SPATIAL PATTERNS IN HARD RED WINTER WHEAT QUALITY AND BASIS

By

YIKUAN CHEN

Bachelor of Science in Agribusiness

Oklahoma State University

Stillwater, OK

2019

Bachelor of Science in Statistics

Oklahoma State University

Stillwater, OK

2020

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
July, 2021

SPATIAL PATTERNS IN HARD RED WINTER
WHEAT QUALITY AND BASIS

Thesis Approved:

Dr. B. Wade Brorsen

Thesis Adviser

Dr. Dayton M. Lambert

Dr. Brian Whitacre

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Dr.Brorsen. Being your student is a little hard, but I learned so many things under your guidance and push. I could not push myself and study so many things without you. Your courses are challenging and sometimes it is difficult for me to understand your course material. It takes me a lot of time to catch up other students in your class, and I learned many things through your courses.

Thank you Davood for helping me with Dr.Brorsen's homework and my thesis. You are an expert in statistics and math.

Thank you, Dr. Lambert, for giving me comments on improving the idea and writing the right words in my research.

I also want to express my thanks to professors and faculties in the Department of Agricultural Economics. I could not get this far without your help and support. I will always remember the 4 years I spent here!

In the end, I want to acknowledge the support from my family and my girlfriend. Without their support, I could not have gone this far and keep moving forward.

Name: YIKUAN CHEN

Date of Degree: JULY, 2021

Title of Study: SPATIAL PATTERNS IN HARD RED WINTER WHEAT QUALITY AND BASIS

Major Field: AGRICULTURAL ECONOMICS

Abstract: Chapter I:

Hard red winter wheat is broadly grown in Great Plains states. Knowing how wheat quality is correlated across space may help wheat buyers know where to find wheat with the characteristics they need. The mean of wheat quality characteristics is estimated for each location first. A variogram is used to represent the spatial correlation of these local expected values. Variograms are also estimated for the residuals by year. Such information could be used to determine how large an area a wheat sample represents. The result shows that expected local wheat quality has a strong spatial correlation even over a large distance. The spatial correlation of residuals changes across years. Thus, the conclusion is that having a survey each year provides new information to wheat buyers. The results can explain why Plains Grains Inc., who provided the data, conducts a wheat quality survey every year.

Chapter II:

This chapter looks at the spatial patterns of hard red winter wheat protein and basis. Additionally, a hedonic model between wheat basis and protein is built to determine if the protein premium varies across space. The spatial regression models are estimated using Bayesian Kriging so that coefficients can vary across space. The theoretical variogram model is fitted in the covariance matrix so we can quantify the spatial variation. Local basis and protein premium are highly correlated across space, and protein premium changes relatively large every year. The hedonic model shows that protein premiums are largest in the western part of the Southern Great Plains. The Pacific Northwest region shows no protein premium, which is presumably because protein premiums are paid directly through price in these areas.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

SPATIAL PATTERNS IN HARD RED WINTER WHEAT QUALITY

1. Introduction

Hard red winter wheat (HRWW) can be used to make bread-use flour and flour suitable for Asian noodles, hard rolls, and flatbreads. HRWW quality information is difficult for millers to acquire because elevators do only limited wheat quality tests due to the time and cost (Regnier 2004). Information about wheat quality is valuable to wheat buyers. Wheat characteristics can directly influence the quality of flour and end products. For example, protein, or gluten, in flour is the key factor to make dough sticky (MacRitchie 1987; Veraverbeke et al. 2002) and gives a smooth texture to the bread. Test weight and percentage of large kernels can give millers an approximation of flour yield. Moisture is an indicator of storage time length. Espinosa and Goodwin (1991) and Roberts (2020) have estimated hedonic models of wheat prices and selected wheat characteristics. Espinosa and Goodwin used Kansas wheat price and quality data and found that wheat buyers and end-users use characteristics other than grading factors to price the wheat. Roberts extended the previous work into the entire HRWW growing region. Plains Grains Inc. (PGI) was formed in 2004 to provide accurate wheat quality data to domestic and international wheat buyers. Thus, this previous research suggests PGI is providing useful information to wheat buyers.

Studies have used various methods to examine the factors affecting wheat quality. Factors that can determine wheat quality includes the wheat producer's choice of wheat variety (Barkley & Porter 1996; Lambert et al. 2003); the weather and soil conditions (Gooding et al. 2003; Johansson et al. 2008; Lee et al. 2013); and the nitrogen level applied in the field (Erekul et al. 2006; Bongiovanni et al. 2007; Zecevic 2010; Meyer-Aurich et al. 2010). Lee et al. utilized a spatial lag model studying the influence of temperature and rainfall on wheat protein and test weight. The spatial lag model improves the model's prediction ability. Meyer-Aurich et al. applied a spatial error model to estimate the wheat yield and protein response to nitrogen applied. Their work provides the support to consider the spatial effect when wheat qualities are involved. However, there are several ways to calculate the weight matrix, so the model's performance partially depends on the weight matrix. A spatial autoregressive model also cannot tell how variables are correlated across space very accurately.

The variogram is used in this work to examine the spatial variation in wheat characteristics. It can give simple results to understand how wheat quality varies across space. PGI breaks the selected wheat characteristics into 3 groups, wheat grading characteristics, kernel quality characteristics, and other wheat characteristics. The grading characteristics include dockage, test weight, damaged kernels, shrunken and broken kernels, and foreign material (PGI 2020). Kernel quality characteristics include total defects, kernel size, thousand kernel weight, and kernel diameter. Other wheat characteristics include protein, ash content, falling number, moisture, and kernel hardness. The main objective in this work is to find if it is necessary to have wheat sampling annually and does PGI provides new information to the market every year. Two specific problems are i) how is local expected wheat quality distributed across space, and ii) how is error distributed across space in each year.

To find out the answers to the questions above, a mixed effect regression model is used in this work. Regression with dummy location variables and year random effects is used to find the mean value of wheat characteristics for each location. Then, variograms are estimated for the means at each location and the residuals for each year. The exponential model and linear model are used for the theoretical variogram. In this way, the spatial variation and correlation in local mean quality value and residuals are estimated separately. The long-term means of each location could be useful to wheat buyers who are contracting purchases in advance of harvest or early in the harvest while quality information is limited. It also provides evidence supporting PGI's surveying work every year.

2. Materials and Methods

2.1 Data

The wheat quality data are PGI's unpublished data. PGI collects wheat samples, tests wheat qualities, and provides this information to the wheat buyers every year. The sampled locations are focused on Great Plains states including Colorado, Kansas, Montana, Nebraska, North Dakota, Oklahoma, South Dakota, Texas, Wyoming, and the Pacific Northwest regions including Oregon, Washington, and Idaho. In the Great Plains States, PGI samples the individual elevators. In the Pacific Northwest, the samples are from regional elevators and so represent a larger area. Local elevators typically sample each load of wheat and after the sample is tested, the wheat is placed into a dump barrel. The PGI representative takes a probe sample from the dump barrel. The samples are sent to the USDA ARS Hard Winter Wheat Quality Lab in Manhattan, Kansas (PGI 2020) to test wheat quality. The grade of sampled wheat is officially declared by the Federal Grain Inspection Service office in Enid, Oklahoma.

The original dataset has 4032 observations. It contains 49 observations of other wheat varieties or missing values in wheat class, 187 observations in the Pacific Northwest region, and 671 observations with missing coordinates. After clearing those observations, the dataset used in this work has 3249 observations. Also, there are 159 locations sampled only once in 8 years, so they were removed to avoid a perfect fit of regression. After removing those observations, the dataset has 3090 observations which cover 8 years from 2012 to 2019 and 430 locations in total. Noted that some quality data are missing for some locations in a certain year, so the observation number may differ according to the wheat quality characteristic. The dataset is also unbalanced within-year and across-year. In each year, the number of sampled elevators in each city differs. Some locations got repeated measures. For example, in 2012, 62 out of 305 sampled locations had more than one sampled elevator. The number of sampled locations also differs across years, and sampled locations change by year.

Latitude and longitude are measured based on city level in the world geodetic system 1984 (WGS84). Single Kernel Characterization System (SKCS) is used to measure kernel hardness and diameters. Other wheat characteristics this work focuses on are dockage, test weight, damaged kernels, shrunken & broken kernels, foreign material, total defects, kernel size large, medium, and small, thousand kernel weight, protein, individual wheat ash, falling number, and moisture. Protein and individual wheat ash measurements are on a 12% moisture basis.

2.2 Mixed Effect Model and Variogram

Since the data involves both spatial and time effects, spatial autocorrelation could come from local values and residuals. A mixed-effect regression model is used to estimate the mean of wheat quality for each location. The model is:

$$y_{it} = \beta_1 + \sum_{j=2}^{L} \beta_j I(location_{it} = locationID_j) + \gamma_t + u_{it} \quad (1)$$

where y_{it} is the i^{th} observation in year t of a wheat characteristic, β's are regression coefficients, $I(location_{it} = locationID_j)$ is an indicator function that equals 1 if the location of the i^{th} observation in year t is equal to the j^{th} location ID and 0 otherwise, $locationID$ is an alphabetically sorted list of cities, $\gamma_t \sim N(0, \sigma_\gamma^2)$ is the random year effect in year t and u_{it} is the error term. Such a mixed model can be estimated using *lme()* in package *nlme* to add random year effects. Restricted maximum likelihood (ReML) is used for the method option. In this way, a block is formed by year (Zhang, 2015). Local wheat quality value, $\hat{y}_j$, can be predicted by:

$$\hat{y}_j = E\left(y \middle| locationID_j\right) = \begin{cases} \beta_1 & , j = 1 \\ \beta_1 + \beta_j & , j = 2, \dots, L \end{cases} \tag{2}$$

and local mean wheat quality $\hat{y}$ is obtained. Then, we need to match the local mean values and coordinates based on city names, and transform such dataset into a spatial object in R by function *coordinates()* in *gstat* package. A projection of longitude and latitude system must be stated by *peoj4string()* to make sure the distance calculated in *variogram()* is the distances on the earth surface.

A variogram measures similarity based on distances (Isaaks 1989). The function for empirical semi-variogram based on data $\{\hat{y}_j, i = 1, \dots, L\}$ is:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{(i,j):h_{ij}=h} \left(\hat{y}_i - \hat{y}_j\right)^2 \tag{3}$$

where $\gamma(h)$ is the estimated semi-variogram at distance h, $n(h)$ is the number of point pairs with distance h, and $\hat{y}_i$ and $\hat{y}_j$ are pair of points at location i and j with distance h. If h_{ij} is small, which means distance is close, then points should be similar to each other, and correspondingly, semi-variogram should be small. The function *variogram()* returns the number of point pairs, distance, and estimated semi-variogram by distance, which are used to estimate a parametric model for the variogram. These values are used to estimate the theoretical variogram. Option width can be used to control the distance intervals. Smaller distance intervals will provide more semi-variograms, but it does not always provide a better plot showing clear patterns of the theoretical variogram. So, the default settings are used since it works well. Under default settings, width is set by the diagonal length of a box that covers all the data spatially divided by 45. In this way, the distances h where semi-variograms are calculated are determined.

Theoretically, a semi-variogram will keep increasing as distance increases. When distance reaches a point, differences between points reach the maximum, and the semi-variogram will reach a plateau and stay there. The exponential variogram model is fitted for wheat dockage, foreign material, and protein. Given the parameters vector $\boldsymbol{\lambda} = \{c_0, c_e, a_e\}$, the exponential model is:

$$(4) \qquad \gamma^*(h|\boldsymbol{\lambda}) = c_0 + c_e(1 - \exp(-h/a_e))$$

where c_0 is the nugget effect, c_e is the partial sill, and a_e is the range.

The empirical variogram of other wheat characteristics never reaches the plateau before 1000 km, and the scatterplot suggests that linear variogram model. We assume that all the linear models will reach the plateau at 1000 km, so given the parameters vector $\boldsymbol{\lambda} = \{c_0, c_1\}$, a linear variogram model is:

$$(5) \qquad \gamma^*(h|\boldsymbol{\lambda}) = \begin{cases} c_0 + c_1 * h, & if\ h \leq 1000 \\ c_0 + 1000c_1, & if\ h > 1000 \end{cases}$$

where c_0 is the nugget effect, and c_1 is the slope.

To estimate the spatial parameters, the weight in least-squares minimization is used:

$$(6) \qquad \min_{\boldsymbol{\lambda}} \sum_{k=1}^{K} n(h_k)\{\gamma(h_k) - \gamma^*(h_k|\boldsymbol{\lambda})\}^2$$

For wheat dockage, foreign material, and protein, such process can be done by function *fit.variogram()* in *gstat*. The model option is set as an exponential model, and fit.method option is set to be 1 so that $n(h)$, the number of point pairs at h distance, is used as the weight in least-squares minimization. Note that *fit.variogram()* only iterates 200 times, so repeating this function using the previous estimation of parameters is necessary to provide better results even though it may not converge at the end. Thus, a general process of estimating parameters in an exponential model is:

i) input initial value $c_0^{(k)}, c_e^{(k)}, a_e^{(k)}, k = 0$

ii) run *fit.variogram()*, obtain results $\widehat{c_0}, \widehat{c_e}, \widehat{a_e}$

iii) $k = k + 1$, let $c_0^{(k)} = \widehat{c_0}, c_e^{(k)} = \widehat{c_e}, a_e^{(k)} = \widehat{a_e}$

iv) repeat step ii) and iii) through $k = 100$

v) output $c_0^{(100)}, c_e^{(100)}, a_e^{(100)}$

For the rest of the wheat characteristics, a linear variogram model can be estimated by function *lm()* in *stats* which is a basic package in R. We can treat semi-variogram as the dependent variable, distance as the independent variable, and the number of point pairs as the weight. Partial sill effect is calculated as $1000c_1$. The variograms for residuals of each wheat characteristic in each year are set to be the exponential model. The estimating process is the same with fitting variogram model for dockage, foreign material, and protein.

3. Results

Range estimates the maximum distance that 2 points are still correlated. For dockage, foreign material, and protein, the estimated range parameters are all below 1000 km (Table 1-2). For foreign material and protein, the fitted variogram is increasing up to 600 km. The variogram of dockage reaches the sill after 175 km which is relatively smaller than the other two. This implies that spatial correlation in dockage decays very fast as distance increases. The plots of the empirical variogram and the fitted variogram for each wheat characteristic are shown in Figure 1-1 and Figure 1-2. Since we have assumed that the linear variogram model for the rest of wheat characteristics reaches the plateau at 1000 km, samples of those wheat characteristics can represent a large area. This result implies that PGI has sampled more than enough locations if their goal is to determine the averages of wheat characteristics over time.

Nugget represents the variability of the data at 0 distance. Measurement errors could result in a relatively large nugget. If the nugget effect is close to the partial sill or bigger than the partial sill, then it implies too much spatial noise or not enough spatial correlation. For foreign material, test weight, shrunken and broken kernels, total defects, the estimated nugget effects are bigger than the partial sill effects (Table 1-2, Table 1-3). The fitted variograms of these wheat characteristics are also more horizontal than others. It indicates that there is not much spatial correlation in these wheat characteristics. SKCS average diameter and individual ash have almost 0 nugget effects and partial sill effects. It means that they probably do not have spatial correlations at all. Other wheat characteristics' nugget effect is smaller than the partial sill effect even though sometimes the difference is small. This implies those wheat characteristics got some spatial correlations. Dockage, foreign material, and protein also have a much smaller estimated distance than other wheat characteristics. So, it implies that the area to have a spatial correlation of these three characteristics is much smaller than others. Due to the estimated distance of 175.39 km, nugget effect of 0.03, and partial sill of 0.05, dockage is locally correlated for sure.

For each wheat characteristic, 8 fitted variograms are estimated by year. They are plotted together to see how the variation changes across years (Figure 1-3). The spatial variation for most wheat characteristics changes in a big way. Individual ash and test weight have a sudden huge change on nugget or partial sill effect in one year. The other 7 fitted lines are almost identical, which implies that there could be a measurement error. The variograms of moisture seem more stable than others because the fitted lines have a similar shape. It implies that the variability of error term for moisture in each year stays relatively stable than others. Other wheat characteristics' fitted variograms do not have a concentrated area as moisture does. For example, the residuals of protein got 2 years with an estimated distance under 400 km while other 6 years over 1000 km. The local variation of the residuals of protein also changes across the year by different nugget levels. Based on these plots, we can conclude that even though error terms have a strong spatial correlation, the variation and the distance to correlate with are changing dramatically across years.

Many factors could make wheat quality deviate from its expected value such as weather, fertilizers, yields, and producer's choices on sub-varieties in HRWW. The variogram is a simple method to extract and summarize the information contained in data. In this work, we find that both expected local wheat quality and error term in each year show a strong correlation. It is good for wheat buyers and PGI researchers. More importantly, how error term is correlated across space changes across the year in a big way. This indicates that wheat buyers cannot simply use past data to predict the current year's wheat quality. Because every year could have a huge change in error term's spatial distribution, and no one can be sure before the quality tests are done. PGI does provide unique, important, and meaningful wheat quality information to the market and PGI needs to do sampling each year.

4. Discussion

In this work, we find that local mean wheat quality and residuals in sampled years have strong spatial correlation. This work may play a part in helping wheat buyers find where to buy wheat with the qualities they seek. Wheat buyers are targeting different niche markets, so sometimes they are very specific about certain quality levels of HRWW. Knowing how the local wheat quality and error distributed in space could give wheat buyers more choices. The result suggests that research that involves wheat qualities may need to consider spatial correlation in them. Omitting the spatial correlation in wheat quality would be a strong restriction in the study of wheat in the future. The results could also be used as prior information for spatial study using Bayesian framework since spatial model under Bayesian framework usually requires a theoretical variogram model and prior probabilities on parameters.

Future improvement that can be done based on this work comes from 3 aspects. The first aspect is to estimate the covariance matrix of local value and error term in a single step. One possible method is through Bayesian Kriging. It is a Gaussian Process, and we can directly set up the covariance matrix into the exponential model. The second aspect is how to improve the optimization speed and overall results. Cressie (1985) provides a better objective function for the weighted least squares method. However, it seems the precision of R would fail in some cases. Thus, another language with higher numerical precision could be a solution to this problem. The third aspect is that some wheat characteristics are not normally distributed. Normalizing it may lead to a better result.

CHAPTER II

QUALITY-ADJUSTED WHEAT BASIS WITH SPATIALLY VARYING PARAMETERS

1. Introduction

There are six wheat types grown in the United States. Hard red winter wheat (HRWW) accounts for about 40% of the wheat production in the U.S., and it is a major export grain of the United States. The objective of this article is to determine how protein and protein premium of hard red winter wheat vary across space. Such information should interest millers, food businesses, and wheat researchers.

HRWW is widely grown in the Great Plains states and northern states. Previous study shows that the performance of wheat flour for bread making depends on gluten (MacRitchie, 1987) because gluten is a key factor to make flour dough sticky and elastic so that bakers can make them into many shapes (Veraverbeke et al. 2002). Higher gluten levels are correlated with stronger dough that will produce chewy bread. The protein level in wheat can provide a rough estimate of the gluten level in the flour made from that wheat. Millers often blend some high protein wheat to reach the high gluten level standard. In Southern Great Plains states, hard red winter wheat (HRWW) generally has a medium to high protein content, making it suitable for all-purpose flour. However, such information is difficult for millers to acquire (Regnier et al. 2004).

Thus, Plains Grain Inc. (PGI) was created to provide information about local wheat quality to the market.

Protein premium has been well examined by many studies (Bale and Ryan, 1977; Espinosa and Goodwin, 1991; Lambert and Wilson, 2003; Roberts 2020). Espinosa and Goodwin estimated a hedonic model with grading characteristics and end-use characteristics. Their study showed that protein level significantly affects the wheat price and milling and baking characteristics. Roberts continued Goodwin's work and expanded the research area to the entire HRWW growing region. He found that premium for end-use characteristics is paid through local basis. Their work assumes that the protein premium is universal constant across space. But the demand for wheat protein may not be uniformly distributed across space. Further, northern states often conduct protein tests and pay protein premiums while southern states do not. So, there must be differences in valuing the wheat protein in basis by location.

To let the protein premiums vary across space, we utilized Bayesian Kriging to estimate the hedonic model between wheat basis and protein. Bayesian Kriging is a spatial smoothing method. Such a technique is a special case of Spatially Varying Coefficient Process (SVCP). It is like geographically weighted regression (GWR) except that Bayesian Kriging sets certain prior probability density functions on spatial parameters where a typical GWR estimates spatial parameters by optimization or assumption. Handcock and Stein (1993) provide an early explanation of Bayesian Kriging. Gelfand et al. (2003) applied SVCP in the study of house prices against some housing characteristics. Their work provided suggestions on handling spatial-temporal data. Wheeler and Calder (2007) compared two major spatial models, GWR and SVCP. They used simulated data and concluded that SVCP provides more flexibility in modeling spatial relationships and gives easier interpretable and accurate results than GWR. Cho (2017) used Bayesian Kriging to estimate mean hay yield for counties in Oklahoma. Park, Brorsen, and Harri (2018) provide a recent application to crop insurance rating. They used the Bayesian Kriging method to smooth the crop yield distributions across counties. The model used here goes beyond these previous works by letting regression parameters vary across space.

This paper addresses two specific questions about wheat protein and basis information in the Great Plains. The first goal is to determine how wheat protein samples are correlated across space. Near places should have similar soil, weather, and, maybe, growing techniques. So, it is reasonable to assume wheat protein in near places should not differ dramatically. Thus, there could be a sample point that might represent its near locations. The second question is how premiums vary across space. Studies have shown there is protein premium and we want to know how it is distributed across space. Low protein wheat is often fed to cattle rather than used for flour and thus its price is expected to be discounted. If protein premium does not vary across location in an obvious difference, then there could be no incentive for farmers to grow high-quality wheat.

To use Bayesian Kriging, a theoretical spatial covariance model is assumed for the local mean values and prior probabilities are set on the spatial parameters. By assuming that the local mean value follows a multivariate normal distribution with a prior mean vector and covariance matrix, a posterior distribution is formed by Hamiltonian Monte Carlo methods (Carpenter 2017). The posterior mean is used as the final estimation for each parameter.

2. Data

2.1 Data Originality

The quality data are the unpublished data provided by PGI. The data set not only has a wheat grade for sampling location but also includes other quality characteristics like protein, moisture, and dockage rate. The original dataset has about 4031 observations over 8 years period from 2012 to 2019. It contains several observations in the Pacific Northwest (PNW) region and for other wheat varieties. Those observations were removed. The number of samples and sampled locations vary in each year in the rest of the data. It is an unbalanced panel dataset.

The sampling method is probe sampling. It is the only effective way of sampling from rest containers by USDA (2016). The probe for grain sampling is a stick with a pointed head and some slots in it. A worker needs to push the probe into the grain bulk, let the slots be filled, then draw it out and empty the slots.

Latitude and longitude are measured based on city level in the world geodetic system 1984 (WGS84), and they are used to calculate distances. Function *geodist()* in R package *geodist* is used to calculate Haversine distance, so it is the distance across the surface of the earth. In this way, the output distance is in meters, then, divided by 1000 gives results in kilometers. The distance matrix is used to calculate the spatial covariance matrix.

Basis data for hard red winter wheat are collected from the Interactive Crop Basis Tool developed by Kansas State University and DTN corporation. The basis is calculated by subtracting the futures price from the cash price. Cash price is the market price for local businesses to buy flour, and the futures price is the nearby futures contract closest to expiration without going into the delivery month. The basis from Kansas State University is weekly average basis. DTN recorded the bid price and local basis based daily. The timing of harvest differs across locations and thus basis can more directly connect the timing of the quality information and its effect on price. Post-harvest basis is used to better reflect the wheat market reaction to wheat quality. Due to the different harvest time windows in each state, July-1 to December-1 is the period to approximately catch-up post-harvest basis for all locations. The local post-harvest basis is calculated by the average post-harvest basis in each location by year. 2047 DTN basis observations and quality data are matched according to year and zip code since the local elevator’s name does not match sometimes, whereas 23 Kansas State University basis observations and quality data are matched based on the local elevator’s name.

2.2 Data Summary

After we match the basis data and quality data based on elevators' names and years, 2070 observations are left, 324 unique locations are sampled. Most basis is negative which is common. But some positive basis implies either outlier of prices or really good quality of wheat in such location (Table1). Figure 1-2 are plotted to show local average post-harvest basis and local average protein spatially. "basis" is calculated as the average of post-harvest basis in each year by location. The same procedure applies to "protein". Most matched samples concentrate in northwest OK, and KS. Some are in MT, ND, and SD. Northwest states got higher average post-harvest basis than other locations. In Great Plains states such as OK and KS, the average post-harvest basis varies around -0.5$/bu. The protein, on the other hand, is different. Wheat samples in Northern states like MT, ND, and SD tend to have higher protein level than OK and KS. Also, wheat protein level in OK and KS varies a lot. Locations in the west OK, north TX and southwest KS tend to have higher protein level. There is no clear pattern for protein level in KS.

3. Procedure and Software

3.1 Covariance Matrix

Bayesian Kriging requires a spatial covariance model. Let $\boldsymbol{\Sigma}$ denotes a $L \times L$ covariance matrix calculated using the exponential function of distance matrix D, range θ and sill ρ:

(1) $$\Sigma_{ij} = \Psi\left(D_{ij}; \rho, \theta\right) = \rho \exp\left(-\frac{D_{ij}}{\theta}\right)$$

where D_{ij} is the element at i^{th} row and j^{th} column. In matrix form:

(2) $$\Sigma = \rho \begin{bmatrix} 1 & \cdots & e^{-\frac{D_{1L}}{\theta}} \\ \vdots & \ddots & \vdots \\ e^{-\frac{D_{L1}}{\theta}} & \cdots & 1 \end{bmatrix}$$

Distance matrix are calculated by longitude and latitude based on the city. Since every location is sampled at least once in eight years, duplicated coordinates need to be removed. The after-matched data are unchanged. We assume that the local means are correlated even if the observations are in different years.

3.2 Bayesian Kriging and Hedonic Model

In our case, the data involves both spatial effect and time effect. Following the model suggestions provided by Gelfand, the following Bayesian Kriging model is used:

(3)
$$Basis_{lt} = ExpBasis(l) + A(t) + \varepsilon_{lt}$$
$$\boldsymbol{ExpBasis} \sim MVN(\boldsymbol{\mu_1}, \boldsymbol{\Sigma_1})$$
$$\boldsymbol{\Sigma}_1 = \Psi(D; \rho_1, \theta_1)$$

where $Basis_{lt}$ is the basis in location l and year t, $ExpBasis(l)$ is the expected local basis at location l, $A(t)$ is the random year effect, ε_{lt} is the independently and identically distributed error term, $\boldsymbol{ExpBasis}$ is a vector containing all expected local basis, and it follows multivariate Gaussian Process with mean vector $\boldsymbol{\mu_1}$ and covariance matrix $\boldsymbol{\Sigma_1}$, and ρ_1, θ_1 are sill and range for local mean basis.

The same process is used for protein level:

(4)
$$Protein_{lt} = ExpProtein(l) + B(t) + \varepsilon_{lt}$$
$$\boldsymbol{ExpProtein} \sim MVN(\boldsymbol{\mu_2}, \boldsymbol{\Sigma_2})$$
$$\boldsymbol{\Sigma}_2 = \Psi(D; \rho_2, \theta_2)$$

where $Protein_{lt}$ is the wheat protein level in location l and year t, $ExpProtein(l)$ is the expected local wheat protein value, $B(t)$ is the random year effect, ε_{lt} is the independently and identically distributed error term, $\boldsymbol{WheatChar}$ is the vector containing all local wheat characteristic value, and it follows multivariate Gaussian Process with mean vector $\boldsymbol{\mu_2}$ and covariance matrix $\boldsymbol{\Sigma_2}$, ρ_2, θ_2 are sill and range for local mean protein level. The method to calculate the covariance matrix is shown in equations (1) and (2).

In Roberts' study (2020), 50% harvest price and 100% harvest price are used to fit two hedonic models with all wheat quality characteristics. His results show that end-use characteristics affect local HRWW prices. It suggests that quality premiums are paid through local basis rather than direct discounts or premiums to producer prices. Thus, basis is used as the dependent variable. The hedonic model, which is also the first layer in Bayesian Hierarchical structure, is shown as follows:

(5) $$y_{lt} = \big(\beta_0(l) + Z_0(t)\big) + \big(\beta_1(l) + Z_1(t)\big)x_{lt} + \varepsilon_{lt}$$

where y_{lt} is the basis in location l and year t, x_{lt} denotes the protein level in location l and year t, β_0 and β_1 are parameters different across location l, Z_0 and Z_1 denote the random year effects, $\varepsilon_{lt} \sim N(0, \sigma_\varepsilon^2)$ denotes the identically and independently distributed error term across year and location.

The second layer is:

(6) $$\boldsymbol{\beta_0} \sim MVN\big(\boldsymbol{\mu_{\beta_0}}, \boldsymbol{\Sigma_{\beta_0}}\big), \quad \boldsymbol{\Sigma_{\beta_0}} = \Psi\big(D_{ij}; \rho_{\beta_0}, \theta_{\beta_0}\big),$$

$$\boldsymbol{\beta_1} \sim MVN\big(\boldsymbol{\mu_{\beta_1}}, \boldsymbol{\Sigma_{\beta_1}}\big), \quad \boldsymbol{\Sigma_{\beta_1}} = \Psi\big(D_{ij}; \rho_{\beta_1}, \theta_{\beta_1}\big),$$

where $\boldsymbol{\beta_0}$ is the vector of intercept parameters for each location, $\boldsymbol{\beta_0} = [\beta_0(1), \beta_0(2), \dots, \beta_0(L)]$, and is assumed to follow an MVN with mean vector $\boldsymbol{\mu_{\beta_0}}$, and spatial covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_0}$ with distance D_{ij}, range parameters θ_{β_0} and sill parameters ρ_{β_0}, $\boldsymbol{\beta_1}$ is the vector of intercept parameters for each location, $\boldsymbol{\beta_1} = [\beta_1(1), \beta_1(2), \dots, \beta_1(L)]$, and is assumed to follow an MVN with mean vector $\boldsymbol{\mu_{\beta_1}}$, and spatial covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_1}$ with distance D_{ij}, range parameters θ_{β_1} and sill parameters ρ_{β_1}, and D_{ij} is the distance between location i and location j.

The third layer is for spatial parameters θ_{β_0}, ρ_{β_0}, θ_{β_1}, and ρ_{β_1}. Though there is no solid proof to back up the prior choice for these parameters, empirically, inverse gamma (IG) is used as a prior probability:

$$\rho_\beta \sim IG(3,1) \tag{7}$$

$$\theta_\beta \sim IG(3,100)$$

Theoretically uniform prior is the best choice, but empirical practices have shown that the boundary of uniform distribution could change the results dramatically. In our case, using a uniform prior would cause the software to crash. Inverse gamma is used because it has a non-negative domain, which is also the theoretical range of spatial parameters.

Random year effects are set to have a normal distribution with mean 0 and variance 10,000 so it is non-informative. A hard sum-to-zero restriction is placed on the random year effect for better convergence.

3.2 Software

R is used as the processing language, and RStudio is the platform to run it. The package we use is Rstan, which enables the stan framework in R. Stan is a probabilistic programming language dealing with Bayesian modeling and inference (Carpenter et al. 2017). It has flexibility with model type and data structure. The disadvantage of it is speed. As the number of parameters increases, it might take a week or even a month to finish sampling.

4. Results

The parameter sill, ρ, measures the maximum variation that a variable can reach across space. It is like variance so the unit of measure will affect its value. The variation of wheat basis across space is 3.91E-04 (Table 2). This is very close to 0 which indicates that basis is very stable across space. The estimated distance is 2.58E+05 km which is a huge distance. This implies that correlation for basis decays very slow as distance increases to cover the entire Great Plains. On the other hand, the sill of wheat protein is 1.66E+01 which implies that protein level has some variation across space. The estimated distance of protein is 1.01E+04 km which is smaller than basis. So, the spatial correlation in protein decays faster than the basis's spatial correlation. The mean basis in northwest locations is a little higher than other locations (Figure 2-1). While as the protein in north and northwest places also tend to have higher protein wheat than Kansas and Oklahoma (Figure 2-2).

The estimated range for β_0 is 4.49E+02 km and $\beta_1 1.46\mathrm{E}+06$ km (Table 2-2). This is huge distance implies almost every sampled location correlated to each other spatially. Estimated β_0 are bounded between -0.0536 and 0.0437. β_1 ranges from -0.0040 to 0.0081. β_1 is the expected protein premium or discount by location. The first quantile of 0.0004 implies at least 75% of sampled locations have protein premium (Table 2-3). The intercept, β_0, in northwest Oklahoma, western Kansas, and southwest Nebraska locations are lower than other places (Figure 2-3). However, the protein premium, β_1, in those locations are much higher than other places (Figure 2-4). The protein premiums in northern locations are very close to 0 even though their protein level is high. This reflects that they pay protein premiums or discount directly rather than basis. Some places in southern Oklahoma and northern Texas tend to receive protein discounts for high protein levels and protein premium with low protein levels. This suggests that the usage of wheat in those places is probably different from other places.

The random year effect of intercept, Z_0, varies more relative to the year effect on the slope, Z_1 (Table 2-4). But Z_0 has little influence on β_0 with absolute value except in the year 2017. Z_1 has 0 estimates in 3 years. In other years, the value is relatively large to the β_1. In 2012, Z_1 could make about 50% of the locations receive protein discount, while in 2017 it brings a lot more protein premium rather than expected level. It suggests that protein premium can change a lot across years. The distribution of predicted basis and actual basis are very similar (Table 2-5), and the predicted basis and actual basis have a high correlation of 0.96 (Figure 2-5). This result indicates that Bayesian Kriging is a powerful model for extracting spatial information in data.

5. Conclusion

This paper studies the spatial correlation of expected basis and protein level. Using Bayesian Kriging, the assumption that parameters are universal constants for a hedonic model is relaxed. By applying Gelfand's (2003) model suggestions dealing with spatial-temporal data, spatial effect and fixed year effect are separated for basis, protein, and protein premium. Spatial parameters, sill and range, are estimated by setting up the kernel function for the spatial correlation matrix.

The estimated spatial parameters of expected wheat basis and protein show strong spatial correlations. For millers and future researchers who need wheat samples, this result indicates that only a few samples would be enough to represent a large area, and the sampling distance could be increased. The spatial difference is not very huge in general. If using the historical data, multiple years should be considered to average out the year effect. The expected local basis is not very different among locations as well. On the other side, high spatial similarities in wheat protein may imply the same wheat varieties or weather conditions. It is the support of the conclusion by Lambert et al. (2003) that farmers are going to grow wheat varieties with certain characteristics that give them the highest profit.

The hedonic model shows that protein premium exists for most locations, and at least 75% of the places pay protein premiums through basis. The year effect could cause a significant change in protein premium given the absolute value of premium being so small. It implies that even though the wheat market has a wheat premium for most of the locations, adjustment by year has a more significant effect, and sometimes year effect could overturn the protein premium into discount. This result aligns with the historical observation from Bale and Ryan (1977). They claimed that protein premium could be influenced by supply and demand in a certain period, and therefore, none or little protein premium is paid in some years. We also find that protein premiums in Northern states are very close to 0$/kg/%. Since we are using the post-harvest basis, those estimates imply that protein premium in those locations is paid directly by price. The contrast of protein premium and protein levels in north Texas and southern Oklahoma implies their wheat might have different usage.

The improvement in this work is to utilize the Bayesian Kriging method on the wheat basis, protein, and protein premium to separate spatial effect and year effect. Compared to the usual OLS output, Bayesian Kriging explained more of the variation in the data. Also, previous studies showed it would produce better estimation than GWR. However, a drawback of this work is only protein is included in the hedonic model, whereas the research results from Roberts suggest that end-use characteristics affect the price. Further improvement on how to accelerate the sampling speed, or how to simplify the sampling process should be made before adding more variables. Because even though Rstan gives users freedom to make all kinds of models, its running speed is very slow. Another improvement that can be done is to stabilize the posterior estimation of spatial parameters. Fuglstad argued that using inverse gamma is practically good but not stable in some cases. Using a penalized prior could stabilize the estimation. Accomplishing such a task in Rstan could bring a much better result.

REFERENCES

Bale, M. D., and Ryan, M. E. 1977. “Wheat Protein Premiums and Price Differentials.” *American Journal of Agricultural Economics* 59(3): 530-532.

Barkley, A. P., and Porter, L. L. 1996. “The determinants of wheat variety selection in Kansas, 1974 to 1993.” *American Journal of Agricultural Economics* 78(1), 202-211.

Bongiovanni, R. G., Robledo, C. W., & Lambert, D. M. 2007. “Economics of site-specific nitrogen management for protein content in wheat.” *Computers and electronics in agriculture* 58(1), 13-24.

Bivand, RS, Pebesma, E, & Gomez-Rubio, V. 2013. *Applied spatial data analysis with R,* Second edition. NY: Springer. Access at: https://asdar-book.org/.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. 2017. “Stan: A Probabilistic Programming Language.” *Journal of Statistical Software* 76(1): 1-32.

Cho, W. 2019. *RI-PRF Crop Insurance Program Design.* Masters Thesis. Oklahoma, Stillwater: Oklahoma State University.

Cressie, N. 1985. “Fitting variogram models by weighted least squares.” *Journal of the International Association for Mathematical Geology* 17(5), 563-586.

Erekul, O., and Köhn, W. 2006. "Effect of weather and soil conditions on yield components and bread‐making quality of winter wheat (Triticum aestivum L.) and winter Triticale (Triticosecale Wittm.) varieties in North‐East Germany." *Journal of Agronomy and Crop Science* 192(6), 452-464.

Espinosa, J. A., and Goodwin, B. K. 1991. "Hedonic Price Estimation for Kansas Wheat Characteristics." *Western Journal of Agricultural Economics* 16(1):72-85.

Fuglstad, G. A., Simpson, D., Lindgren, F., & Rue, H. 2015. "Interpretable priors for hyperparameters for Gaussian random fields." *arXiv preprint arXiv:1503.00256.*

Gelfand, A. E., Kim, H. J., Sirmans, C. F., & Banerjee, S. 2003. "Spatial Modeling with Spatially Varying Coefficient Processes." *Journal of the American Statistical Association* 98(462): 387-396.

Gooding, M. J., Ellis, R. H., Shewry, P. R., & Schofield, J. D. 2003. "Effects of restricted water availability and increased temperature on the grain filling, drying and quality of winter wheat." *Journal of Cereal Science* 37(3), 295-309.

Handcock, M. S. and Stein, M. L. 1993. "A Bayesian Analysis of Kriging." *Technometrics* 35(4):403-410.

Isaaks, E. H. and Srivastava, R. M. 1989. An Introduction to Applied Geostatistics. New York, New York: Oxford University Press, Inc.

Johansson, E., Prieto‐Linde, M. L., & Gissén, C. 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 the Science of Food and Agriculture* 88(11), 2011-2018.

Lambert, D. K., and Wilson, W. W. 2003. "Valuing varieties with imperfect output quality measurement." *American Journal of Agricultural Economics* 85(1), 95-107.

Lee, B. H., Kenkel, P., & Brorsen, B. W. 2013. "Pre-harvest forecasting of county wheat yield and wheat quality using weather information." *Agricultural and forest meteorology* 168, 26-35.

MacRitchie, F. 1987. "Evaluation of Contributions from Wheat Protein Fractions to Dough Mixing and Breadmaking." *Journal of Cereal Science* 6(3): 259-268.

Meyer-Aurich, A., Griffin, T. W., Herbst, R., Giebel, A., & Muhammad, N. 2010. "Spatial econometric analysis of a field-scale site-specific nitrogen fertilizer experiment on wheat (Triticum aestuvum L.) yield and quality." *Computers and electronics in agriculture* 74(1), 73-79.

Park, E., Brorsen, B. W., & Harri, A. 2018. "Using Bayesian Kriging for Spatial Smoothing in Crop Insurance Rating." *American Journal of Agricultural Economics* 101: 330-351.

Pebesma, EJ. 2004. "Multivariable geostatistics in S: the gstat package." *Computers & Geosciences* 30, 683-691.

Pinheiro, J, Bates, D, DebRoy, S, Sarkar, D, & R Core Team. 2021. *nlme: Linear and Nonlinear Mixed Effects Models.* R package version 3.1-152. Access at: https://CRAN.R-project.org/package=nlme.

Plains Grains, Inc. 2020. "Hard Red Winter Wheat 2020 Regional Quality Survey."

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

Roberts, S. J., 2020. *The Role of Quality Characteristics in Pricing Hard Red Winter Wheat.* Master Thesis. Nebraska, Lincoln: University of Nebraska.

USDA. 2016. "Practical Procedures for Sampling Grain at Farm Sites and Remote Locations." Access at: https://www.ams.usda.gov/sites/default/files/media/PracticalSamplingProcedures2017.pdf.

Veraverbeke, W. S., and Delcour, J. A. 2002. "Wheat Protein Composition and Properties of Wheat Glutenin in Relation to Breadmaking Functionality." *Critical Reviews in Food Science and Nutrition* 42(3): 179-208.

Wilson, W. W., and Dahl, B. L. 1999. "Quality Uncertainty in International Grain Market: Analytical and Competitive Issues." *Review of Agricultural Economics* 21: 209-224.

Wheeler, D. C., and Calder, C. A. 2007. "An Assessment of Coefficient Accuracy in Linear Regression Models with Spatially Varying Coefficients." *Journal of Geographical Systems* 9(2): 145-166.

Zecevic, V., Knezevic, D., Boskovic, J., Micanovic, D., & Dozet, G. 2010. "Effect of nitrogen fertilization on winter wheat quality." *Cereal Research Communications* 38(2), 243-249.

Zhang, X. 2015. "A tutorial on restricted maximum likelihood estimation in linear regression and linear mixed-effects model." Access at: http://statdb1.uos.ac.kr/teaching/multi-grad/ReML.pdf.

Figure 1-4. Map of Local Mean by Wheat Characteristics.

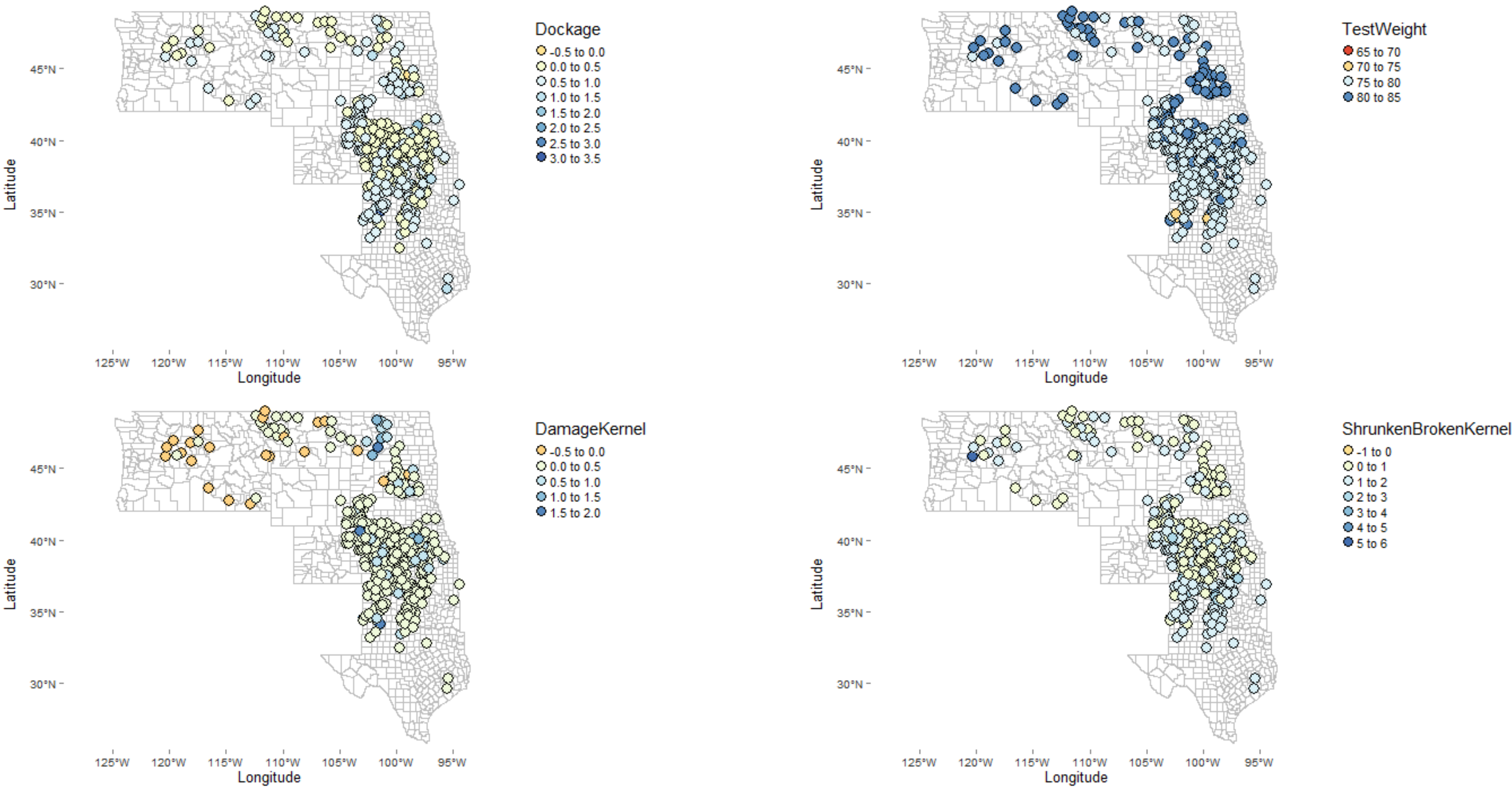


Figure 1-4. Continue.

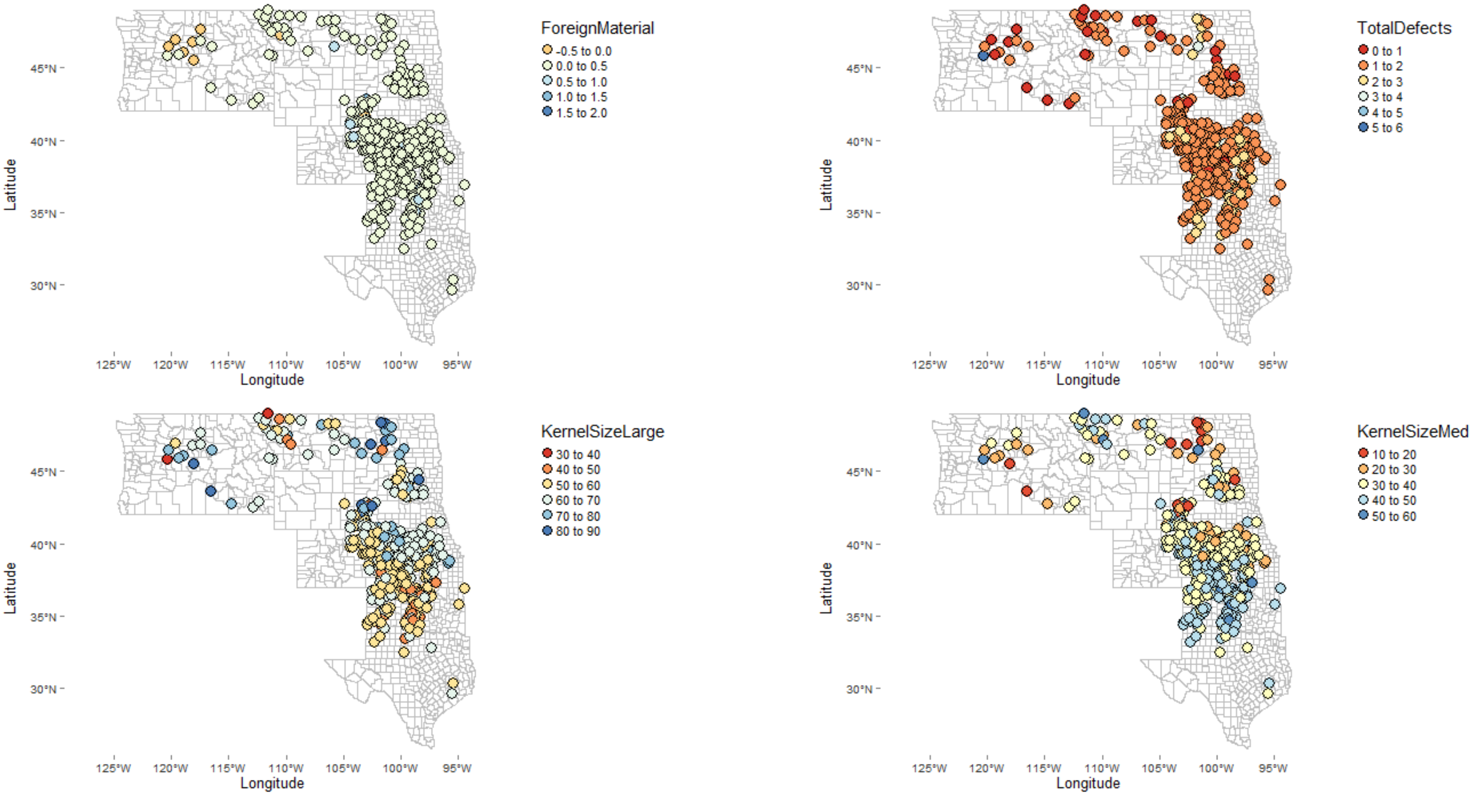


Figure 1-4. Continue.

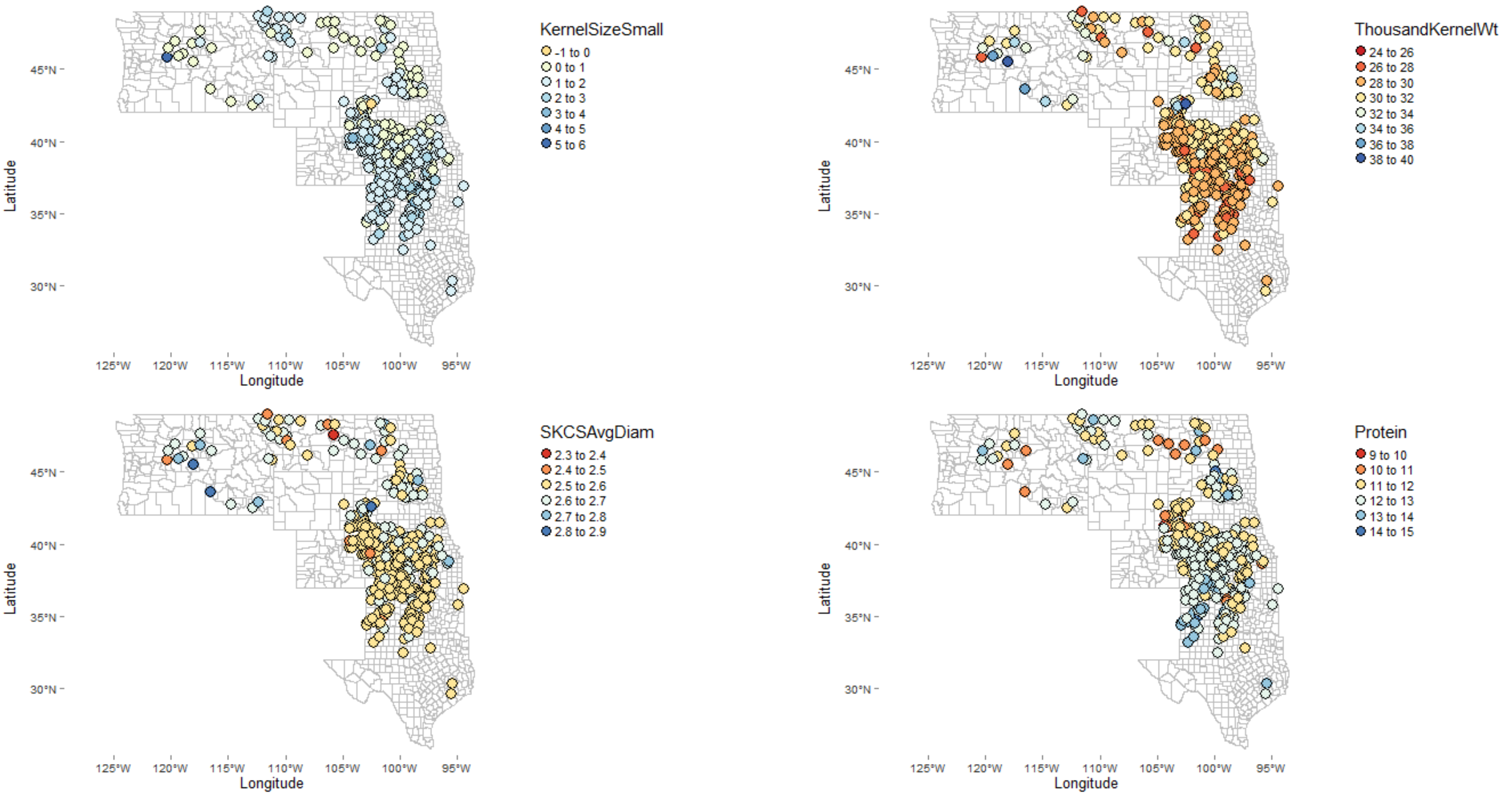


Figure 1-4. Continue.

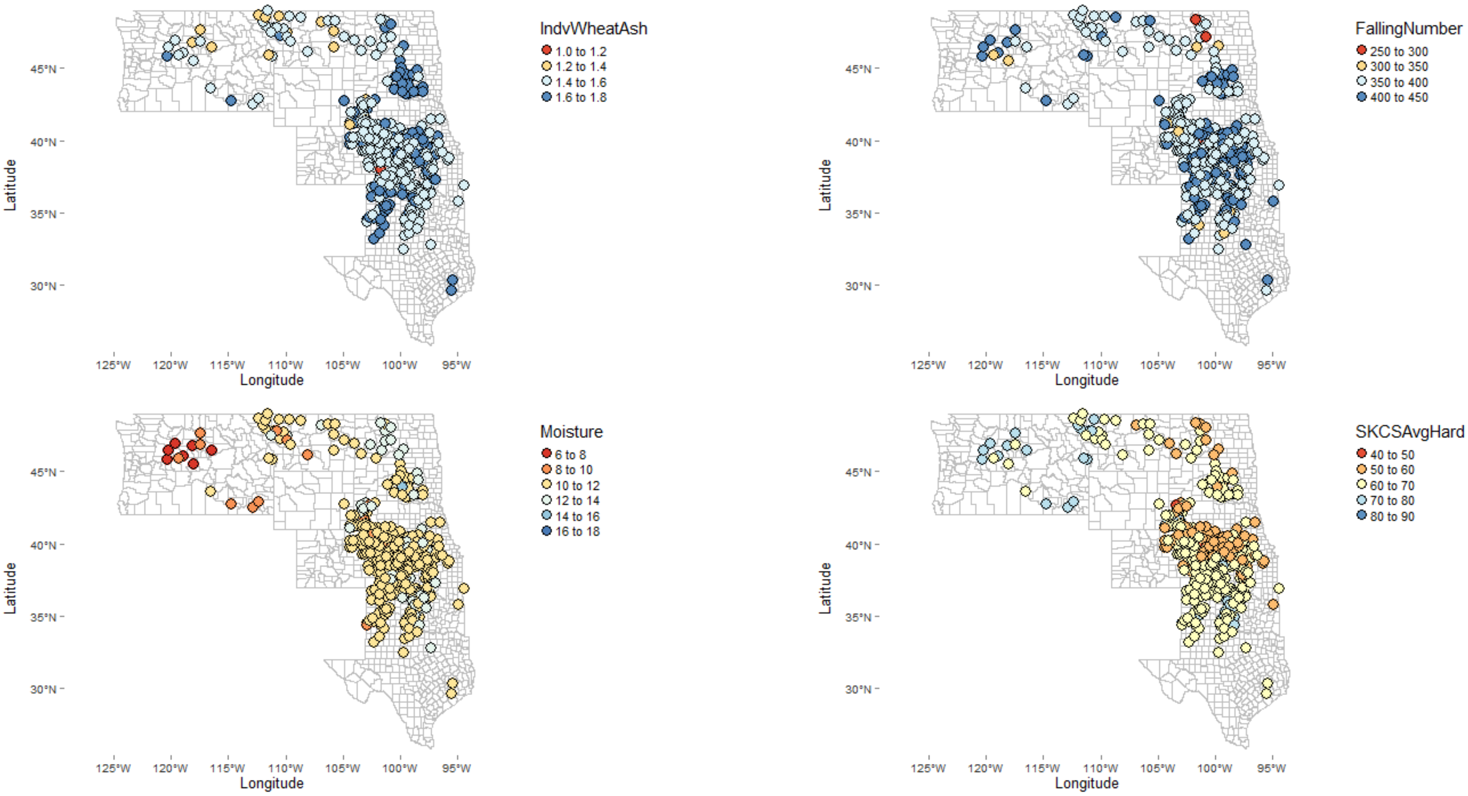


Table 1-1. Summary Statistics of Wheat Characteristics.

	Unit	Count	Mean	SD	Min	Max
Grading Characteristics						
Dockage	%	3132	0.55	0.52	0.00	7.00
Test weight	kg/hl	3133	60.35	1.76	52.10	65.40
Damage kernel	%	3129	0.27	0.50	0.00	10.60
Shrunken & broken kernel	%	3134	1.11	0.80	0.00	13.20
Foreign material	%	3131	0.15	0.28	0.00	6.70
Total defects	%	3122	1.52	0.95	0.10	13.40
Kernel Characteristics						
Kernel size large	%	3072	59.89	16.17	0.35	98.10
Kernel size medium	%	3072	38.71	15.36	1.85	96.35
Kernel size small	%	3072	1.41	1.19	0.00	11.10
Thousand kernel weights	g	3136	29.88	3.44	19.67	47.64
SKCS average diameters	mm	3136	2.58	0.12	2.24	3.21
Other Wheat Characteristics						
Protein	12% mb	3137	12.17	1.40	7.60	17.50
Individual wheat ash	12% mb	3135	1.56	0.13	0.00	2.10
Falling number	sec	3133	391.76	42.10	98.00	598.00
Moisture	%	3137	11.29	1.40	1.23	19.20
SKCS average hardness	-20-120	3072	63.62	9.89	27.33	94.71

Note: SKCS hardness is an index. Below 50 is soft grain. Above 50 is hard grain.

Table 1-2. Estimated Spatial Parameters of Wheat Characteristics with Exponential Model.

	Nugget	Partial Sill	Range
Dockage	0.03	0.05	175.39
Foreign material	0.02	0.01	497.68
Protein	0.19	0.48	531.13

Note: Unit of the range is km. Nugget and partial sill are similar to variance. Its unit depends on the unit of parameters. Squaring the corresponding unit in Table 1-1 will give the unit of nugget and partial sill. For example, the unit for nugget and partial sill of dockage is $\%^2$.

Table 1-3. Estimated Spatial Parameters of Wheat Characteristics with Linear Model.

	Nugget	Partial Sill
Test weight	1.2353	1.1445
Damaged kernels	0.0197	0.0663
Shrunken & broken kernel	0.1178	0.0411
Total defects	0.1608	0.1317
Kernel size large	22.8816	92.1343
Kernel size medium	20.5398	83.0046
Kernel size small	0.1991	0.2808
Thousand kernel weights	0.5033	3.7690
SKCS average diameters	0.0010	0.0032
Individual wheat ash	0.0022	0.0062
Falling number	233.3085	415.9014
Moisture	0.2340	1.0564
SKCS average hardness	12.1769	27.8823

Note: The variogram of these wheat characteristics does not have a plateau before 1000 km. Thus, the range is fixed at 1000 km and partial sill is calculated based on that.

Table 2-1. Summary Statistics of Matched Wheat Protein and Basis.

	Unit	Count	Mean	SD	Min	Max
Basis	$/kg	2070	-0.02	0.01	-0.06	0.04
Protein	%	2070	12.07	1.38	7.70	17.20

Note: Basis comes from Kansas State University and DTN corporation. The average post-harvest basis from July-1 to December-1 is calculated for each location by year.

Table 2-2. Spatial Parameters for Local Basis and Wheat Characteristics

	ρ		θ	
	Mean	SE	Mean	SE
Basis	3.91E-04	0.00E+00	2.58E+05	2.07E+03
Protein	1.66E+01	2.40E-01	1.01E+04	1.79E+02
β_0	6.88E-04	7.35E-03	4.49E+02	1.86E+02
β_1	2.43E-04	2.94E-03	1.46E+06	1.58E+05

Note: Unit of θ is km. ρ is similar to variance. Its unit depends on the unit of parameters. Squaring the corresponding unit in table 4 will give the unit of ρ for basis and wheat quality. ρ for β_1 is $(\$/kg/\%)^2$.

Table 2-3. Summary of Sampled $\boldsymbol{\beta_0}, \boldsymbol{\beta_1}$.

	Unit	Min	1st Qu.	Median	Mean	3rd Qu.	Max
$\boldsymbol{\beta_0}$	$/kg	-0.0536	-0.0404	-0.0349	-0.0334	-0.0301	0.0437
$\boldsymbol{\beta_1}$	$/kg/%	-0.0040	0.0004	0.0011	0.0011	0.0015	0.0081

Note: Bayesian output is sampled distribution, the mean value for each sampled distribution is treated as the estimated value and summarized in this table.

Table 2-4. Statistics of Sampled Year Effect in Hedonic Model.

Year	Z_0		Z_1	
	Mean	SE	Mean	SE
2012	0.0125	0.0033	-0.0011	0.0004
2013	0.0037	0.0033	0.0007	0.0004
2014	-0.0051	0.0033	0.0007	0.0004
2015	0.0037	0.0040	0.0000	0.0004
2016	-0.0048	0.0033	-0.0015	0.0004
2017	-0.0253	0.0033	0.0015	0.0004
2018	0.0070	0.0033	0.0000	0.0004
2019	0.0084	0.0037	0.0000	0.0004

Note: A hard sum-to-zero is placed on both parameters. The values are not sum-to-zero in the table due to the rounding problems.

Table 2-5. Comparison between Predicted Basis and Actual Basis.

	Min	1st Qu.	Median	Mean	3rd Qu.	Max
Predicted	-0.0026	-0.0014	-0.0009	-0.0010	-0.0006	0.0018
Actual	-0.0026	-0.0014	-0.0009	-0.0010	-0.0005	0.0016

Figure 1-1. Plots of Exponential Variogram for Local Mean of Wheat Characteristics.

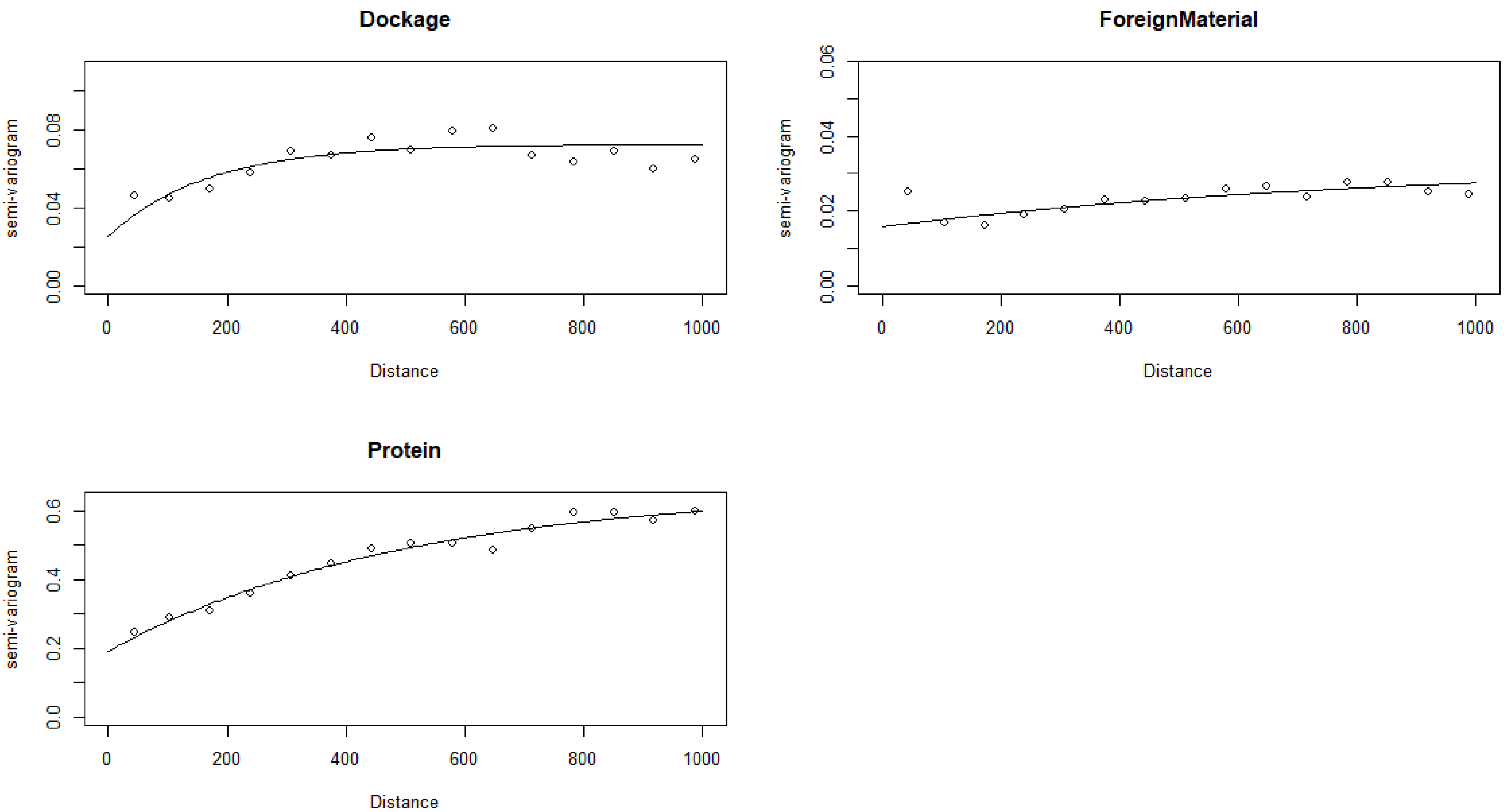


Figure 1-2. Plots of Linear Variogram for Local Mean of Wheat Characteristics.

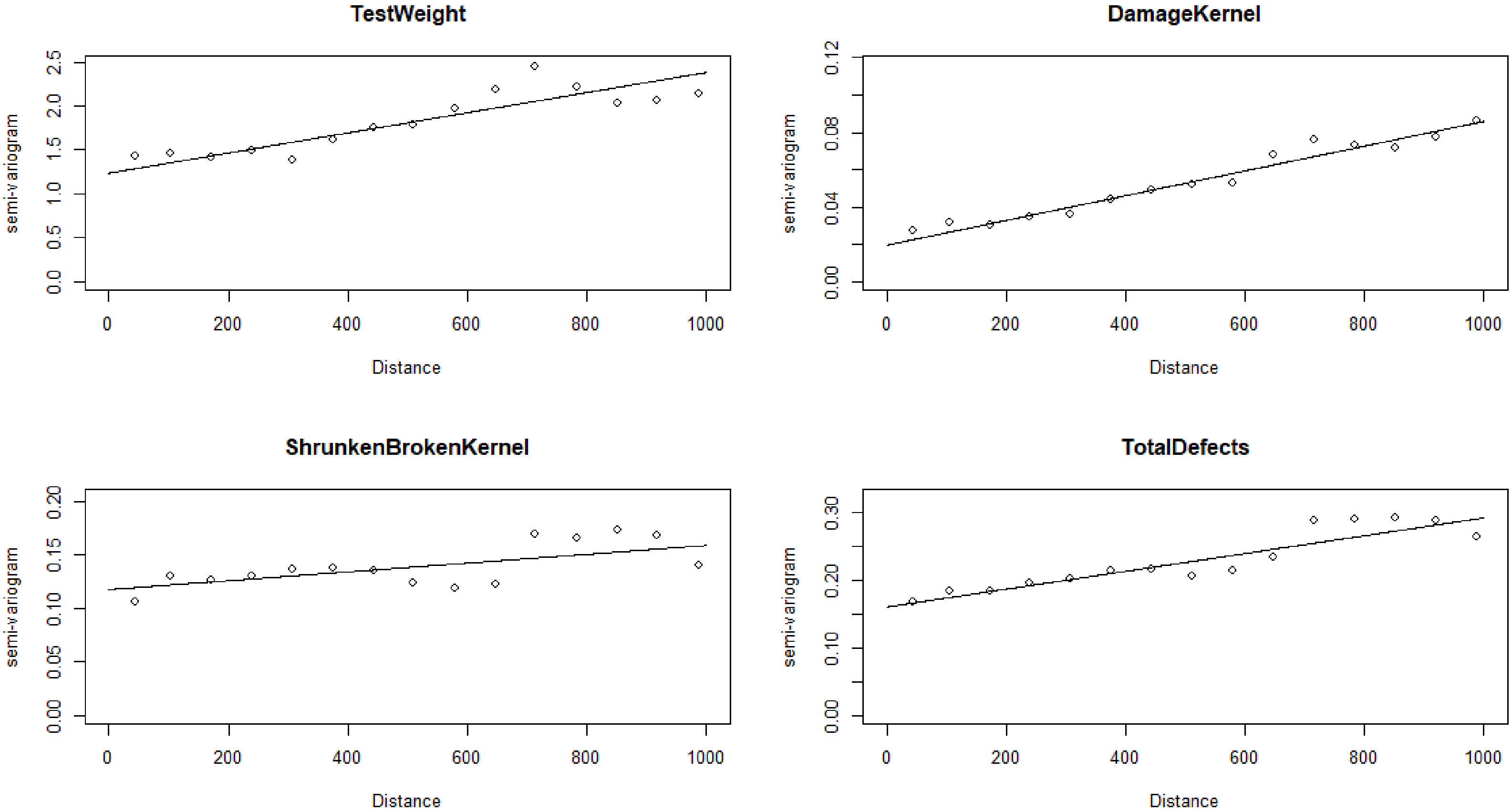


Figure 1-2. Continue.

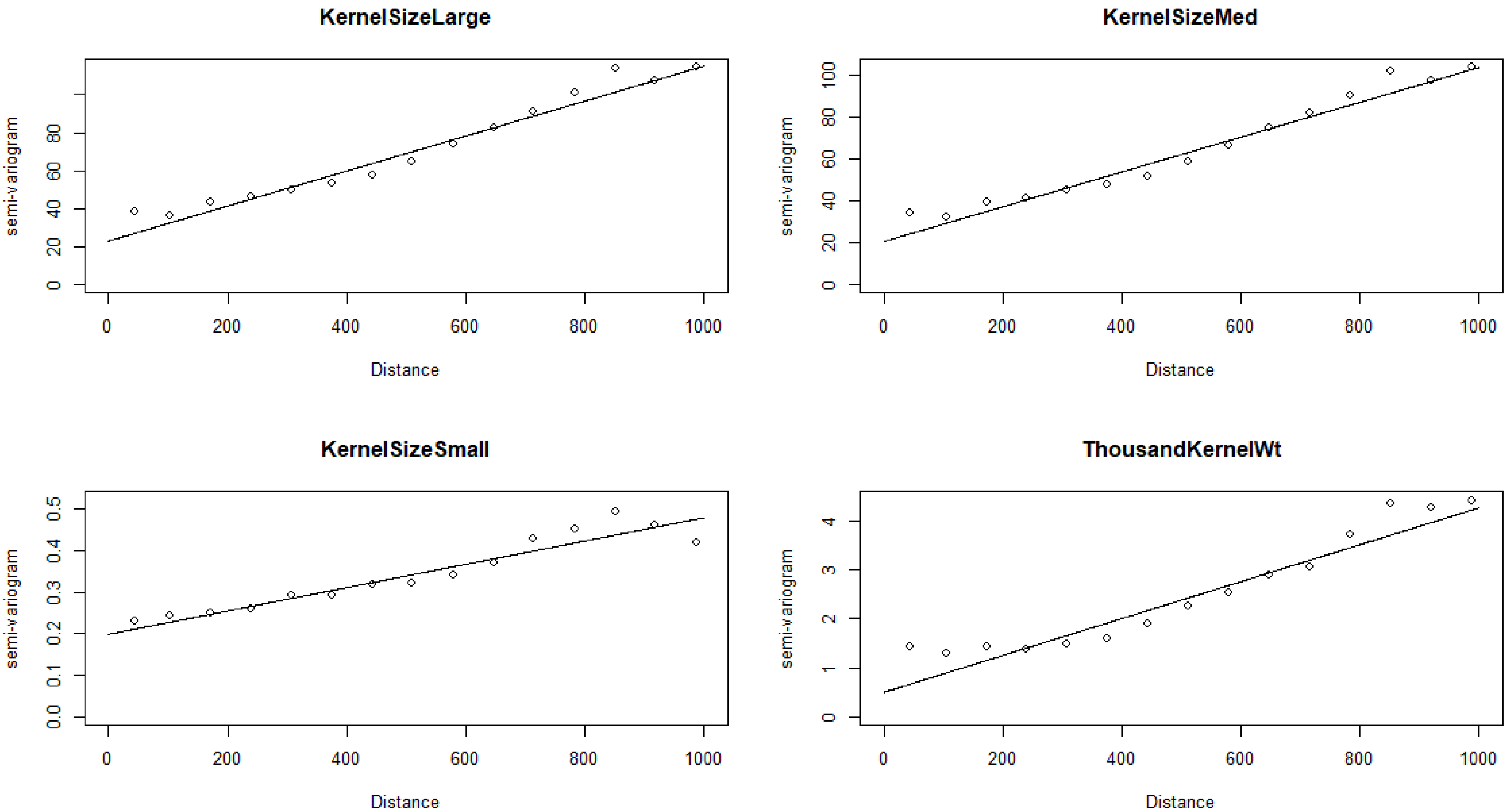


Figure 1-2. Continue

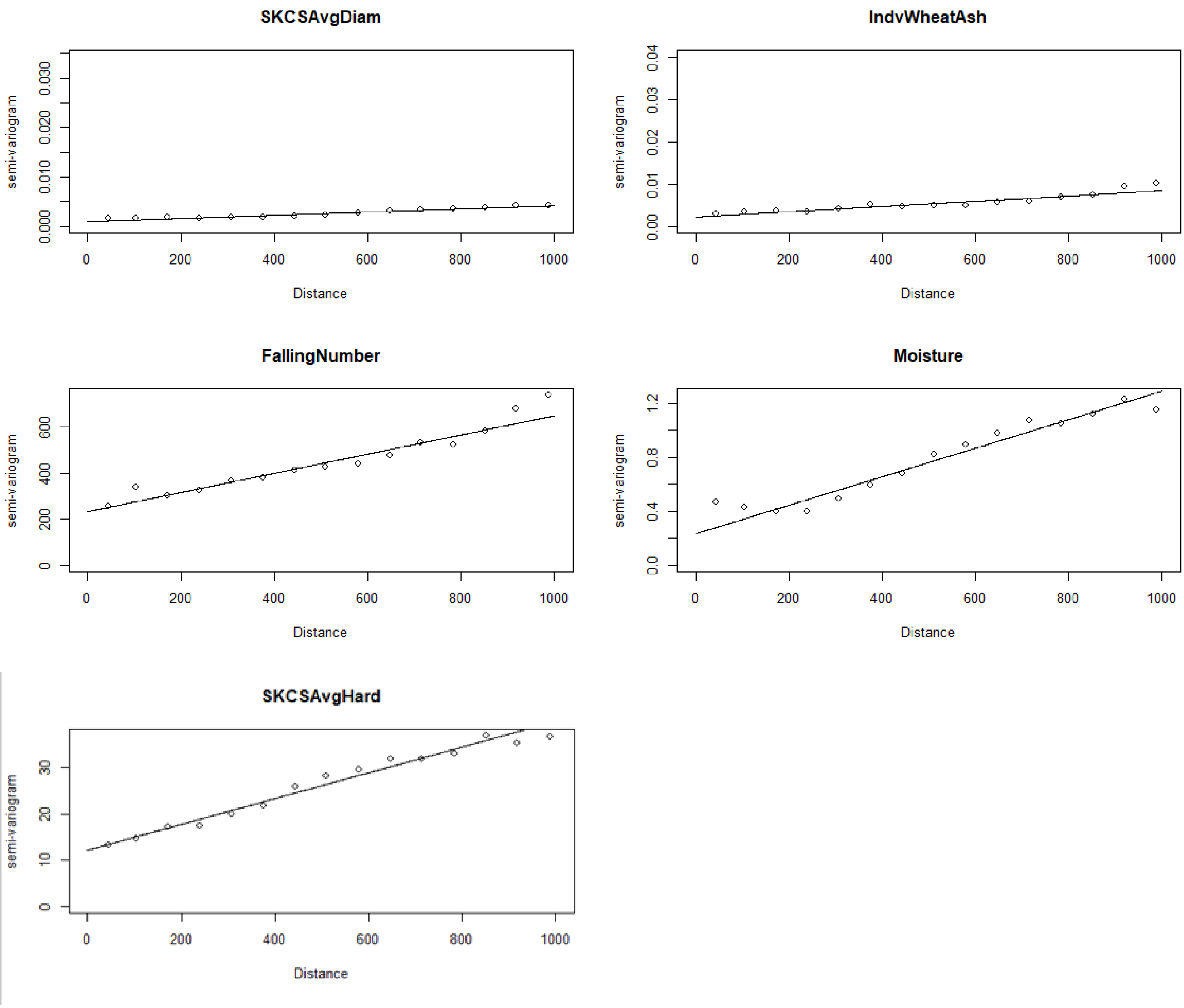


Figure 1-3. Fitted Variograms for Residuals by Characteristics.

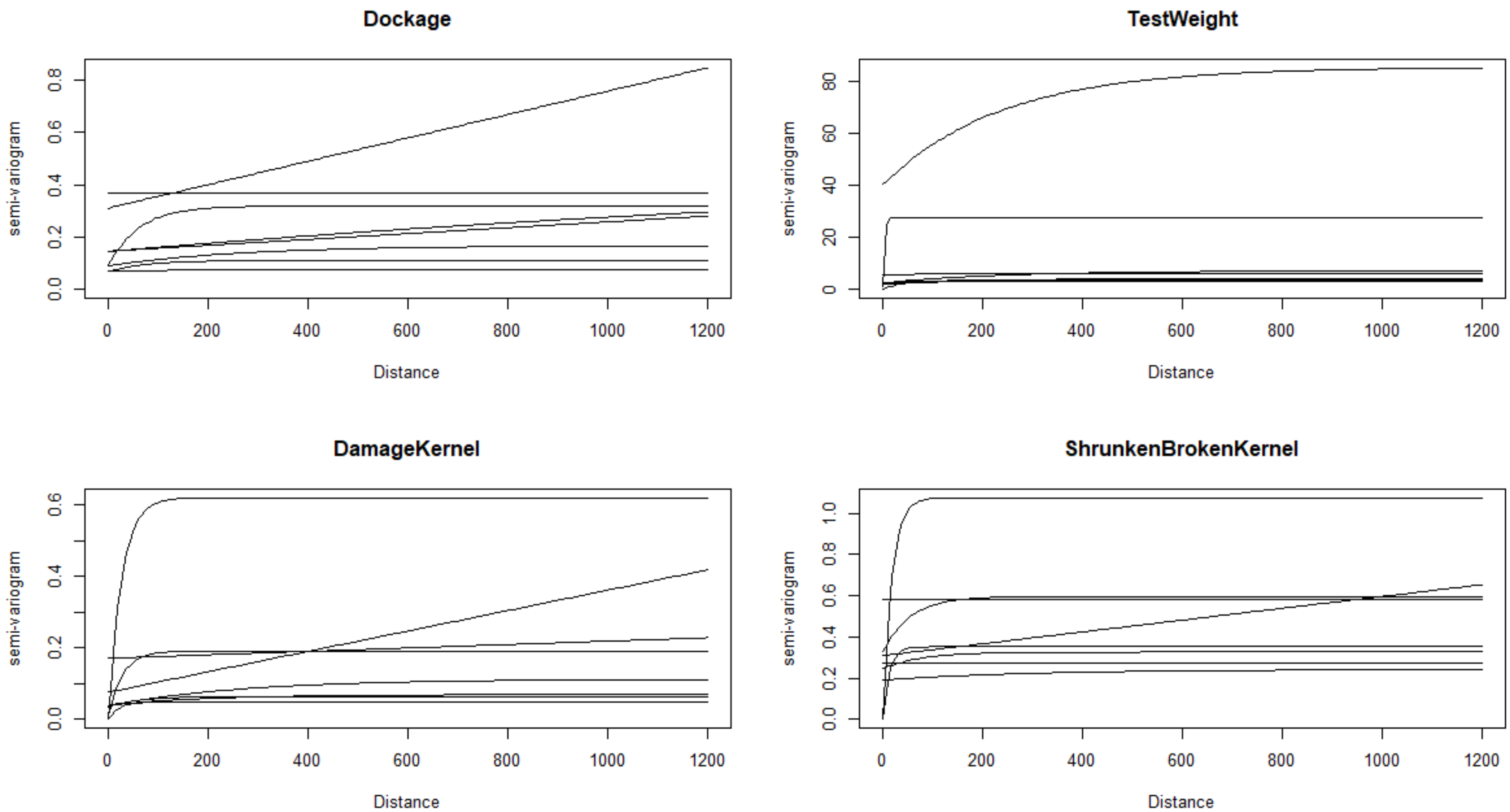


Figure 1-3. Continue.

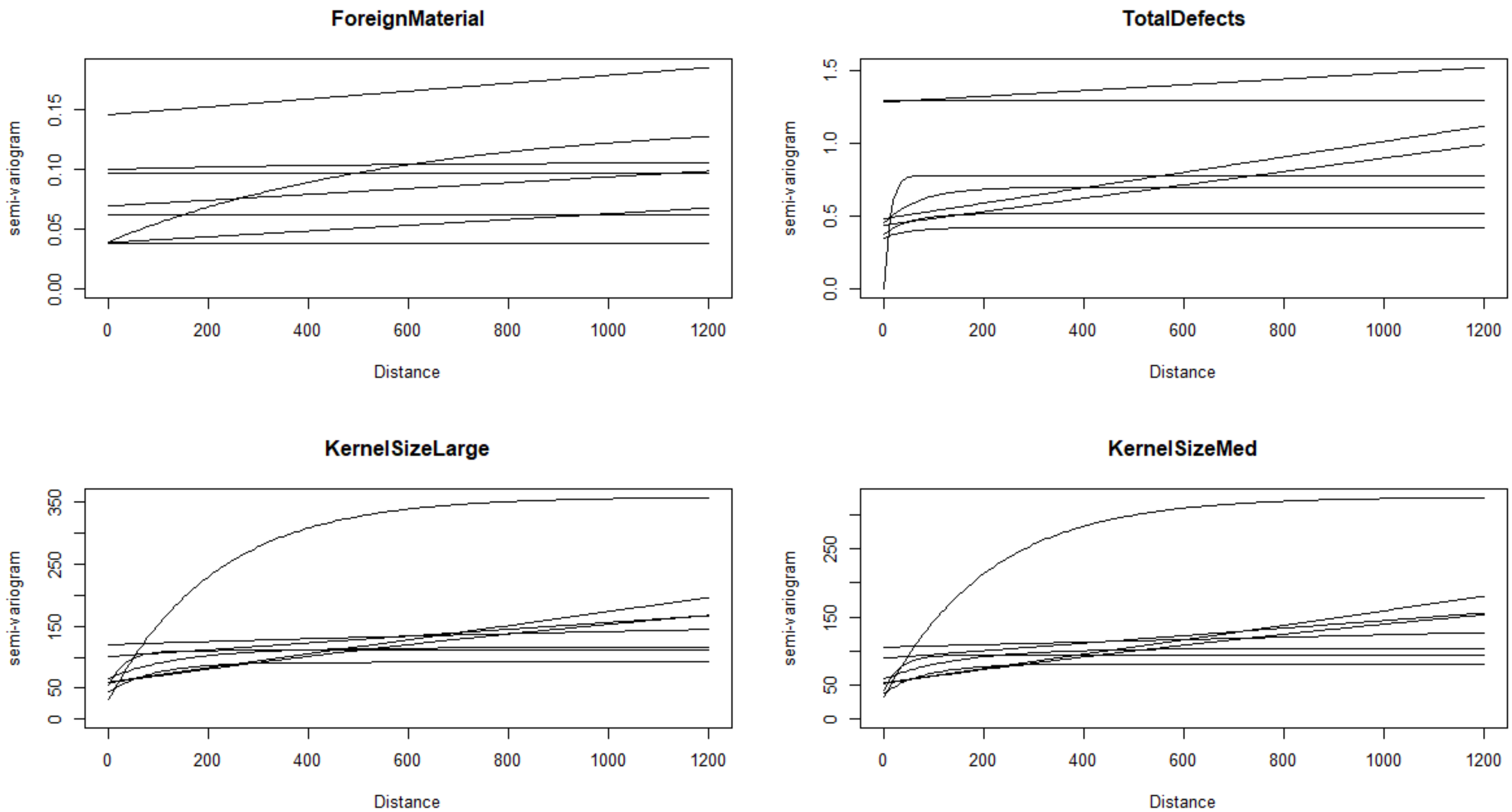


Figure 1-3. Continue.

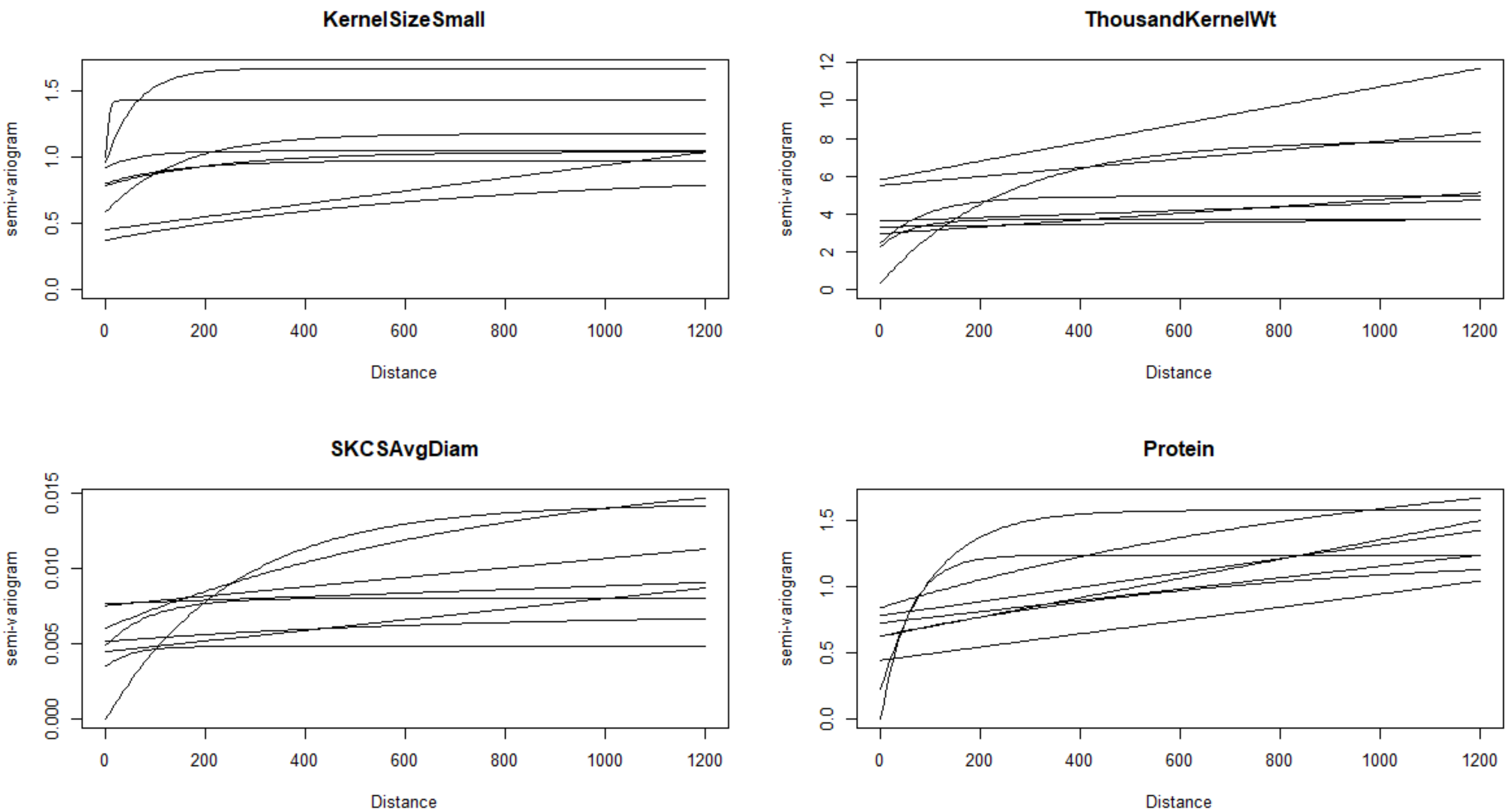


Figure 1-3. Continue

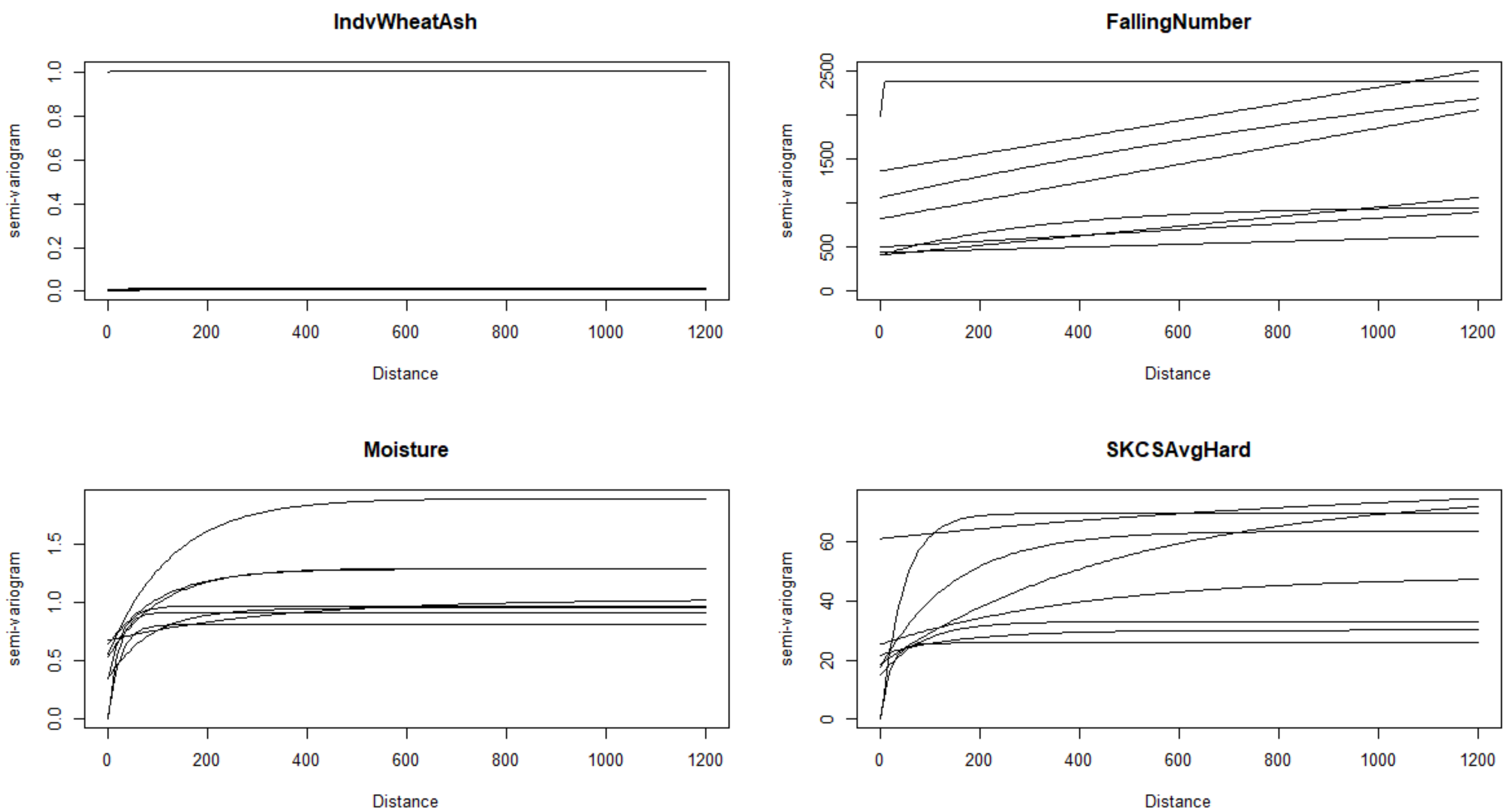


Figure 2-1. Average Post-harvest Basis ($/kg).

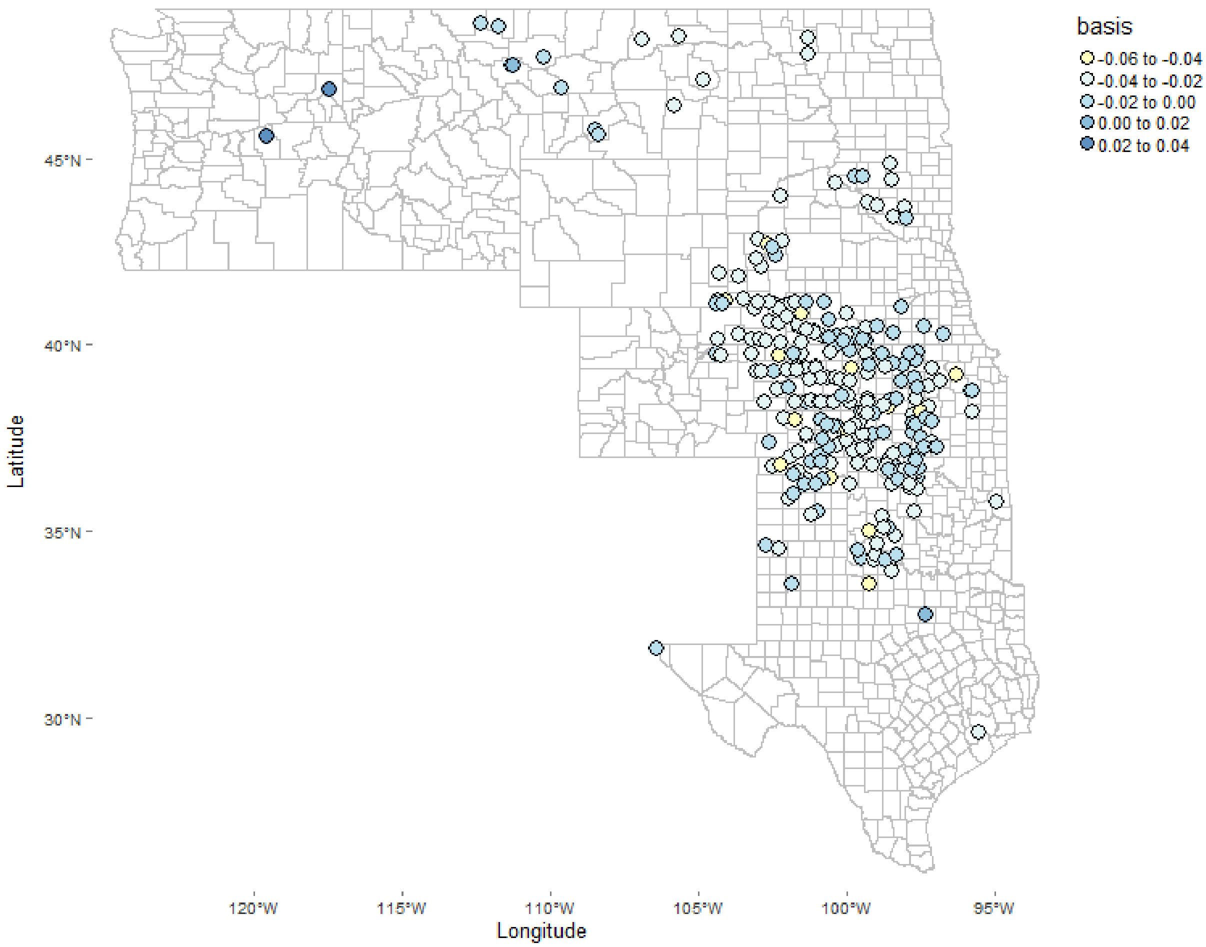
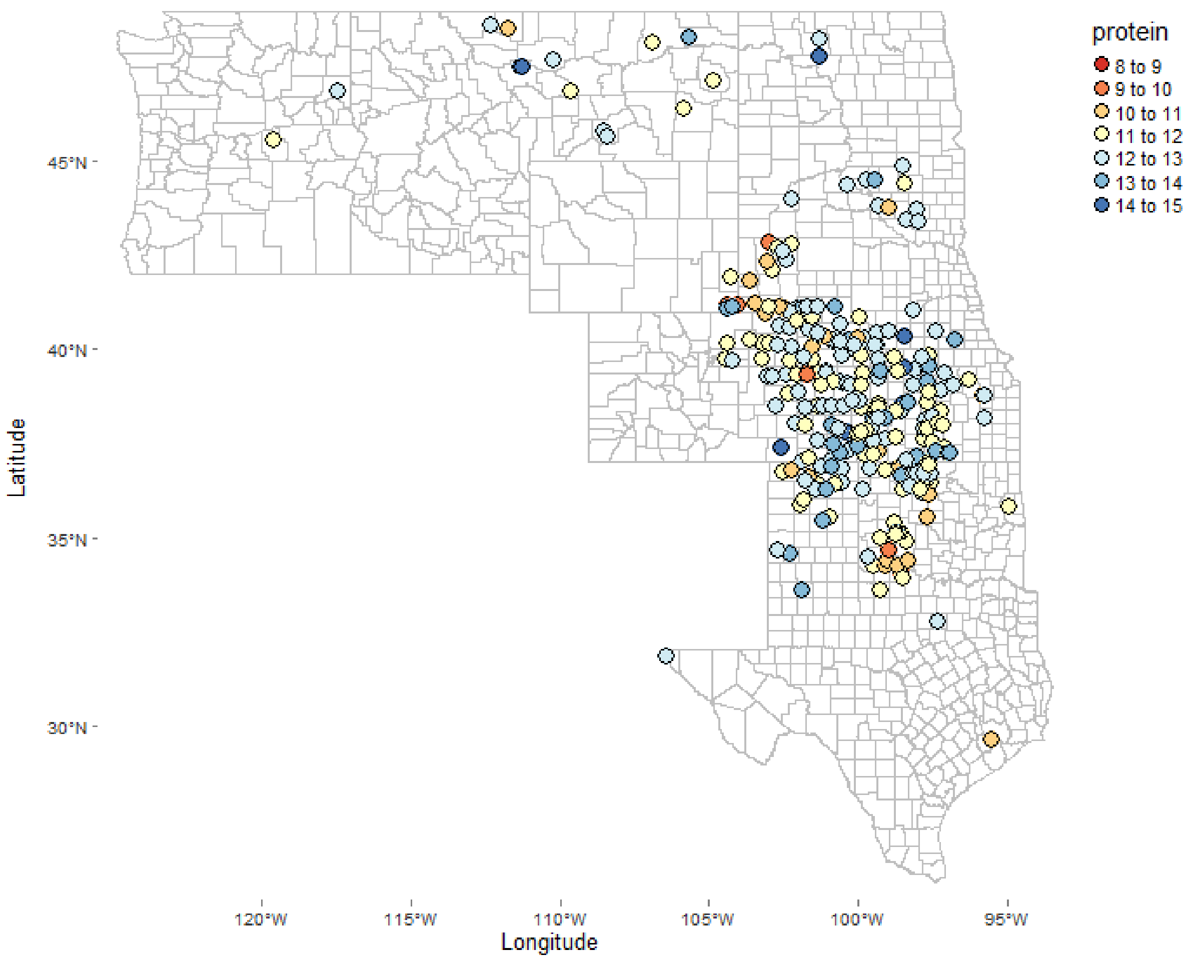


Figure 2-2. Average Protein Level (%).



*Average protein level for each location is calculated as average of recorded protein level.

Figure 2-3. Estimated Mean β_0 by Location.

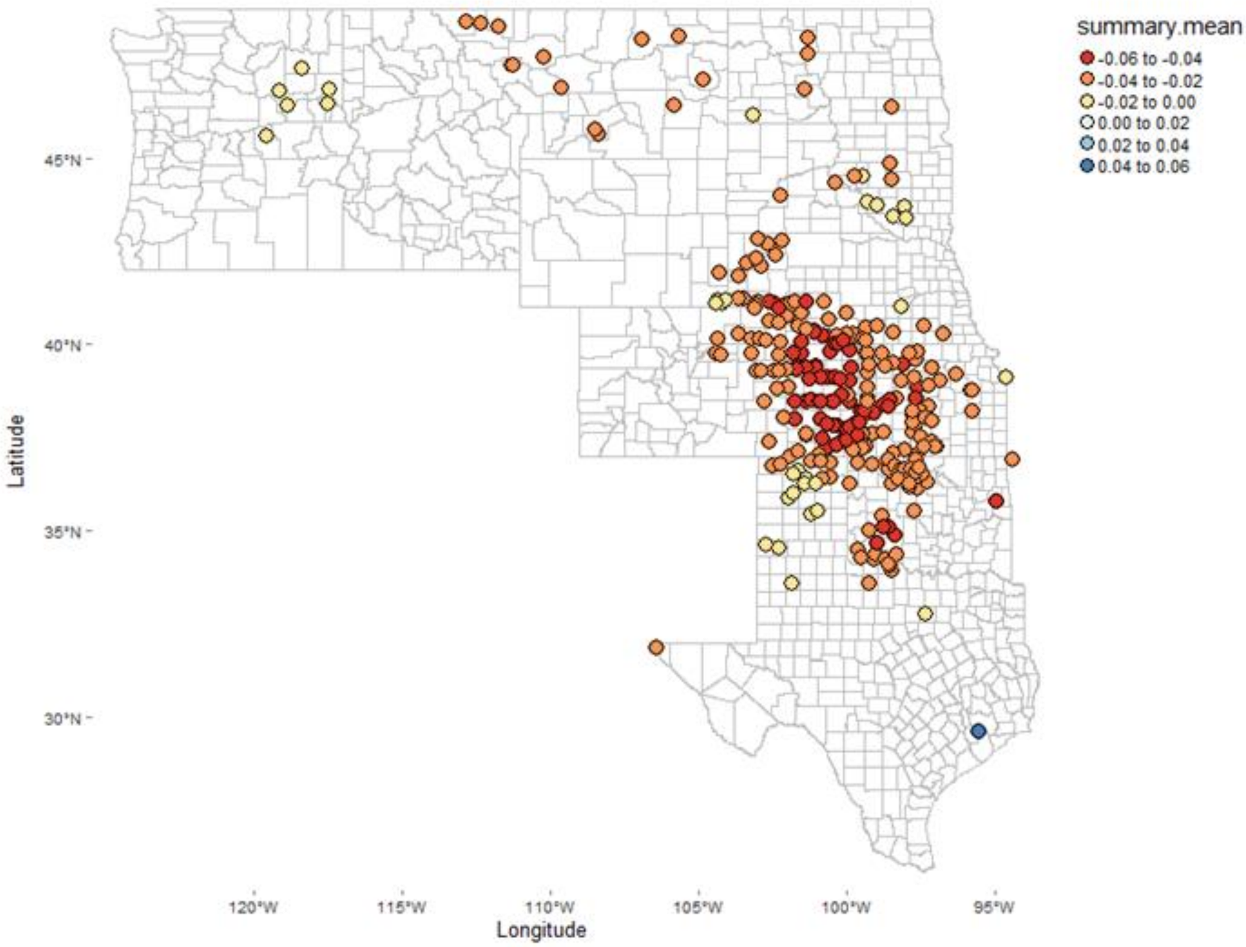


Figure 2-4. Estimated Mean β_1 by Location.

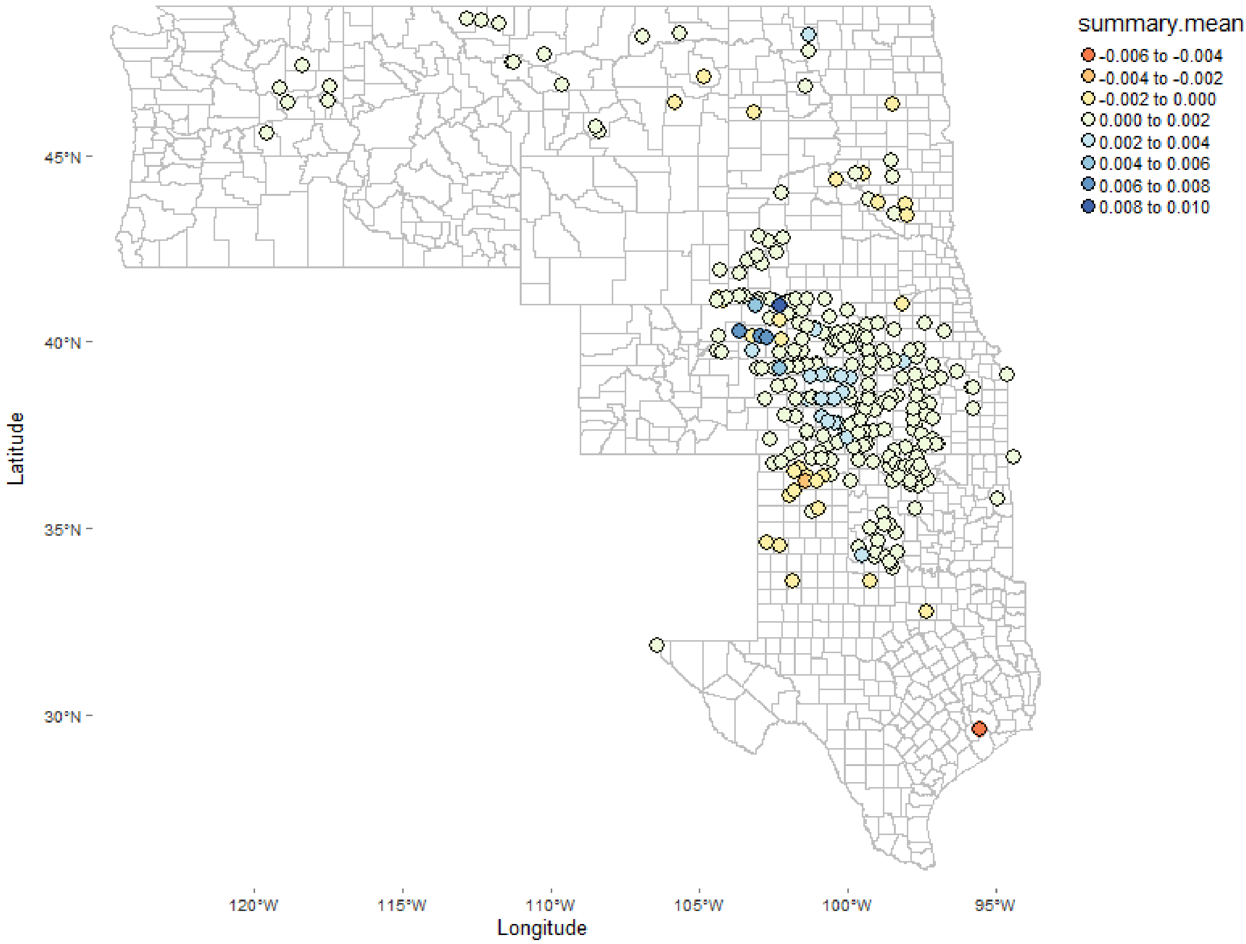
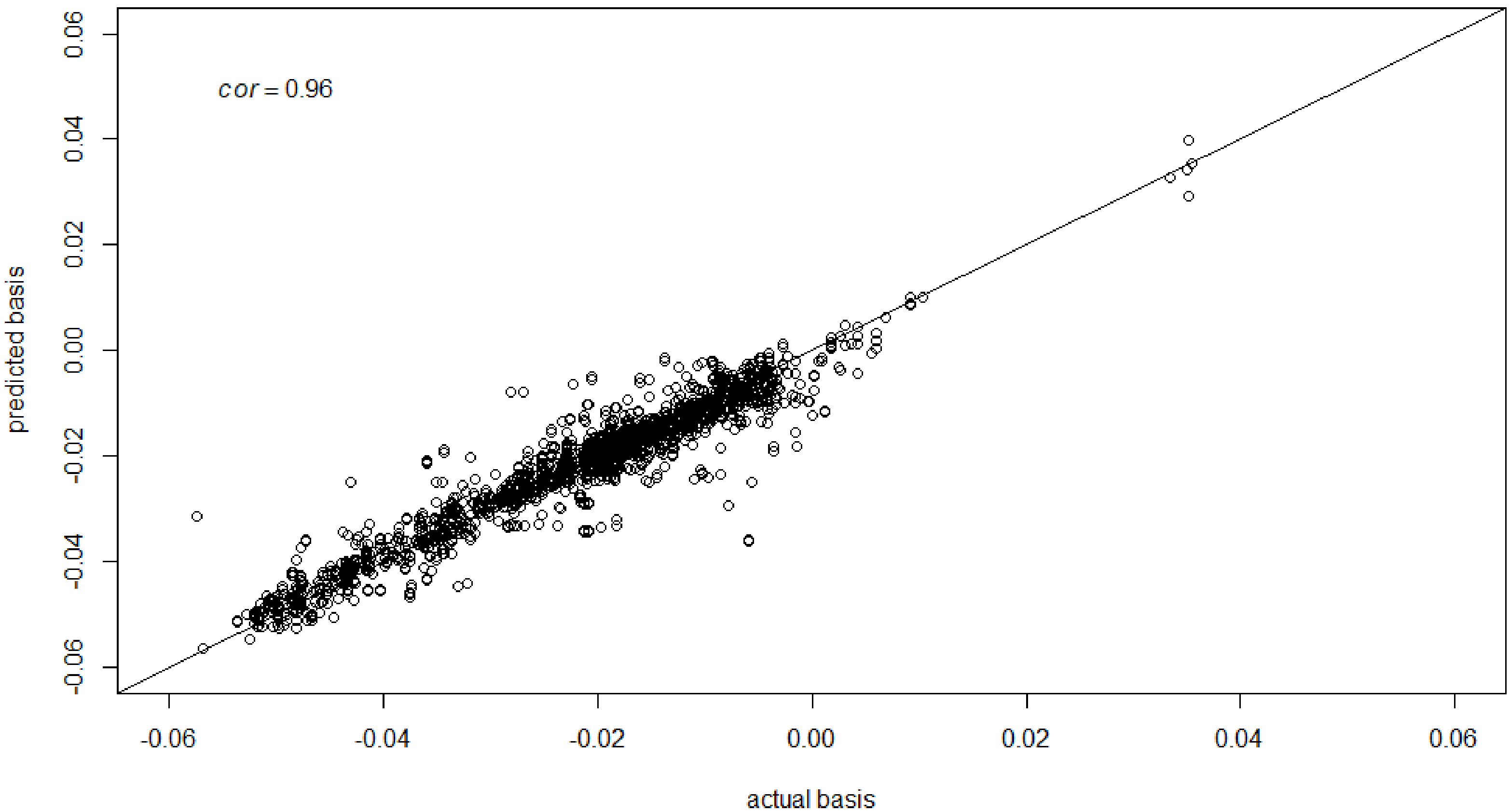


Figure 2-5. Actual Basis vs. Predicted Basis



VITA

YIKUAN CHEN

Candidate for the Degree of

Master of Science

Thesis: SPATIAL PATTERNS IN HARD RED WINTER WHEAT QUALITY AND BASIS

Major Field: Agricultural Economics

Biographical:

Education:

Completed the requirements for the Master of Science in Agricultural Economics at Oklahoma State University, Stillwater, Oklahoma in July, 2021.

Completed the requirements for the Bachelor of Science in Statistics at Oklahoma State University, Stillwater, Oklahoma in 2020.

Completed the requirements for the Bachelor of Science in Agribusiness at Oklahoma State University, Stillwater, Oklahoma in 2019.

Experience:

Graduate Research Assistant
Department of Agricultural Economics, Oklahoma State University, 2020-2021