

Kansas State University Libraries

## New Prairie Press

---

Conference on Applied Statistics in Agriculture

2002 - 14th Annual Conference Proceedings

---


# ESTIMATING THE CONTRIBUTION OF SPECIFIC WHEAT VARIETIES TO STATE-LEVEL PRODUCTION

Bingxin Yu

Richard K. Perrin

David B. Marx

Follow this and additional works at: <https://newprairiepress.org/agstatconference>

 Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

---

### Recommended Citation

Yu, Bingxin; Perrin, Richard K.; and Marx, David B. (2002). "ESTIMATING THE CONTRIBUTION OF SPECIFIC WHEAT VARIETIES TO STATE-LEVEL PRODUCTION," *Conference on Applied Statistics in Agriculture*.  
<https://doi.org/10.4148/2475-7772.1204>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact [cads@k-state.edu](mailto:cads@k-state.edu).

## ESTIMATING THE CONTRIBUTION OF SPECIFIC WHEAT VARIETIES TO STATE-LEVEL PRODUCTION

Bingxin Yu

*103A Filley Hall, East Campus  
Department of Agricultural Economics  
University of Nebraska  
Lincoln, NE 68583-0922*

Richard K. Perrin

*314 Filley Hall, East Campus  
Department of Agricultural Economics  
University of Nebraska  
Lincoln, NE 68583-0922*

David B. Marx

*103 Miller Hall, East Campus  
Department of Biometry  
University of Nebraska  
Lincoln, NE 68583-0712*

## ESTIMATING THE CONTRIBUTION OF SPECIFIC WHEAT VARIETIES TO STATE-LEVEL PRODUCTION

Bingxin Yu, Richard K. Perrin and David B. Marx

### ABSTRACT

This paper addresses measurement issues involved in **estimating the benefits from a university wheat breeding program**. The conceptual foundation is to estimate the spatial distribution of farm-level yields for varieties in each year, based on Nebraska winter wheat variety test data from 1972 to 2001, with 9-20 locations per year. We can then estimate the state average yield difference that would occur between the portfolios of varieties that farmers actually grew, and the portfolio of varieties excluding those varieties from the university breeding program. This paper reports estimates of the statewide annual yields of individual varieties using a **standard fixed effect model**, and with anisotropic power as **spatial covariance structure**. The results indicates that UNL-developed varieties added significant production increases until the mid 1990's when their portfolio advantage diminished due to the introduction of several high yield private varieties. The lump sum direct producer benefit of the Nebraska varieties is estimated to be about \$7 million per year in Nebraska during the 1972-2001 period. Incorporating the spatial structure on variety test data, our estimated producer benefit are about 20% higher than if spatial structure had been ignored.

### Key Words

agricultural research, spatial structure, semivariogram, anisotropic power.

## Section One Introduction

Land-grant universities provided the majority of agricultural research up until the late twentieth century, when advances in molecular biology and new intellectual property rights created incentives for a boom in privately-funded agricultural research. After the implementation of the Plant Variety Protection Act in 1970 and the 1980 Supreme Court *Chakrabarty* decision that initiated patenting of living material, it is now possible for private agricultural researchers to acquire property rights to varieties they developed and thus appropriate benefits from them. Nevertheless, the cost of obtaining and enforcing these property rights still inhibits private research on some crops, leaving opportunities for publicly-funded research programs that can create marginal public benefits in excess of the research costs (Perrin, Poor and Coyne, 2001). University research administrators now search for publicly-funded research programs that are complementary to, rather than competitive with, the private research effort.

A university breeding program can have a number of objectives in addition to releasing new varieties. The wheat breeding program at Nebraska, for example, also intends to develop useful germplasm for others to use, to develop new breeding methods, and to train students in crop breeding methods. Thus a measure of the direct farmer benefits of released varieties is only a partial measure of the benefits of this program. The examination of direct farmer benefits in Nebraska is further limited as a measure of benefits because of the spillovers of benefits to out-state users of the varieties, and because the UNL breeding program benefits from spill-ins from research efforts elsewhere. As stated by Alston, Norton and Pardey (1995), a complete economic evaluation of the benefits from a research program should include

- 1) an evaluation of spillover benefits,
- 2) an enumeration of the time path of costs, and
- 3) a dynamic framework reflecting the human capital accumulated during the course of new variety development.

The purpose of this study is thus quite limited in scope, namely, estimating just the direct producer benefits in Nebraska associated with varieties released by the wheat breeding program at University of Nebraska-Lincoln (UNL). The distribution of winter wheat production of Nebraska in 2000 is illustrated in Figure 1.

The conceptual foundation of this research is that the fundamental payoff from wheat breeding research stems from the value of the resulting increase in wheat production. Here we estimate the hypothetical average yield that Nebraska farmers would have obtained had the university breeding program not existed, so that this hypothetical average yield can be compared with the production achieved in reality. This estimation problem is approached by estimating the average yield for each variety in each year, based on Nebraska winter wheat variety tests. We then estimate the statewide average yield difference that would occur between the portfolio of varieties that farmers actually grew and the remaining portfolio of varieties excluding those from the UNL breeding program.

To come up with a single dollar figure representing the direct producer benefits from UNL-developed varieties, the year-by-year stream of benefits through time is first estimated, and then adapted to present value at a given point in time using standard time discounting procedure.

## Section Two Spatial Structure

Accurate cultivar evaluation and the ability to differentiate among varieties are critical to any plant breeding program. Usually breeding programs rely on replicated field trials involving relatively small plots at different locations. The statistical validity of the field trials depends essentially on the assumption of homogeneous variance across locations, which might be hundreds of miles apart. Failure of this assumption may result in erroneous conclusions (Stroup, Baenziger and Mulitze, 1994).

Natural endowments, such as soil fertility, previous land use, and various weather-related conditions like average precipitation and temperature, are considered as some of the primary influences on agricultural output. However, spatial homogeneity of those factors among locations is extremely uncommon in field experiments. In our case, experiment plots are located in different counties across Nebraska, and in general such fields exhibit a high degree of spatial heterogeneity in soil fertility and average precipitation. Therefore, spatial distribution is introduced to account for the difference in natural endowments among locations. First, we need to characterize the spatial structure by estimating covariance parameters.

In any spatial model, measurements of a certain attribute  $V(x)$  at a specific location, as well as the vector of coordinates,  $x$ , should be included. To estimate the spatial variability structure for this study, the single measurement is the average annual county-level wheat yield (on non-irrigated, continuously-cropped land) for the 91 counties in Nebraska over the 46-year period, 1956-2001, which was obtained from National Agricultural Statistics Service, USDA. Longitude and latitude at the center of each county are taken as initial coordinates (americasroof.com). The geographical coordinates are then calculated into distance, with origin at the center of Nebraska and distances (in kilometers) determined in the direction of west-east and south-north. The objective of this univariate spatial model is to describe and quantify the relation between yields in any two locations (counties).

The relationship between measurements made at various locations can be displayed by using a semivariogram, which is to measure the average dissimilarity between non-located data based on spatial separation distance (Goovaerts, 1997). For all pairs of measurement ( $V(x), V(x+h)$ ) at locations separated by a given distance  $h$  in a particular direction, the semivariogram is computed as half the expected squared difference between measurements separated by distance  $h$ :

$$\gamma(h) = \frac{1}{2N(h)} \sum [V(x) - V(x+h)]^2 \quad (1)$$

where  $N(h)$  is the number of data pairs within a radius of 400 kilometers and in 4 direction categories. The semivariogram value increases as the points spread out farther with increasing separation distance  $h$ , implying the increasing dissimilarity between measurements farther apart.

A basic transition semivariogram model consists of three parts: nugget effects, sill and range (Figure 2). Nugget effect is the vertical jump from the value of zero at the origin to the value of the semivariogram at an extremely small separation distance. The plateau the semivariogram reaches is called the sill. The distance at which the semivariogram value reaches about 95% of the sill is the range.

Often we expect to see the spatial variability (dissimilarity), expressed by semivariogram, the same in all directions. If direction does not affect the semivariogram, this is called isotropy or omnidirectional. If not, then we have anisotropy. An anisotropy is said to be geometric when the directional semivariograms have the same shape and sill but different range values (see Figure 3).

A simple spatial correlation model is specified as

$$y_i = \mu + e_i \tag{2}$$

where  $y_i$  is the annual wheat yield of the  $i$ th county,

$\mu$  is the grand mean for all counties in Nebraska,

$e_i$  is the specific yield effect for the  $i$ th county.

In a standard linear model analysis, the specific county yield effects may be assumed to be a randomly-generated error term that is identically and independently distributed with a constant variance of  $\sigma^2$ . This corresponds to the assumption of no spatial trends. Under the existence of spatial structure, the error terms are assumed to be distributed according to some spatial correlation models describing the local trends. The basic idea behind the spatial structure is that adjacent counties tend to be more alike whereas more distant counties tend to be less alike if local trends exist.

Typically, the covariance of spatial correlation model is assumed to be a function of the distance between location (in our case, county)  $i$  and  $j$ . Let  $h$  be the distance between location  $i$  and  $j$ . A model with a nugget effect has the general form

$$\text{Var}(e_i) = \sigma^2 + \sigma_1^2 \tag{3}$$

$$\text{Cov}(e_i, e_j) = \sigma^2 [f(d_{ij})] \tag{4}$$

where  $f(d_{ij})$  is a function of distance between county  $i$  and  $j$ , with vector  $d_{ij}$  accounting for both distance and direction. The parameters  $\sigma_1^2$  and  $\sigma^2 + \sigma_1^2$  correspond to the nugget and sill in semivariogram respectively. Parameter  $\rho$  denotes the range of semivariogram (Littell, Milliken, Stroup and Wolfinger, 1996). In a spatial correlation model, the range is the distance at which correlation between observations is effectively zero, and the sill is the covariance at distance zero.

Figure 3 shows the experimental semivariogram computed from the average annual continuous cropping dryland winter wheat yield data, with an angular tolerance of 22.5 degrees (area within the rays from origin which are 22.5 degrees from rays at the four directions). Directions are expressed as starting from north and proceeding clockwise at 0, 45, 90, and 135 degrees. From the graph we can conclude that the semivariogram is almost linear and geometric anisotropy exists.

In SAS, two spatial covariance structures are available: anisotropic exponential and anisotropic power. The anisotropic power model is chosen to represent the spatial covariance structure because it requires fewer parameters and can be made to approximate a linear model.

No sill exists for a power model. To solve this problem, the  $(i, j)$ th element of the covariance matrix is defined as

$$\text{cov}(e_i, e_j) = \sigma^2 \rho_1^{d(i,j,west-east)} \rho_2^{d(i,j,south-north)} \quad (5)$$

where  $\sigma^2$  is model variance,

$\rho_1$  is the correlation for west-east direction,

$\rho_2$  is the correlation for south-north direction, and

$d(i,j,k)$  is the absolute distance between the  $k$ th coordinate,  $k = \text{west-east, south-north}$  of the  $i$ th and  $j$ th county.

The estimated parameters for anisotropic power model  $\sigma^2$ ,  $\rho_1$  and  $\rho_2$  are 7.8887, 0.9971 and 0.9963 respectively. The estimated nugget effect  $\sigma_1^2$  is 0.7088. For the regressions presented later, the element of covariance matrix for this specific anisotropic power spatial structure can be expressed as

$$\text{var}(e_i) = 7.8887 + 0.7088 \quad (6)$$

$$\text{cov}(e_i, e_j) = 7.8887 * (0.9971)^{d(i,j,west-east)} * (0.9963)^{d(i,j,south-north)} \quad (7)$$

with  $d(i,j,k)$  as defined before. We will use this covariance structure to describe the spatial distribution of winter wheat varieties in Nebraska.

### Section Three Estimation of Portfolios of Wheat Yield

#### 1. Estimation of Wheat Yields

The fundamental payoff from wheat breeding research is the value of the resulting increase in wheat production. This requires estimation of the hypothetical average yields that farmers would have obtained had the breeding program not existed, so that this hypothetical average yield can be compared with the production achieved in reality. We approach this estimation problem by first estimating the farm-level yields for varieties in each year under the anisotropic power spatial structure estimated in section two. It is assumed that the year-to-year observations are independent of each other and hence no correlation between any two years.

The *Nebraska Variety Tests of Fall-sown Small Grains* reports winter wheat variety trials conducted throughout Nebraska from 1972 to 2001, with 9 to 20 locations per year and 10 to 40 varieties per location. Entries include released varieties, hybrids and promising experimental strains from UNL, surrounding states and private breeders. Trials were located on research centers or private farms and all tests were of nursery type consisting of plots approximately 40 square feet in size although plot size varies with location. Plots at each location were replicated 3 to 6 times.

For this research, the categorical regression model is expressed as:

$$\text{yield}_{ijkl} = \mu + \text{year}_i + \text{variety}_j + \text{crop}_k + \text{irrigate}_l + e_{ijkl} \quad (8)$$

where  $yield_{ijkl}$  is the yield of the  $j$ th variety in  $i$ th year,

$\mu$  is the overall average yield for all varieties,

$year_i$  is the additional yield for the  $i$ th year,  $i = 1972, 1973, \dots, 2001$ ,

variety  $j$  is the additional yield for the  $j$ th variety,  $j = 1, 2, \dots, 63$ ,

$crop_k$  is the additional yield for the  $k$ th type of prior crop (including fallow and ecofallow),  $k = 1, 2, \dots, 10$ ,

$irrigate_l$  is the additional yield for irrigated plots, and it is dummy variable taking value one when the plot was irrigated,  $l = 1, 2$ ,

$e_{ijkl}$  is the error term and its distribution depends on the model definition for covariance structure.

First, a standard fixed effect model is estimated, which assumes the identically independently normal distribution of error term both across location and over time. The fitness statistics are summarized in Table 1. Similarly, the fitness statistics of the same regression function, imposing the anisotropic power spatial covariance structure derived from the 46-year average yield of non-irrigated, continuous cropping winter wheat in section two, are also included on the second column of Table 1. According to the information criteria, the spatial model is more plausible since it interprets the wheat yield better. The estimated yield of varieties in different locations for each year is available upon request.

## 2. Estimation of Hypothetical Portfolio of Wheat Varieties

Next, data on percentage of planted acreage to each winter wheat variety is used to obtain the portfolios of varieties planted by farmers. This information was collected from the yearly publication of National Agricultural Statistics Service, USDA except for year 1992. This wheat variety data is provided through cooperative response of about 1,000 Nebraska wheat producers who responded to the yearly *Fall Acreage and Production Survey*. The wheat variety survey summary includes reports from producers representing approximately 8-16% of the state's total planted acreage. About 15-20 varieties are included in each year. In most cases, these varieties account for more than 90% of total acreage planted.

Using percentage of planted acreage to each winter wheat variety as weights, we estimate the statewide average yield that would occur for the portfolios of varieties with and without those from the breeding program being examined. The annual estimated statewide average yields including and excluding varieties from the UNL breeding program are illustrated in Figure 4. For comparison, we also show the statewide average yield from National Agricultural Statistics Service, USDA. Except for the mid-1970's to early-1980's, wheat varieties developed by the breeding program in University of Nebraska-Lincoln contributed substantially to production increases up until mid-1990's, with the yield margin as high as 4 bushels per acre (a 8% increase



in state wheat production) in 1986-1988. Afterwards, this portfolio advantage over the hypothetical group excluding UNL varieties diminished and finally was negative. This happened due to the introduction of several high-yield private varieties, for example, AgriPro Ogallala.

### 3. Estimation of Benefit from Breeding Program

We present here the value of direct producer benefits from the Nebraska wheat varieties, based on our estimates of the present value of yearly producer benefits. The estimates of year-to-year benefit are generated for year 1972 to 2001, and are then translated into present values as of the end of year 2001.

The value of the producer benefit is calculated from planted acreage and adjusted price that Nebraskan producers received from 1972 to 2001. The planted acreage and price data are summarized in Table 2. The value of producer benefit (column 4) is computed as the product of yield difference between portfolios with and without Nebraska varieties, planted acreage and adjusted price. Thus the yearly producer benefits are expressed in the value of 2001 dollars. Figure 5 depicts the developing pattern of the year-to-year producer benefits.

Estimated producer benefit streams such as those in Table 2 are commonly converted to lump sum equivalents at some arbitrary point in time, usually the present. This is accomplished by compounding past amounts to the present using compound interest formula. While the compounding formula is clear-cut, there is no straightforward criterion for selecting a discount rate appropriate for government projects. It is generally accepted that this rate should lie between the before-tax and after-tax rates of return on investments of comparable riskiness, but these conceptual rates are not unique (Perrin, Poor and Coyne, 2001). Alston, Norton and Pardey (1995) suggested real rates of 3-5 percent, while Brent (1996) argued that the several relevant social discount rates in use that range from 3% to 12% were proper for the calculation of agricultural research. Real discount rates of 5% or 10% seem to provide a credible range for present study.

In the last three columns of Table 2, the producer benefit of each year is brought to the present value of year 2001, first without any discount rate (a simple sum of benefits), and then using discount rate of 5% and 10%. These figures represent the direct economic value in Nebraska of the wheat varieties developed by the breeding program at University of Nebraska-Lincoln, expressed in terms of an equivalent amount of cash received as a lump sum at the end of year 2001.

The estimated 2001 value of the total producer benefit in Nebraska ranges from \$200 (simple sum) to \$1156 million (high discount rate). If converted into annuity value, the varieties developed by UNL create about \$7 million (in 2001 value) yearly since 1972.

For the purpose of comparison, results from the standard fixed effect model are also included in the last two rows of Table 2. In general, the spatial structure effectively corrects the bias from a mistaken hypothesis of homogeneous variance. If we had used the standard model rather than the spatial model, we would have underestimated the benefit by about 20% (depending upon the discount rate used).

### Section Four Summary

We estimate that it has generated direct benefit worth about \$7 million every year (at 5% discount rate) to winter wheat producers in Nebraska during the 1972-2001 period. We rejected the hypothesis of no spatial structure based on a continuous cropping, dryland commercial yield by county in Nebraska for the period 1956-2001. Incorporating this spatial structure on variety test data, our estimated producer direct benefits are about 20% higher than if the spatial structure had been ignored.

## References

- Alston, J. M., G. W. Norton and P. G. Pardey. *Science Under Scarcity: Principles and Practice for Agricultural Research Evaluation and Priority Setting*. Cornell University Press, Ithaca and London, 1995.
- Brent, Robert J. *Applied Cost-Benefit Analysis*. Edward Elgar, Brookfield, VT, 1996.
- Consumer price index. Federal Reserve Bank of Minneapolis.  
<http://minneapolisfed.org/economy/calc/hist1913.html>, last visited April 17, 2002.
- Goovaerts, Pierre. *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, 1997.
- Littell, Ramon C., George A. Milliken, Walter W. Stroup and Russell D. Wolfinger. Chapter 9, *SAS System for Mixed Models*. Cary, NC: SAS Institute Inc., 1996.
- National Agricultural Statistics Service, U.S. Department of Agriculture, *Nebraska Wheat Varieties*, 1972-1991, 1993-2001.
- National Agricultural Statistics Service, U.S. Department of Agriculture,  
<http://www.nass.usda.gov:81/ipedb/>, Last visited March 1, 2002.
- Perrin, Richard K. Intellectual Property Rights and Developing Country Agriculture. *Agricultural Economics*. 21(3), pp. 221-229, December 1999.
- Perrin, Richard K., Joan Poor and Dermot Coyne. Economic Impact of The Chase Variety of Pinto Bean. University of Nebraska Cooperative Extension RB 338, 2001.
- Spatial Information. Americans Roof,  
<http://www.americasroof.com/counties/nebraska.shtml>. Last visited April 14, 2002.
- Stroup, Walter W., P. Stephen Baenziger and Dieter K. Mulitze. Removing Spatial Variation from Wheat Yield Trials: A Comparison of Methods. *Crop Science*. 34, pp 62-66, January-February 1994.
- University of Nebraska-Lincoln, Institute of Agricultural and Natural Resources, Agricultural Research Division, Cooperative Extension, *Nebraska Variety Tests of Fall-sown Small Grains*, 1972-2001.

### Winter Wheat: Production Nebraska, 2000

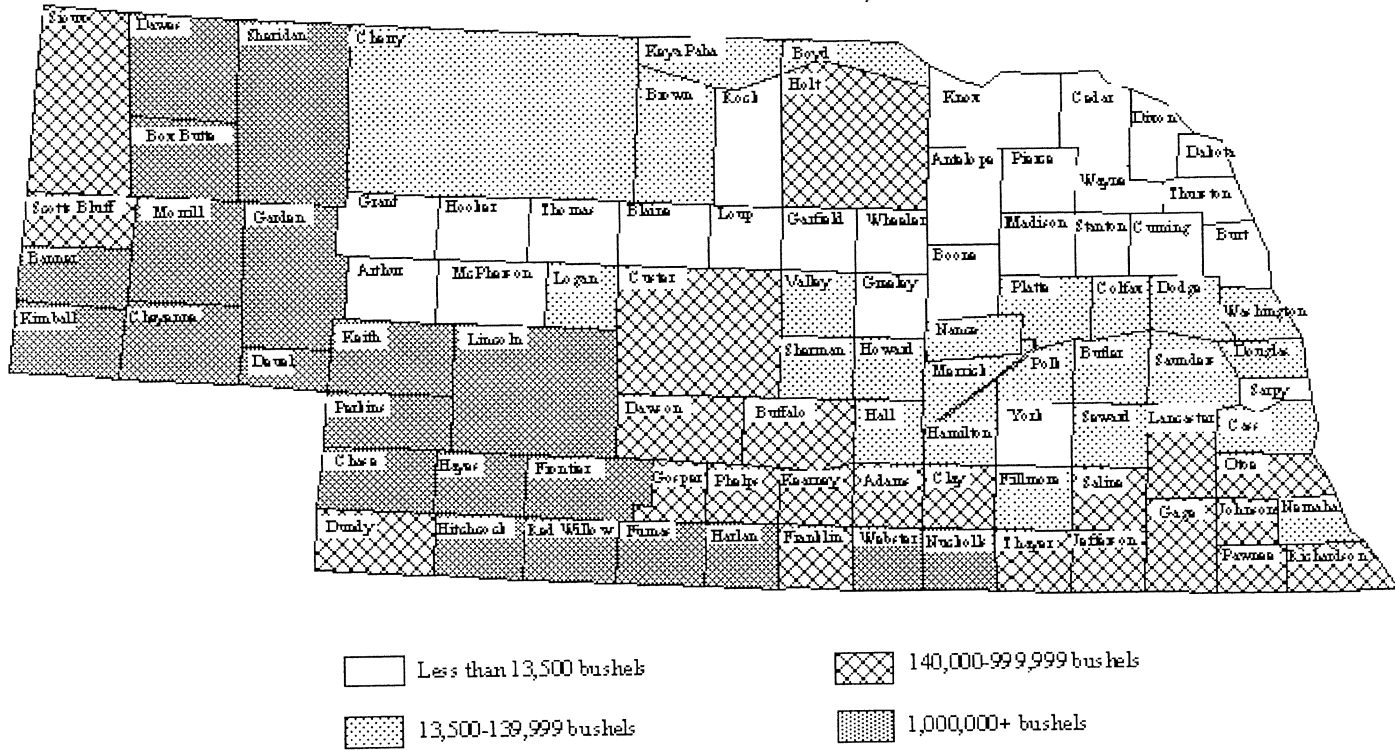


Figure 1. Winter wheat production in Nebraska, 2000.

Source: NASS, USDA, 2002.

Figure 2. Semivariogram for spherical model.

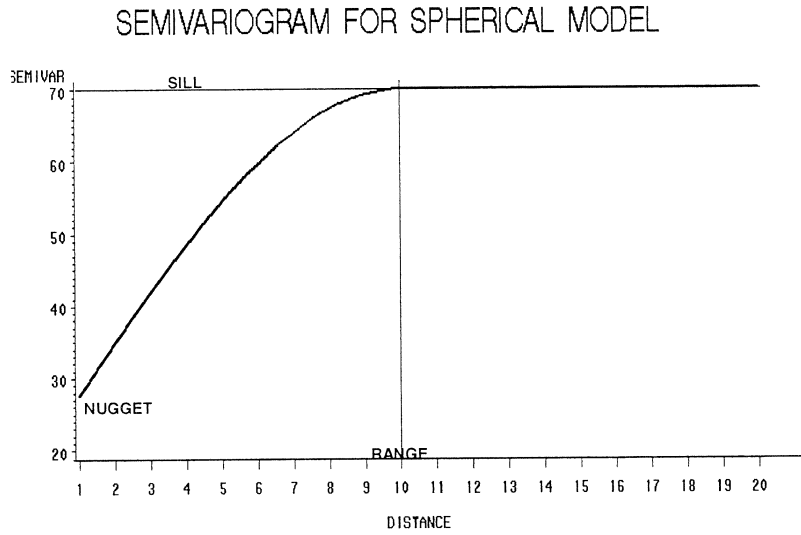


Figure 3. Anisotropy in different directions.

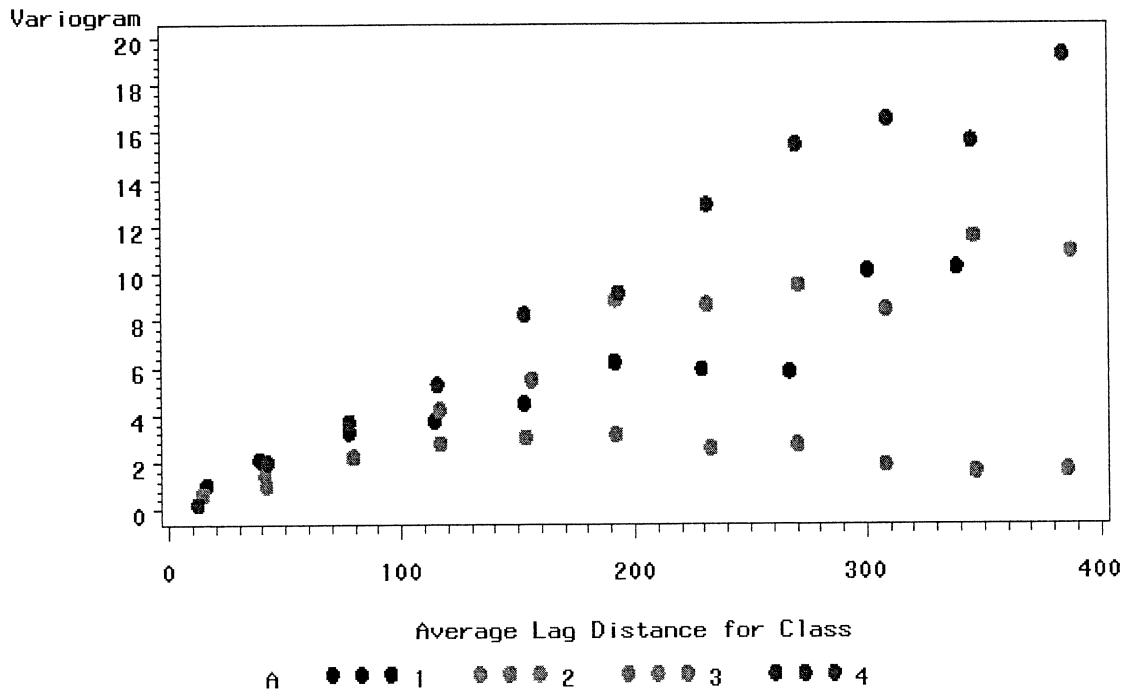


Figure 4. Difference in portfolio of varieties.

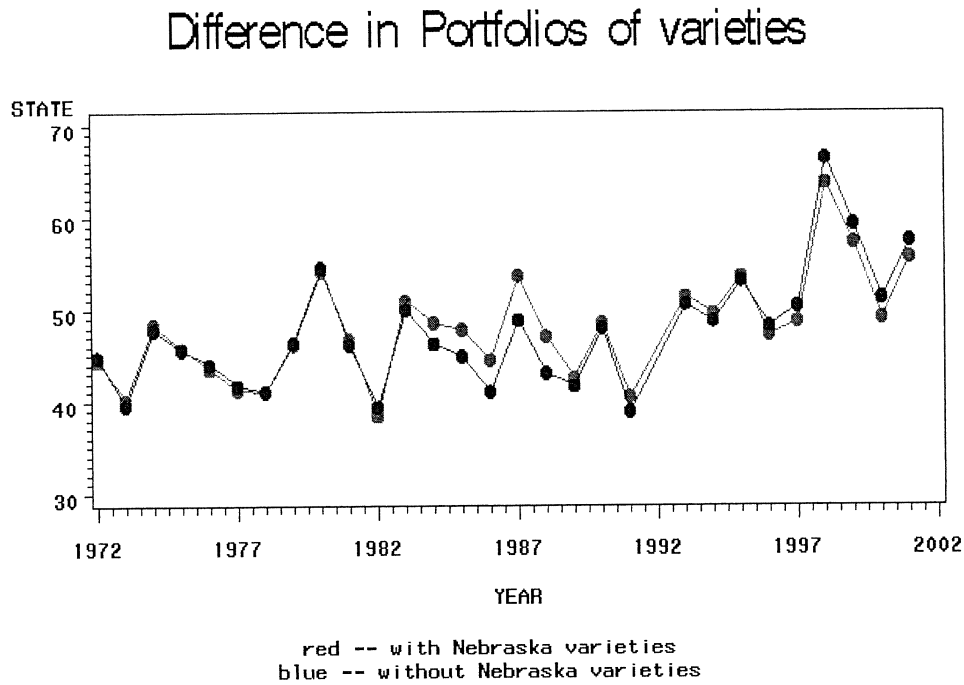


Figure 5. Benefit from UNL wheat breeding program, 1972-2001.

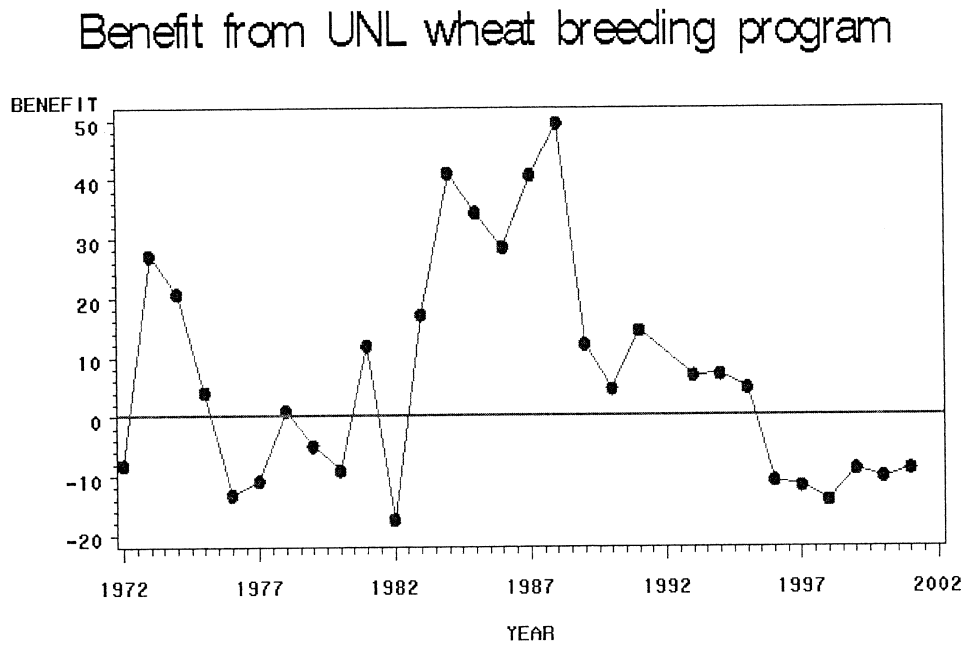


Table 1. The comparison of fitness statistics between regular model and spatial model.

	Standard Model	Spatial Model	Information Criterion
Res Log Likelihood	-25507.0	-25118.9	bigger
Akaike's Information Criterion	-25508.0	-25122.9	bigger
Schwarz's Bayesian Criterion	-25511.3	-25136.4	bigger
-2 Res Log Likelihood	51013.9	50237.9	smaller

Applied Statistics in Agriculture

Table 2. Planted acreage and price of winter wheat that producer received in Nebraska.

year	planted	adjusted	benefit	present value (\$millions)		
	acreage	price				
	(1,000 acres)	(\$/bu)	(\$millions)	0%	5%	10%
	(1)	(3)	(4)=(1)*(3)*diff			
1972	2742	7.38	-8.16	-8.16	-33.58	-129.42
1973	2800	15.17	27.20	27.20	106.64	392.28
1974	3000	13.69	20.74	20.74	77.43	271.88
1975	3200	11.17	4.07	4.07	14.48	48.54
1976	3400	7.41	-13.19	-13.19	-44.67	-142.93
1977	3300	6.49	-10.85	-10.85	-35.00	-106.91
1978	2900	7.58	0.94	0.94	2.89	8.41
1979	3000	8.98	-4.98	-4.98	-14.57	-40.53
1980	3000	8.21	-9.06	-9.06	-25.24	-67.05
1981	3000	7.17	11.92	11.92	31.62	80.18
1982	3050	6.15	-17.37	-17.37	-43.89	-106.23
1983	2800	5.96	17.14	17.14	41.24	95.27
1984	3200	5.61	41.06	41.06	94.10	207.52
1985	2600	4.59	34.41	34.41	75.11	158.10
1986	2300	3.61	28.49	28.49	59.24	119.03
1987	2200	3.82	40.71	40.71	80.60	154.59
1988	2300	5.48	49.52	49.52	93.37	170.94
1989	2550	5.36	12.07	12.07	21.68	37.89
1990	2450	3.43	4.60	4.60	7.86	13.11
1991	2350	3.92	14.43	14.43	23.51	37.44
1993	2350	3.73	6.84	6.84	10.10	14.66
1994	2200	4.05	7.05	7.05	9.93	13.75
1995	2150	5.30	4.75	4.75	6.37	8.42
1996	2300	4.85	-10.87	-10.87	-13.87	-17.50
1997	2000	3.53	-11.76	-11.76	-14.29	-17.21
1998	1900	2.76	-14.16	-14.16	-16.40	-18.85
1999	1900	2.34	-8.96	-8.96	-9.88	-10.84
2000	1750	2.68	-10.19	-10.19	-10.70	-11.21
2001	1750	2.80	-8.85	-8.85	-8.85	-8.85
spatial	sum of present value (\$millions)			197.53	485.21	1154.46
model	annuity (\$millions/yr)			6.81	7.48	7.12
standard	sum of present value (\$millions)			168.73	401.65	809.77
model	annuity (\$millions/yr)			5.82	6.19	4.99