood), smaller is better. LR test was also used to test for spatial lag effect in the models. The null hypothesis of no spatial lag effect ($\rho = 0$) was rejected. The estimated coefficients indicate how weather variables affect wheat yield. The weather variables were all significant at a critical level of 5% for all yield models. Precipitation has a positive relation with yield; while, squared

---

[9] However, proc MIXED procedure does not provide for nonnormal residuals. Hence, in order to handle nonnormality and heteroskedasticity of residuals proc GLIMMIX procedure in SAS (SAS Institute Inc. 2004) was used. If the EMPIRICAL option (FIRORES) is specified, the procedure provides MacKinnon and White (1985)'s heteroscedasticity-consistent covariance matrix estimators (HCMM) to estimate standard errors. The GMM procedures in GLIMMX only give OLS parameter estimates and those are not efficient. We use proc MIXED and correct for heteroskedasticity to increase efficiency. Our standard errors are not adjusted for nonnormality, but that is of less concern here since our objective is forecasting.

precipitation and temperature are negatively related to yield. This is consistent with Yang et al. (1992). Finally, spatial yield response model's log likelihood statistic indicated the accuracy of the yield response models could be significantly improved by adding the spatially lagged dependent variable (appendix table 3).

Table 2. Yield Model and Spatial Yield Model Estimates, 1994-2009

Variable	Yield Response				Spatial Yield Response			
	Linear		Cobb-Douglas		Linear		Cobb-Douglas	
	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
Intercept	98.621	<.0001	12.732	<.0001	66.790	0.0012	9.879	0.0004
Precipitation	0.569	<.0001	0.379	<.0001	0.301	0.0018	0.164	0.0196
Precipitation^2	-0.008	<.0001	-0.049	<.0001	-0.005	0.0002	-0.025	0.0225
Temperature	-1.610	<.0001	-2.605	<.0001	-1.362	<.0001	-2.547	<.0001
Spatial lag					0.790	0.0002	0.915	<.0001
-2 Log Likelihood	6496.2		-696.8		6481.5		-717.4	

Note: A first-order and second-order spatial weight matrices were used as instruments for the spatial lag term as WX, W^2X.

To measure readily how weather variables affect yield, the elasticity for weather variables was calculated. Table 3 shows the estimated coefficients in Cobb-Douglas form are elasticities. Therefore, precipitation elasticity at mean precipitation is calculated as 0.067 to 0.12 for yield response models and 0.069 to 0.144 for spatial yield response models, that is, as the precipitation is increased by 1%, the average yield level would be expected to rise by 0.067% to 0.12% and 0.069 % to 0.144% in the yield response models and the spatial response models respectively. Temperature elasticity was measured as -2.6 to -2.74 in yield models, that is,1% rise in temperature decreases the average yield level by 2.6% to 2.74%, however, for spatial yield models, temperature elasticity was estimated as -2.32 to -2.55, so we cannot decide which variables more affect yield level of wheat due to the units on variables are arbitrary.

Table 3. The Elasticity of Weather Variables

	Yield Response		Spatial Yield Response	
	Linear	Cobb-Douglas	Linear	Cobb-Douglas
Precipitation	0.116	0.067	0.144	0.069
Temperature	-2.738	-2.605	-2.316	-2.547

Forecast error statistics for all yield models are summarized in table 4. The calculated statistics showed forecasts from the yield response model were similar or slightly less accurate than spatial yield response models. The linear model was slightly less accurate out of sample just as it was in sample. The models for yield performed better relative to the benchmark 5year year average.

Table 4. Out of Sample Forecast Error Statistics for Yield Models, 2010

Forecast Errors	Average (2005-2009) wo/weather effects	Yield Response Linear	Cobb-Douglas	Spatial Yield Response Linear	Cobb-Douglas
RMSE	7.436	4.420	4.087	4.172	4.035
MAE	6.136	3.410	3.146	3.247	3.095
Theil U_1	0.0081	0.00379	0.00370	0.00364	0.00364

Note: County level data were not yet available, so these evaluations were done with the available district data.

Protein Response Model and Weight Response Model

The estimated quality response models of equations (5) and (5.1) for wheat characteristics: protein and test weight level are reported in table 5. Yield and weather variables were all significant at the 5% level. Precipitation and maximum temperature positively affect protein and test weight. Minimum temperature is negatively related with test weight. High yield reduces protein, while, yield positively influenced test weight. These relationships between weather variables and wheat quality are consistent with the findings of Johansson and Svensson (1998), and Smith and Gooding (1999), who found warm temperature affects crude protein positively and precipitation at the end of the season has significant positive correlation with protein concentration. For test weight temperature positively influences test weight and rainfall also is partially associated with test weight. Even though the spatial lag term was not significant

using Wald test, the more reliable likelihood ratio test does reject the null hypothesis of no

spatial lag.

Table 5. Protein Model and Test Weight Model Estimates, 2004-2009

| | Protein (12%, mb) | | | | Test Weight (lb/bu) | |
| | No-Spatial | | Spatial | | No-Spatial | |
Variable	Coeff.	p-value	Coeff.	p-value	Coeff.	p-value
Intercept	2.986	0.0587	2.708	0.1417	35.075	0.0058
Yield	-0.103	<.0001	-0.096	<.0001	0.1152	<.0001
Precipitation	0.014	0.0006	0.017	0.0009	0.0533	0.0003
Max. temp.	0.149	<.0001	0.130	<.0001	0.4651	0.0001
Min. temp.					-0.2922	0.0108
Spatial lag			0.138	0.3843		
-2 Log Likelihood	1045.5		1019.1		1482.9	

Note: A first and second-order spatial weight matrices were used as instruments for the spatial lag term as *WX, W²X.*

The estimated results express the tradeoffs between quality of wheat and yield using the

calculated elasticity (figure16). The elasticity is similar to Dahl and Wilson (1997)'s trade-off

coefficients, which take the derivative with respect to protein and then multiply by average

yields[10]. It indicates how much protein levels affect yield, while, this study is about how much

yield affects quality levels. In other words, yield elasticity indicates that increases in yield lead

to lower protein levels and higher test weight. Table 6 shows a 1% rise in yield decreases

average protein by 0.25% and increases average test weight by 0.28%. Precipitation affects

protein and test weight, however, maximum and minimum temperatures have a stronger

influence on quality levels. Meanwhile, the spatial protein model is more sensitive to weather

conditions than the no spatial protein model. Additionally, the study treated crop year as a

random effect and the counties as fixed effects and the estimated results showed specific region

or crop year of wheat also affects the quality level (appendix table 4).

---

[10] The relationship between protein and yield was estimated using $yield = f(protein, location, year)$, and trade-off coefficients $= \frac{\partial yield}{\partial protein} * \overline{yield}$ (Dahl and Wilson, 1997).

Table 6. The Elasticity of Weather Variables for Quality Model

| | Protein response | | Test Weight Response |
	No-Spatial	Spatial	No-Spatial
Yield	-0.247	-0.266	0.275
Precipitation	0.039	0.052	0.144
Max. Temperature	0.924	0.940	2.747
Min. Temperature			-1.434

Forecasting accuracy of the quality models was evaluated over the 2010 crop year. Table 7 shows the forecast error statistics with protein and test weight response models. Firstly, the study estimated the general quality models combining predicted yield from Yield Response Models. Secondly, the spatial quality models were estimated by using predicted yields. The forecast error values indicate that the accuracy of the quality models can be improved by adding predicted yield from the spatial protein response model rather than that of the no spatial protein response model. Also RMSE values showed that the forecasting performance of the models combining spatially predicted yield were improved. Additionally, average values (2004-2009) were used as benchmark forecasts and as expected, all forecast error values for average values were slightly higher than those of the weather models. The models for protein performed better relative to the benchmark six year average than did the test weight models.

Table 7. Out of Sample Forecast Error Statistics of Wheat Quality, 2010

| | Average (2004-2009) | 2010 | | | |
| | | No-Spatial | | Spatial | |
	wo/weather effects	W/ spatially predicted yield	W/ generally predicted yield	W/ spatially predicted yield	W/ generally predicted yield
Protein					
RMSE	1.080	0.727	0.731	0.702	0.717
MAE	0.870	0.588	0.594	0.582	0.583
Theil U_1	0.0074	0.00530	0.00530	0.005135	0.005197
Test Weight					
RMSE	1.495	1.305	1.361		
MAE	1.196	1.117	1.174		
Theil U_1	0.00041	0.000357	0.000372		

6. Summary and Conclusion

The study developed wheat regression models to account for the effect of weather on wheat yield, protein, and test weight and to forecast wheat yield and the two wheat quality measures. The explanatory variables included precipitation and temperature for growing periods that correspond to biological wheat development stages. The models included county fixed effects, crop year random effects, and a spatial lag effect. Yield and quality level are strongly influenced by weather variables. For yield, precipitation has a positive relation with yield, while, precipitation squared and temperature are negatively related to yield. Precipitation and maximum temperature positively affect protein and test weight. Minimum temperature is negatively related with test weight. Yield affects negatively protein level, while, yield positively influenced test weight. In the forecast evaluation, the forecasting ability of both yield and protein models was enhanced by adding the spatial lag effect. Out of sample forecasting tests showed the developed models are more accurate than using a benchmark six-year average.

Reference

Aitken, Y. 1974. *Flowering Time, Climate and Genotype*. Melbourne, Melbourne University Press.

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht (the Netherlands): Kluwer Academic Publishers.

Anselin, L., and A.K. Bera. 1998. Spatial Dependence in Linear Regression Models with an Introduction to Spatial Economics. In Ullah, A. and D.E.A. Giles, editors, *Handbook of Applied Economic Statistics*, pages 237-289. New York, Marcel Dekker, Inc.

Anselin, L., J. Gallo, and H. Jayet. 2008. Spatial Panel Econometrics. In Mayas, L., P. Sevestre, editors, *The Econometrics of Panel Data*, pages 625-660. New York, Springer-Verlag.

ArcGIS. 2006. ArcMap under Microsoft Windows. Redlands, CA, USA, ESRI.

Acevedo, E., P. Silva, and H. Silva. 2002. Wheat Growth and Physiology. In Curtis, B.C., S. Rajaram, and H.G. Macpherson, editors, *Bread Wheat Improvement and Production*, pages 39-70. Roma, Italy, FAO.

Bannayan, M., N.M.J. Crout, and G. Hoogenboom. 2003."Application of the CERES-Wheat Model for Within-Season Prediction of Winter Wheat Yield in the United Kingdom." *Agronomy Journal* 95:114-125.

Beach, R., A. Thomson, and B. McCarl. 2010. "Climate Change Impacts on U.S. Agriculture." Paper Presented IATRC Public Trade Policy Research and Analysis 'Climate Change in World Agriculture: Mitigation, Adaptation, Trade and Food Security' Symposium, Universität Hohenheim, Stuttgart, Germany, 27 June-29 June 2010.

Bechter, D.M., and J.L. Rutner. 1978."Forecasting with Statistical Models and a Case Study of Retail Sales." *Economic Review(Kansas City)* 63(3):3-11.

Bongiovanni, R. and J. Lowenberg-Deboer. 2001. "Precision Agriculture: Economics of Nitrogen Mannagement in Corn Using Site-Specific Crop Response Estimates from a Spatial Regression Model." Paper Presented at American Agricultural Economics Association Annual Meeting, 5-8 August.

Britt, M.L., O.A. Ramirez, and C.E. Carpio. 2002. "Effects of Quality Considerations and Climate Weather Information on the Management and Profitability of Cotton Production in the Texas High Plains." *Journal of Agricultural and Applied Economics* 34(3):561-583.

Center for Spatially Integrated Social Science. 2004. GeoDa 0.95-I Release Notes. Urbana. IL

Chambers, R.G. 1988. *Applied Production Analysis*. Cambridge, England: Cambridge University Press.

Chang, C.C. 2002. "The Potential Impact of Climate Change on Taiwan's Agriculture." *Agricultural Economics* 27:51-64.

Claborn, W.B. 1998. "Testing and Calibration of the CERES-Wheat Model in Oklahoma." M.S. Dissertation, Oklahoma State University, Stillwater, Oklahoma.

Dahl, B.L., and W.W. Wilson. 1997. "Factors Affecting the Supply of High Quality Spring Wheats: Comparisions Between the United States and Canada. " Agricultural Economics Report No. 374-S, Department of Agricultural Economics, North Dakota State University at Fargo, North Dakota.

Dixon, B.L., S.E. Hollinger, P. Garcia, and V. Tirupattur. 1994. "Estimating Corn Yield Response Models to Predict Impact of Climate Change." *Journal of Agricultural and Resource Economics* 19(1):58-68.

Duchon, C.E. 1986. "Corn Yield Prediction Using Climatology." *Journal of Climate and Applied Meteorology* 25:581-590.

Eastering, W.E., A. Weiss, C.J. Hays, and L.O. Mearns. 1998. "Spatial Scales of Climate Information for Simulating Wheat and Maize Productivity: The Case of the US Great Plains." *Agricultural and Forest Meteorology* 90:51-63.

FAO, 2002. *Bread Wheat Improvement and Production*. Roma, Italy : FAO.

Graybosch, R.A., C.J. Peterson, P.S. Baenziger, and D.R. Shelton. 1995. "Environmental Modification of Hard Red Winter Wheat Flour Protein Composition." *Journal of Cereal Science* 22:45-51.

Greene, W.H. 2008. *Econometric Analysis, Sixth Ed.* Upper Saddle River, NJ: Prentice Hall.

Griffith, D.A. 1987. *Spatial Autocorrelation.* Washington, D.C.: Associatioin of American Geographers.

Hall, H.H. 1998. "Choosing an Empirical Production Function: Theory, Nonnested Hypotheses, Costs of Specifications." Agricultural Economics Research Report No. 59, Department of Agricultural Economics, University of Kentucky at Lexington, Kentucky.

Hansen, L. 1991. "Farmer Response to Changes in Climate: The Case of Corn Production." *Journal of Agricultural Economics Research* 43:18-25.

Irwin, S. H., D. L. Good, and M.A. Tannura 2008. "Forming Expectations about 2008 U.S. Corn and Soybean Yields-Application of Crop Weather Models That Incorporate Planting Progress. " Marketing and Outlook Research Report 2008-01, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, Illinois.

Integrated Pest Management Center (IPM Center). 2005. Crop Profile for Wheat in Oklahoma. Retrieved Novermber 1, 2010, From:http://www.ipmcenters.org/cropprofiles/docs/okwheat.pdf.

Johansson, E. and G. Svensson. 1998. "Variation in Bread-Making Quality: Effects of Weather Parameters on Protein Concentration and Quality in Some Swedish Wheat Cultivars Grown During the Period 1975-1996." *Journal of Science Food Agriculture* 78:109-118.

Just, R.E., and G.C. Rausser. 1981. "Commodity Price Forecasting with Large-Scale Econometric Models and the Futures Market." *American Journal of Agricultural Economics* 63:197-208.

Kafumann, R.K., and S. E. Snell. 1997. "A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants." *American Journal of Agricultural Economics* 79:178-190.

Johansson, E., M.L. Prietolinde, and C. Gissen. 2008. "Influences of Weather, Cultivar and Fertilizer Rate on Grain Protein Polymer Accumulation in Field-Grown Winter Wheat, and Relations to Grain Water Content and Falling Number." *Journal of Science Food Agriculture* 88:2011-2018.

Jones, C.A., and J.R. Kiniry. 1986. *CERES-Maize: A simulation model of maize growth and development.* College Station, Texas, Texas A&M Univ..

Kelejian, H.H., and D.P. Robinson. 1993. "A Suggested Method of Estimation for Spatial Interdependent Models with Autocorrelated Errors, and an Application to a County Expenditure Model." *Regional Science* 72(3):297-312.

Kandiannan, K., R. Karithikeyan, R. Krishnan, C. Kailasam, and T.N. Balasubramanian. 2002. "A Crop-Weather Model for Prediction of Rice Yield Using an Empirical- Statistical Tehnique." *Journal of Agronomy and Crop Science* 188:59-62.

Kim, M.K., and A. Pang. 2009. "Climate Change Impact on Rice Yield and Production Risk." *Journal of Rural Development* 32(2):17-29.

Lambert, D.M., and J. Lowenberg-Deboer. 2003. "Spatial Regression Models for Yield Monitor Data: Case Study from Argentina." Paper Presented American Agricultural Economics Association Annual Meeting, Montreal, Canada, 27-30 July.

Land, K.C., and G. Deane. 1992. "On the Large-Sample Estimation of Regression Models with Spatial- or Network-Effects Terms: A Two-Stage Least Squares Approach." *Sociological Methodology* 22:221-248.

Luper, C. 2005. "Crop Profile for Wheat in Oklahoma" USDA Regional IPM Centers Information System, accessible at http://www.ipmcenters.org/CropProfiles/docs/okwheat.pdf

MacKinnon, J.G. and H. White. 1985."Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properies." *Journal of Econometrics* 29:305-325.

Mass, S.J. 1982. "Forecasting Yields Using Weather-Related Indices." *USDA Staff Report No. AGES820317*. Washington DC: U.S. Department of Agriculture.

McGuirk, A.M., P. Driscoll, and J. Alwang. 1995. "Misspecification Testing: A Comprehensive Approach." *American Journal of Agricultural Economics* 75:1044-1055.

Miralles, D. J., and G. A. Slafer. 1999. Wheat Development. In Satorre, E.H., and G.A. Slafer, editors, *Wheat: Ecology and Physiology of Yield Determination*, pages 13-43. Food Products Press, New York.

National Agricultural Statistics Service. Crop Production Report. Washington DC: U.S. Department of Agriculture, Jan 1994-Jan 2010. Available at < http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1047>.

Oklahoma Department of Agriculture. 2000. *OKLAHOMA AGRICULTURAL STATISTICS 1999*. Oklahoma City, OK.

Oklahoma Mesonet. County Climate Data. Retrieved Jun 1, 2010 from :<http://climate.mesonet.org/county_climate/Products/QuickFacts/>.

Oklahoma Mesonet. Premium Data Available only to Subscribed Members. Retrieved Jun 1, 2010, From: http://www.mesonet.org/premium/statistics.html.

Plains Grains, Inc (PGI),Wheat Quality Data Available only to Private Contracted Members. Retrieved 2004-2010.

Plains Grains, Inc (PGI). 2009. "2009 Hard Red Winter Wheat Qualtiy Report." The Annual Report, PGI, OK. Available at <http://www.plainsgrains.org/pdfs/PGI_HRW_2009.pdf>.

Quantitative Micro Software. 2000. *Eviews 4.0 User's Guide*. Irvine, CA, USA: Quantitative Micro Software.

Regnier, S., R. Holcomb, and P. Rayas-Durate. 2007. *"Relating Wheat Quality to End-Product Quality."* FAPC-129, Oklahoma State University, Stillwater, Oklahoma.

SAS Institute. 2004. The GAM Procedure. *SAS 9.2 Help and Documentation*. Cary. NC

SAS Institute. 2004. The GLIMMIX Procedure. SAS 9.2 Help and Documentation. Cary. NC

SAS Institute. 2004. The IML Procedure. *SAS 9.2 Help and Documentation*. Cary. NC

SAS Institute. 2004. The MIXED Procedure. *SAS 9.2 Help and Documentation*. Cary. NC

Smith, G.P., and M.J. Gooding. 1999. "Models of Wheat Grain Considering Climate Cultivar and Nitrogen Effects." *Agricultural and Forest Meteorology* 94:159-170.

Stone, P. J., and R. Savin. 1999. Grain Quality and Its Physiological Determinants. In Satorre, E.H., and G.A. Slafer, editors, *Wheat: Ecology and Physiology of Yield Determination*, pages 114-122. New York, Food Products Press.

Tannura, M. A., S. H. Irwin, and D. L. Good. 2008. "Weather, Technology, and Corn and Soybean Yields in the U.S. Corn Belt. " Marketing and Outlook Research Report 2008-01, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, Illinois.

Tsvetsinskaya, E.A., L.O. Mearns, T. Mavromatis, W. Gao, L. McDaniel, and M.W. Downton. 2003. "The Effect of Spatial Scale of Climate Change Scenarios on Simulated Maize, Winter Wheat, and Rice Production in the Southeastern United States." *Climatic Change* 60:37-71.

Wang, E., J.R. Williams, and B.B. Little. 2006. "Prediction of Weather Event Associated Crop Yield Losses in Kansas." Paper Presented Sourthern Agricultural Economics Association Annual Meeting, Orlando, Florida, 5-8 February.

Walker, G.K. 1989. "Model for Operational Forecasting of Western Canada Wheat Yield." *Agricultural and Forest Meteorology* 44:339-351.

Weiss, A., C.J. Hays, J. Won. 2003. "Assessing Winter Wheat Responses to Climate Change Scenarios: A Simulation Study in the US Great Plains." *Climatic Change* 58:119-147.

Woolfolk, C.W., W.R. Raun, G.V. Johnson, W.E. Thomason, R.W. Mullen, K.J. Wynn, and K.W. Freeman. 2002."Influence of Late-Season Foliar Nitrogen Applications on Yield and Grain Nitrogen in Winter Wheat." *Agronomy Journal* 94:429-434.

Yang, S.R., W.W. Koo, and W.W. Wilson. 1992. "Heteroskedasticity in Crop Yield Models." *Journal of Agricultural and Resource Economics* 17(1):103-109.

Appendix 1. The Estimation Procedures of Spatial Lag Model

In equation (2), Anselin and Bera (1998) point out the dependent variable may be influenced by spatially lagged dependent variables (Wy) in the other locations. The spatially lagged term is also not only correlated with the same location's error term, but also the other locations' error terms. The consequence is violation that error terms are assumed to be independent and identically distributed, therefore the OLS estimates of a spatial lag model will be biased and inconsistent (Land and Deane, 1992). Instead of OLS estimation, alternative estimation methods are needed to estimate consistent estimators and the procedure follows. Equation (2) can be represented as a reduced form; a function of explanatory variables and error terms at all locations by using an inverse matrix[11]

$$\text{(A1)} \quad \mathbf{y} = (\mathbf{I} - \rho \mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho \mathbf{W})^{-1}\boldsymbol{\varepsilon}$$

under general assumptions, $\boldsymbol{\varepsilon}$ is independent of $\mathbf{X}$, and $E(\boldsymbol{\varepsilon}|\mathbf{X}) = 0$, therefore (A1) is rewritten as conditional mean form as

$$\text{(A2)} \quad E(\mathbf{y}|\mathbf{X}) = (\mathbf{I} - \rho \mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}$$

and, specifically (A2) can be written as

$$\text{(A3)} \quad \widehat{\mathbf{y}} = (\mathbf{I} - \rho \mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}$$

$$= \mathbf{X}\boldsymbol{\beta} + W\mathbf{X}(\rho \boldsymbol{\beta}) + \ \dots$$

and then empirically to estimate (2) using (A1) - (A3), Anselin (1988), Land and Deane (1992), and Kelejian and Robinson (1993) suggested instrumental variables approach to estimation and the study selected Kelejian and Robinson (1993)'s instrumental variables and two-stage least squares estimate (2SLS) approach. Hence, equation (2) can be rewritten as

---

[11] Inverse matrix can be expanded as $(\mathbf{I} - \rho \mathbf{W})^{-1} = (1 + \rho \mathbf{W} + \rho^2 \mathbf{W}^2 + \rho^3 \mathbf{W}^3 + \cdots)$

$$\text{(A4)} \quad \mathbf{y} = \rho \boldsymbol{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{u}$$

where $\mathbf{Z} = (\boldsymbol{W}\mathbf{y}, \mathbf{X})$ and $\delta' = (\rho', \ \beta')$, $E(\mathbf{u}|\mathbf{X}) = 0$, and δ is a two-stage least squares estimator and those can be expressed using (A2) and (A3) as

$$\text{(A5)} \quad E(\boldsymbol{W}\mathbf{y}|\mathbf{X}) = \boldsymbol{W}(\mathrm{I} - \rho\boldsymbol{W})^{-1}\mathbf{X}\boldsymbol{\beta}$$

and specifically (A5) can be rewritten as

$$\text{(A6)} \quad \widehat{\boldsymbol{W}}\mathbf{y} = \boldsymbol{W}(\mathrm{I} - \rho\boldsymbol{W})^{-1}\mathbf{X}\boldsymbol{\beta}$$

$$= \boldsymbol{W}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{W}^2\mathbf{X}(\rho\boldsymbol{\beta}) + \ \dots$$

where $\boldsymbol{W}\mathbf{X}$, $\boldsymbol{W}^2\mathbf{X}$ ... are instrumental variables, and thus, the 2SLS estimator ($\hat{\delta}$) is expressed as

$$\text{(A7)} \quad \hat{\delta} = (\widehat{\mathbf{Z}}' \, \widehat{\mathbf{Z}})^{-1} \, \widehat{\mathbf{Z}}\mathbf{y}$$

where $\widehat{\mathbf{Z}} = (\widehat{\boldsymbol{W}}\mathbf{y}, \mathbf{X})$, indicating $\widehat{\boldsymbol{W}}\mathbf{y} = \boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})\boldsymbol{H}'\boldsymbol{W}\mathbf{y}$ and $\mathbf{H} = (\mathbf{X}, \boldsymbol{W}\mathbf{X})$ as a matrix of instruments, and then a matrix of fitted values $\widehat{\boldsymbol{Z}}$ can be rewritten as

$$\widehat{\boldsymbol{Z}} = \boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})\boldsymbol{H}'\boldsymbol{Z}$$

and also (A7) can be expressed as

$$\text{(A8)} \quad \hat{\delta} = \boldsymbol{Z}'\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})\boldsymbol{H}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})\boldsymbol{H}'\mathbf{y} \,.$$

In addition, the variance-covariance matrix is

$$\mathrm{Var}(\hat{\delta}) = \sigma^2 \ [\boldsymbol{Z}'\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{H})\boldsymbol{H}'\boldsymbol{Z}]^{-1}$$

where σ^2 is residual variance and can be estimated from $\hat{\sigma}^2 = 1/n(\ \mathbf{y} - \widehat{\rho\boldsymbol{W}\mathbf{y}} - \mathbf{X}\hat{\delta})'(\ \mathbf{y} - \widehat{\rho\boldsymbol{W}\mathbf{y}} - \mathbf{X}\hat{\delta})$, thus, we can obtain consistent estimates and standard errors of spatial lag model.

In summary, since OLS estimation in spatial lag model leads to biased and inconsistent estimators, as the alternative method, 2SLS estimation was adopted here and empirically the procedures as follows.

In the first stage, regress equation (A5)'s the right hand side spatially lagged dependent variable $W\mathbf{y}$ on all instrumental variables WX, W^2X… and draw out the estimated $\widehat{W}\mathbf{y}$ of equation (A6), in other words, $\widehat{W}\mathbf{y}$ is obtained from that regress $W\mathbf{y}$ on $\mathbf{H} = (\mathbf{X}, W\mathbf{X})$ and then in the second stage, replace $W\mathbf{y}$ in equation (2) with the estimated $\widehat{W}\mathbf{y}$ in the first stage, then rewrite (2) as

$$(A9)\ \ \mathbf{y} = \rho\,\widehat{W}\mathbf{y} + \mathbf{X\beta} + \mathbf{\varepsilon}$$

and finally, regress (A9) on $\widehat{W}\boldsymbol{y}$ and X and then obtain the 2SLS estimator ($\hat{\delta}$) of (A7) and (A8).

The estimators is consistent and uncorrelated with error term ($\mathbf{\varepsilon}$) and independent variables ($\mathbf{X}$).

Appendix 2. The Calculation Procedures of Elasticity of Weather Variables

As is well known, the elasticity of dependent variable (y) with respect to independents

variables (X_i) is equal to the slope value of the first derivative of dependent variable with respect

to independent variables (dy/dX_i) multiplied by mean ($\bar{X}_i/\bar{y}$) and the elasticity of equation(1)

can be written as

$$(B1) \quad \epsilon_i = \frac{dy}{dX_i}\frac{\bar{X}_i}{\bar{y}} = \beta_i \frac{\bar{X}_i}{\bar{y}}$$

and using equation (A1), the elasticity of equation (2) spatial lag model can be defined as

$$(B2) \quad \epsilon_{spatial\ i} = \frac{dy}{dX_i}\frac{\bar{X}_i}{\bar{y}} = \frac{\beta_i}{(1-\rho W)}\frac{\bar{X}_i}{\bar{y}}$$

and $(1-\rho W)$ is similar to including long run coefficient in long run elasticity, on the other

hand, in the Cobb-Douglas (CD) case[12], (B1) and (B2) can be demonstrated by total

differentiation:

$$(B3) \quad \frac{\partial \ln y}{\partial y}dy = \beta_1 \frac{\partial \ln X_1}{\partial X_1}dX_1 + \cdots + \beta_i \frac{\partial \ln X_i}{\partial X_i}dX_i$$

$$\frac{1}{y}dy = \beta_1 \frac{1}{X_1}dX_1 + \cdots + \beta_i \frac{1}{X_i}dX_i$$

$$\epsilon_i{}^{CD} = \frac{dy}{dX_i}\frac{X_i}{y} = \beta_i$$

and in the spatial model,

$$(B4) \quad \frac{\partial \ln y}{\partial y}dy = (1-\rho W)^{-1}\beta_1 \frac{\partial \ln X_1}{\partial X_1}dX_1 + \cdots + (1-\rho W)^{-1}\beta_i \frac{\partial \ln X_i}{\partial X_i}dX_i$$

---

[12] For general CD form, $\ln y = (\ln X)\boldsymbol{\beta} + e$, and for spatial CD form, $\ln y = \rho W \ln y + (\ln X)\boldsymbol{\beta} + e$

$$\frac{1}{y}\mathrm{d}y = (1-\rho W)^{-1}\beta_1\frac{1}{X_1}\mathrm{d}X_1 + \cdots + (1-\rho W)^{-1}\beta_i\frac{1}{X_i}\mathrm{d}X_i$$

$$\epsilon_i^{CD-spatial} = \frac{\mathrm{d}y}{\mathrm{d}X_i}\frac{X_i}{y} = (1-\rho W)^{-1}\beta_i$$

Appendix Table 1. Correlation Coefficients among Yield, Protein, Weight and Weather Variables

	Yield	p-value	Protein	p-value	Weight	p-value
Yield	1.00	<.0001	-0.47	<.0001	0.15	0.0019
Precipitation (in)						
February	0.49	<.0001	-0.52	<.0001	0.06	0.1912
March	0.04	0.4098	-0.20	<.0001	-0.15	0.0014
April	-0.05	0.3255	-0.16	0.0005	0.08	0.1028
May	-0.22	<.0001	-0.09	0.0613	0.10	0.0406
Avg. temperature (˚F)						
February	-0.43	<.0001	0.06	0.1734	-0.05	0.2891
March	-0.31	<.0001	-0.02	0.6924	-0.12	0.0094
April	-0.42	<.0001	0.34	<.0001	0.20	<.0001
May	-0.02	0.6899	0.00	0.9968	0.06	0.1994
Max. temperature (˚F)						
February	-0.50	<.0001	0.29	<.0001	0.08	0.0815
March	-0.35	<.0001	0.10	0.0396	-0.07	0.1247
April	-0.38	<.0001	0.45	<.0001	0.27	<.0001
May	0.19	<.0001	0.08	0.0742	0.13	0.0071
Min. temperature (˚F)						
February	-0.14	0.0029	-0.22	<.0001	-0.13	0.0045
March	-0.25	<.0001	-0.11	0.0174	-0.18	0.0001
April	-0.35	<.0001	0.11	0.0171	0.10	0.0417
May	-0.20	<.0001	-0.16	0.0004	-0.07	0.1368
Growing degree days						
February	-0.55	<.0001	0.07	0.1295	-0.12	0.0136
March	-0.32	<.0001	-0.01	0.8454	-0.13	0.0069
April	-0.39	<.0001	0.32	<.0001	0.22	<.0001
May	0.02	0.7532	-0.06	0.1994	0.06	0.2246
Solar radiation (MJ m-2d-1)						
February	-0.50	<.0001	0.35	<.0001	0.14	0.0027
March	0.51	<.0001	-0.12	0.0129	0.13	0.0056
April	0.16	0.0013	0.22	<.0001	0.30	<.0001
May	0.40	<.0001	0.06	0.177	0.28	<.0001

Appendix Table 2. Diagnostic Test Statistics: Yield, Protein, and Test Weight Response Models

	Method	Normality Shapiro-Wilk (1) W	p-value	Normality Shapiro-Wilk (2) W	p-value	Heteroskedasticity Likelihood ratio X^2
Yield response model[a]	Linear	0.9958	0.0047**			0.5
	Cobb	0.9903	0.0001**	0.9884	0.0001**	19.1**
Spatial yield response model[a]	Linear	0.9965	0.0157*			0.1
	Cobb	0.9895	0.0001**	0.9883	0.0001**	16.1**
Quality response model[b]	Protein	0.9919	0.0231*	0.9890	0.0033**	27.7**
	Weight	0.9933	0.0614	0.9912	0.0150*	22.2**
Spatial quality model[b]	Protein	0.9989	0.0030**	0.9856	0.0004**	26.6**

Notes: *(*) significant at 1% (5%), X^2 critical value(X^2=7.82) at the 5% level, X^2 critical value(X^2=9.47) at the 5% level, and (1) and (2) indicate normality tests of standardized residuals before and after correction for heteroskedasticity, respectively.

Appendix Table 3. Yield Model and Spatial Yield Model Estimates, 1994-2009

	Yield Response Model		Spatial Yield Response Model	
	Linear	Cobb-Douglas	Linear	Cobb-Douglas
Variable	F Value	F Value	F Value	F Value
Precipitation	79.26	43.94	14.48	20.97
Precipitation^2	67.36	22.52	9.83	5.46
Average temperature	29.98	23.72	14.27	5.23
County fixed effect	3.8	3.41	3.71	3.25
Spatial lag			20.67	23.19
-2 Log Likelihood	6496.2	-696.8	6481.5	-717.4

Notes: Values of precipitation variable was scaled by multiplying one hundred due to unit of rainfall.

Appendix Table 4. Protein Model and Test Weight Model Estimates, 2004-2009

	Protein Model				Test Weight Model	
	No-Spatial		Spatial		No-Spatial	
Variable	F Value	Pr > F	F Value	Pr > F	F Value	Pr > F
Yield (bu/acre)	281.44	<.0001	110.94	<.0001	40.06	<.0001
Precipitation (in)	11.87	0.0006	11.25	0.0009	13.62	0.0003
Max. temperature (°F)	101.61	<.0001	34.5	<.0001	15.28	0.0001
Min. temperature (°F)					6.56	0.0108
County fixed effect	8.9	<.0001	7.25	<.0001	2.76	<.0001
Spatial lag			0.76	0.3843		
-2 Log Likelihood	1045.5		1019.1		1482.9	

Appendix Table 5. Descriptive Statistics for Yield Model Variables, 1994-2009

Variable	N	Mean	SD	Minimum	Maximum
Year	1232	2002	4.61	1994	2009
Yield	1097	30.06	7.87	6.50	53.10
Precipitation	1218	27.02	12.09	0.47	83.47
Avg. temp.(Feb.-April)	1225	51.13	2.58	42.48	57.14
Precipitation^2	1218	876.31	769.19	0.22	6966.68
Log yield	1097	3.36	0.29	1.87	3.97
Log precipitation	1218	3.16	0.59	-0.76	4.42
Log Avg. temp. (Feb.-April)	1225	3.93	0.05	3.75	4.05
Log precipitation^2	1218	10.36	3.21	0.02	19.58

Appendix Table 6. Descriptive Statistics for Quality Model Variables, 2004-2009

Variable	N	Mean	SD	Minimum	Maximum
Year	457	2006.6	1.67	2004	2009
Longitude	457	-98.72	1.32	-102.50	-94.80
Latitude	457	36.10	0.74	34.17	36.90
Protein	457	12.18	1.40	8.90	16.00
Test weight	453	59.79	1.96	52.80	64.60
Yield	421	29.45	7.80	6.50	46.00
Precipitation	454	27.40	13.54	3.40	65.97
Max. temp. (Mar-May)	456	71.80	2.60	65.98	80.47
Min. temp. (Mar-May)	456	46.27	3.36	35.68	55.43
Max. temp. (April-May)	456	75.81	3.52	69.05	85.88
Min. temp. (April-May)	456	50.41	3.10	39.78	59.53

Appendix Table 5. Comparing Actual and Prediction Values for Quality Models

County No.	Protein (12% mb)				Test weight (lb/bu)		
	2010 Actual Value	Average (2004-2009)	Predict Value	Spatial Predict Value	2010 Actual Value	Average (2004-2009)	Predict Value
2	11.9	12.6	11.7	11.5	60.1	60.0	59.2
4	12.6	14.1	13.3	13.5	61.5	59.5	59.1
6	12.2	11.7	11.3	11.3	62.7	60.2	60.5
8	11.3	13.1	11.5	11.2	61.8	58.9	60.1
9	11.1	12.2	10.3	10.3	61.2	60.5	62.5
13	12.4	13.5	12.8	12.5	61.8	60.0	58.9
17	12.1	12.0	11.4	11.3	60.0	60.1	60.6
20	11.4	12.1	11.6	11.0	62.7	60.2	61.5
22	11.5	10.4	10.0	10.0	60.0	58.2	60.3
23	12.2	12.2	12.1	11.9	60.4	58.5	58.7
24	12.3	11.9	11.2	11.3	59.2	59.5	59.3
27	12.2	11.8	11.7	11.6	59.6	59.1	58.5
28	11.8	13.7	12.1	12.0	62.2	61.0	61.2
29	11.2	12.1	12.4	12.5	63.3	60.8	62.1
30	11.9	13.1	12.2	12.0	60.2	59.7	58.7
33	11.6	12.6	12.3	12.1	61.7	61.6	60.9
38	11.5	11.9	11.4	11.2	62.2	60.0	60.6
42	11.5	12.0	11.6	11.6	60.2	59.4	59.4
44	12.0	12.1	11.4	11.4	60.0	59.4	59.7
52	12.3	12.2	11.6	11.6	57.2	58.6	57.9
70	11.7	13.2	12.6	12.3	61.0	59.9	60.2
71	11.6	12.6	11.6	11.5	60.9	61.2	61.2
75	10.0	12.7	11.4	10.9	62.2	60.4	61.3
76	12.5	11.8	11.5	11.5	60.6	60.4	59.6
77	11.8	12.1	11.7	11.4	59.3	59.0	58.9
Avg.	11.8	12.4	11.7	11.6	60.9	59.8	60.0
RMSE		1.080	0.727	0.702		1.495	1.305
MAE		0.870	0.588	0.582		1.196	1.117

Figure 1. Wheat Growth and Development Stages

Schematic diagram of wheat growth and development stages, periods of initiation or growth of specific organs and periods of different components of grain yield
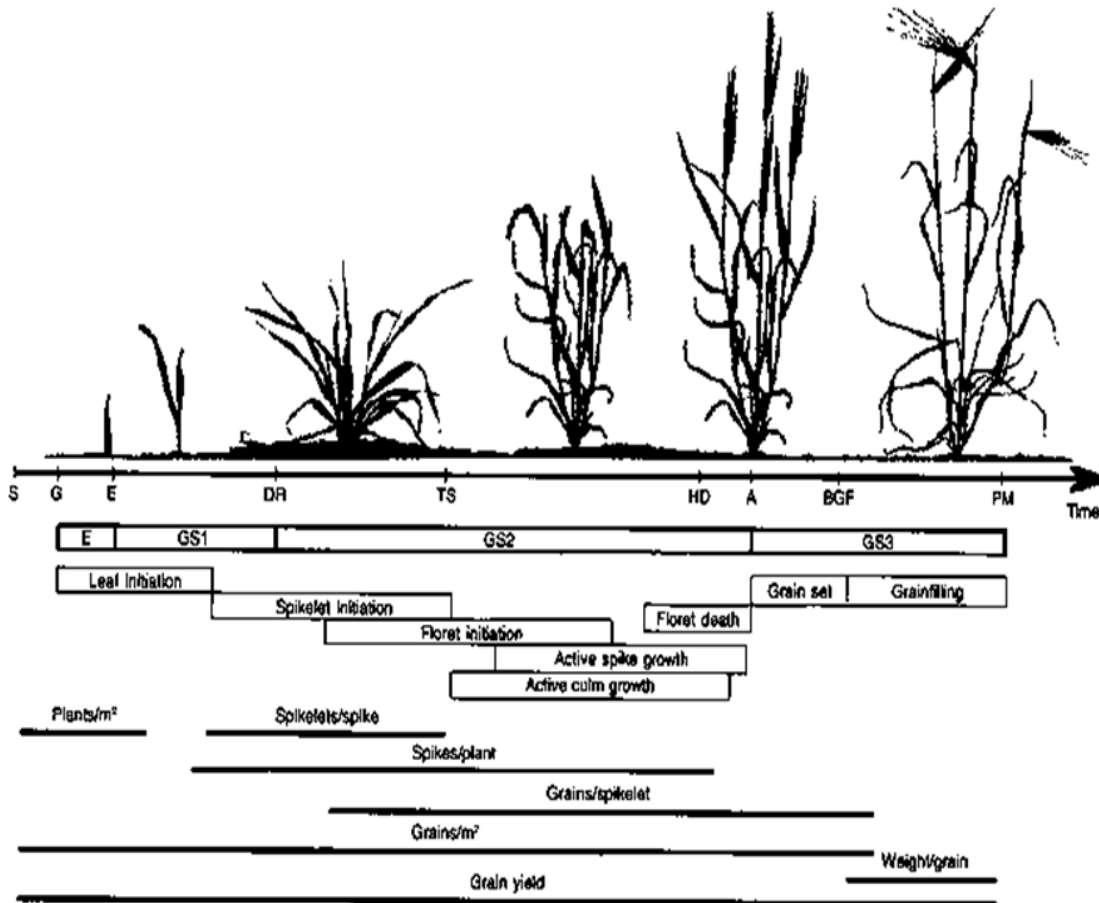
S=sowing; G=germination; E=emergence; DR=double ridge appearance; TS=terminal spikelet initiation; HD=heading; A=anthesis; BGF=beginning of grainfilling period; PM=physiological maturity; GS=growth stage

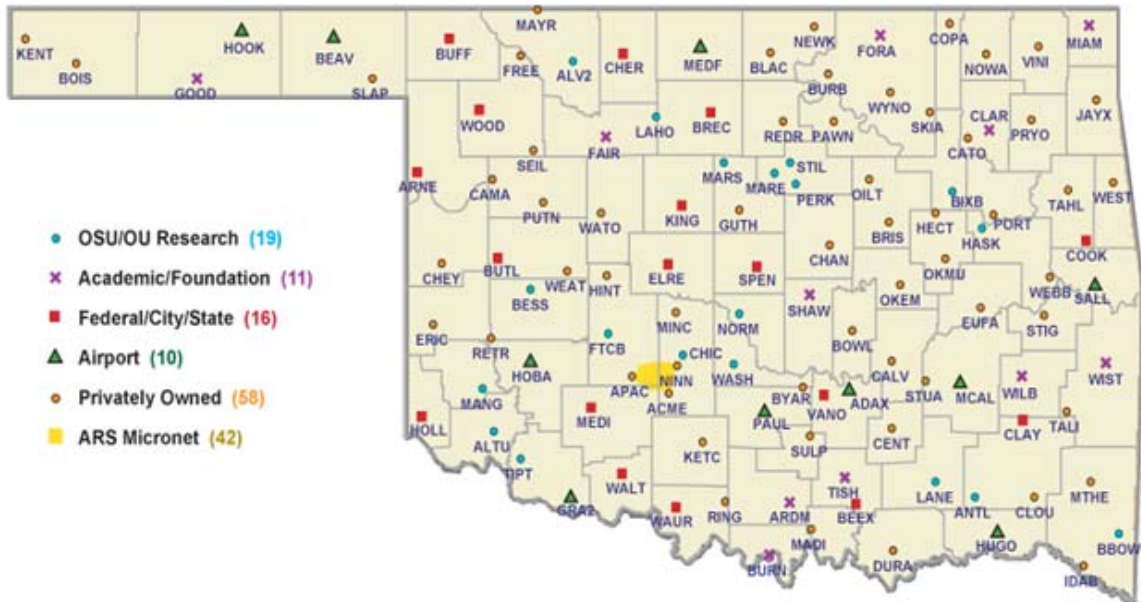Source: Adapted from Slafer and Rawson, 1994

Figure 2. The Study Area of Grainshed Basis
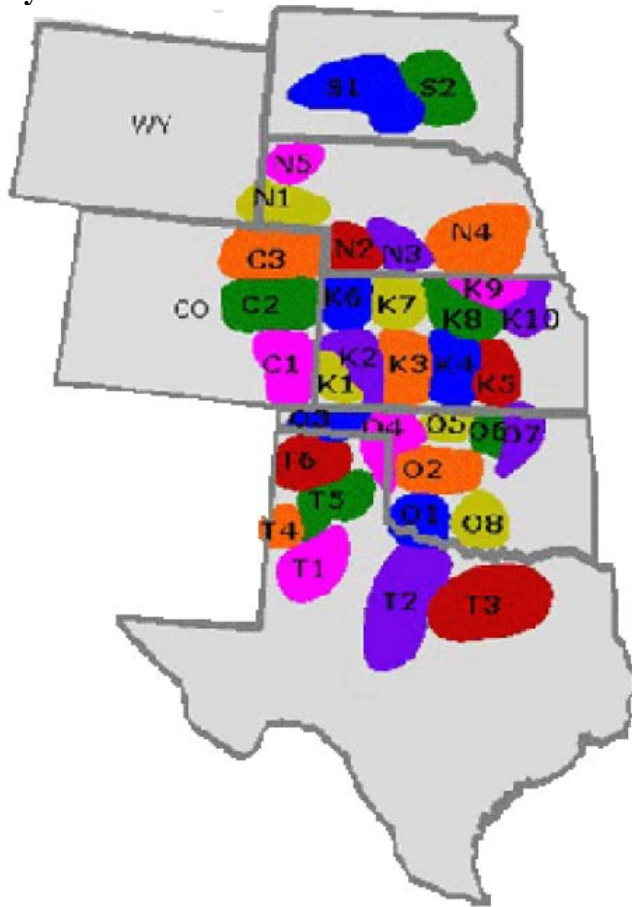
Source: 2009 Oklahoma wheat quality report. (2010)

Figure 3. Location of Oklahoma Mesonet Stations and Elevators
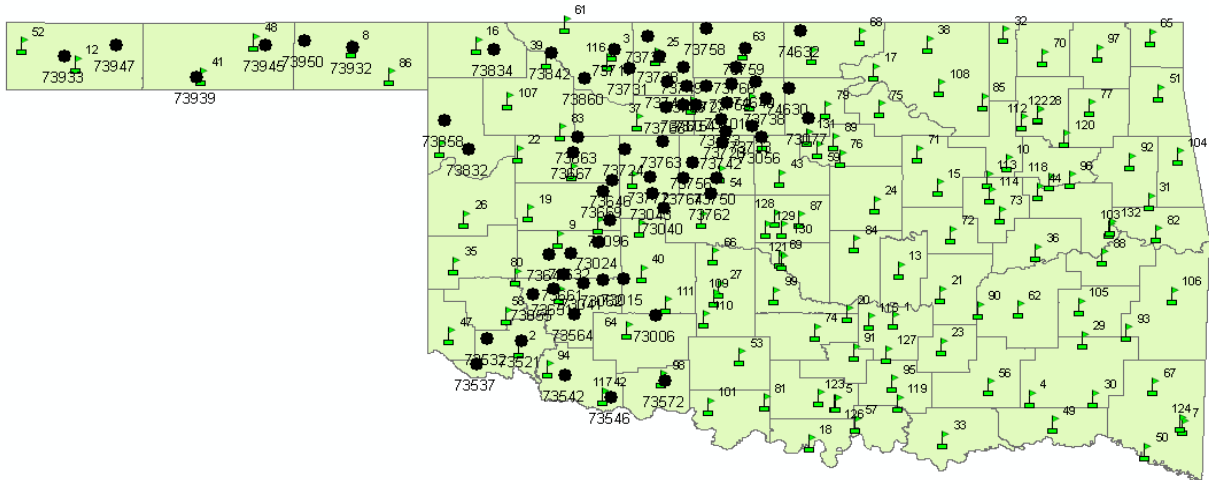
Notes: the study matched elevators' quality data(points) with Mesonet stations' weather data (flags) on each the closest distance basis.

Source: Oklahoma Mesonet

Figure 4. Plot Yield and Precipitation
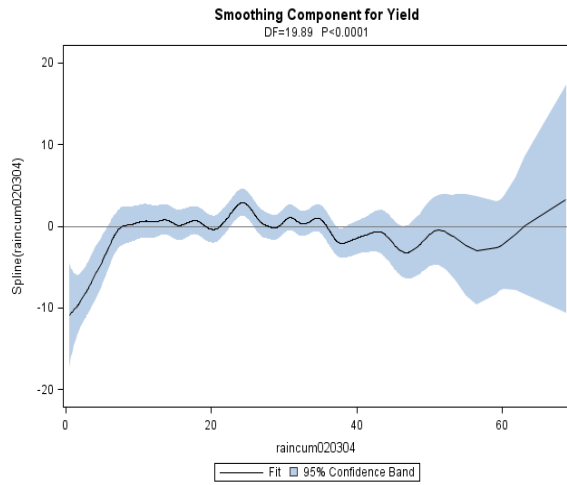
Figure 5. Plot Yield and Avg. Temp.
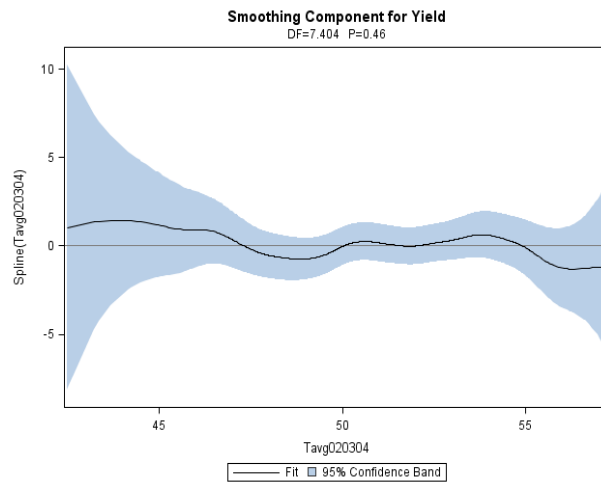
Figure 6. Plot Protein and Precipitation
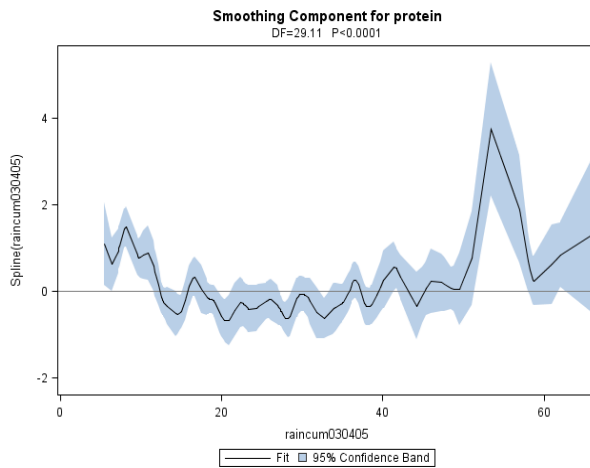
Figure 7. Plot Protein and Max. Temp.
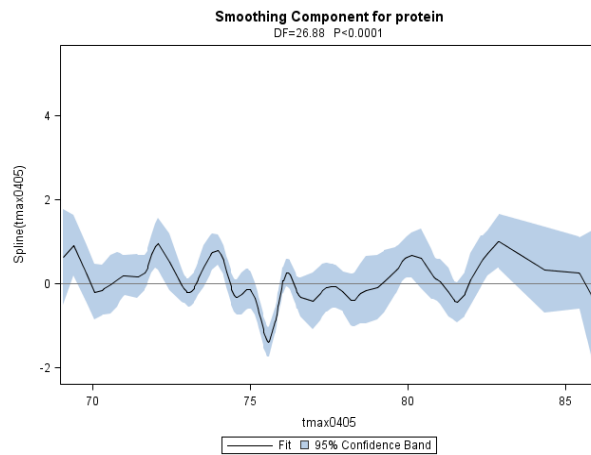
Figure 8. Plot Protein and Yield
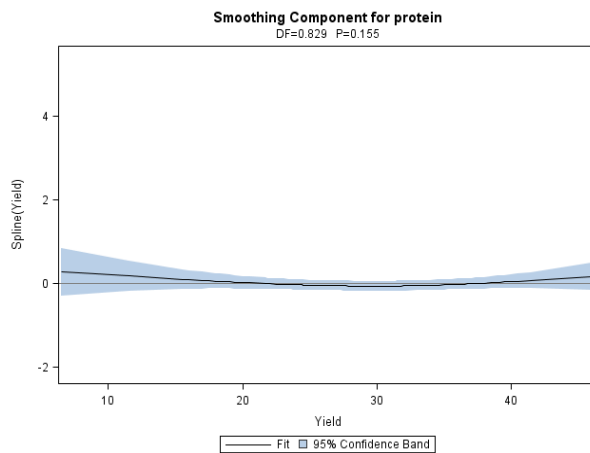
Figure 9. Plot Weight and Precipitation
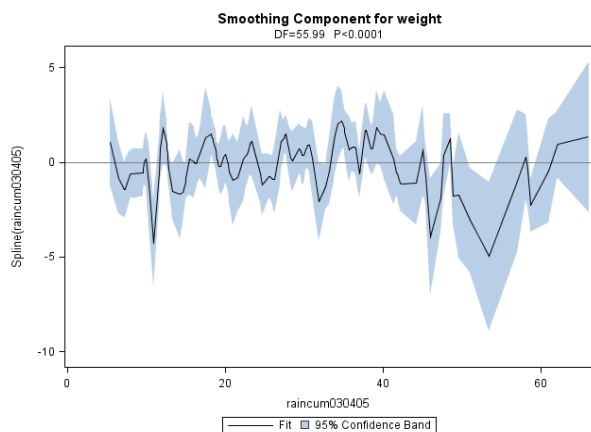
Figure 10. Plot Weight and Maximum Temp. Figure 11. Plot Weight and Minimum Temp.
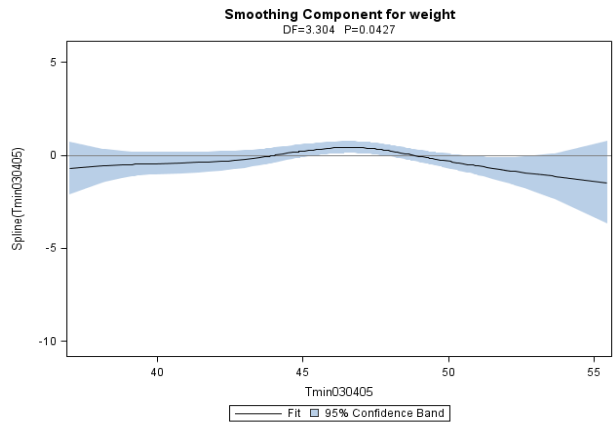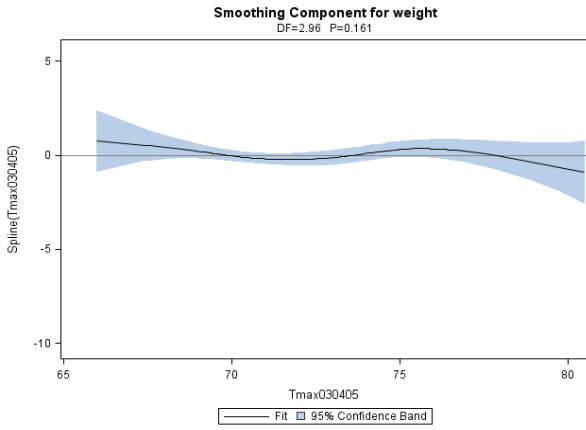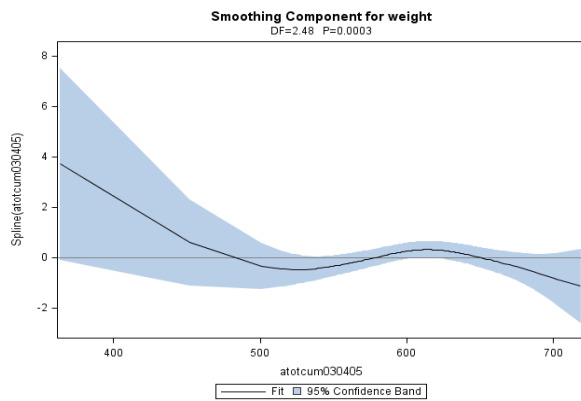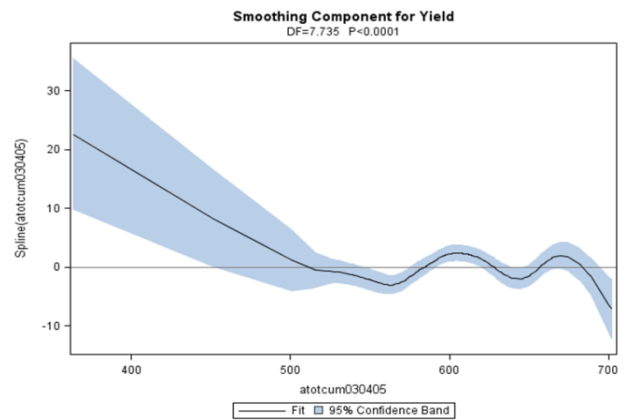

Figure 12. Plot Weight and Yield   Figure 13. Plot Protein and Solar Radiation
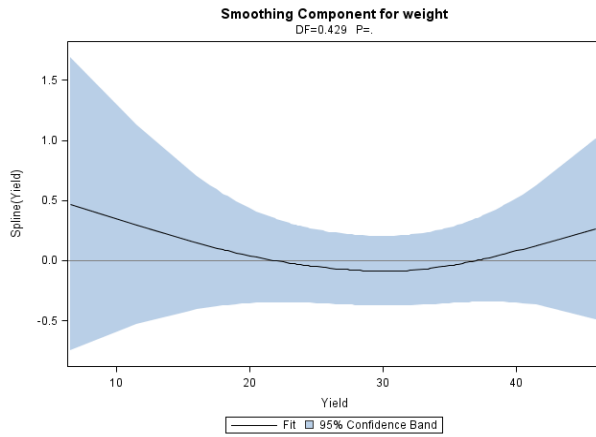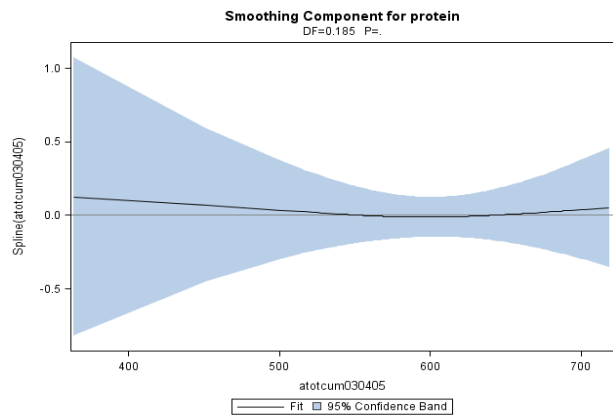
Figure 14. Plot Weight and Solar Radiation   Figure 15. Plot Yield and Solar Radiation

Figure 16. Calculated Relationship between Yield and Protein and Test Weight Using Elasticity