

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1998 - 10th Annual Conference Proceedings

HURST PHENOMENON AND FRACTAL DIMENSIONS IN LONG-TERM YIELD DATA

Susanne Aref

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

Aref, Susanne (1998). "HURST PHENOMENON AND FRACTAL DIMENSIONS IN LONG-TERM YIELD DATA," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1275>

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.

HURST PHENOMENON AND FRACTAL DIMENSIONS IN LONG-TERM YIELD DATA

Susanne Aref

Department of Statistics

University of Illinois

A fractal dimension may be thought of as a measure of randomness. Fractal dimensions based on semivariograms have been used to determine degree of randomness in yearly crop yields. Through rescaled range analysis Hurst exponents also define fractal dimensions. This method of obtaining fractal dimensions gives more reasonable and sensitive measures than the semivariogram method. To address the inherent randomness due to yearly variations, global trends in yield must be removed before either method is applied. After detrending, a fractal dimension obtained from semivariogram is usually that of a random process. The Hurst method yields an exponent H , which results in a fractal dimension $D = 2-H$. The Hurst exponent $H = 0.5$ corresponds to a completely random system, $H = 1$ to a completely deterministic system. Natural processes such as river discharges, temperatures, precipitation, and tree rings have Hurst exponents about 0.72. The Hurst phenomenon is the occurrence of H greater than 0.5 corresponding to persistent Brownian motions, rather than equal to 0.5 corresponding to a random process. There is not much auto-correlation in the detrended crop yields, but the Hurst exponents from detrended yearly crop yields of Illinois soybean and wheat and of US soybean, wheat, and cotton are mainly between 0.5 and 1 suggesting long-term dependence similar to that of other natural processes. Illinois and US detrended yearly corn yield have exponents less than 0.5, corresponding to anti-persistent Brownian motions. Corn data from the Morrow Plots also have Hurst exponents less than 0.5 for two plots that were either over-fertilized or previously not treated, while an optimally treated (properly fertilized and previously manured) plot had an exponent greater than 0.5.

INTRODUCTION

Weather plays a major role in the amount of crop yields in any given year. Certainly crop yields depend on temperature and precipitation but also on other conditions such as management practices. Temporal variation raises the following questions:

- 1) How random is the variation from year to year in different crops?
- 2) Does randomness of yearly crop yields at different scales (United States, Illinois, and long-term experiments) exhibit the same pattern of randomness?
- 3) Can controlled long-term experiments such as the Morrow Plots identify management practices that produce less randomness?

The use of fractal dimensions may be a tool to address the degree of randomness in crop yields. These dimensions are defined in several different ways. Only the Hausdorff-Besicovitch dimension for curves in the plane will be considered here. For further discussion of fractal dimensions see Mandelbrot (1982) or Feder (1988). Berry and Lewis (1980) suggested to use semivariograms (see Clark (1979) for a discussion of semivariograms) to obtain fractal dimensions. Their method was based on the assumption of a stationary system. The slope, β , from the regression of $\log(\text{semivariogram})$ on $\log(\text{lag})$ was used to obtain the Hausdorff-Besicovitch fractal dimension, D , through the relationship, $\beta = 4 - 2D$.

Semivariograms based on yearly yields have been used to determine and compare fractal dimensions of ten crop yields in the US between 1930-1990 (Eghball and Power 1995). The semivariograms for the ten crop yields show drifts corresponding to trends in the data. The positive correlations between crop yields and year are mentioned, but the data are not detrended since "long-term variation would be removed and the detrended data resemble a data set with complete domination of short-term variation". However, if the data is not detrended, global trends will dominate and small-scale variation will not be addressed. Furthermore, the stationarity assumption for the use of a semivariogram to obtain a fractal dimension is violated.

The semivariogram method was also used for crop yields between 1953-1993 (Eghball et al., 1997), where again a positive trend is present in the crop yields. Large-scale variation may be modeled by monotonely increasing functions (for example simple polynomials) during periods where there are positive correlations between yield and year, such as in the periods 1930-1990 and 1953-1993. Residuals from the regressions have constant means (of 0) and therefore will not violate the assumption of stationarity. Unfortunately, detrending the data include the removal of any trends that may occur due to randomness of the time series. These trends cannot be separated out.

Crop yields between 1974-1992 were used to study whether fractal dimensions in long-term experiments can be used to differentiate between management practices (Eghball and Varvel, 1997). The raw data from the period between 1974 and 1992 used in Eghball and Varvel (1997) appear to be slightly positively correlated with year as well.

After detrending crop yield (or if there is no trend) a horizontal line, resulting in a horizontal log/log plot usually best describes the corresponding semivariogram. The fractal dimension obtained from such a semivariogram is that of a random process. If the data is not detrended the slope obtained is mainly a description of the drift.

The alternative method to obtain a fractal dimension for (detrended) crop yield data suggested in this paper is based on Hurst's rescaled range analysis (Hurst, 1951; Hurst et al., 1965). The technique produces a set of points corresponding to a period of a certain length. Each point at a period length is calculated from the whole data set much like values in a semivariogram. The log/log plot of the points versus period length results in a fairly straight line for which the slope is used to determine the fractal dimension through the equation $D = 2 - H$ (Fig. 1). Hurst used the technique on very long data records (Hurst et al., 1965), which is not possible for yield records where only about 130 years of data are available. Furthermore, the nature of developing technology has had such an impact on a crop like corn that it is more appropriate to only consider crop yields from 1970 and on.

Semivariograms and rescaled ranges were obtained using SAS macros (SAS Institute, 1989).

HURST EXPONENTS

Rescaled Range Analysis

Hurst originally devised rescaled range analysis to determine the size of a proposed reservoir on the Nile. To estimate a proper storage capacity of the prospective reservoir, Hurst used long-term records of annual discharges, ξ . The mean discharge $E_{m,k}\xi$ over the period from $m(k-1)+1$ to mk years is

$$E_{m,k}\xi = 1/m \sum_{i=m(k-1)+1}^{mk} \xi(i) \quad k = 1, \dots, a$$

where
 years

$$\begin{aligned} n &= \text{number of years} & m &= \text{length of interval of} \\ a &= \text{number of intervals} & am &\simeq n \\ \xi(i) &= \text{annual river discharge} & i &= 1, \dots, n \end{aligned}$$

The standard deviation $S_{m,k}$ of $\xi(i)$ for the period from $m(k-1)+1$ to mk , is defined as

$$S_{m,k} = 1/(m-1) \sum_{i=m(k-1)+1}^{mk} (\xi(i) - E_{m,k}\xi)^2$$

The average standard deviation S_m of ξ for intervals of length m is the mean of $S_{m,k}$

The discharges $\xi(i)$ were used to find the accumulated deviations $X_m(j)$ from the mean discharge $E_{m,k}\xi$ for each year j between $m(k-1)+1$ to mk .

$$X_m(j) = \sum_{i=m(k-1)+1}^j (\xi(i) - E_{m,k}\xi) \quad j = m(k-1)+1, \dots, mk$$

Note that $X_m(mk) = 0$ (Fig. 2.).

The range $R_{m,k}$ of the deviations $X_m(j)$ for the period from $m(k-1)+1$ to mk , is defined as

$$R_{m,k} = \max_{j=m(k-1)+1}^{mk} X_m(j) - \min_{j=m(k-1)+1}^{mk} X_m(j).$$

The average R_m over the intervals is the mean capacity that the reservoir must have in order to never overflow or empty in an m -year period.

The rescaled range for a period of length m is the dimensionless quantity R_m/S_m , (R_m adjusted for the average standard deviation S_m). For a more detailed discussion see Chap. 8 in Feder (1988).

Brownian motion

For Brownian motions R_m/S_m is proportional to the squareroot of the number of years, which is the exponent for a completely random series (Hurst, 1951; Feller, 1951). Hurst discovered that in the natural time series of yearly discharges of the Nile, R/S was proportional to the number of years to a power H (for Hurst exponent), where H was not equal to 0.5 as he expected but was greater than 0.5. A scatter plot of yearly discharges did not appear to have trends or cyclic periods.

The Hurst exponent H for fractal Brownian motion processes is between 0 and 1, corresponding to a fractal dimension from 2 to 1. For $H = 0.5$ the process is completely random, while $H > 0.5$ means the process has some dependencies so that positive steps make the process tend to have more positive steps and negative steps make the process tend to have more negative steps. Such a process is called persistent and agrees with complete dependency for $H = 1$. In the case where $H < 0.5$ the process is called anti-persistent. Here positive increments will make the process tend to have more negative increments and vice versa.

Natural time series

Hurst also calculated exponents for other natural time series such as other river discharges, river levels, temperatures, precipitation, and tree rings, where exponents mainly fall between 0.5 and 1 with a mean value of 0.72. This occurrence is often referred to as the Hurst phenomenon. It suggests that there is some degree of randomness but also a deterministic element, which may be thought of as a 'reservoir' effect. For an actual reservoir it is clear that high levels in the reservoir one year will tend to promote high levels in the reservoir the following year and similarly for low levels. Tree rings exhibit the same 'reservoir' effect, which may be explained by the fact that temperature and precipitation have average exponents of 0.68 and 0.70, respectively.

HURST EXPONENTS IN CROP YIELDS

It is not necessary to remove trends in the yield data in order to use Hurst's method as it was when using semivariograms to obtain fractal dimensions. However, if there is a positive trend in yearly crop yields and this is not removed, an exponent of about 1 will appear, resulting in a fractal dimension of 1 indicating a completely deterministic process. An existing positive trend in the data will simply show that the process is deterministic. The point of interest here is to identify the small-scale variation.

US and Illinois Crop Yields

Table 1 shows the Hurst exponents and fractal dimensions for US and Illinois corn, soybean, and wheat yield residuals, and also US cotton yield residuals and Morrow Plots corn yields and yield residuals. Only data between 1970 and 1997 were considered since technology

changes in this period are more homogeneous than in earlier periods. The changes in technology since 1970 have resulted in rather linear patterns in crop yield trends. Yields were detrended using simple linear regressions. Slopes, standardized slopes, and R^2 are shown in Table 1. The last two columns of Table 1 contain the semivariogram slope and corresponding fractal dimensions (almost equal to 0 and 2, respectively).

The plot of yearly US crop yields in Fig. 3 shows the positive trends for the period considered 1970-1997. The actual yield variation is much larger in corn than the other crops but that is not the concern of the rescaled range analysis - it is how the corn yield varies. The standardized yield residuals are on the same scale in Fig. 4 and residuals corresponding to the close yields in a row are easier to distinguish. It does appear that corn crop residuals have fewer runs in the same direction. Soybean, wheat and cotton yields have Hurst exponents of about 0.70, with soybean and wheat in the same order for Illinois and US yields. Corn yields show anti-persistent behavior with exponents that are less than 0.50. Illinois corn yield shows less anti-persistence than US yield with an exponent of 0.40 compared to the exponent of 0.33 for US yield.

Morrow Plots' Corn Yield

Table 1 also contains Hurst exponents for corn yield from the Morrow Plots. These plots at University of Illinois were started in 1876 and are the oldest field experiment in the US. Only the M-NPK plot yield had a slight positive trend in the period between 1970 and 1997 and was detrended. The other plots do not have an increasing trend for the period between 1970 and 1997, so the raw yield data were used. The treatment plots that are comparable to today's farming practices are labeled U-NPK, M-NPK, and H-NPK.

All three plots receive commercial fertilizer, nitrogen, phosphate, and potassium (NPK). U-NPK and M-NPK plots have received University of Illinois recommended fertilizer levels based on soil tests since 1955, while the H-NPK plot has received 1.5 times the recommended fertilizer levels since 1967. The U-NPK plot was untreated from 1876 until 1955, while the M-NPK and H-NPK plots were untreated from 1876 to 1903. These plots were then treated according to the Illinois system of permanent soil fertility, which includes a treatment of manure, lime, and phosphate, from 1904 to 1954 for the M-NPK plot and from 1904 to 1966 for the H-NPK plot (Aref and Wander, 1998). The Hurst exponents for the U-NPK and H-NPK plots are slightly higher than the Illinois exponent but still lower than 0.50 at 0.45 and 0.44, respectively, while the exponent for the M-NPK plot at 0.58 is greater than 0.50 (still 0.10 less than exponents for the other crop yields). Previous management practices in the Morrow Plots may explain the rather different results for U-NPK and M-NPK plots, which are now treated similarly. Over-fertilization may explain the rather different results for M-NPK and H-NPK plots, which were treated similarly until 1955.

The two long-range time series available in the Morrow Plots experiment produced exponents from a 94-year period that agreed with the short-term exponents. The plots are U and M in Table 1. The U plots did not receive any inputs for 122 years and the M plot was treated according to the Illinois system of permanent soil fertility for 94 years. The short-term period of

28 years showed persistent behavior of both the U plot and the M plot time series possibly due to much lower yields (than the NPK plots) and thus not responding as much to weather in the absence of commercial fertilizer. Though the soil organic matter (SOM) are at different levels in the U and the M plots with the higher SOM level in the M plot supporting a much higher yield, the yield trends look almost parallel (Aref and Wander, 1998). The similarity of the Hurst exponents is then not surprising. When using the 94 years yield time series available for the M plots, the Hurst coefficients did not change much for either the U plots or the M plot. $H = 0.61$ for U plots and $H = 0.58$ for the M plot as compared to the short-term time series values of $H = 0.66$ for the U plots and $H = 0.65$ for the M plot (Table 1).

CONCLUSION

Fractal dimensions derived using Hurst exponents (or the exponents themselves) are useful in describing degree of randomness of yield time series. They confirm knowledge of randomness of crop variation and provide a measure of variation. One problem is that the rescaled range analysis was developed for very long-term time series. The rescaled range analysis has also been used for tree ring records, which are at about the same length as yield records. Here the method was used for only the last 28 years of yield records, because of major changes in technology.

To validate the appropriateness of fractal dimension as a measure of randomness exponents were compared at different scales, US, Illinois, and Morrow Plots yield. The US and Illinois exponents for corn, soybean, and wheat yield residuals were comparable with slightly higher exponents at the Illinois scale.

The treatment plots in the Morrow Plots experiment with exponents closest to the Illinois corn yield exponent were the U-NPK and H-NPK plots. The former plot received no inputs for 80 years thus being depleted in SOM before current inputs of University of Illinois recommended fertilization levels based on soil tests was started in 1955. The latter plot received no inputs for the initial 30 years of its existence and was then was treated according to the Illinois system of permanent soil fertility for 60 years before a super-fertilization scheme was applied from 1967 to 1997. The exponents for the U-NPK and H-NPK plots were slightly higher than the Illinois corn yield exponent but still less than 0.5.

The remaining plot that corresponds to recommended farming practices at any time is the M-NPK plot. Apart from the initial 30 years of depletion, the plot was treated according to the Illinois system of permanent soil fertility until 1955, when the same current practice as on the U-NPK plot was started. The M-NPK plot exponent was greater than 0.5, which was greater than the Illinois and US exponents for corn, possibly indicating that there may be depletion of SOM as in the U-NPK plot and/or over-fertilization as in the H-NPK plot happening on US farmland.

Applying rescaled range analysis to the short-term data may not be valid in terms of the actual size of the exponents, though the long-term plots in the Morrow Plots experiment produced similar coefficients for the 94-year period as compared to the 28-year period. Further

REFERENCES

- Aref, S. and M.M. Wander. 1998. *Advances in Agron.* 62:153-197
- Berry, M.V. and Z.V. Lewis. 1980. On the Weierstrass-Mandelbrot Fractal Function. *Proc. R. Soc. Lond. A.* 370:459-484
- Clark, I. 1979. *Practical geostatistics.* Appl. Sci. Publ., London.
- Eghball, B. and J.F. Power. 1995. Fractal description of temporal yield variability of 10 crops in the United States. *Agron. J.* 87:152-156
- Eghball, B., J.D. Binford, J.F. Power, D.D. Baltensperger, and F.N. Anderson. 1997. Maize temporal yield variability under long-term manure and fertilizer application: Fractal analysis. *Soil Sci. Soc. Am. J.* 59:1360-1364
- Eghball, B. and G.E. Varvel. 1997. Fractal analysis of temporal yield variability of crop sequences: Implication for site-specific management. *Agron. J.* 89:851-855
- Feder, J. 1988. *Fractals.* Plenum Press. New York.
- Feller, W. 1951. The Asymptotic Distribution of the Range of Partial Sums of Independent Random Variables. *Ann. of Math. Stat.* 22: 427-432.
- Hurst, H.E. 1951. Long-term Storage Capacity of Reservoirs. *Trans. Amer. Soc. Civ. Engrs.* 116:770-808.
- Hurst, H.E., R.P. Black and Y.M. Simaika. 1965. *Long-term Storage.* Constable, London.
- Mandelbrot, B.B. 1982. *The fractal geometry of nature.* W.H. Freeman, San Francisco.
- SAS Institute. 1989. *SAS Stats User Guide Vol I & II.* SAS Inst. Cary, NC.

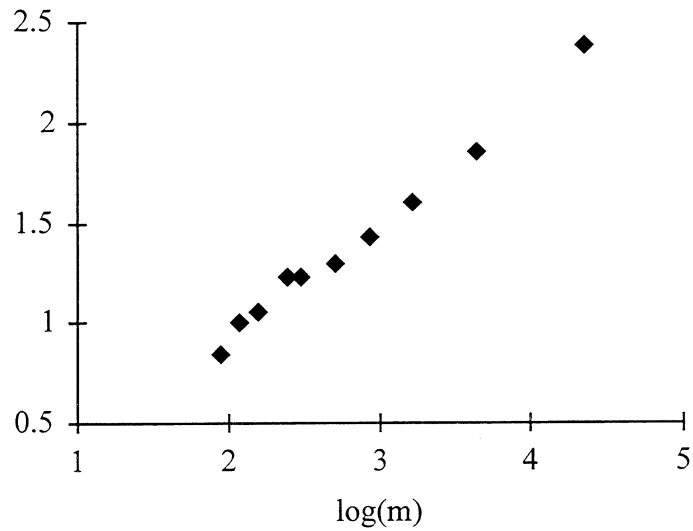


Figure 1. Example of the rescaled range and period relationship.

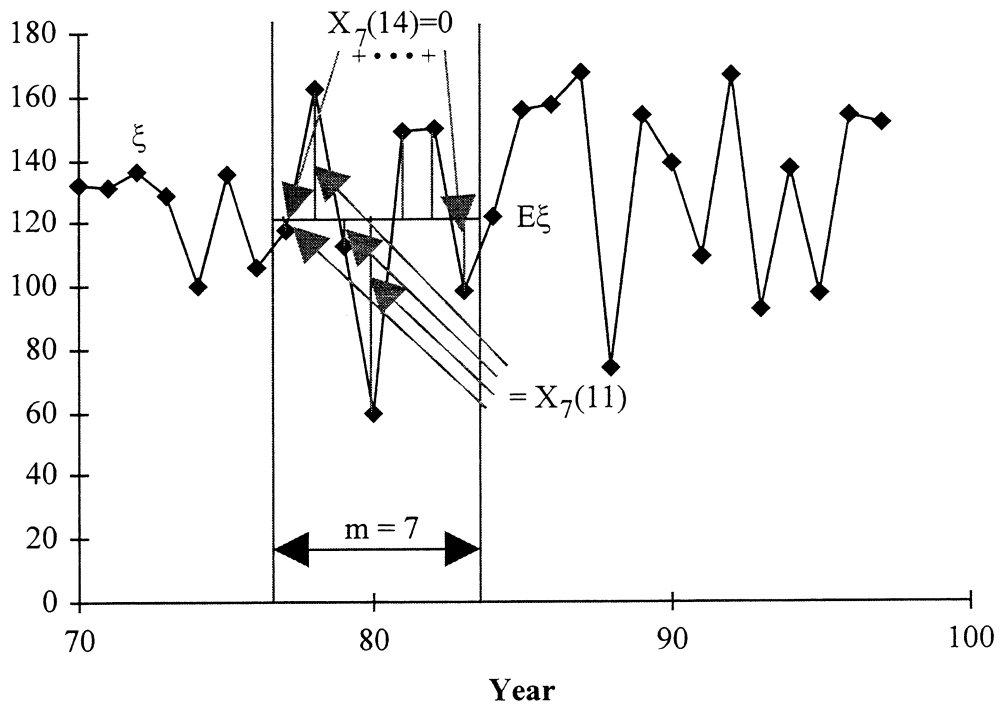


Figure 2. The construction of accumulated deviations $X_m(j)$ of corn yield residuals ξ from $E\xi$.

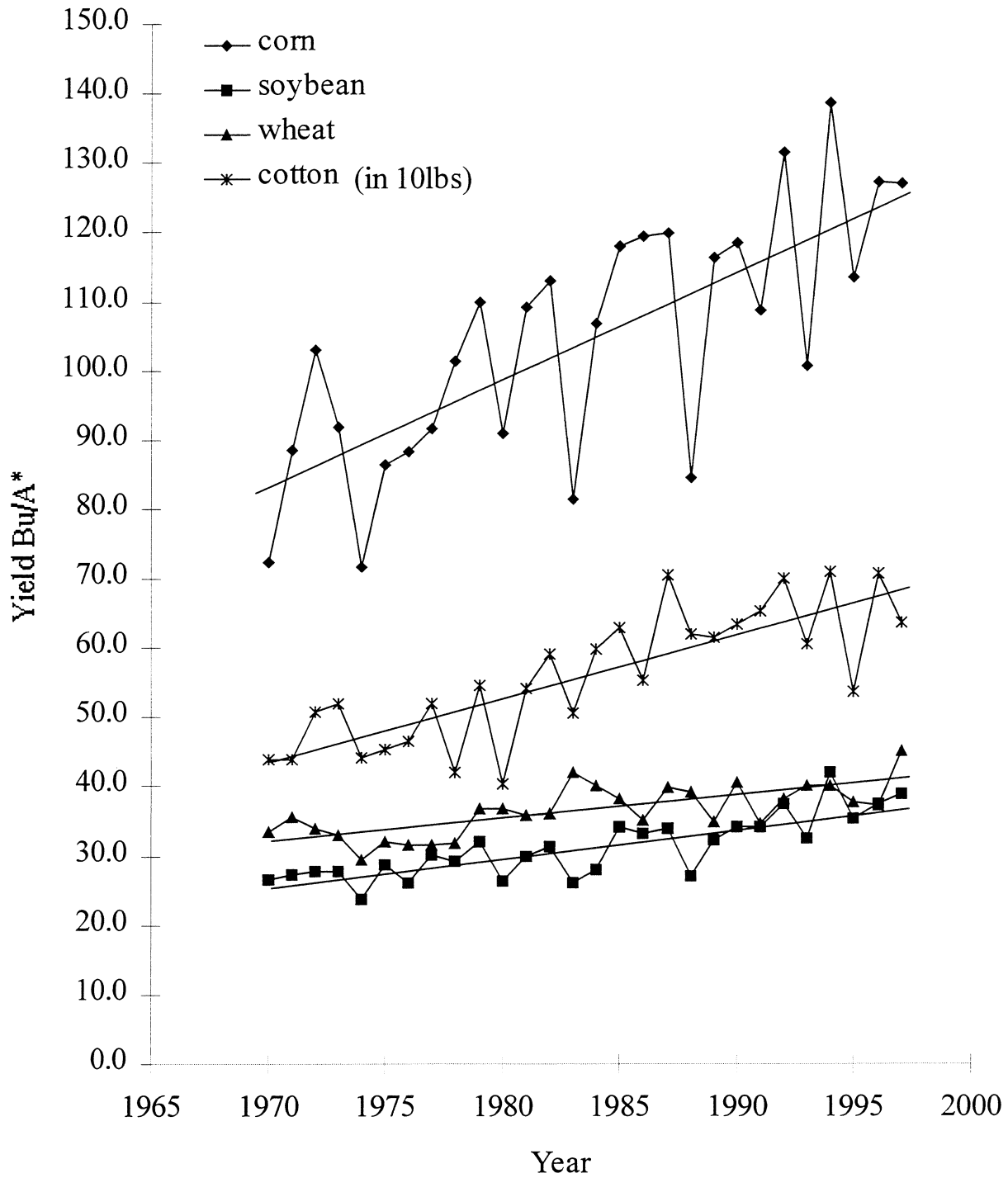


Figure 3. Yields of corn, soybean, wheat, and cotton in the United States, 1970-1997.

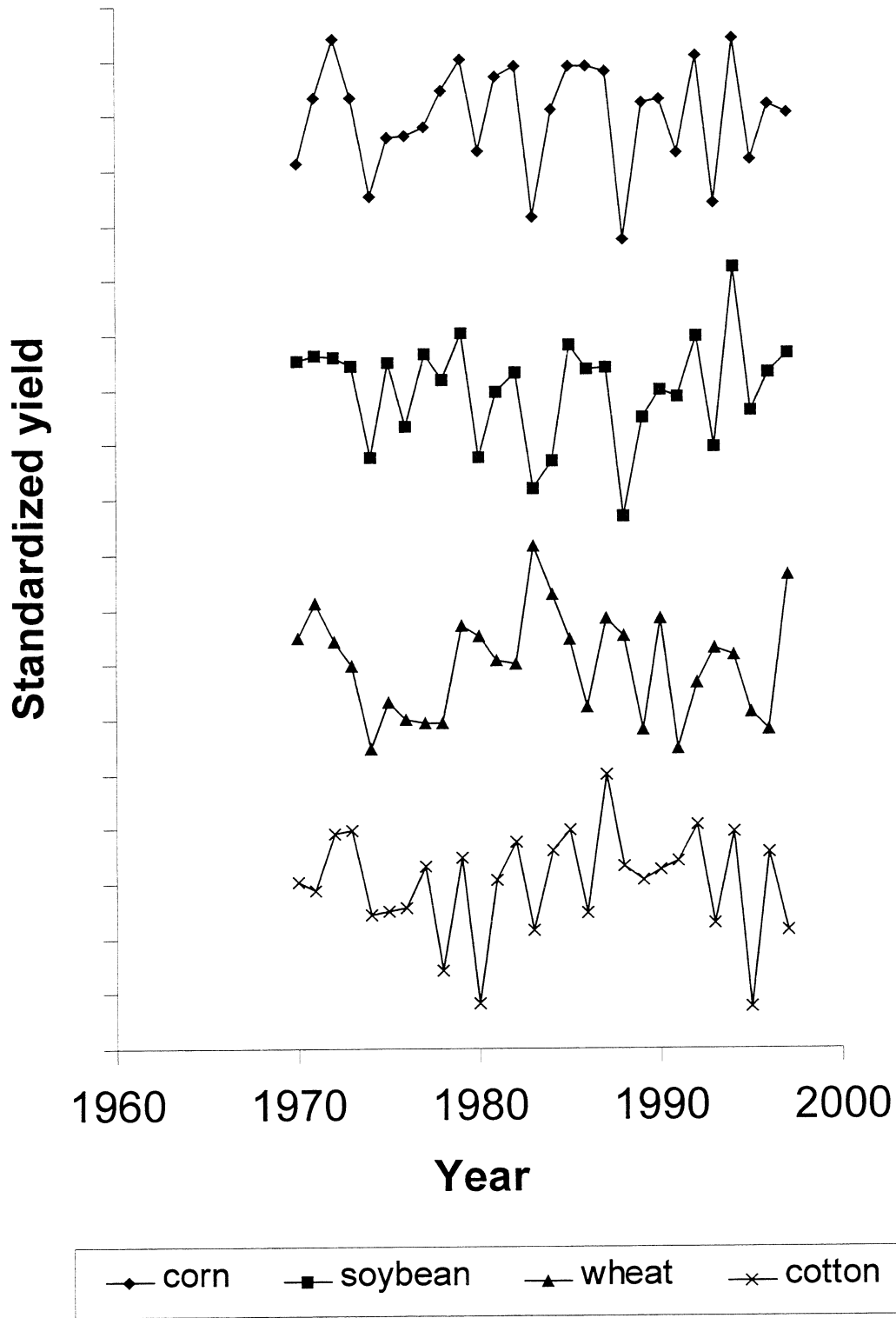


Figure 4. Standardized yield residuals for US crops.

Table 1. Rescaled range analysis and semivariogram of residual crop yield from simple linear regressions.

System	Crop	Rescaled range analysis		Yield regression			Semivariogram	
		H	D=2-H	slope	std slope	R ²	β	D=2-β/2
US	Corn	0.33	1.67	1.61	0.74	0.55	-0.07	2.03
	Soybean	0.67	1.33	0.44	0.81	0.66	0.08	1.96
	Wheat	0.71	1.29	0.31	0.70	0.49	-0.09	2.05
	Cotton	0.72	1.28	9.19	0.80	0.64	0.06	1.97
Illinois	Corn	0.40	1.60	1.42	0.55	0.30	-0.08	2.04
	Soybean	0.73	1.27	0.54	0.65	0.42	0.08	1.96
	Wheat	0.75	1.25	0.40	0.42	0.18	0.13	1.94
Morrow Plots								
M-NPK	Corn	0.58	1.42	1.35	0.33	0.11	-0.09	2.04
U-NPK*	Corn	0.45	1.55	0.41	0.12	0.01	-0.07	2.04
H-NPK*	Corn	0.44	1.56	0.52	0.10	0.01	-0.08	2.04
U*	Corn	0.66	1.34	-0.07	-0.05	0.00	-0.06	2.03
M*	Corn	0.65	1.35	-0.02	-0.01	0.04	-0.09	2.04

* Yield was not detrended since the slope was not significant.