YRB 82-11

PAPER NO. 82-4516

LARGE AREA CROP YIELD MODELING

by

F. N. Kogan
University of Missouri
Columbia, MO.

For presentation at the 1982 Winter Meeting
AMERICAN SOCIETY OF AGRICULTURAL ENGINEERS

Palmer House, Chicago, Illinois
December 14-17, 1982

SUMMARY:

Weather indices were designed for large area barley yield modeling in the USSR. These indices were based on quantitative evaluation of the physiological response of barley yield to weather. Independent verification of models using indices showed fairly reliable results in predicting average yields for large areas with a good lead time of prediction.

American Society of Agricultural Engineers

St. Joseph, Michigan 49085

ACKNOWLEDGEMENTS

This work is supported by a cooperative research agreement with the Statistical Reporting Service of the USDA. The author has benefitted from very useful discussions and comments on the topic of this paper given by Wendell W. Wilson and other specialists with whom the author is working in the Yield Evaluation Section (USDA).

LARGE AREA CROP YIELD MODELING

F. N. Kogan[1/]

Abstract. Weather indices were designed for large area barley yield modeling in the USSR. These indices were based on quantitative evaluation of the physiological response of barley yield to weather. Independent verification of models using indices showed fairly reliable results in predicting average yields for large areas with a good lead time of prediction.

Introduction

Presently, the multiple efforts of scientists in different fields of biological science are directed to solving the problem of accurate estimates of agricultural production for large areas. These estimates, especially those made well in advance of harvest, acquire particular importance in the present tough situation with food supply and demand in the world.

Among different approaches for modeling, the analogue or regression approach has showed fairly extensive application for large area estimate of crop production (Thompson, 1969; Kogan, 1977). Analogue modeling is a type of simulation which is based on the statistical description of the interaction between crop and environment. This is extracted from the historical data through the application of general knowledge of physiology, climate, soil, technology and also statistical tools.

The analogue models normally satisfy almost all requirements of "crop yield model test and evaluation criteria" (Wilson and others, 1980) developed to estimate the performance of a model. They are simple, can be developed in a relatively short period, and are not costly. They are also fairly accurate, reliable and provide a good lead time for timely assessments of crop production. At the same time, lack of historical data and the large number of factors affecting crop yield limit the accuracy of these models. To improve the accuracy of crop-weather models, weather indices were designed and used in this study for large area yield estimates.

Techniques for Crop-Environment Modeling and Historical Data

Normally, two components (long and short term) define yield variability (Obukhov, 1949; Thompson, 1969). The long term component represented in the form of trend in yield series reflects mainly technological changes in growing crops and changes in climate. These changes usually occur in a smooth manner. Trend is normally approximated by a function of time in a yield series. The short term component reflects year-to-year fluctuation in yields as a result

1/ F. N. Kogan is a Research Associate in the Department of Atmospheric Science, University of Missouri, Columbia, Missouri. Through a cooperative research agreement, he works in the Yield Evaluation Section (Columbia, Missouri) of the Statistical Reporting Service, USDA.

of changes in weather. This component expressed in the form of a deviation of yield from trend can be approximated by the intensity of weather patterns. Following this principal the equation for yield modeling used in this study was written in form:

$$Y = F\,[f_1(TR),\ f_2(W)] \tag{1}$$
$$f_1(TR) = g_1(t) + e_{TR}$$
$$f_2(W) = g_2(P,T) + e_W$$

where Y is the reported yield, f_1 is a function of the yield's dependence on trend (TR) and f_2 is a function expressing the dependence of departures from the trend function (e_{TR}) on weather (W). Trend is usually approximated as a function of year number $g_1(t)$. Departures from trend are approximated as a function of derived weather indices g_2 (P,T) based on precipitation (P) and temperature (T): e_W are errors connected with estimating this function. Equation 1 differs from the commonly used form of the yield relationship to trend and weather which can be written as:

$$Y = a_o + a_1 TR + \sum_{i=2}^{K} a_i\, W_i + e \tag{2}$$

where a_o, a_1, a_2, ... , a_k are regression coefficients and e is the model error.

The trend estimate in equation (1) was carried out with the application of least square regression techniques. For a better approximation of the weather induced fluctuation of yield in the last few years the following approach was applied. If the yields of the last several years in a yield series were approximately equally and uniformly distributed on both sides of the trend, then these years should be included in the data for trend specification. Otherwise these years should not be included in the calculation. In such a case the trend would be projected based on data prior to these years. For a yield-series longer than 20-25 years, a second and sometimes a third degree of polynominal approximation might be considered the best from the standpoint of the long-range relationship among technology, climate and cropping power. In our study, a second degree trend approximation was used.

The weather dependent function (f_2) in equation (1) was calculated as a ratio of actual to trend estimated yields. Weather was represented in the form of some indices. These indices expressed the aggregated effect of weather on a crop over time provided that:

* the influence of weather on yield over time is not uniform;
* weather has some signal useful for prediction long before harvest;
* two major weather parameters (precipitation and temperature) show some collinearity in their influence on yield.

Taking all these statements into consideration, the values of the departure of yield from trend were modeled as a function of combined effect of both precipitation and temperature over the entire period of growing grain and also over the pregrowing period. For this purpose and also to eliminate collinearity, two aggregations of the weather variables were computed based on the following equations:

First aggregation:
$$IV_i = \sum_{j=1}^{19} W_{ij} V_{ij} \quad (3)$$

$$W_{ij} = R_{ij} / \sum_{j=1}^{19} |R_{ij}| \quad (4)$$

Second aggregation:
$$IV = \sum_{i=1}^{2} W_i * IV'_i \quad (5)$$

$$W_i = R_i \ / \sum_{i-1}^{2} |R_i| \ , \quad (6)$$

where, for the first aggregation, IV_i is index weather variable i aggregated over time (19 months), V_{ij} is weather variable i for month j, W_{ij} is weight of weather variable i for month j, and r_{ij} is is the Pearson correlation coefficient between departure of yield from trend and weather variable i in month j. For the second aggregation, IV is index variable double weighted over time and over the weather variables (precipitation and temperature), IV'_i is the standardized index weather variable i, W_i is the weight of standardized index weather variable i, and R_i is the Pearson correlation coefficient between departure of yield from trend and standardized index weather variable i.

The first aggregation of weather over time involves the calculation of index precipitation (IP) and index temperature (IT) based on equation (3) and (4) for the entire growing and pregrowing seasons (19 months, from January of the previous year (number 1) to July of the current year (Number 19). The second aggregation involves combining the standardized IP and IT variables into an index-PT variable based on equations (5) and (6).

Models were developed for five economic regions of the USSR where barley is the prevalent crop. These regions are located in different climatic and soil zones. Barley normally occupies an area from 1.2 million hectares in the smallest region (Volgo-Vyatka) to 5.6 million hectares in the largest one (Kazakh). For each region, average yearly barley yields for the period 1945-1978 were used. Mean monthly temperature and total monthly precipitation averaged over the territory of each region for the same period were used as the major weather elements.

Results and Discussion

Barley yields in the five economic regions from 1945 through 1978 are shown in Figure 1. Estimates for regression coefficients and some statistics for equations of trend are given in Table 1. The model fitted for trend is $\hat{Y}_t = a_o + a_1 t + a_2 t^2$ where $\hat{Y}_t$ is the estimated yield, t = year - 1944 and a_o, a_1 and a_2 are parameters estimated by least squares methods.

The analysis shows that in regions located in areas with adequate water supply trend explains 76 to 88 percent of yield variance. In regions with a deficit of water, the proportion of potentially weather related yield variance increases

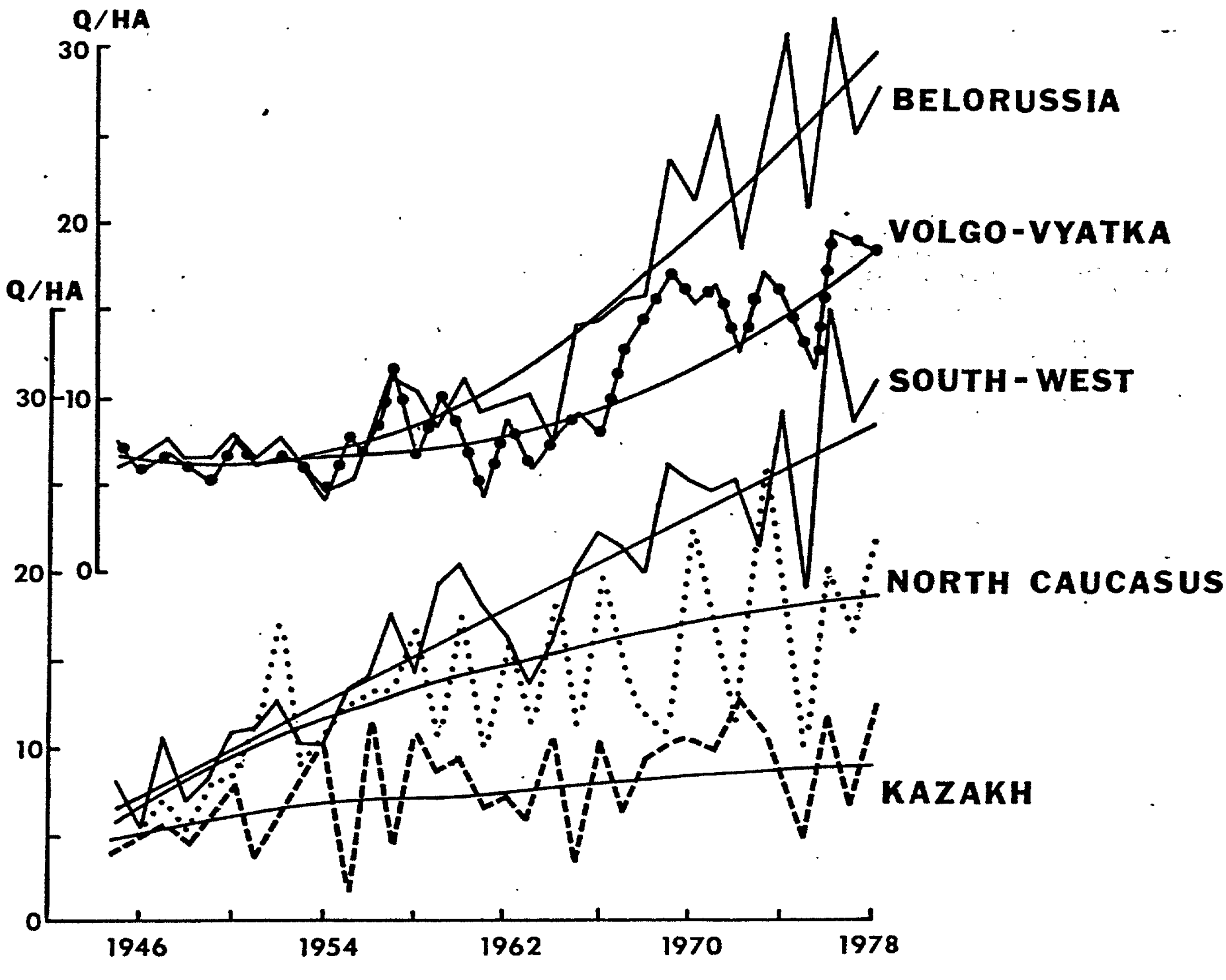

Figure 1. Barley yield and trend.

considerably and the share of trend-related yield variance constitutes less than 50 percent. Thus, when modeling the response of yield to environment, both components (trend and weather) must be taken into consideration.

Table 1. Estimates for regression coefficients and some statistics for equations of the trend shown in Figure 1

Economic Region	Water Supply	Estimates for regression coefficients			Percentage of yield variance explained by:	
		A_o	A_1	A_2	trend alone: (technology,) climate	remainder (weather,) other
South-West	Adequate	6.460	.566	.0033	85	15
Belorussia	Adequate	6.994	-.302	.0293	88	12
Volgo-Vyatka	Adequate	6.694	-.188	.0164	76	24
North Caucasus	Deficient	5.912	.576	-.0065	50	50
Kazakh	Deficient	4.546	.247	-.0032	25	75

Trend in yield series can be mostly explained by the changes in applied technology for growing crops and by the type of climate in a particular area which conditions the technology effect on the crop. As seen in Figure 1, this conditioned effect was expressed in different levels of barley yield in regions by the end of the period (1978) and non-linearity of trend over the entire period. As a result of this climate conditioning, in water deficit regions the applied technology and natural resources were not entirely compatible. Accordingly, increases in barley yield have been limited in the second part of the period. In regions with adequate water supply, natural resource factors have been more complementary to the improved technology and yield levels have responded with an increased growth rate.

Using estimates presented in Table 1 the trend for each region was calculated and a ratio of departures of the reported barley yield from this trend were computed. These departures were correlated with monthly precipitation and temperature of the harvest and previous years. These correlation coefficients for two of the regions are shown in Figure 2. This figure shows the effect of weather on barley yields is not uniform over time in magnitude and in sign; almost every month has some information which can be used as a signal for barley yield assessment; there are some periods of the year when the direction of the response of barley yield to weather is fairly similar even for regions with different climatic and soils conditions.

Correlation coefficients for barley yield departures from trend with weather were further used to "weight" the influence of weather of a particular month on barley yield based on equation (4) and to aggregate monthly precipitation and temperature over the harvest and previous year periods into index-variables based on equation (3). The second aggregation was then carried out based on equations (5) and (6) to compute the combined PT index. Estimates for regression coefficients for equations describing the relation between departures of barley yield from trend and index-PT are shown in Table 2.

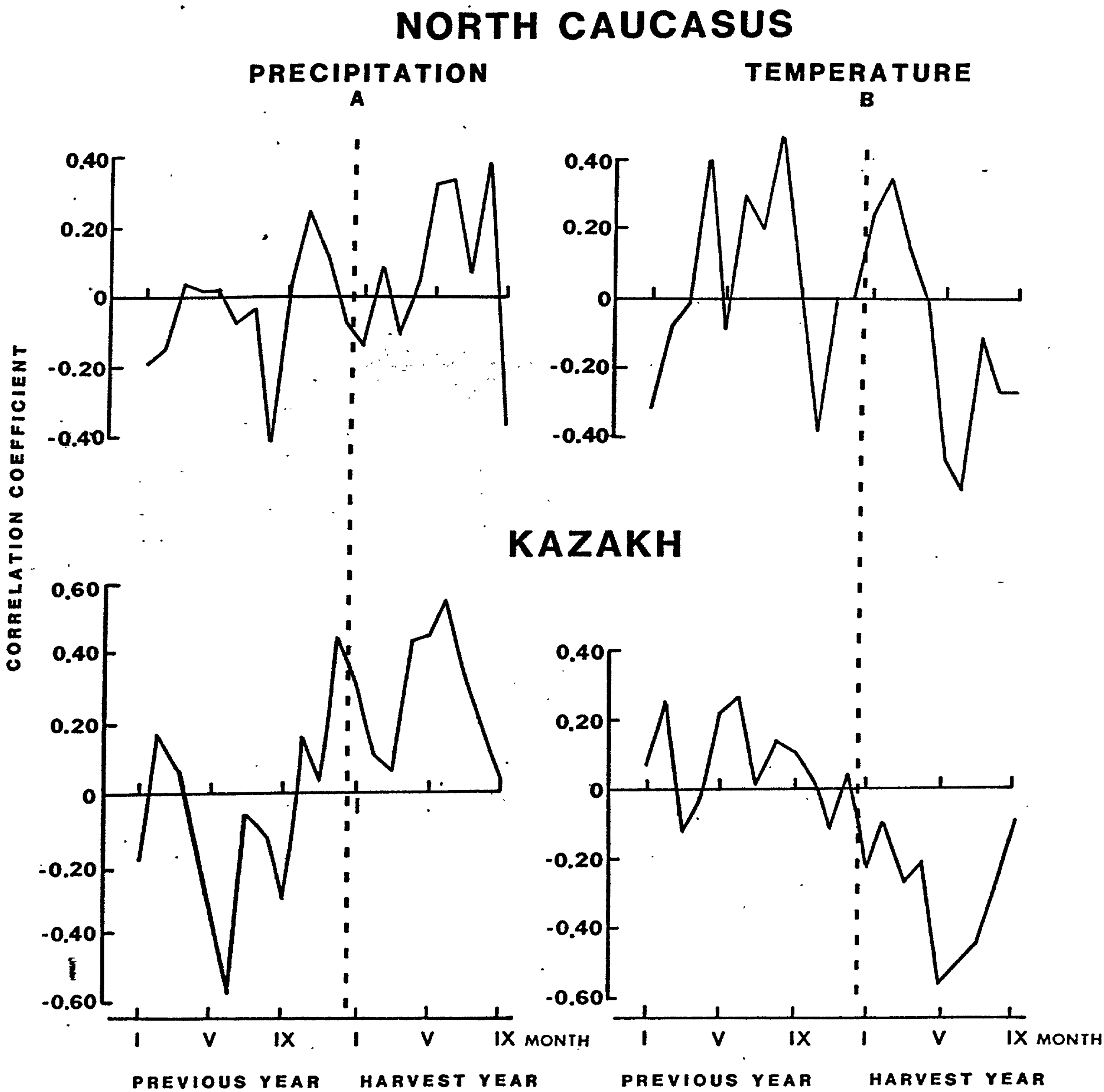

Figure 2. Correlation coefficients of barley yield departures from trend with precipitation and temperature in North Caucasus and Kazakh economic regions.

Table 2. Estimates for regression coefficients and some statistics for equations relating departures of barley yields from trend with index-PT

Economic region	Estimates of regression coefficients			R^2	Condition index
	Intercept	PT	PT^2		
South-West	100.013	14.213	-0.759	0.55	1.77
Belorussia	99.159	16.324	1.471	0.43	1.94
Volgo-Vyatka	99.610	17.802	4.302	0.48	2.30
North Caucasus	96.164	22.235	4.580	0.75	2.50
Kazakh	104.871	30.852	-5.550	0.78	2.24

According to the values of R^2, simulated models explain 43 to 78 percent of weather-related barley yield variance. This portion is larger in regions with shortage of water and smaller in regions with adequate water supply. Taking into consideration both simulated components trend and weather, the models explained 90 to 95 percent of barley yield variance in a dependent test.

These models were also verified in an independent test. For this purpose the bootstrap technique was used. This technique consists of developing a model for an earlier base period and applying this model to the data of the following year. The period of 1971-1978 was used for independent tests in all five regions. The results of these tests are shown in Figure 3. As seen in most of the cases, simulated yield corresponded very well to reported yield. The most important fact is that regions with different climatic and soil conditions and also with different levels of barley yield showed good accuracy in the independently simulated barley yields. In addition to this, it is necessary to mention that the developed models are not costly to operate and also to develop. They are consistent with current scientific knowledge concerning weather-crop-technology relationships for large areas. The lead time of a yield estimate, based on actual weather value data, is 2-3 months ahead of the barley harvest. These models are easy to understand. The models were developed in such a way that their redevelopment can begin as soon as the new weather and crop data are available.

References

Kogan, F. N., 1977. Cereal production in the Chernozem Zone of the USSR and the possibility for its prediction based on weather information. Trudy GMTS No. 174:3-23 (in Russian).

Obukhov, V. M., 1949. Yield and meteorological factors (in Russian). Gosplanizdat, Moscow: 349.

Thompson, L. M., 1969. Weather and technology in the production of wheat in the United States. Journal of Soil and Water Conservation, 23:219-224.

Wilson, W. W., T. L. Barnett, S. K. LeDuc, F. B. Warren, 1980. Crop yield model test and evaluation criteria. AgRISTARS Yield Model Development Project Document YMD-1-1-2.

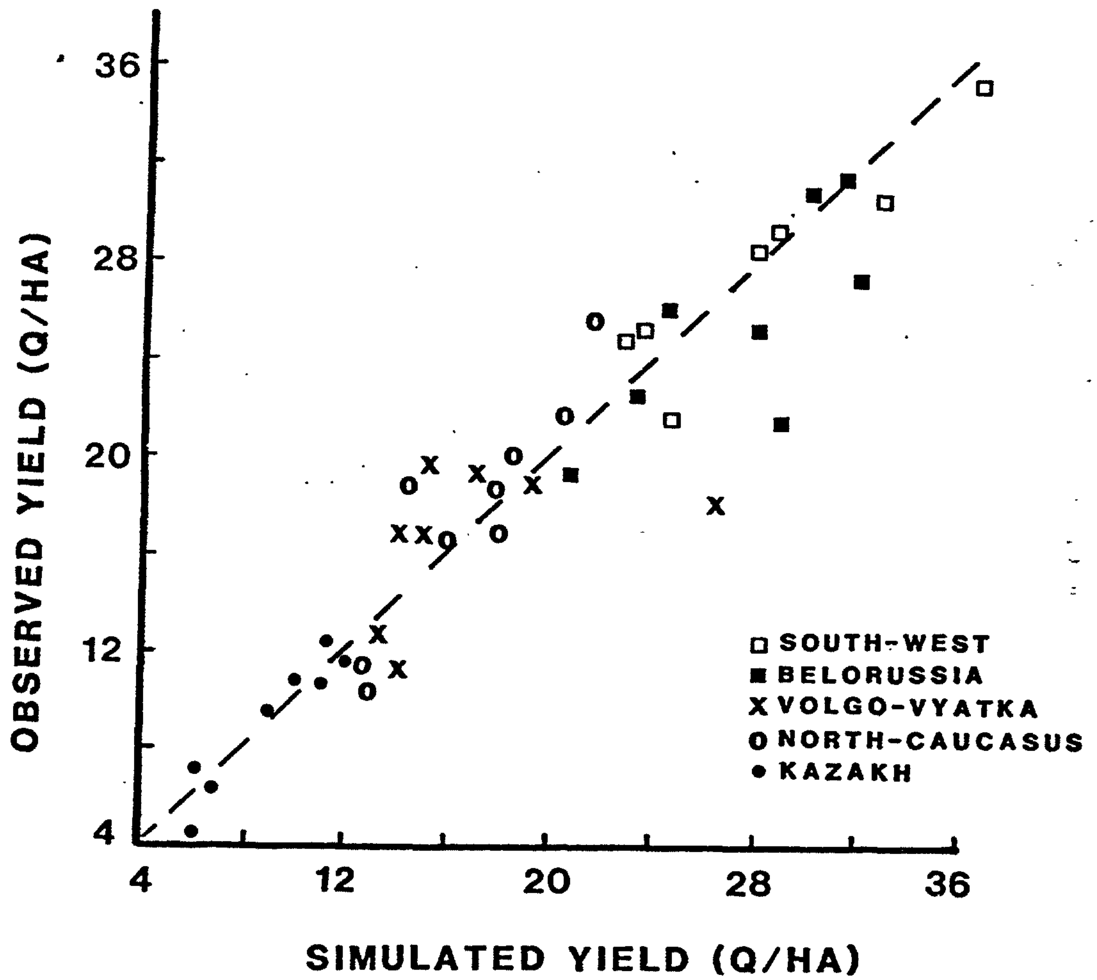

Figure 3. Correlation between observed and independently simulated barley yields.