A COMPARISON OF CLASSICAL STATISTICS AND GEOSTATISTICS FOR ESTIMATING SOIL SURFACE TEMPERATURE

By

PARICHEHR HEMYARI

Bachelor of Science
Pahlavi University
Shiraz, Iran
1977

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1980

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
December, 1984

A COMPARISON OF CLASSICAL STATISTICS AND
GEOSTATISTICS FOR ESTIMATING SOIL
SURFACE TEMPERATURE

Thesis Approved:

D L Nofziger
Thesis Adviser

Marvin S. Keener

J M Stone

Laval M. Verhalen

Norman N. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

Much appreciation and gratitude is extended to all of those who assisted in this research:

To my thesis adviser, Dr. D. L. Nofziger, to whom I am deeply indebted for his guidance, assistance, and friendship.

To the other members of my committee: Dr. J. F. Stone, Dr. L. M. Verhalen, and Dr. M. S. Keener for their suggestions and constructive criticisms.

To Mr. H. R. Gray, Mr. J. R. Williams, Mr. B. A. Bittle, Mr. K. W. Grace, and Mr. D. A. Stone for their technical assistance in accomplishing research goals.

To Dr. P. W. Santelmann and the Oklahoma State University Agronomy Department for my research assistantship and the necessary research funds.

Finally, my deepest and sincerest appreciation to my family in Iran, and also to my family away from home, Mr. and Mrs. H. R. Gray.

This acknowledgment would not be complete without a very special thanks to God our Lord for keeping me in faith and giving me strength throughout my life.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure | Page

CHAPTER I

INTRODUCTION

Many soil parameters vary over distance and time. However, it is not practical or economical to measure the values of those parameters everywhere in the field. One must take measurements at several selected locations and then use an estimation procedure to predict the values of those parameters at locations not sampled. To obtain accurate estimates a "good" estimation method must be utilized. A good method is usually consistent, unbiased, and efficient. It should provide reliable estimates and a method for determining estimation variance. That variance is a measure of the precision with which the actual values can be estimated. It should be as small as possible.

Several estimation methods from classical statistics are available for use. The least-squares method has been used in soil science for many years. It provides estimated values at unsampled locations as well as an estimation variance.

In recent years geostatistical estimation methods have been used for analyzing spatial variability of soil parameters and for preparing isarithmic maps of soil properties. Frequently, measurements taken close together in space or in time give values of approximately the same magnitude while measurements farther apart tend to give values differing by a greater amount. In other words, measurements close together are spacially dependent; and measurements farther apart are spacially independent. Geostatistical estimation methods take advantage of spatial dependence.

The ultimate test of any estimation method is its ability to reliably predict values at unsampled points in space or in time. To evaluate different estimation methods, measured values must be compared to the predicted values for each method. The primary purpose of this research was to compare the geostatistical estimation method known as "kriging" with the classical estimation method.

The objectives of this research were as follows:

1. To determine variation in soil temperature at the 5-cm depth over distance and time by means of semi-variograms and classical statistics,
2. To compare measured temperatures with temperatures estimated by simple kriging,
3. To compare measured temperatures with temperatures estimated by the least-squares method,
4. To compare the actual estimation variance and predicted estimation variance for simple kriging and for least-squares method, and
5. To compare the actual estimation variance for simple kriging with the actual estimation variance for the least-squares method.

CHAPTER II

LITERATURE REVIEW

A. Introduction to Geostatistcs

The field of geostatistics was developed by George Matheron and his coworkers at the Morphological Mathematical Center at Fontainebleau, France about 20 years ago. Geostatistics has been used extensively by South African and French mining engineers. Mining engineers are particularly interested in optimizing sampling patterns and estimation methods. They want to estimate the amount of minerals in ore deposits precisely because overestimating or underestimating them can have serious economic consequences. They can not collect too many samples for improving the estimation process because each sample costs considerable expense and labor (Clark, 1979).

The term "geostatistics" designates the statistical study of a natural phenomenon which is characterized by the distribution of one or more variables in space or time (Journel and Huijbregts, 1978). Geostatistics is based on the concept that a sample value is expected to be affected by its position and its proximity to neighboring positions (Clark, 1979).

B. Geostatistical Assumptions Versus Classical Statistical Assumptions

In classical statistics, the variance is assumed to be totally random. In geostatistics, the variance is assumed to be partly random and partly spatial. In classical statistics, all the samples are assumed to come from one distribution (Steel and Torrie, 1980). This is referred to as the "stationarity assumption". Geostatistics accepts the concept that each point in the field represents a sample from some distribution, but the distribution at any one point may differ completely from that at all other points in its shape, mean, and variance. Differences in sample values that are the same distance apart define a distribution. Geostatistics assumes that these differences in sample values come from a single distribution. In other words, the distribution of differences in sample values separated by a specified distance is assumed to be the same over the entire field. This is referred to as a "quasi-stationarity assumption" (Clark, 1980).

Most distributions are described by their mean and variance. In classical statistics, the mean of the distribution is an estimate of the sample value at each point (Steel and Torrie, 1980). In geostatistics, the mean of the distribution is an estimate of the differences in sample values separated by a specified distance. Geostatistics assumes that the mean of the differences in sample values is zero (Clark, 1980).

C. Geostatistical Stages

Geostatistics consists of two stages known as semi-variogram construction and kriging (Clark, 1980). In geostatistics, if the sample values are highly correlated, the random variance of the distribution of differences in sample values is relatively small. If the sample values are not correlated, this variance is larger. This variance is a measure of similarity, on the average, between points a given distance apart. Half of this variance is called "semi-variance". The graph of semi-variance versus distance or time is called a "semi-variogram".

Figure 1 shows a typical semi-variogram. The horizontal axis shows the distance between samples, i.e., the "lag distance". The vertical axis shows the semi-variance. Typically, the semi-variance increases initially with distance and then flattens out. The semi-variance at the point that it becomes flat is equal to the variance of sample values in classical statistics. The distance at this point is called "range of influence". In the range of influence the total variance is divided into random and spatial components. In this range the spatial variance is subtracted from total variance so the random variance is less than that in classical statistics. This range is important in selecting a sampling pattern. If it is large, then samples should be taken at relatively large intervals. If it is small, then samples should be taken at relatively small intervals (Clark, 1980).

Semi-variance is one-half of the sum of squares of differences in sample values separated by a specific distance divided by the number of pairs (Burgess and Webster, 1980). The following formula can be used to calculate semi-variance:

$$S(H) = \frac{1}{2(N-H)} \sum_{I=1}^{N} (T_I - T_{I+H})^2 \qquad (1)$$

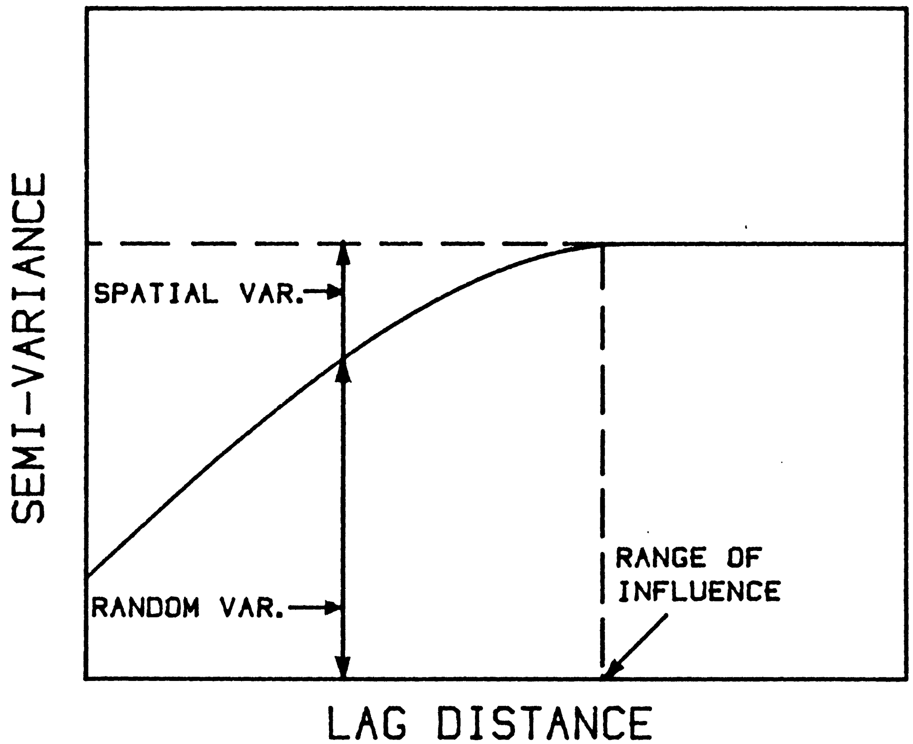

Figure 1. A Typical Semi-variogram.

where S(H) is the semi-variance, H is 1,2,3,... (number of lags between the samples), N is the number of samples, and $T_I - T_{I+H}$ is the difference in sample values separated by lag H (Clark, 1980). A lag is an interval in time or distance. The semi-variance is calculated for different lags and a semi-variogram is constructed by plotting the semi-variance as a function of lag distance or lag time. A mathematical model is then fitted to the semi-variogram. Some of the theoretical models that have been used to fit semi-variograms are linear, spherical, exponential, and Gaussian (Journel and Huijbregts, 1978). Figure 2 shows some of these models. (These examples simply illustrate the shapes of different models. As they are drawn, they do not describe the same data set). Semi-variogram construction is a critical stage in geostatistics because the model chosen to fit into the semi-variogram will be used throughout the second stage or kriging process and it will affect all subsequent results (Clark, 1980).

The essence of a "good" estimation method is not simply to produce a number, but it is also to give some estimate of the amount by which the actual value may vary from that estimate (Clark, 1980). The estimation variance is a measure of the extent to which an estimate approaches its actual value. In addition, a good estimation method is usually consistent, unbiased, and efficient; and it yields a minimum estimation variance. An estimation method is consistent if the probability of the estimate to be the same as the actual value approaches one when number of samples approaches infinity. An estimation method is unbiased if the expectation of the estimate is equal to the actual value. An estimation method is efficient if it is mathematically simple and not time consuming (Mikhail and Ackermann, 1976).

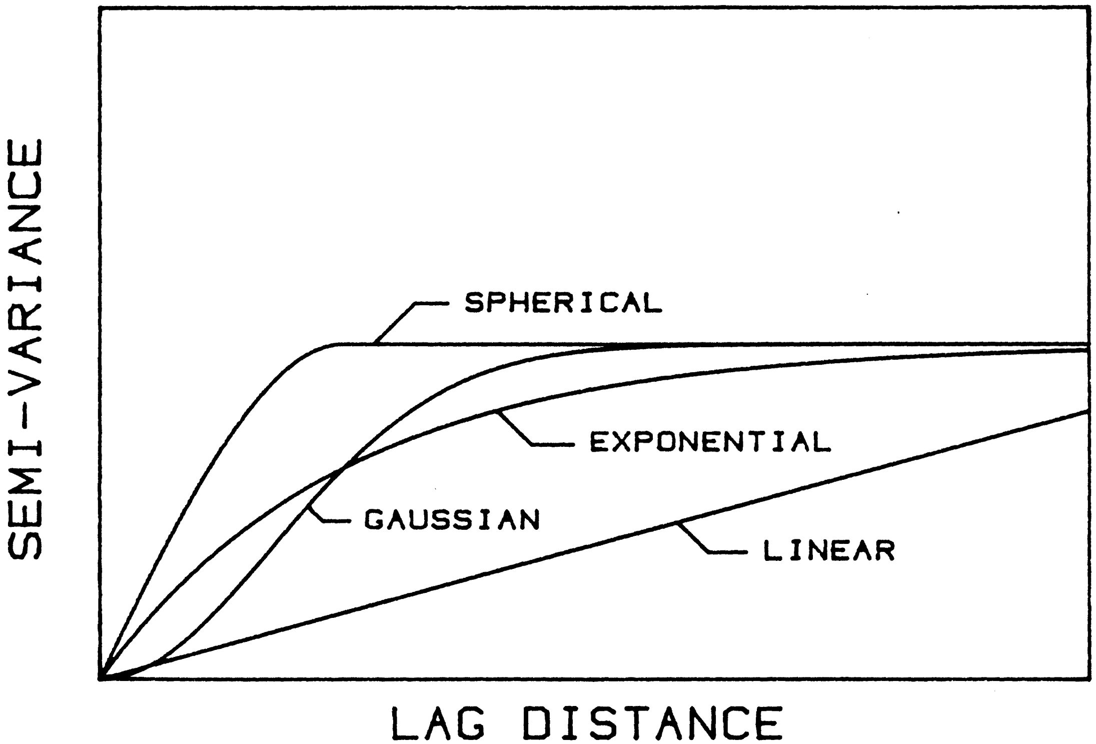

Figure 2. Four Different Semi-variogram Models.

Geostatistical estimation methods are based on the study of spatial variability as reflected in the semi-variogram. Kriging is a form of weighted local averaging. It is optimal in the sense that it provides estimates of values at unsampled locations without bias and with minimum and known variance (Webster and Burgess, 1980). In kriging, a set of weights must be found. When these weights are multiplied by the measured values, one obtains an estimate such that the error associated with this estimate is less than that for any other set of linear weights (Journel and Huijbregts, 1978). The following matrix equation has been used to calculate the set of weights W_i for i=1,2,3,...,n where n is the number of measured values used in the estimation process:

$$[A]\,\vec{W} = \vec{B} \tag{2}$$

where

$$[A] = \begin{bmatrix} S(X_1,X_1) & S(X_2,X_1) & \cdots & S(X_n,X_1) & 1 \\ S(X_1,X_2) & S(X_2,X_2) & \cdots & S(X_n,X_2) & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ S(X_1,X_n) & S(X_2,X_n) & \cdots & S(X_n,X_n) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \tag{3}$$

$$\vec{B} = \begin{bmatrix} S(X_1,X_0) \\ S(X_2,X_0) \\ \vdots \\ S(X_n,X_0) \\ 1 \end{bmatrix} \quad (4) \qquad \vec{W} = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \\ \mu \end{bmatrix} \quad (5)$$

Here $S(X_i, X_j)$ is the value of the semi-variance when the lag distance is $|X_i - X_j|$, and μ is a Lagrange multiplier.

The predicted estimation variance σ^2 by kriging is given by:

$$\sigma_k^2 = \vec{B}^T \vec{W} \tag{6}$$

where $\vec{B}^T$ is the transpose of vector $\vec{B}$ and $\vec{W}$ is defined in equation (5).

The predicted estimation variance for the least-squares method (Steel and Torrie, 1980) is given by:

$$\sigma^2 = S^2_{T.X_i} \left(1 + \frac{1}{N} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{N} (X_i - \bar{X})^2}\right) \tag{7}$$

where σ^2 is the predicted estimation variance for the estimated temperature T at a distance X_0, $S_{T.X_i}$ is the standard error of the estimates, N is number of samples, X_0 is the position at which temperature T is estimated, X_i for i=1,2,3,...,N are positions at which temperatures were measured.

In "simple" kriging, the value of a soil property is estimated at one point. In "block" kriging, value of the soil property is estimated over an area rather than at a point. In "universal" kriging, the value of the soil property is estimated for a volume of soil (Burgess and Webster, 1980; Webster and Burgess, 1980).

D. Geostatistical Applications in Agricultural Research

Burgess and Webster (1980) applied geostatistics to three sets of data from detailed soil surveys in Central Wales and Norfolk. They observed that sodium content at Plas Gogerddan varies isotropically with a linear semi-variogram. They used simple kriging and produced a map with intricate isarithms and fairly large estimation variance due to large random variance. The estimation variance for the central portion of the field was as high as 10.72 which was quite large when compared to sodium content values in the range of 15 to 30 meq/10 kg. The stone content of soil on the same land varied anisotropically with a linear semi-variogram. Again the estimation error was quite large. At Hole Farm, Norfolk, the depth to sand and gravel varied isotropically, but with a spherical semi-variogram. This semi-variogram was used for kriging, and an isarithmic map was produced from kriged values.

Vieira et al. (1981) studied the spatial variability of field-measured infiltration rate using geostatistics. They used a variogram constructed from 1,280 measured values of infiltration rate to krige 800 additional values. They observed that the kriging estimates were exceptionally good because the linear correlation coefficient for the measured and estimated values was 0.96, the mean estimation error was not significantly different from zero, and the estimation variance was relatively small.

Uehara (1982) observed that semi-variograms of exchangeable sodium percentage (ESP) showed a spatial relationship between samples taken in a distance of 3.5 to 4.0 km on the Kenana sugar project, Sudan. Semi-variograms were used to krige ESP in a grid pattern along the field. The estimation variance of kriged values increased only slightly using 56% of the samples compared to kriging based on all the samples. The mean estimation variance was 10.5 when 100% of the samples were used in the analysis and it was 13.1 when 56% of the samples were used. They concluded that geostatistics can help soil survey by obtaining similar results with fewer samples.

Vauclin et al. (1982) studied the spatial variability of soil surface temperature along two transects of a bare field at the University of California at Davis. Soil surface temperatures were correlated over space. Temperature measurements were taken 1 m apart along the transects using two infrared thermometers. Measurements with both thermometers were taken for 3 consecutive days between 1230 to 1330. Semi-variograms were constructed for almost half the length of the transects. These semi-variograms show a random variance and sills. All the semi-variances became constant after a range of influence of at least 8 m. Linear models were fitted into the semi-variograms.

Sometimes, it is possible to take advantage of one variable which has been sampled sufficiently to provide estimates of another variable which has not. In this case, the cross correlation or the cross semi-variogram between the variables must be calculated and cokriging must be used to obtain estimates of the variable not sampled sufficiently (Journel and Huijbregts, 1978). Vauclin et al. (1983) studied spatial variability of sand, silt, and clay contents, available water content (AWC), and water stored at 1/3 bar (pF2.5) by using classical statistics and geostatistics. Samples were taken within a 70 X 40 m area with nodes in a 10-m square grid. Sample means, variance, and coefficients of variation for all variables were determined using classical statistics. Linear correlations between available water content and textural components, and between water stored at 1/3 bar and textural components were established by assuming that all the samples were independent. The highest correlation was found between available water content and sand. No significant correlation was found between water stored at 1/3 bar and either the silt or clay content. Semi-variograms for all the variables and cross semi-variograms for the spatial correlation between available water content, water stored at 1/3 bar, and sand content values were used to krige and cokrige additional values of available water content and water stored at 1/3 bar every 5 m. Although the variables were found to be normally distributed over the field, the use of semi-variogram showed that the samples were autocorrelated within distances ranging from 26 m for water stored at 1/3 bar to 50 m for silt content. Mean values for available water content and water stored at 1/3 bar were 11.53 and 22.74%, respectively. Estimation variances for the kriged values of available water content and water stored at 1/3 bar at the center of the field were 4.06 and 10.25, respectively. The kriged and cokriged values were compared to the actual measured values, and the advantage of cokriging over kriging was demonstrated by comparing the estimation variances at the estimated points. For a limited number of samples, cokriging could be a promising tool to provide unbiased estimates at unrecorded points and also to provide a minimum estimation variance.

Palumbo and Khaleel (1983) used kriging to estimate transmissivity values (amount of water obtainable from an aquifer under a unit hydraulic gradient) in the Santa Fe aquifer in Mesilla Bolson, New Mexico. They applied kriging to 141 transmissivity values to evaluate transmissivity distribution and produced contour maps of estimated transmissivity values and associated estimation variances. An exponential model was fitted into the variogram. The range was 3 miles, and the average variance was 2.74 with a mean of 8.65 gpd/ft. Kriged estimates were generally lower than estimates based on available transmissivity maps.

Russo (1983) used geostatistics to analyze the spatial variability of two measured soil hydraulic parameters. One of those parameters was saturated hydraulic conductivity (K_s). A spherical and a linear variogram were used to calculate kriging estimates and estimation variances of log K_s from 31 observed values at the nodes of a 10 X 10 m square grid. The two variograms resulted in kriging estimates which were not significantly different.

Tabor et al. (1984) studied the spatial variability of nitrate in irrigated cotton (Gossypium Hirsutum L.) petioles. They observed that petiole nitrates were sometimes spatially dependent in seven commercial fields. The variograms, kriged maps of petiole nitrates, and map of kriging variance were constructed. The map of kriging variance showed that kriging variances were higher for estimated points on the border of the plots and for points further from the sampled points. They also showed differences along the row from that across rows.

None of these researchers compared kriging with the least-squares estimation method to find out if kriging had any advantage or disadvantage.

CHAPTER III

METHODS AND MATERIALS

The study site was located at the Agronomy Research Station at Perkins Oklahoma. Soil type was a Teller sandy loam (Udic Argiustolls). The soil had been tilled and subjected to rainfall. It was bare of vegetation when this experiment was conducted.

Temperature readings were taken at 96 equally spaced locations along a transect 192 m long. For this purpose, a Campbell Scientific, Model CR7 data logger with 98 channels was used. (Two of the channels were used for recording time and reference temperature). Thermocouple wire connected each channel to each sampled location. Thermocouples were placed 5 cm below the soil surface and 2 m apart along the transect. Temperature at all locations were recorded at 5 minute intervals for 10 days from 22 June through 2 July 1983. Data were transferred to a cassette-tape recorder in the field and then to a "NorthStar" computer system. Approximately 27,000 temperature readings were taken each day.

Temperature semi-variograms over distance were constructed for every half hour of each day. Measured values of temperature every 6 m were used to construct semi-variograms over distance. Thirty-two measured values of temperature were used each time. Each semi-variogram was constructed using the first 15 values of semi-variance or the first 15 lags. A linear model $S(X_i,X_j)=C + D|X_i-X_j|$ was fitted to each semi-variogram.

Temperature values were estimated at 25 points using simple kriging. After kriging, the residuals or the differences between the measured and the estimated values were determined. The variance of those residuals was calculated as the actual estimation variance. The predicted and actual estimation variances were calculated every half hour for 10 consecutive days.

The least-squares method was used to predict temperature at the same 25 points. A polynomial of the fifth order was fitted to 32 measured values of soil temperature along the transect. The predicted estimation variance corresponding to each estimated value of soil temperature was calculated using equation (7). Because the estimation variance for the least-squares method changed with position, the mean estimation variance for 25 estimated values was calculated simply by taking the average of the 25 calculated estimation variances for the estimated values. This was deemed reasonable since the change in predicted estimation variance with position was less than 5.5%. The actual estimation variance was obtained by calculating the variance of the residuals as described for kriging.

Temperature semi-variograms over time were constructed for 10 locations (1, 10, 20, . . ., and 90). Measured values of temperature every hour for 10 consecutive days were used to construct semi-variogram over time. Two-hundred-forthy measured values of temperature were used at each location. Two-hundred-thirty-nine semi-variances were calculated using equation (1). Each semi-variogram was constructed using the first 12 values of semi-variance for the first 12 lags in time. A linear model $S(X_i,X_j)=D|X_i-X_j|$ was used.

Semi-variograms over time show a trend in temperature values. The temperature values were estimated by simple kriging at 223 different times for two cases. In the first case, trend was not considered; and in the second case, it was considered. In the first case, the temperature semi-variograms over time were used for the kriging process; and equation (6) was used to calculate the predicted estimation variance. To remove the trend, 10 polynomials of order eight were fitted into 240 measured values of the soil temperature at each location. Each polynomial was fitted into 24 measured values of temperature for one day. The residuals or the differences between the measured and estimated values from the polynomials were then calculated. The semi-variogram of the residuals was constructed at each location using the first 60 lags or hours. A linear model was fitted into the semi-variogram of the residuals at each location. Each semi-variogram of the residuals was used to calculate the kriging estimates of the soil temperature at 223 specific times. The predicted and actual estimation variances were calculated at all 10 locations after the trend was removed.

The least-squares method was used to calculate temperature estimates at the same 223 points estimated by simple kriging. Ten polynomials of order eight were fitted into measured values of soil temperature every hour for 10 consecutive days. The predicted estimation variance corresponding to each estimated value of soil temperature was calculated using equation (7). In this case, σ^2 is the estimation variance for estimated temperature T at time X_0, X_i's are times at which temperatures were measured, and $\bar{X}$ is the mean of times at which temperatures were measured. The mean predicted estimation variance for 223 values was calculated simply by taking the average of the 223 predicted estimation variances corresponding to estimated values. The difference between the maximum and minimum predicted estimation variances was less than 12%. The actual estimation variance was obtained by calculating the variance of the residuals or the difference between the actual measured values and the estimated values by the least-squares method.

CHAPTER IV

RESULTS AND DISCUSSION

A. Analytical Evaluation of Simple Kriging for Values on a Transect

Simple kriging for values measured along a transect when a linear semi-variogram model $S(X_i,X_j)=C + D|X_i-X_j|$ is used results in linear interpolation between closest neighbors. This is proven below for the general case where n measured values are used to estimate value at location X. Equation (2) for a linear semi-variogram model yields:

$$[A]\,\vec{W} = \vec{B}$$

where [A] is given by:

$$\begin{bmatrix} C+D|X_1-X_1| & C+D|X_1-X_2| \ldots & C+D|X_1-X_k| & C+D|X_1-X_{k+1}| & \ldots & C+D|X_1-X_n| & 1 \\ C+D|X_2-X_1| & C+D|X_2-X_2| \ldots & C+D|X_2-X_k| & C+D|X_2-X_{k+1}| & \ldots & C+D|X_2-X_n| & 1 \\ \vdots & \vdots \quad \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C+D|X_n-X_1| & C+D|X_n-X_2| \ldots & C+D|X_n-X_k| & C+D|X_n-X_{k+1}| & \ldots & C+D|X_n-X_n| & 1 \\ 1 & 1 \quad \ldots & 1 & 1 & \ldots & 1 & 0 \end{bmatrix}$$

and vectors $\vec{W}$ and $\vec{B}$ are give by:

$$\vec{W} = \begin{bmatrix} W_1 \\ W_2 \\ \cdot \\ \cdot \\ \cdot \\ W_k \\ W_{k+1} \\ \cdot \\ \cdot \\ \cdot \\ W_n \\ \mu \end{bmatrix} \qquad \vec{B} = \begin{bmatrix} C+D|X_1-X| \\ C+D|X_2-X| \\ \cdot \\ \cdot \\ \cdot \\ C+D|X_k-X| \\ C+D|X_{k+1}-X| \\ \cdot \\ \cdot \\ \cdot \\ C+D|X_n-X| \\ 1 \end{bmatrix}$$

$|X_i-X_j|$ is 0 for i=j, and $|X_i-X_j|$ is distance or time between measured values at points X_i and X_j for $i \neq j$.

Case 1: If X_i is ordered such that X_i increases as i increases from 1 to n and such that $X_k < X < X_{k+1}$ where X is the position at which the value is to be estimated, the solution is:

$$\vec{W} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{X_{k+1} - X}{X_{k+1} - X_k} \\ \dfrac{X - X_k}{X_{k+1} - X_k} \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 0 \end{bmatrix} \tag{8}$$

Proof: Multiplying matrix [A] by vector $\vec{W}$ yields:

$$[A]\vec{W} = \begin{bmatrix} (C+D|X_1-X_k|)W_k + (C+D|X_1-X_{k+1}|)W_{k+1} \\ (C+D|X_2-X_k|)W_k + (C+D|X_2-X_{k+1}|)W_{k+1} \\ \cdot \\ \cdot \\ \cdot \\ W_k + W_{k+1} \end{bmatrix} \tag{9}$$

Rearranging the terms in equation (9) yields:

$$[A]\vec{W} = \begin{bmatrix} C(W_k + W_{k+1}) + D(X_1)(W_k + W_{k+1}) - D(X_kW_k + X_{k+1}W_{k+1}) \\ C(W_k + W_{k+1}) + D(X_2)(W_k + W_{k+1}) - D(X_kW_k + X_{k+1}W_{k+1}) \\ \cdot \\ \cdot \\ \cdot \\ W_k + W_{k+1} \end{bmatrix}$$

Substituting values for W_k and W_{k+1} into the above equation yields:

$$[A]\vec{W} = \begin{bmatrix} C + D(X_1) - D(X) \\ C + D(X_2) - D(X) \\ \cdot \\ \cdot \\ C + D(X_k) - D(X) \\ \cdot \\ \cdot \\ C + D(X_n) - D(X) \\ 1 \end{bmatrix} = \begin{bmatrix} C + D|X_1-X| \\ C + D|X_2-X| \\ \cdot \\ \cdot \\ C + D|X_k-X| \\ \cdot \\ \cdot \\ C + D|X_n-X| \\ 1 \end{bmatrix} \tag{10}$$

The vector on the right side of equation (10) is equal to vector $\vec{B}$.

Case 2: If the value to be estimated is on either end of the transect, then the weights are given by:

$$\vec{W} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ D|X_1 - X| \end{bmatrix} \qquad \text{for } X < X_1$$

and

$$\vec{W} = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \\ D|X_n - X| \end{bmatrix} \qquad \text{for } X > X_n$$

Proof: Multiplying matrix [A] by vector $\vec{W}$ yields:

$$[A]\,\vec{W} = \begin{bmatrix} C + D|X_1 - X_1| + D|X_1 - X| \\ C + D|X_2 - X_1| + D|X_1 - X| \\ \cdot \\ \cdot \\ \cdot \\ C + D|X_n - X_1| + D|X_1 - X| \\ 1 \end{bmatrix} = \begin{bmatrix} C + D|X_1 - X| \\ C + D|X_2 - X| \\ \cdot \\ \cdot \\ \cdot \\ C + D|X_n - X| \\ 1 \end{bmatrix} \tag{11}$$

and

$$[A]\,\vec{W} = \begin{bmatrix} C + D|X_1-X_n| + D|X_n-X| \\ C + D|X_2-X_n| + D|X_n-X| \\ \cdot \\ \cdot \\ \cdot \\ C + D|X_n - X_n| + D|X_n-X| \\ 1 \end{bmatrix} = \begin{bmatrix} C + D|X_1-X| \\ C + D|X_1-X| \\ \cdot \\ \cdot \\ \cdot \\ C + D|X_n-X| \\ 1 \end{bmatrix} \tag{12}$$

The vector on the right side of equations (11) and (12) is the same as vector $\vec{B}$. As a result, the estimated value at any point outside of the measured range is the same as the closest measured value at either end of the transect.

Thus, values of the weights can be calculated by knowing the distance or time between the measured points and the position of the estimated point with respect to its closest measured points without solving the kriging system of equation (2). Note that the above results do not depend upon the values of the coefficients C and D in the linear model (assuming D is not zero). Thus, the value estimated by kriging is independent of the slope and intercept in the linear model.

These weights can then be inserted into equation (6) for the estimation variance σ_k^2. This results in:

$$\sigma_k^2 = C + 2D\ \left(\frac{(X-X_k)(X_{k+1}-X)}{X_{k+1} - X_k} \right) \qquad \text{for } X_k < X < X_{k+1} \tag{13}$$

$$\sigma_k^2 = C + 2D\ (X_1-X) \qquad \text{for } X < X_1 \tag{14}$$

$$\sigma_k^2 = C + 2D\ (X_n-X) \qquad \text{for } X > X_n \tag{15}$$

Thus, the predicted estimation variance can be simply calculated by knowing the values of the coefficients in linear semi-variogram model, the distance between the measured points, and the position of the estimated point with respect to its closest neighbors.

Because kriging is dependent upon the semi-variogram model, the question can be asked, "How sensitive are kriged values and the estimation variance to changes in the parameters of the semi-variogram model?" As shown above, for a linear semi-variogram, the kriged value is totally independent of the values of the intercept and slope. However, the estimation variance is linearly dependent on these parameters.

B. Variation of Temperature Over Distance

Figures 3 through 8 show soil temperature along the transect for 10 consecutive days. Starting time was 1800 on June 22 and ending time was 1200 on July 2. The weather was cloudy with rain on June 25-29. Rainfall amounts were 1.1, 2.1, 1.9, 1.1, and 1.9 cm on June 25, 26, 27, 28, and 29, respectively. The figures show soil temperatures at 0600, 1200, 1800, and 2400 for all days. The variation of soil temperature along the transect was high at high temperatures on sunny days. The variation of temperature along the transect was higher at 1200 and 1800 than at 2400 and 2600. Table I shows the sample mean, sample variance, and coefficient of variation for soil temperature along the transect every 6 hours.

Figures 9 through 14 show the temperature semi-variograms over distance for each day. In each figure, the horizontal axis shows distance between sampling points (or the lag distance); and the verti-

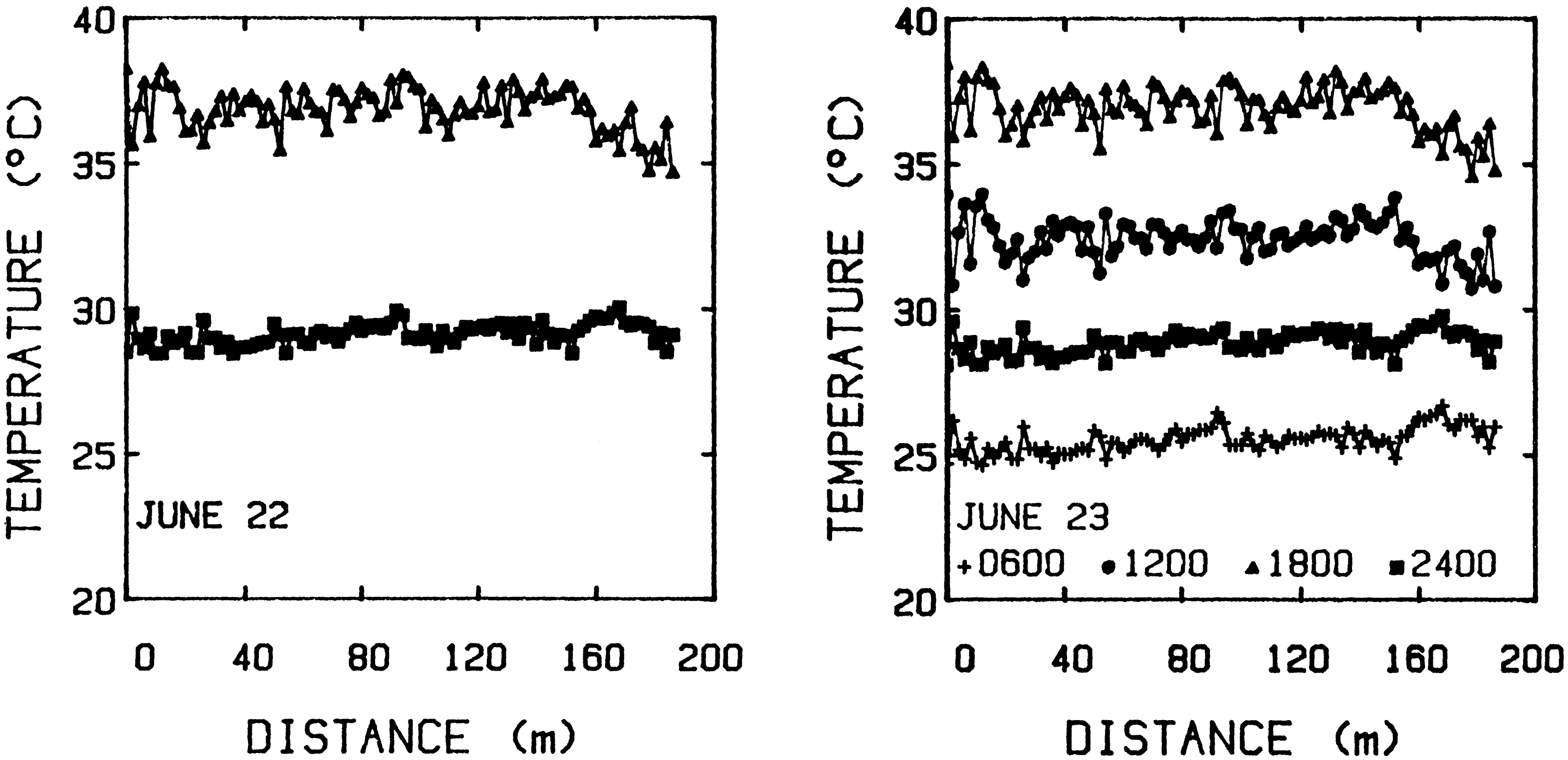

Figure 3. Temperature Across the Transect on June 22 and 23.

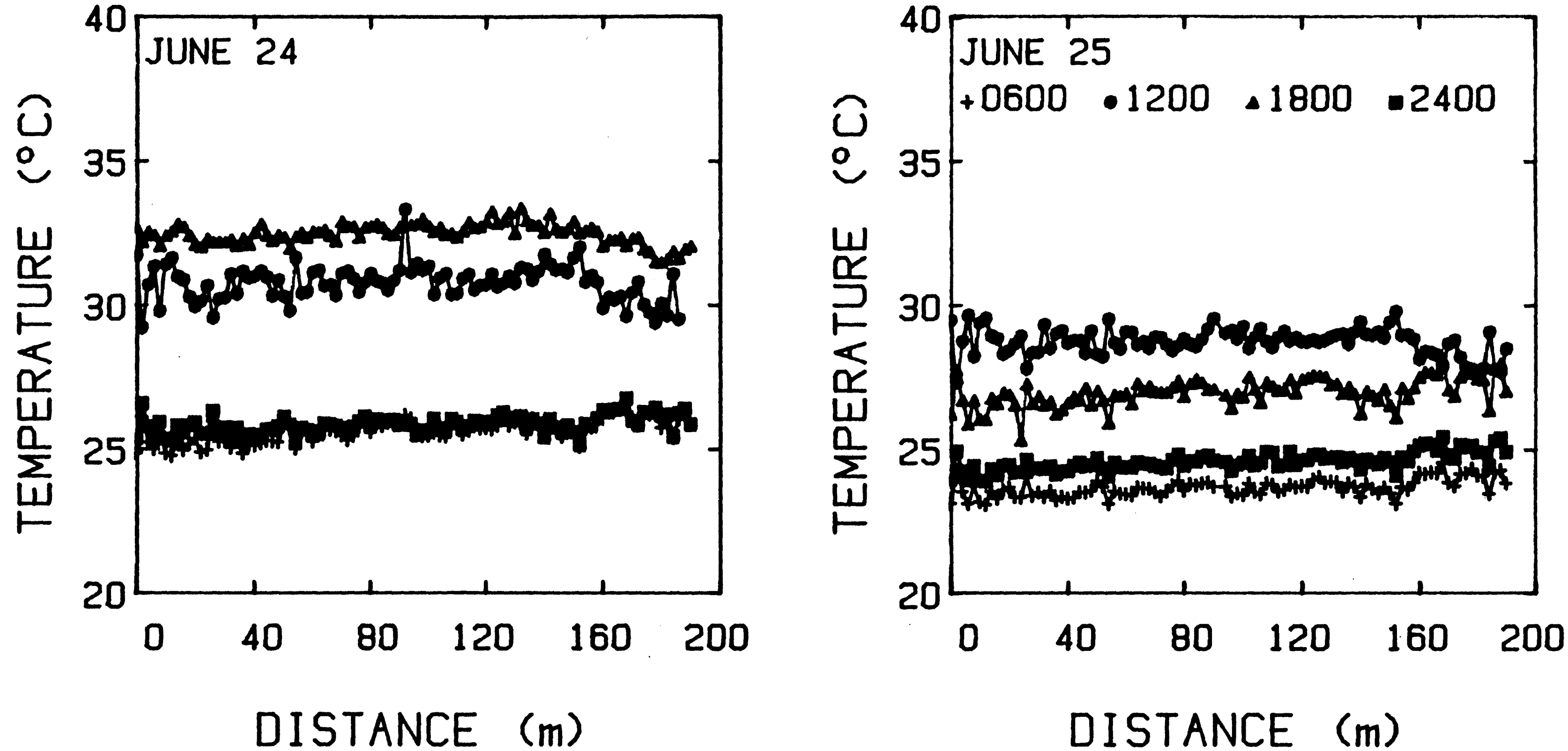

Figure 4. Temperature Across the Transect on June 24 and 25.

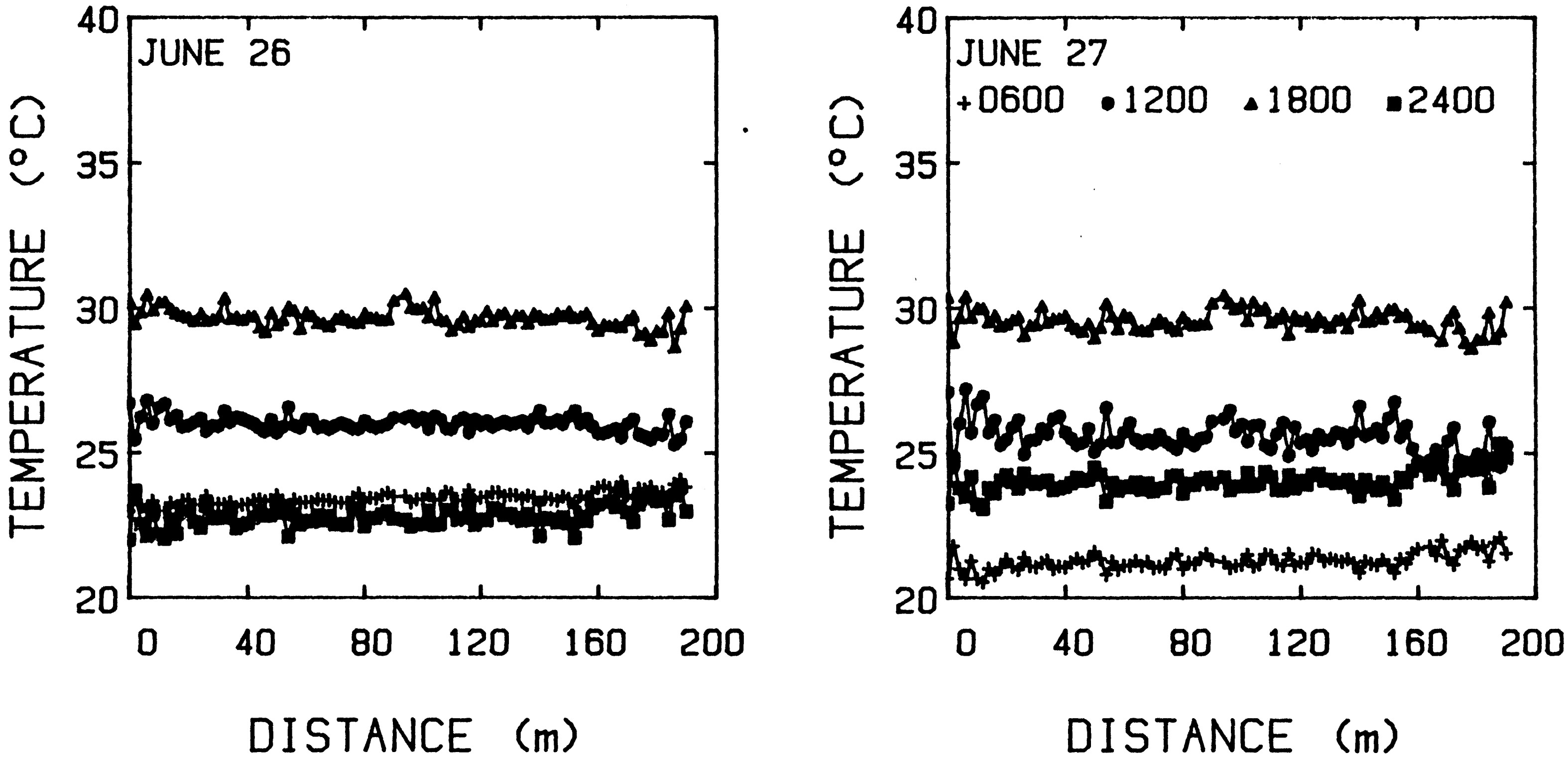

Figure 5. Temperature Across the Transect on June 26 and 27.

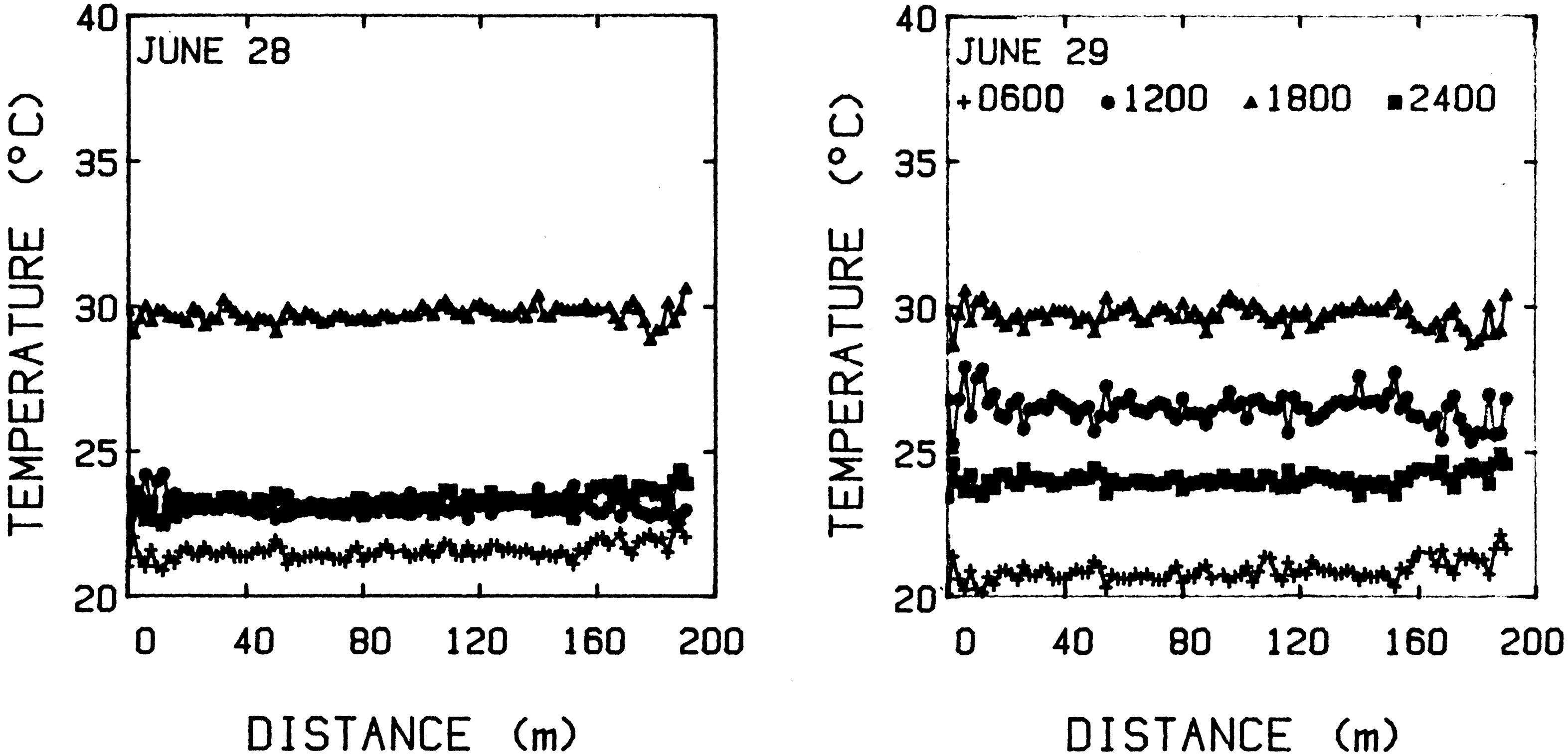

Figure 6. Temperature Across the Transect on June 28 and 29.

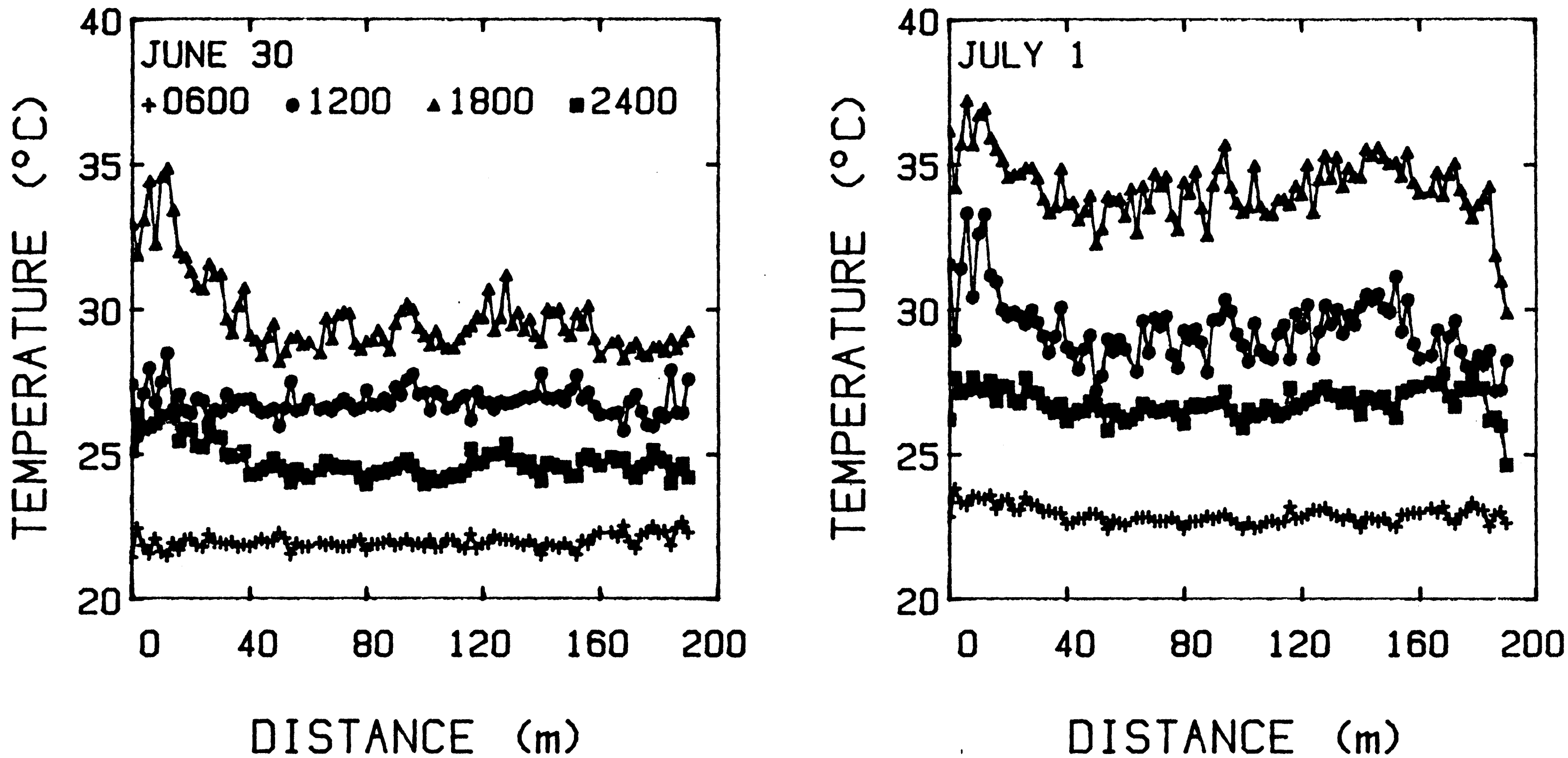

Figure 7. Temperature Across the Transect on June 30 and July 1.

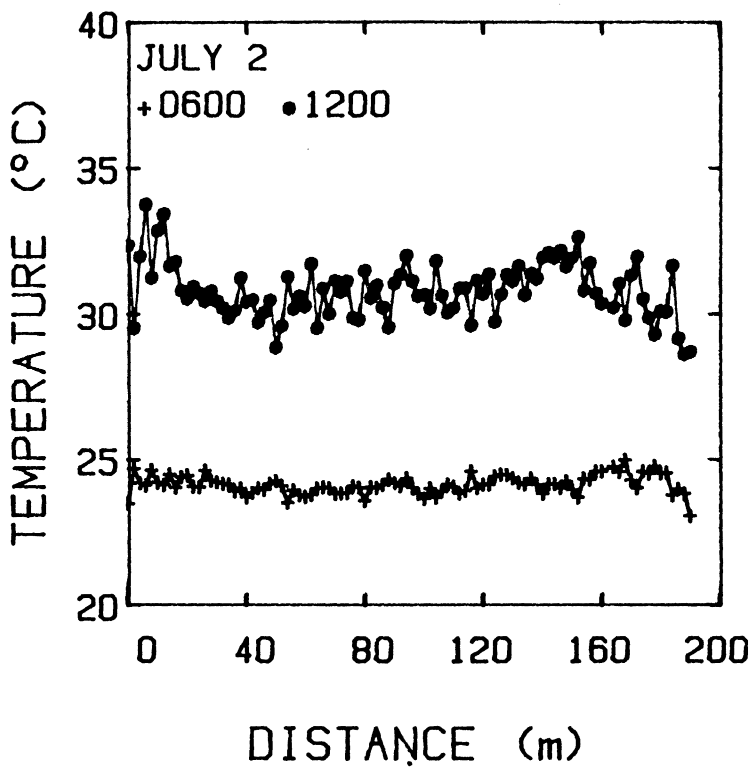

Figure 8. Temperature Across the Transect on July 2.

TABLE I

SAMPLE MEAN, VARIANCE, AND COEFFICIENT OF VARIATION FOR SOIL TEMPERATURES ALONG THE TRANSECT

Time	Mean	Variance	Coef. Var.
June 22			
1800	36.86	0.59	2.09
2400	29.14	0.14	1.30
June 23			
0600	25.55	0.19	1.73
1200	32.41	0.49	2.16
1800	36.92	0.62	2.16
2400	28.84	0.15	1.35
June 24			
0600	25.63	0.17	1.61
1200	30.73	0.34	1.87
1800	32.46	0.14	1.13
2400	25.86	0.10	1.22
June 25			
0600	23.67	0.09	1.31
1200	28.74	0.22	1.65
1800	26.98	0.25	1.85
2400	24.59	0.10	1.30
June 26			
0600	23.44	0.05	0.94
1200	26.01	0.07	1.03
1800	29.66	0.10	1.08
2400	22.80	0.14	1.62
June 27			
0600	21.26	0.08	1.36
1200	25.58	0.31	2.19
1800	29.57	0.14	1.29
2400	24.02	0.14	1.54

TABLE I (Continued)

Time	Mean	Variance	Coef. Var.
June 28			
0600	21.56	0.08	1.35
1200	23.18	0.09	1.29
1800	29.74	0.08	0.94
2400	23.24	0.11	1.42
June 29			
0600	20.88	0.12	1.68
1200	26.51	0.25	1.89
1800	29.72	0.14	1.28
2400	24.05	0.07	1.12
June 30			
0600	21.97	0.05	1.05
1200	26.82	0.22	1.75
1800	29.77	2.02	4.77
2400	24.78	0.34	2.34
July 1			
0600	22.91	0.08	1.27
1200	29.32	1.35	3.96
1800	34.23	1.28	3.30
2400	26.77	0.24	1.83
July 2			
0600	24.14	0.10	1.33
1200	30.80	0.92	3.12

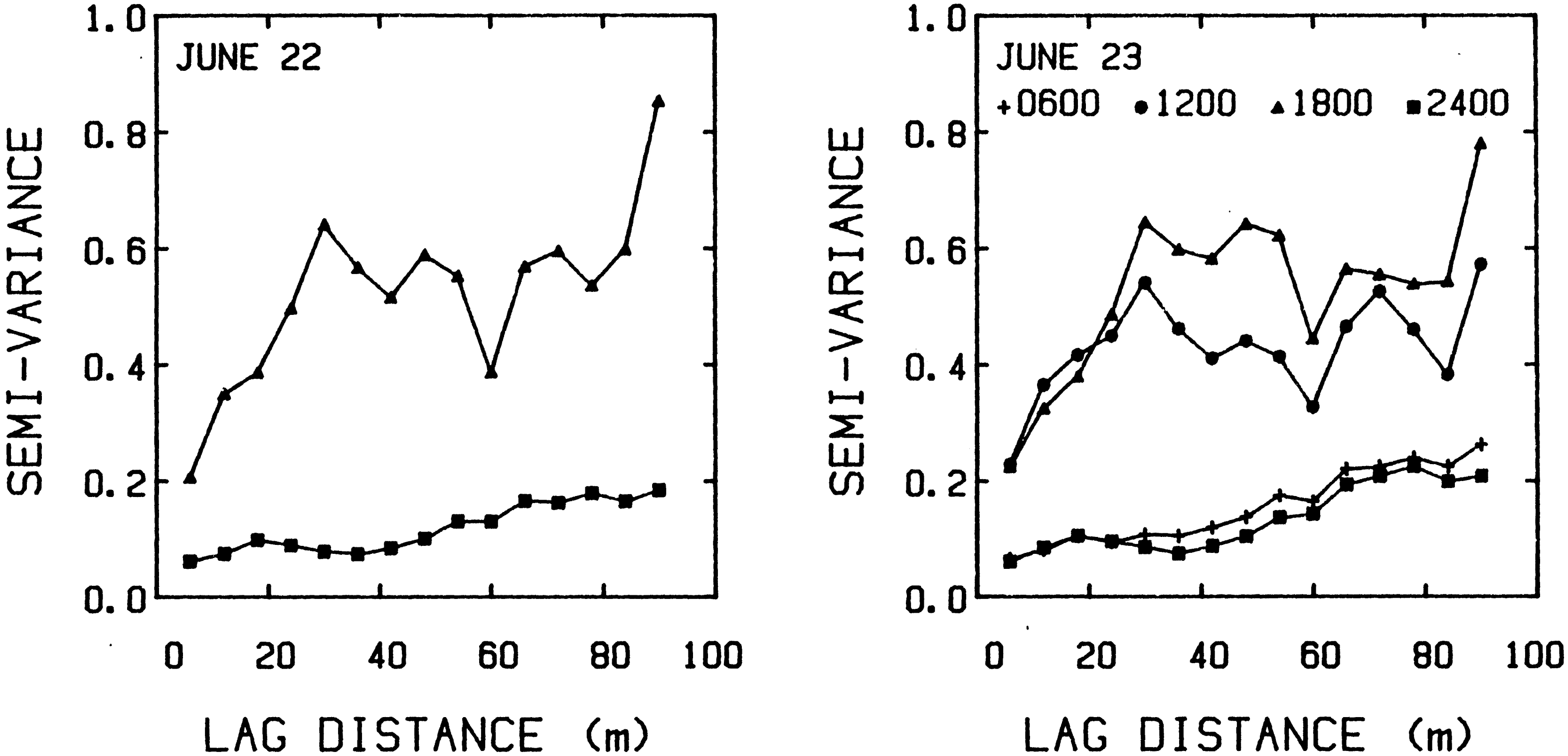

Figure 9. Temperature Semi-variogram Over Distance on June 22 and 23.

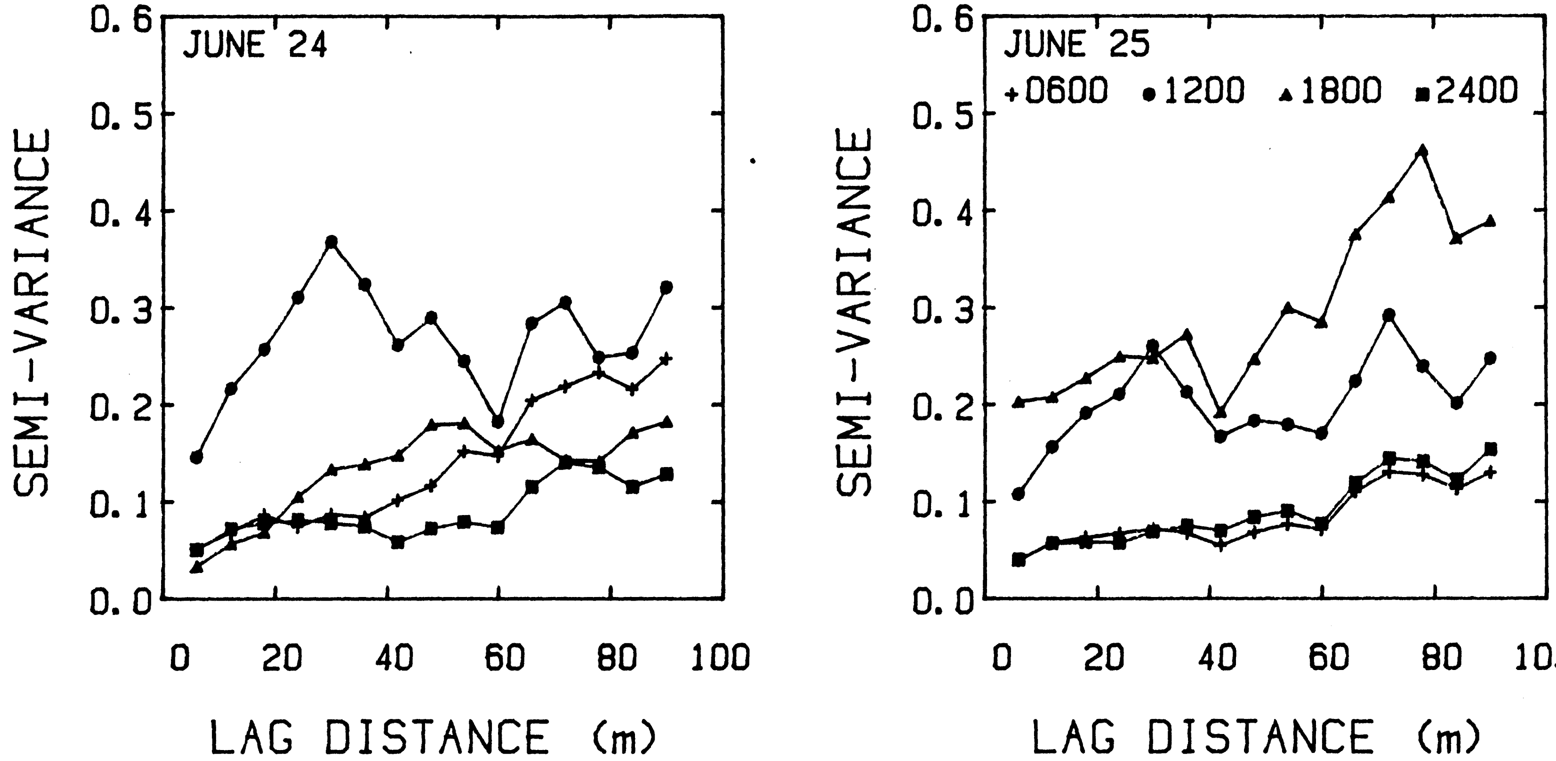

Figure 10. Temperature Semi-variogram Over Distance on June 24 and 25.

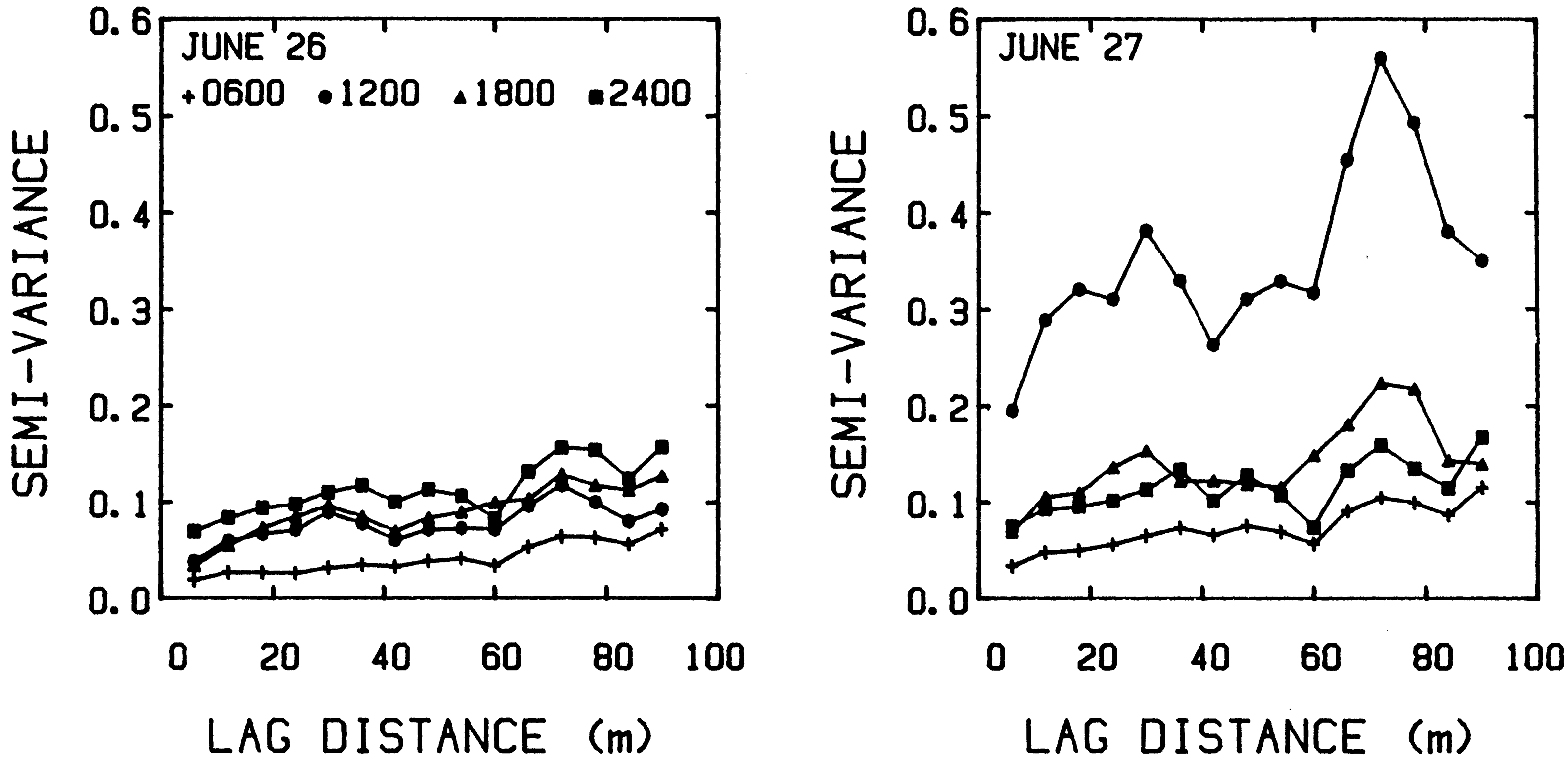

Figure 11. Temperature Semi-variogram Over Distance on June 26 and 27.

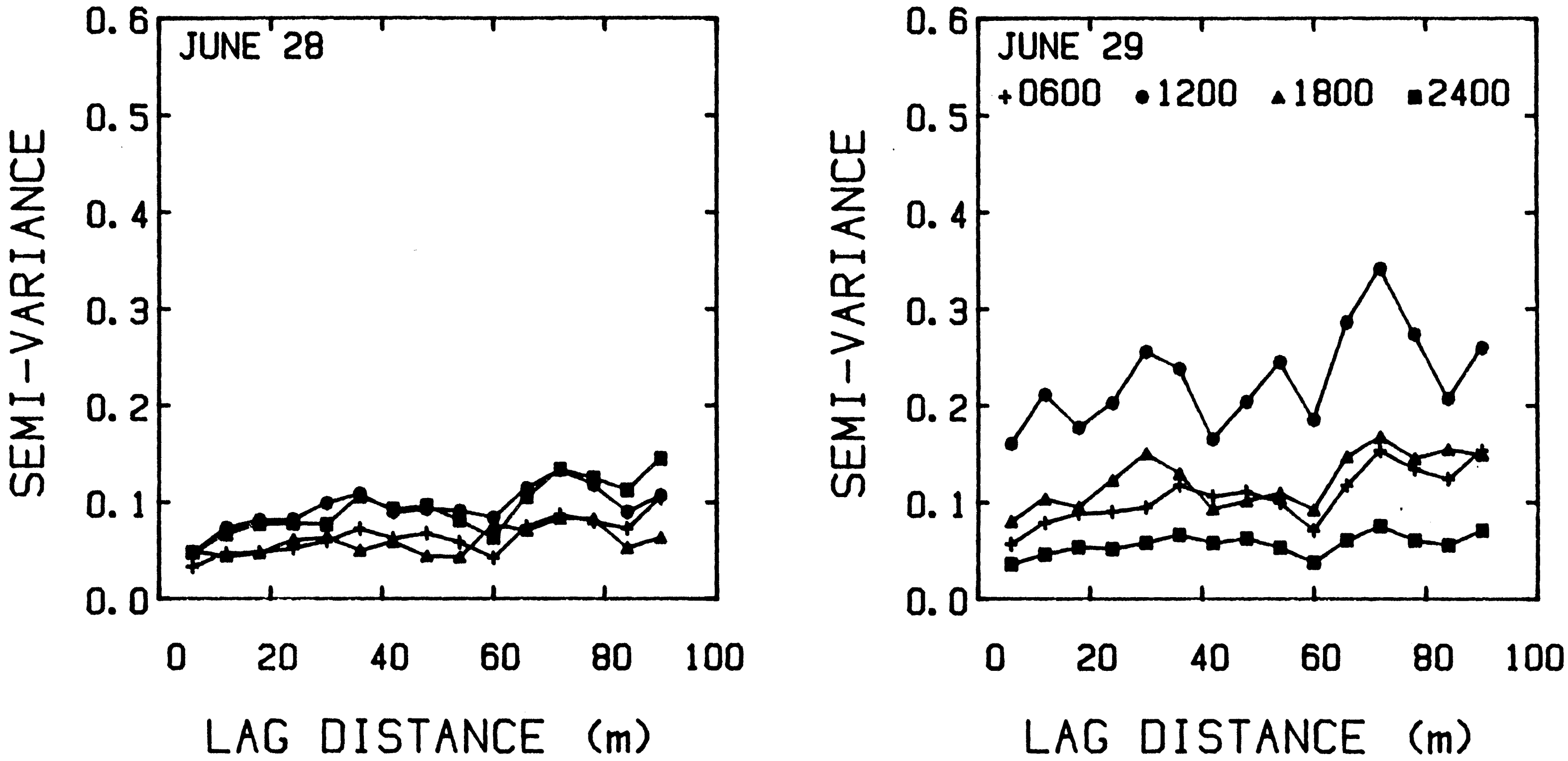

Figure 12. Temperature Semi-variogram Over Distance on June 28 and 29.

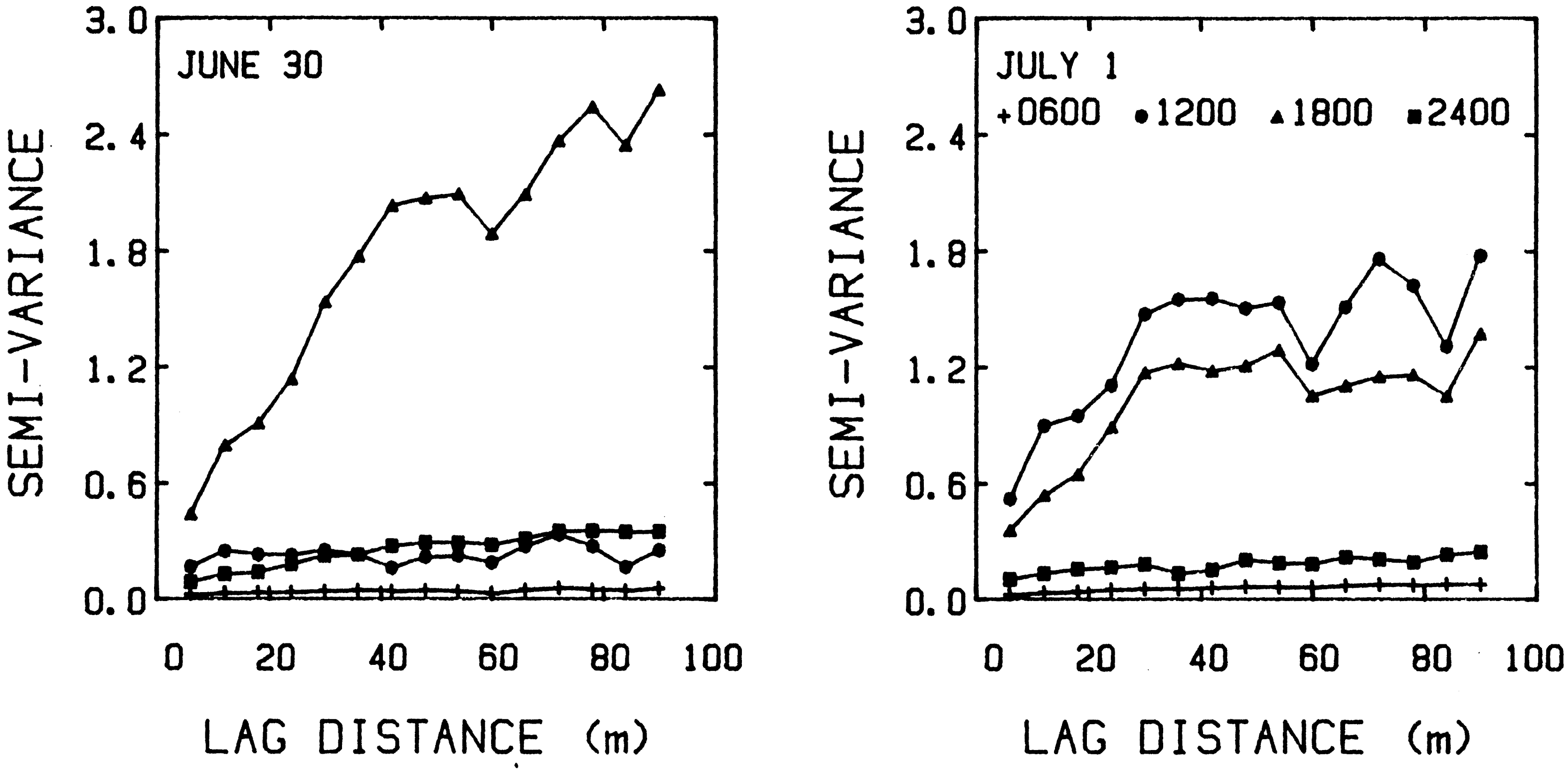

Figure 13. Temperature Semi-variogram Over Distance on June 30 and July 1.

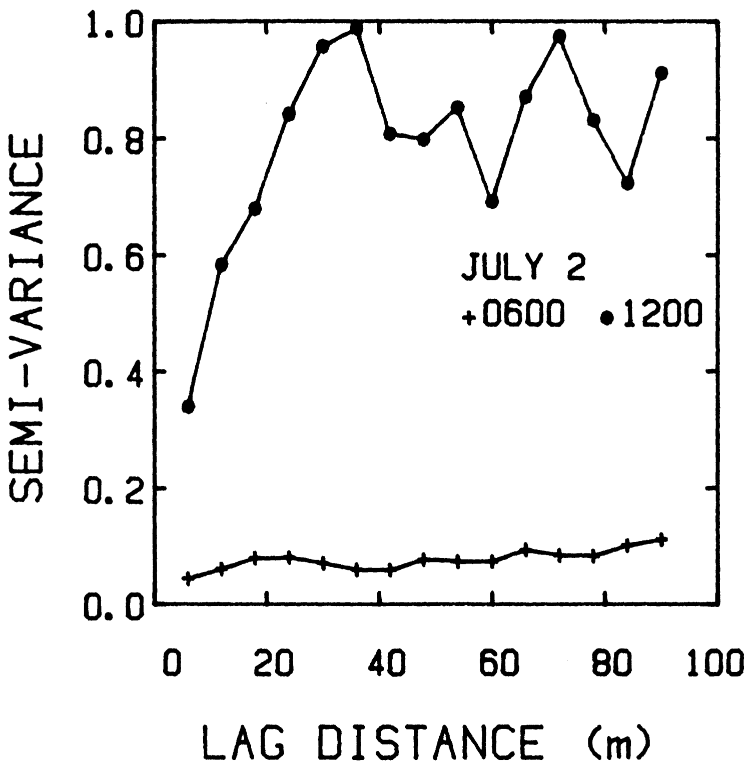

Figure 14. Temperature Semi-variogram
Over Distance on July 2.

cal axis shows the semi-variance. (Note that vertical scales are different). The semi-variance increases with distance to a distance of pproximately 30 m. After 30 m the changes in semi-variance were small and the semi-variance was similar to sample variance in Table I. At night, the semi-variance increased to a distance greater than 30 m. Semi-variance during the day were higher than those at night. Soil temperature values were more correlated at night than during the day. The linear semi-variogram model was fitted to these data. Table II includes the coefficients of the linear semi-variogram model.

The semi-variances at 2400 and 0600 were smaller than those at 1800 and 1200 at any distance for all sunny days and when the soil was dry. On sunny days, the semi-variance was higher than on rainy days at any time and at the majority of distances. On June 23 the semi-variances at any distance at 0600, 1200, 1800, and 2400 were higher than semi-variances at the same distances and times on June 30 when the soil was wet.

One of the assumptions of simple kriging is that no general trend exists in measured values, i.e., the mean of differences in measured values is zero for all lags. Figure 15 shows the mean of differences in temperature values for different lag distances. The mean of differences in temperature values is less than 0.8 °C for the lags up to 96 meters. This trend was assumed to be negligible.

Figures 16 shows the measured temperature, temperature estimated by kriging, and temperature estimated by the least-squares method at 1200 on June 27. In this figure the horizontal axis shows the distance, and the vertical axis shows the soil temperature. The estimated values by least-squares are much smoother than those estimated by kriging. This

TABLE II

VALUES OF THE COEFFICIENTS C AND D IN THE LINEAR SEMI-VARIOGRAM MODEL $S(X_i, X_j) = C + D|X_i - X_j|$

Time	Coefficient	0600	1200	1800	2400
June 22					
	C			0.330	0.050
	D			0.004	0.002
June 23					
	C	0.040	0.360	0.360	0.040
	D	0.002	0.002	0.003	0.002
June 24					
	C	0.020	0.240	0.060	0.050
	D	0.002	0.001	0.001	0.001
June 25					
	C	0.040	0.160	0.160	0.030
	D	0.001	0.001	0.003	0.001
June 26					
	C	0.010	0.050	0.050	0.070
	D	0.001	0.001	0.001	0.001
June 27					
	C	0.040	0.250	0.090	0.080
	D	0.001	0.002	0.001	0.001
June 28					
	C	0.040	0.070	0.050	0.050
	D	0.001	0.001	0.000	0.001
June 29					
	C	0.070	0.180	0.090	0.050
	D	0.001	0.001	0.001	0.000
June 30					
	C	0.030	0.210	0.640	0.110
	D	0.000	0.001	0.024	0.003

TABLE II (Continued)

Time	Coefficients	0600	1200	1800	2400
July 1					
	C	0.030	0.880	0.690	0.120
	D	0.001	0.010	0.008	0.001
July 2					
	C	0.050	0.640		
	D	0.001	0.003		

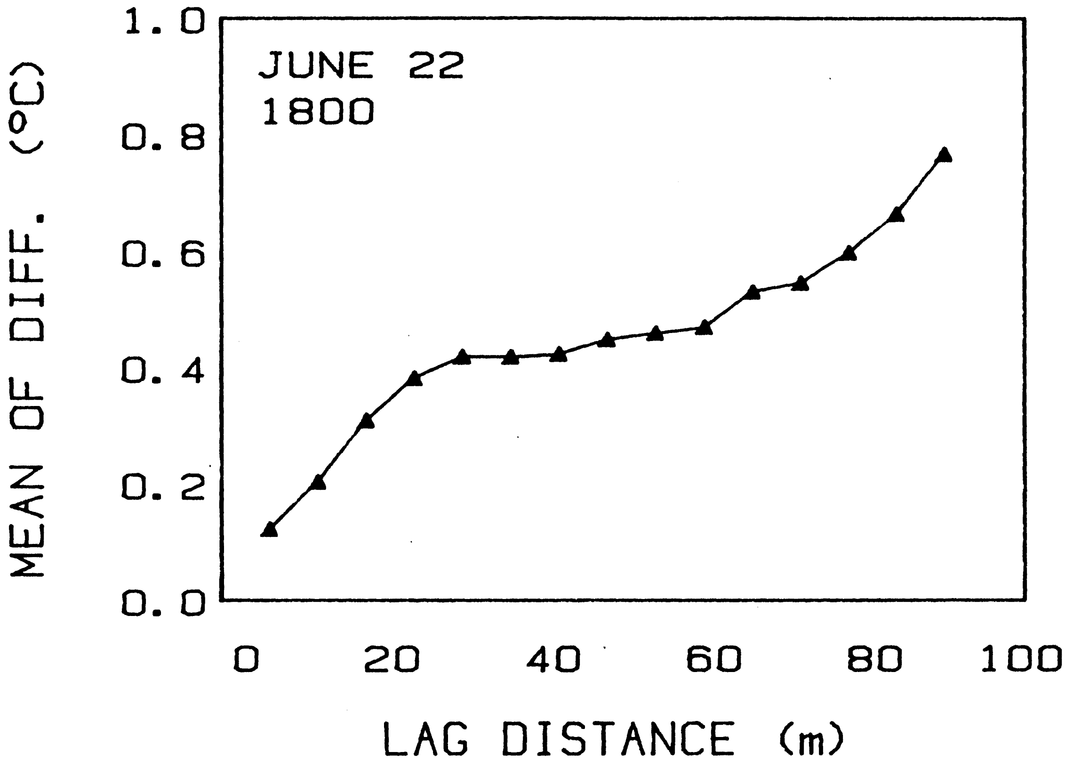

Figure 15. Mean of Differences in Measured Temperatures Versus lag Distance at 1800 on June 22.

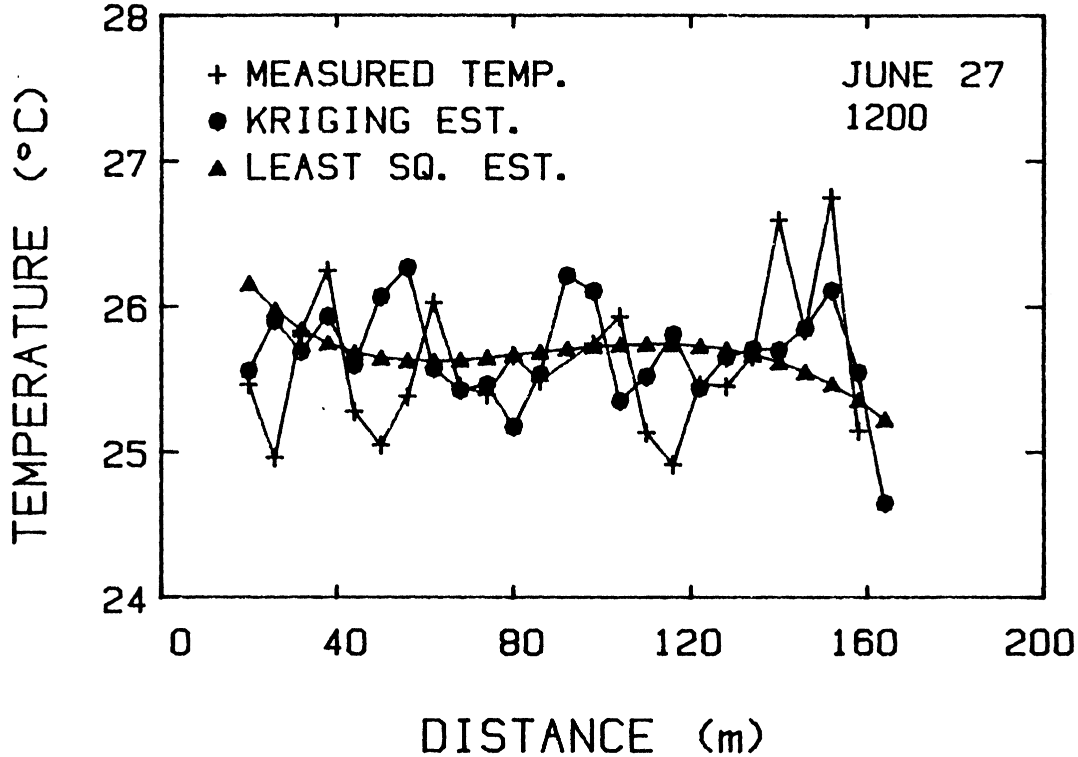

Figure 16. Measured Temperature, Temperature Estimated by Kriging, and Temperature Estimated by the Least-squares Techniques on June 27.

was observed at all times and for all days. In kriging the estimated values at the measured points are exactly the same as the measured values (or values that have been used to construct the semi-variogram) but this is not true for the least-squares method. Figures 17 and 18 show the kriging and least-squares residuals at four times on June 26. The residuals for both methods were approximately the same. These data are representative of those calculated for other times.

The residuals described above indicate that both estimation methods produce comparable results. Another way of evaluating the methods is by comparing their actual estimation variance or the variances of the residuals for each method. The actual variances were calculated for soil temperatures recorded every half hour from 2400 on June 22 untill 1200 on July 2. Four-hundred-fifty-seven actual variances were calculated for each method. Figure 19 shows the distributions of actual variances for kriging and least-squares. The two distribution are approximately the same. Figure 20 shows the distribution of the differences between the least-squares actual variances and kriging actual variances. This figure shows that kriging actual variances were usually slightly smaller than the least-squares variances.

The results discussed above are for actual estimation variances. Both estimation methods also provide a theoretical means for calculating estimation variances. Figure 21 shows the predicted estimation variance versus actual estimation variance for the kriging and least-squares techniques. This figure shows that the estimation variance predicted by kriging tended to overestimate the actual variance while those predicted by least-squares tended to underestimate the actual variance. Still, the agreement appeared to be relatively good.

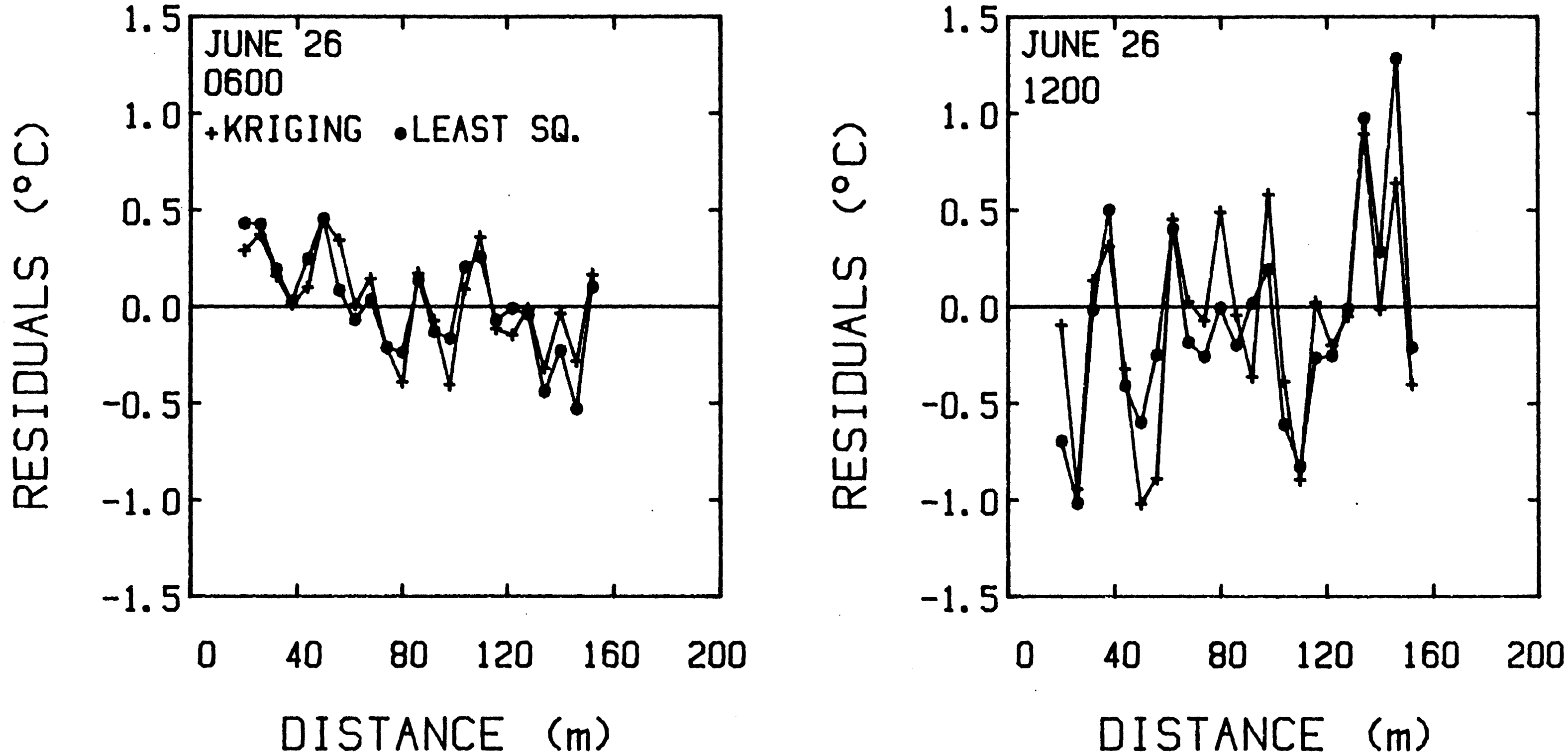

Figure 17. Kriging and Least-squares Residuals at 0600 and 1200 on June 26.

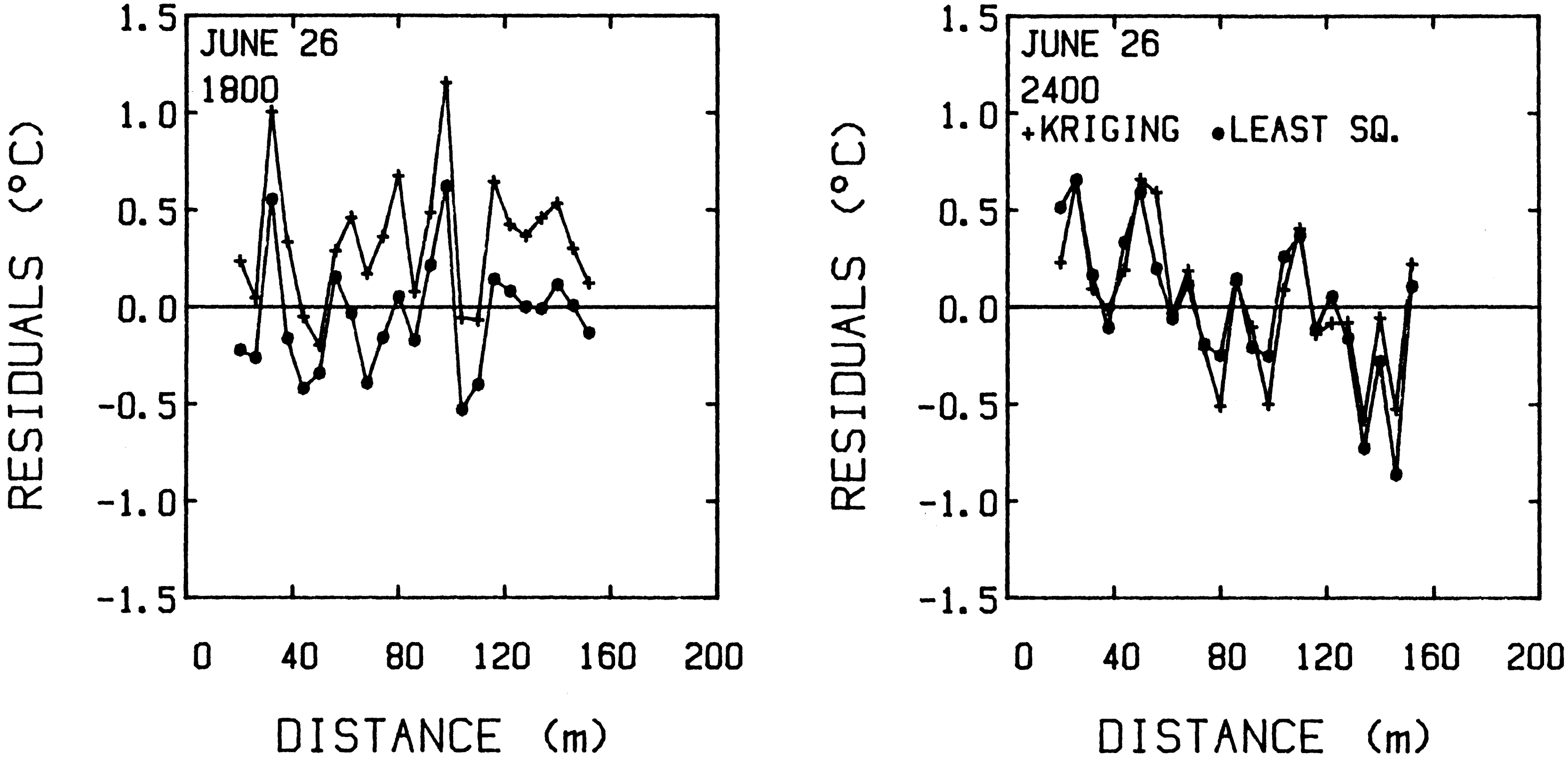

Figure 18. Kriging and Least-squares Residuals at 1800 and 2400 on June 26.

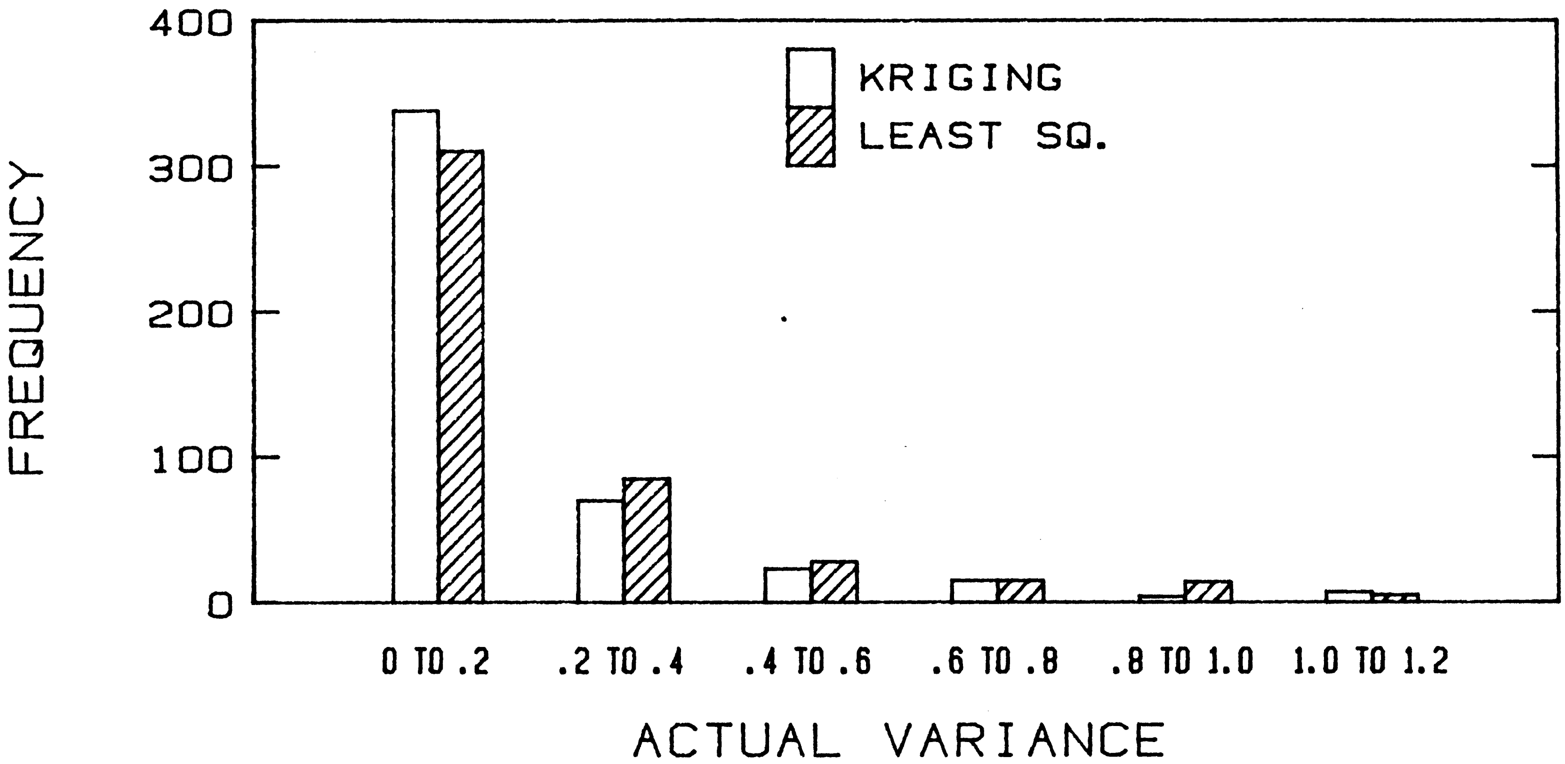

Figure 19. Distribution of Actual Variances for Kriging and Least-squares.

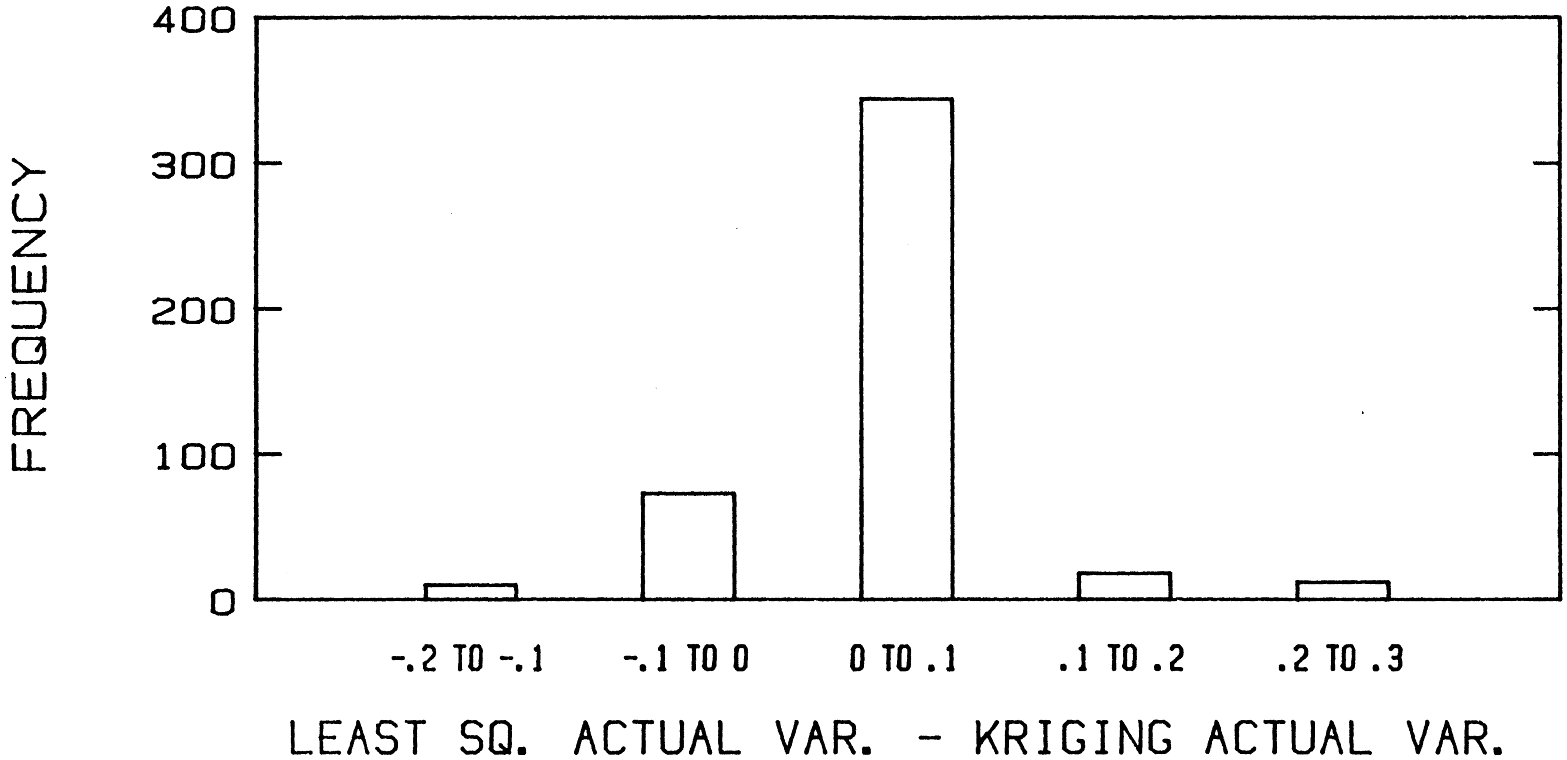

Figure 20. Distribution of Least-squares Actual Variance Minus Kriging Actual Variance.

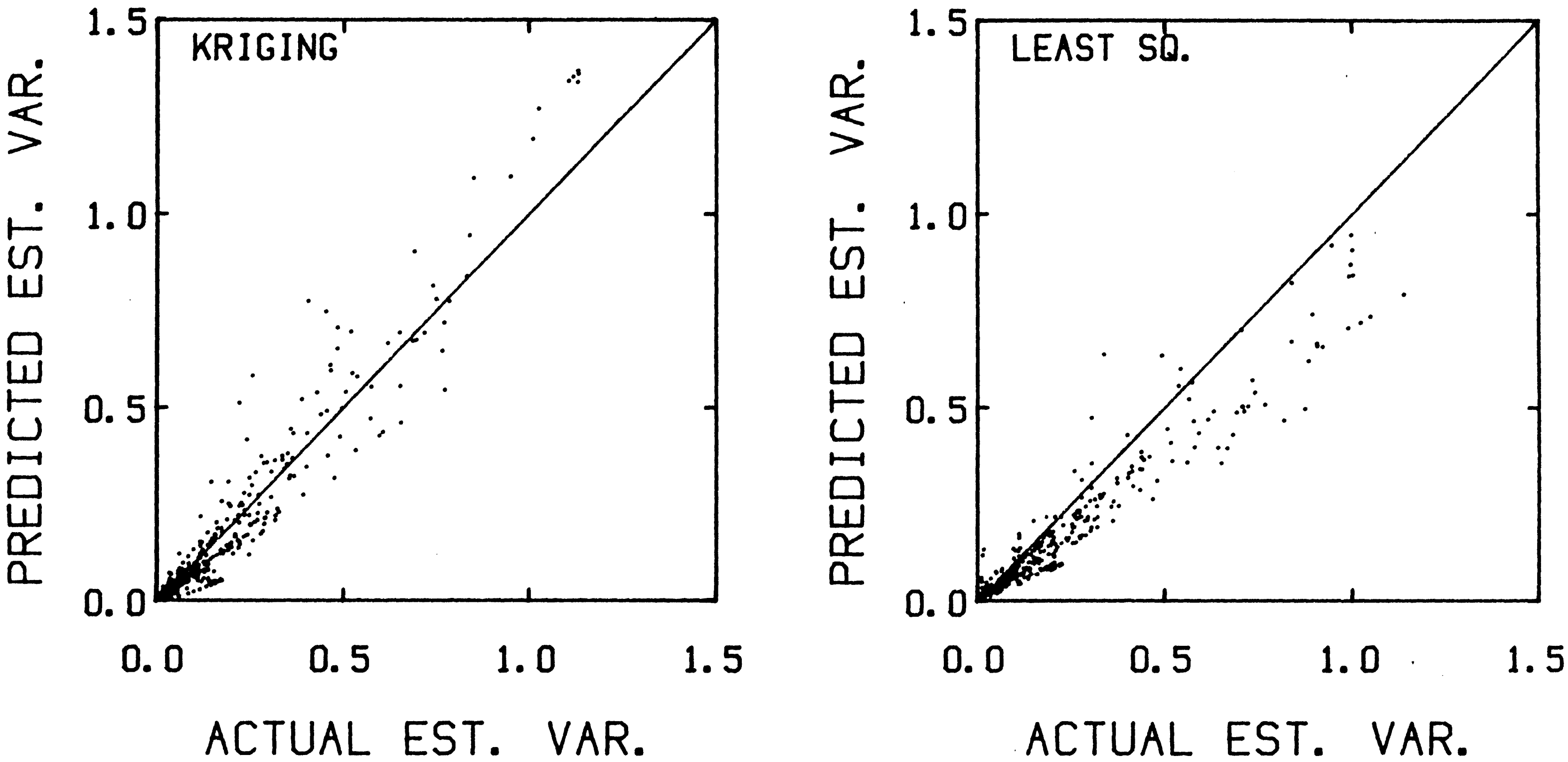

Figure 21. Predicted Estimation Variance Versus Actual Estimation Variance for Kriging and Least-squares.

Figure 22 shows the actual variance and the predicted estimation variance by kriging for every half hour. Starting time was 1800 on June 22 and the ending time was 1200 July 2. At low estimation variances, the actual variances were often somewhat greater than the predicted variances. The predicted and actual estimation variances were greatest about 1500 every day.

Figure 23 shows information similar to that in Figure 22 for the least-squares method. The predicted and actual estimation variances were again maximum at about 1500 every day. At low estimation variances, the actual variances were usually greater than the predicted variances. This figure was very similar to the previous one for kriging.

C. Variation of Temperature Over Time

Figure 24 shows the soil temperature as a function of time. Starting time was 2400 on June 22 and ending time was 1200 on July 2. The two curves in this figure show the temperature fluctuations at two typical, locations. The changes in temperature over time were gradual at all locations.

Figure 25 shows the temperature semi-variogram over time for half of the lags (or 120 hours) at location 1. This semi-variogram shows a polynomial-type trend. To avoid the trend the first 12 lags were used to construct a semi-variogram. Figure 26 shows temperature semi-variograms over time for the two locations; similar semi-variograms were obtained for other locations. These semi-variograms show parabolic behavior near the origin. It appears that the Gaussian model would be a reasonable one to fit into these semi-variograms. However, the Gaussian model results in ill-conditioned matrices which can not be reliably

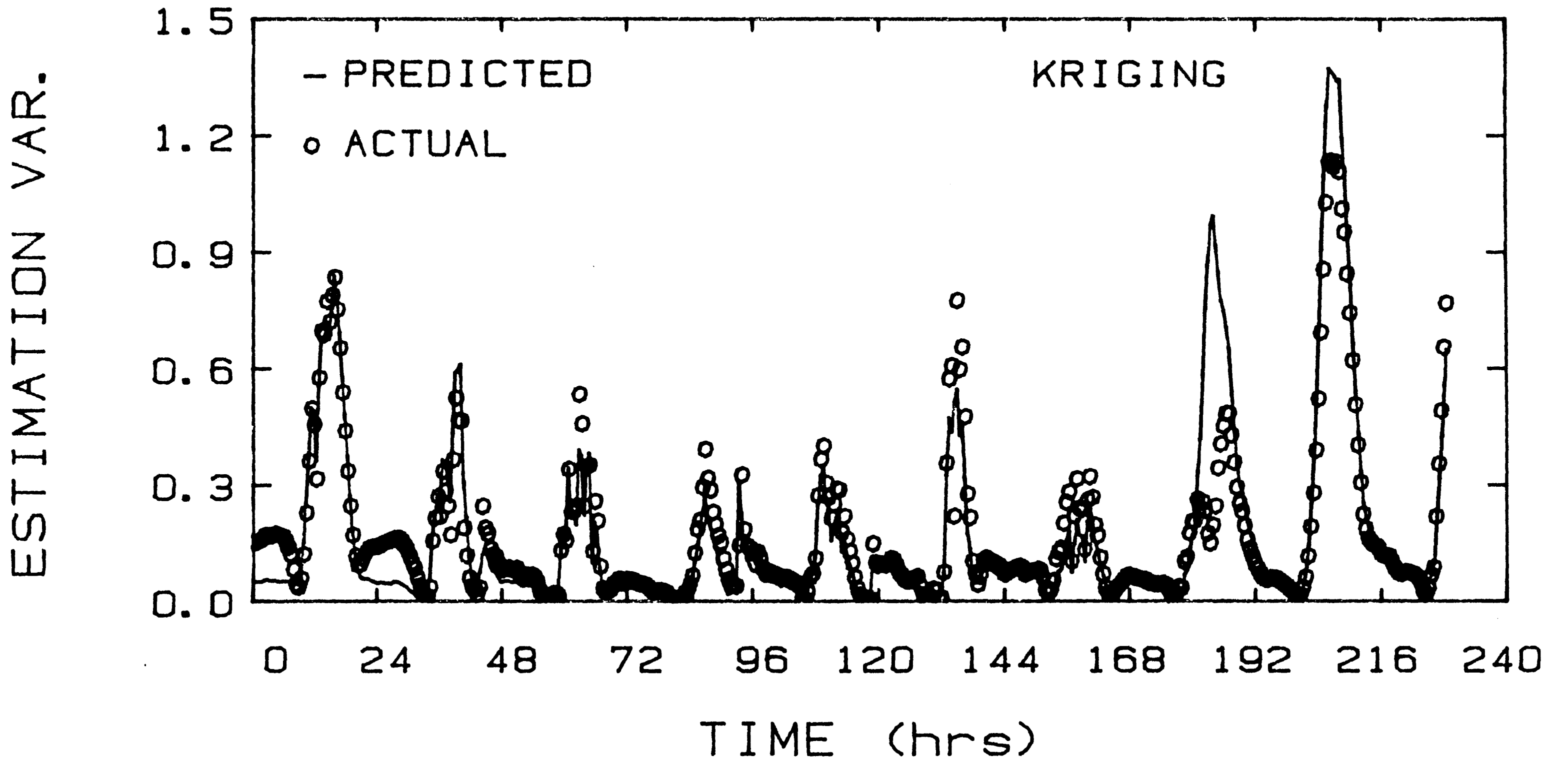

Figure 22. Predicted and Actual Estimation Variances for Kriging.

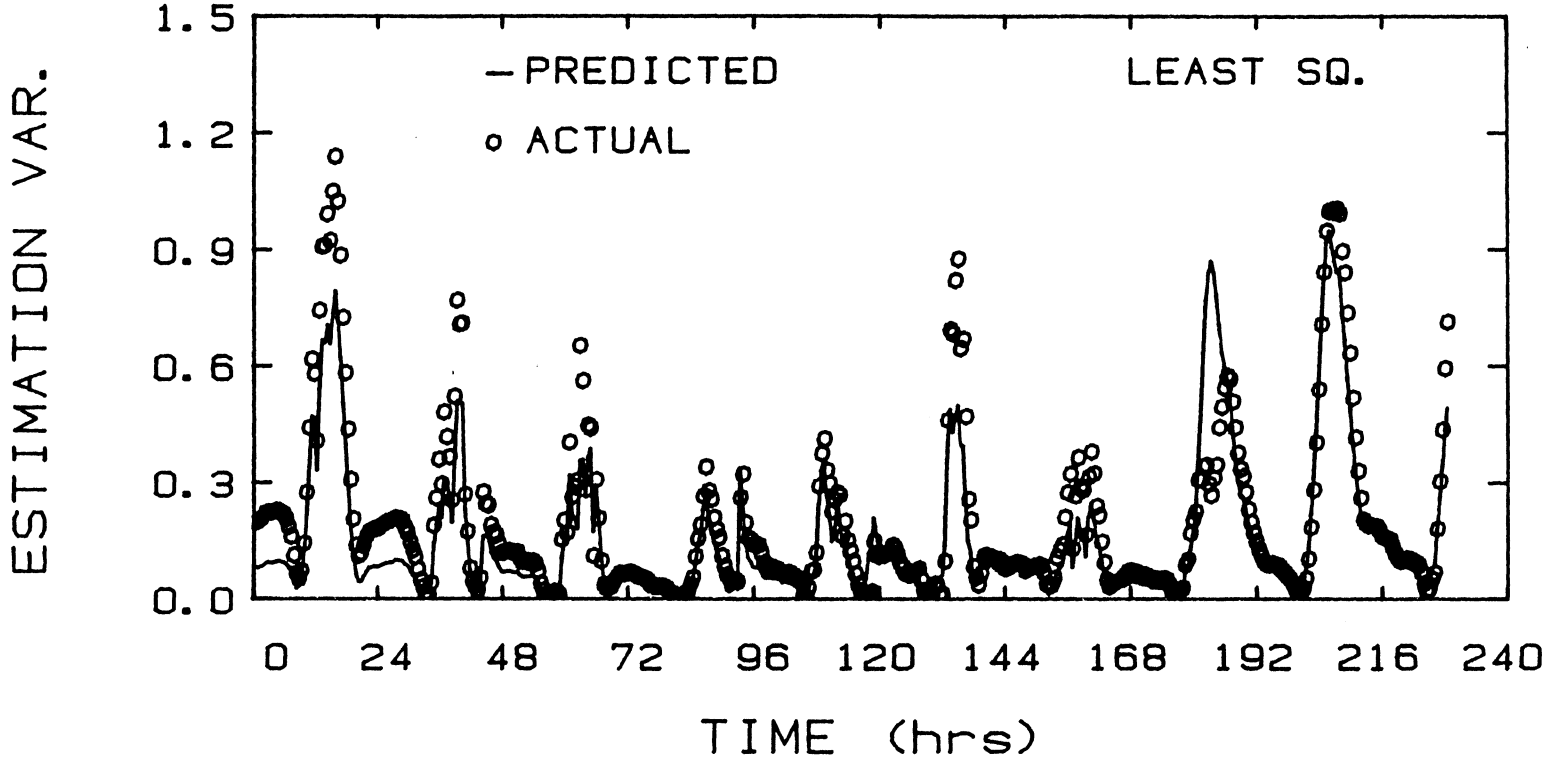

Figure 23. Predicted and Actual Estimation Variance for the Least-squares Techniques.

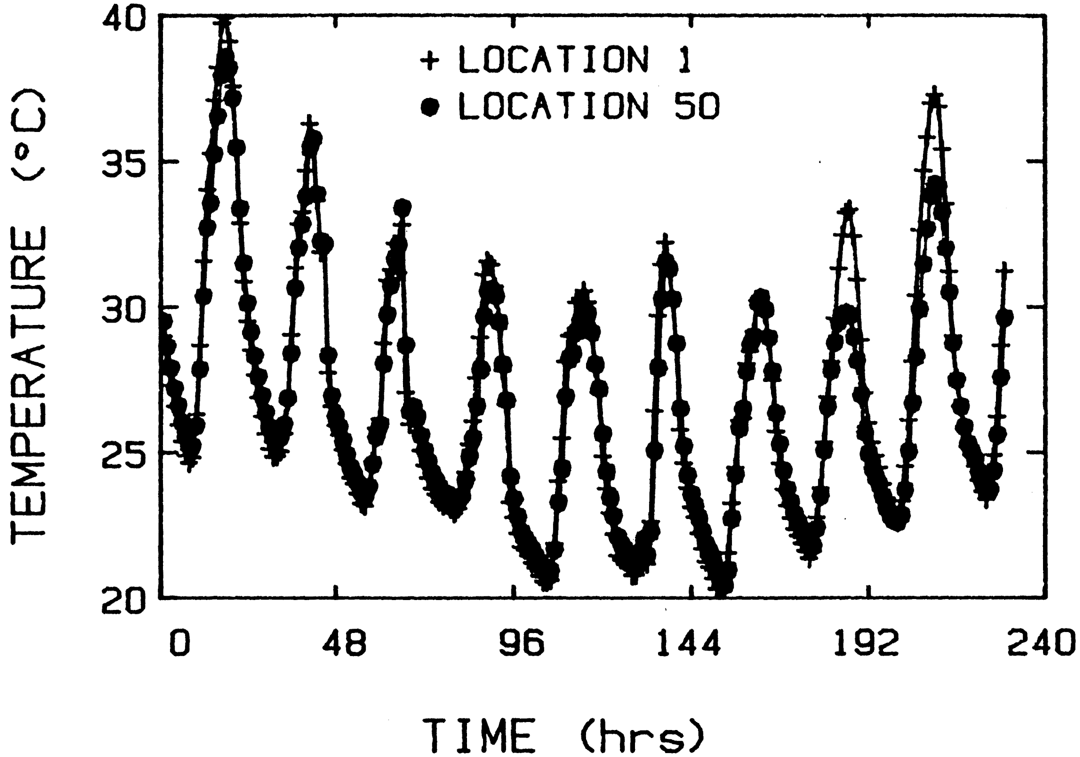

Figure 24. Variation of Temperature Over Time on June 22 Through July 2.

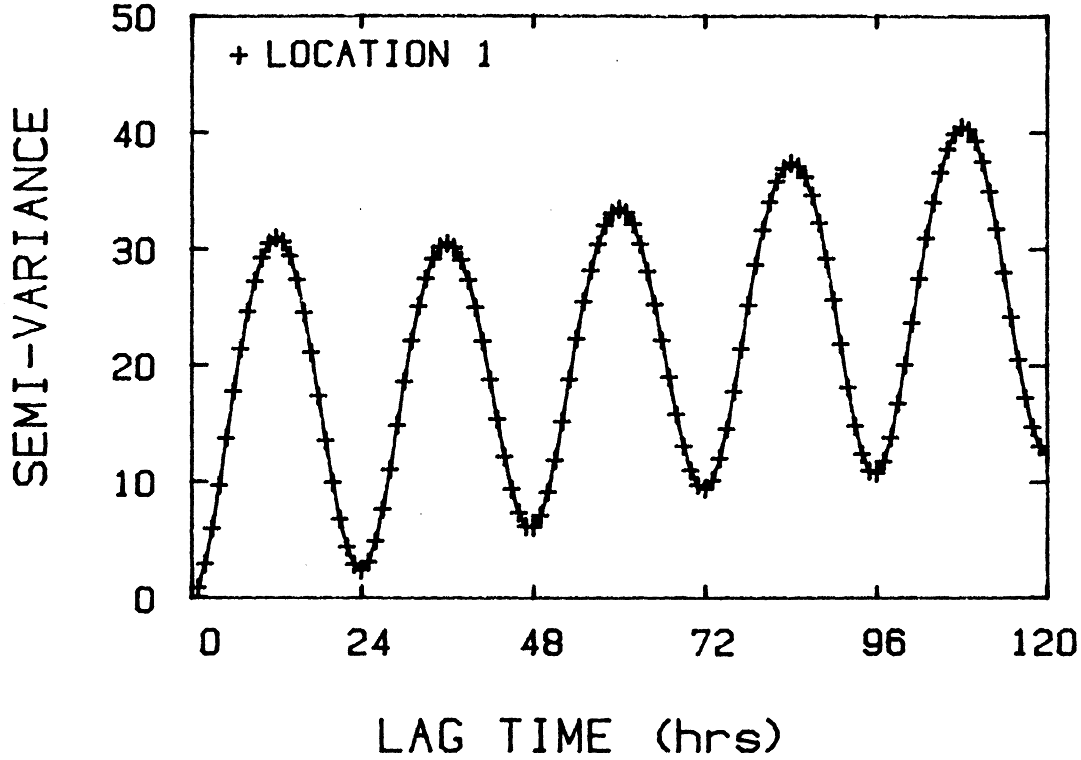

Figure 25. Temperature Semi-variogram for Lags up to 120 Hours at Location 1.

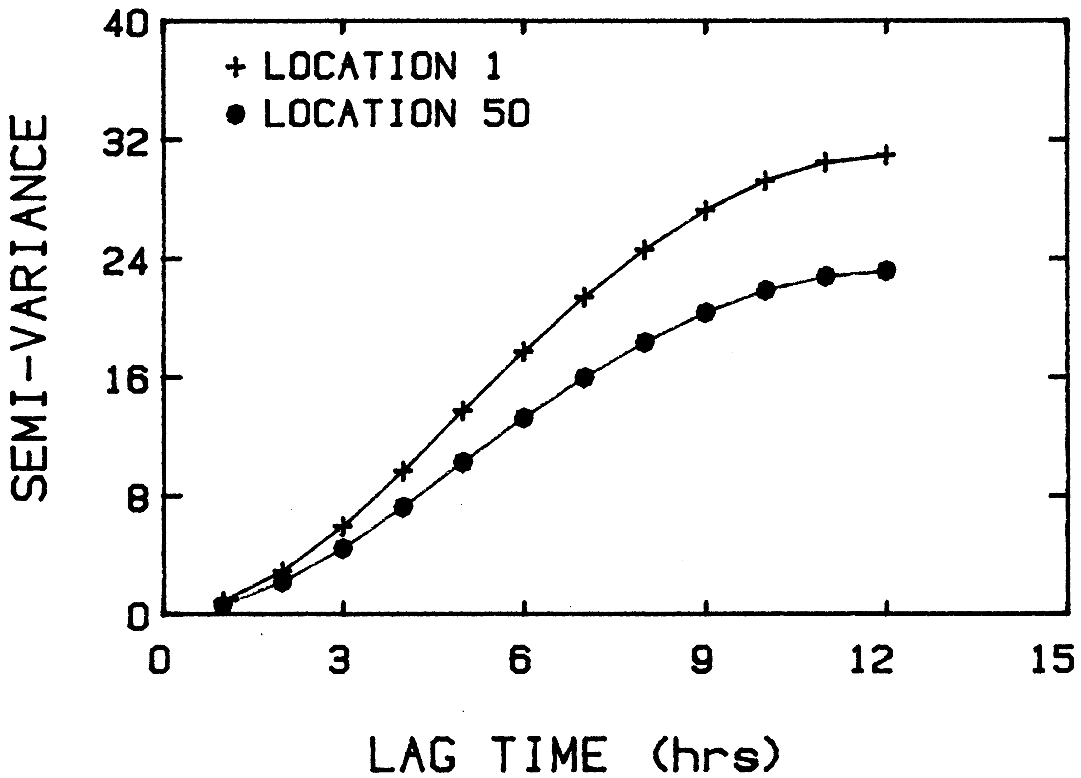

Figure 26. Temperature Semi-variograms for Lags up to 12 Hours at Location 1 and 50.

solved. The accuracy of the estimated values by kriging was only 1 to 2 digits when the Gaussian model was used with 14-digit floating point arithmetic. To avoid these ill-conditioned matrices, linear models were fitted into the semi-variograms and used for kriging. The temperature values were estimated at 223 times by simple kriging. The predicted and actual estimation variances were calculated for the estimated values. The kriged values and the estimated values by least-squares closely approximated the measured temperatures. The estimated values by both methods have a relative error of less than 5%. Figure 27 shows the kriging and least-squares residuals. This figure shows that kriging residuals were generally smaller than those from least-squares. These predictions resulted in low actual variances. However, the predicted estimation variances for kriging were 30 to 47 times the actual estimation variances. Figure 28 shows the predicted versus the actual estimation variance. The reason for the poor agreement in Figure 28 was investigated. Figure 29 shows the mean of the differences in temperature values for different lag times. The mean of the differences in temperature values was less than 0.3 °C for the lag times up to 12 hours. Although this trend was less than that for distance, the data were analyzed again. To remove the trend, polynomials of order eight were fitted into measured values of soil temperature for each 24-hour period. Figure 30 shows a typical polynomial fitted for location 1. Starting time was 2400 on June 22 and the ending time was 2400 on June 23. Similar polynomials were fitted into measured values of temperature at all other locations and days. Regression coefficients were between 0.995 to 0.997 for all 10 locations. The residuals from those polynomials were then calculated. The semi-variogram of the residuals was

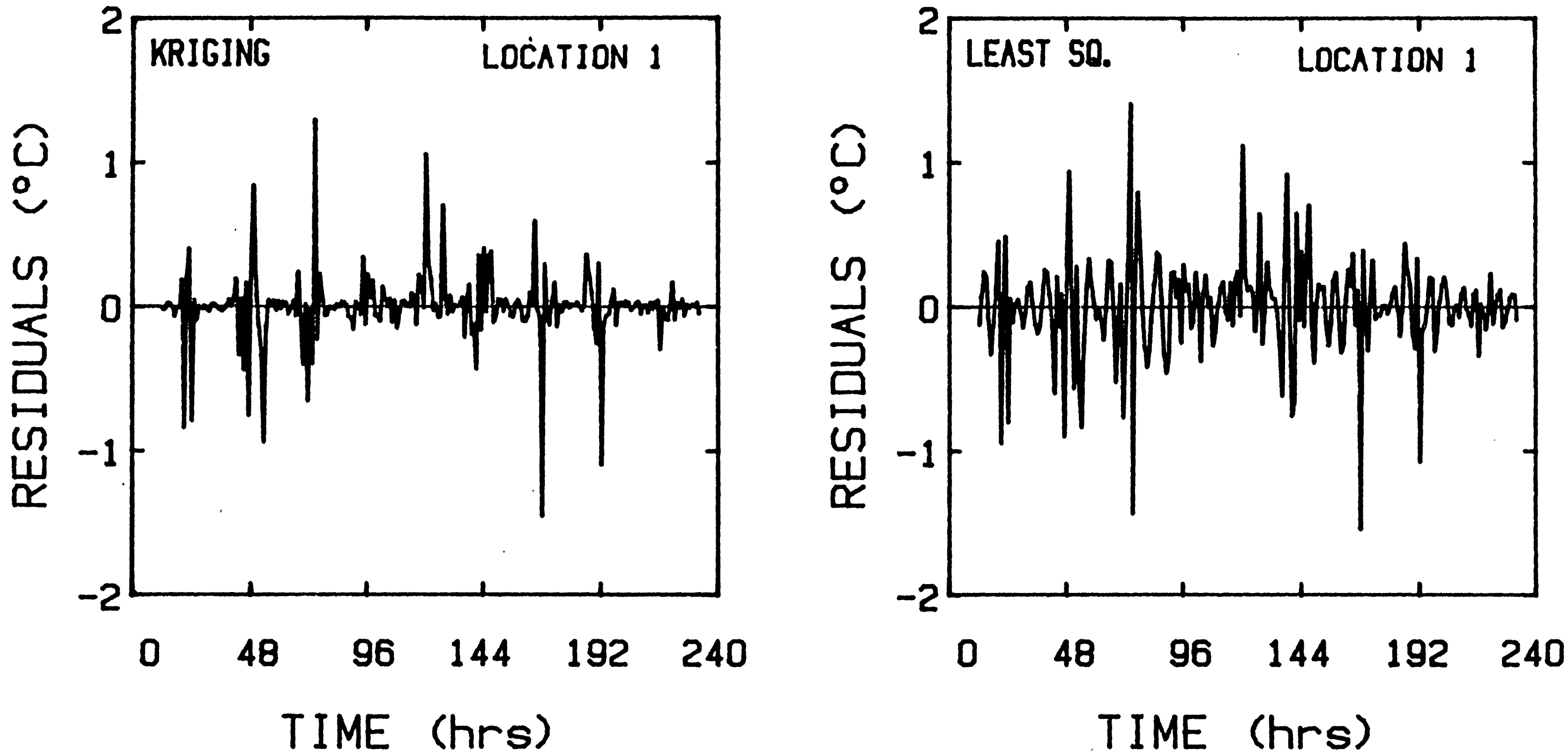

Figure 27. Kriging and Least-squares Residuals at Location 1.

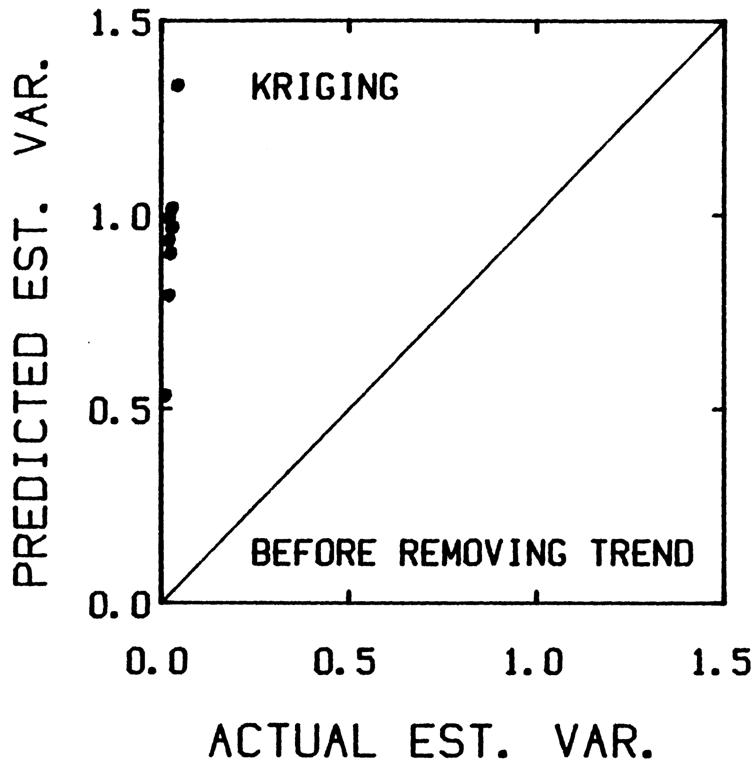

Figure 28. Predicted Estimation Variance Versus Actual Estimation Variance for Kriging Before Removing the Trend.

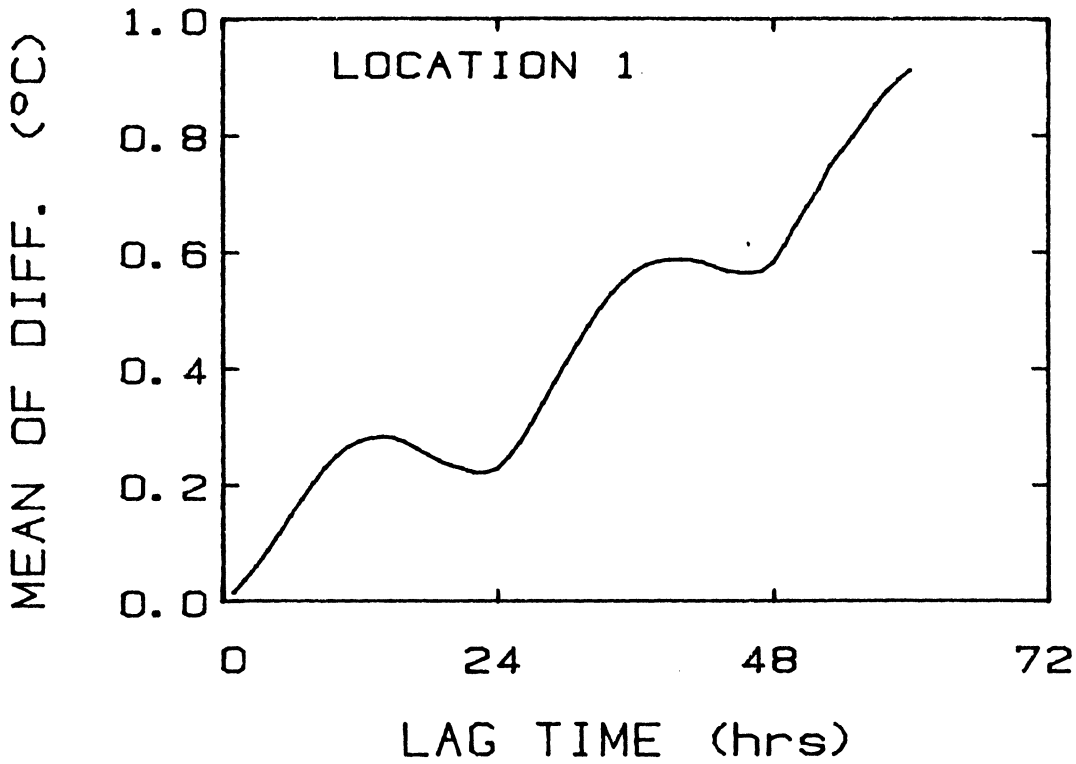

Figure 29. Mean of Differences in Temperature Values Versus lag Time at Location 1.

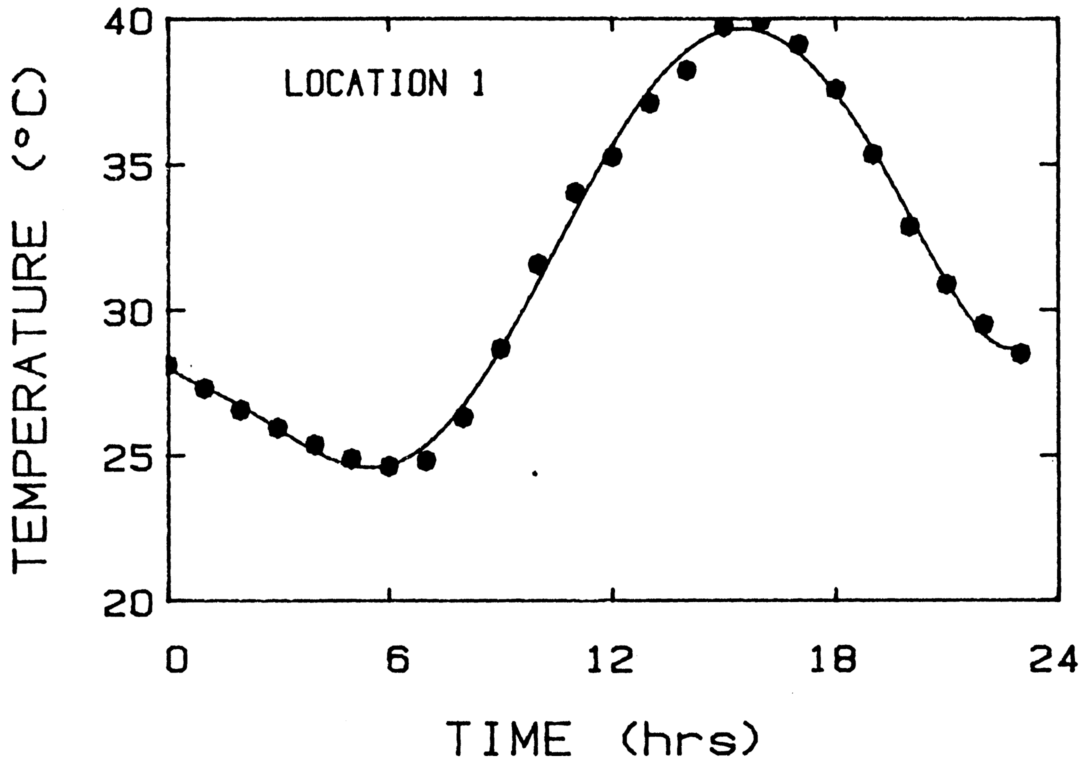

Figure 30. Fitted Polynomial Into Measured Temperature at Location 1.

constructed over one fourth of the lags (or for 60 hours). Figure 31 shows the semi-variograms of residuals at location 1. Note that the structure of the semi-variogram disappeared when the trend was removed. In fact, no real evidence exist of any temporal component to this variance. Similar semi-variograms were obtained at other locations.

The predicted and actual estimation variances were calculated after removing the trend from the data. Figure 32 shows the predicted versus actual estimation variance after removing the trend. The predicted estimation variances were 1.85 to 2.70 times the actual estimation variances. Comparing Figures 32 and 28 shows that removing the trend in the data dramatically reduced the predicted estimation variances at all locations. It appears that adjusting the trend in this manner increases the reliability of the predicted estimation variance.

Figure 33 shows the predicted estimation variance versus the actual estimation variance for the least-squares approach. The predicted estimation variance was 0.98 to 1.70 times the actual estimation variance at any location. These results indicate that the predicted estimation variances for both methods overestimated the actual variance by about the same amount.

Figure 34 shows the actual estimation variance for least-squares versus the actual estimation variance for kriging. The kriging variance was 60 to 67% less than the least-squares actual variance at all locations. As a result, the temperatures estimated by kriging more closely approximated the actual temperatures. However, both methods provide very low variances.

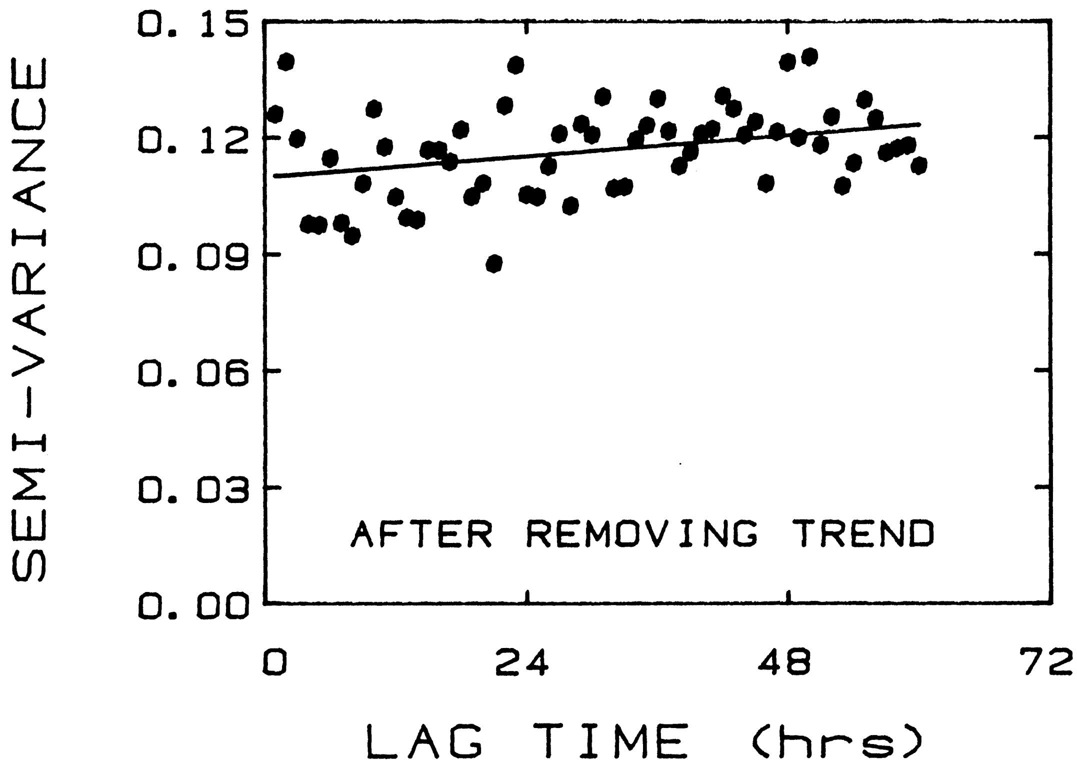

Figure 31. Semi-variogram of Residuals After Removing the Trend at Location 1.

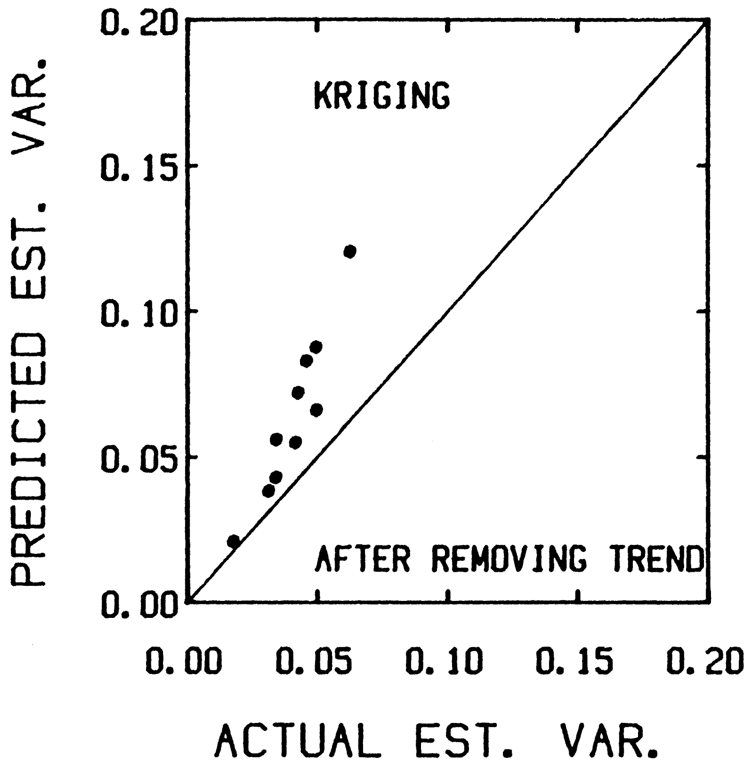

Figure 32. Predicted and Actual Estimation Variance for Kriging After Removing the Trend.

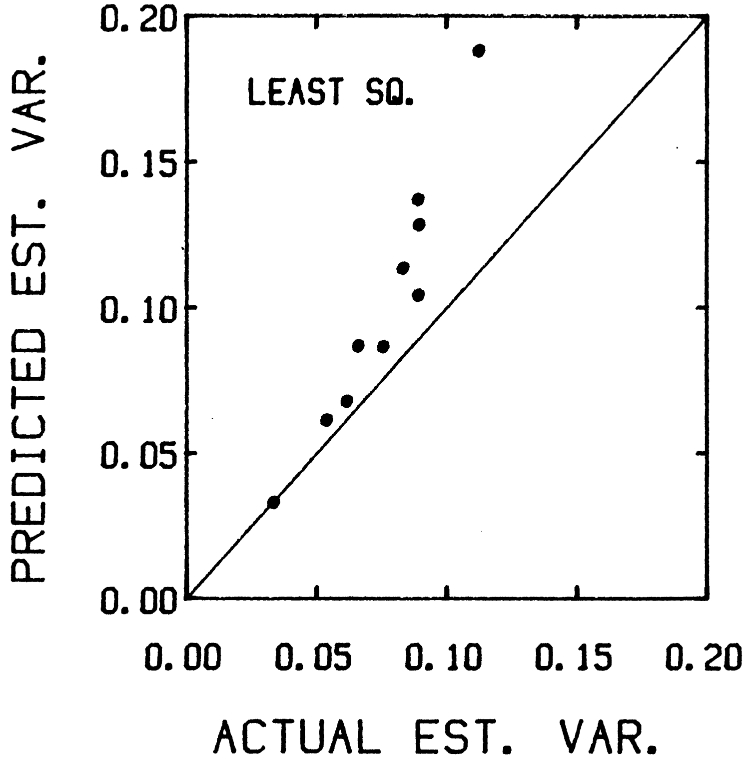

Figure 33. Predicted Estimation Variance Versus Actual Estimation Variance for the Least-squares Techniques.

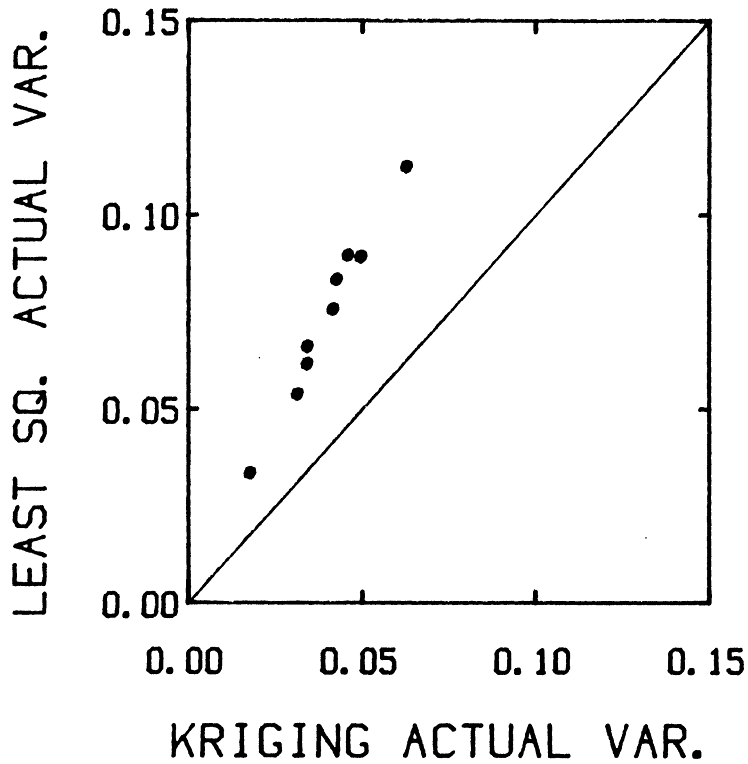

Figure 34. Least-squares Actual Estimation Variance Versus Kriging Actual Estimation Variance.

CHAPTER V

SUMMARY

It was shown that simple kriging based on a linear semi-variogram for values measured along a transect results in linear interpolation between the closest neighboring points. This is true whether the points are uniformly spaced or non-uniformly spaced. The estimated value for a point at either end of the transect is equal to that of its closest neighbor. The values of the weights can be calculated by knowing the positions of the estimated point and its one or two neighboring measured points. The kriged value is independent of the intercept and slope of the linear semi-variogram model. The predicted estimation variance can be calculated from values of the coefficients in linear semi-variogram model, and the position of the estimated point and its closest neighbors. Estimation variance varies linearly with coefficients in linear semi-variogram model.

The variation of soil temperature along the transect increased as temperature increased when the soil was dry. The temperature semi-variograms over distance show that temperature values were spacially dependent to a distance of approximately 30 m. The temperature values estimated by kriging were slightly more reliable than the temperatures estimated by least-squares. The values estimated by least-squares showed more gradual changes than did the estimated values by kriging. The actual variances for kriging were less than the actual variance for the least-squares method in most cases. The actual estimation variances for both methods ranged from 0 to 1.2 °C^2 with approximately 90% of the values less than 0.4 °C^2. Differences in variances for the two methods were less than 0.1 °C^2 for more than 90% of the cases. The estimation variance predicted by kriging overestimated the actual estimation variance, and the estimation variance predicted by least-squares underestimated actual variance.

The variation of soil temperature over time was very gradual. The temperature semi-variograms over time show that temperature values were temporally dependent in each 12-hour period. The temperatures estimated by kriging at any time were slightly more reliable than the temperatures estimated by least-squares. The actual variance for estimated values over time by kriging ranged from 0.02 to 0.06 °C^2 while the values for least-squares ranged from 0.03 to 0.12 °C^2. The predicted estimation variance (for values over time) calculated by kriging greatly overestimated (30 to 47 times) the actual variance when the trend was ignored. When the trend was removed the predicted estimation variance was 1.85 to 2.7 times the actual variance. For least-squares, predicted estimation variance was 0.90 to 1.70 times the actual variance.

LITERATURE CITED

Burgess, T. M., and R. Webster. 1980. Optimal interpolation and isarithmic mapping of soil properties. I. The semi-variogram and punctual kriging. J. Soil Sci. 31:315-331.

Clark, I. 1979. Practical geostatistics. Applied Sci. Publ., Ltd., London.

Clark, I. 1980. The semi-variogram. p. 29-60. In Geostatistics. McGraw-Hill Book Co., New York.

Journel, A. G., and C. J. Huijbregts. 1978. Mining geostatistics. Academic Press, London.

Mikhail, E. M., and F. Ackermann. 1976. Observations and least squares. IEP-A Dun-Donnelley Publ., New York.

Palumbo, M. R., and R. Khaleel. 1983. Kriged estimation of transmissivity in the Mesilla Bolson, New Mexico. Water Resources Bull. 19:929-936.

Russo, D. 1983. A geostatistical approach to the trickle irrigation design in heterogeneous soil. 1. Theory. Water Resources Res. 19:632-642.

Steel, R. G. D., and J. H. Torrie. 1980. Principles and procedures of statistics: A biometrical approach. McGraw-Hill Book Co., New York.

Tabor, J. A., A. W. Warrick, D. A. Pennington, and D. E. Myers. 1984. Spatial variability of nitrate in irrigated cotton: I. Petioles. Soil Sci. Soc. Am. J. 48:602-607.

Uehara, G. Personal Interview. Auburn, Alabama, March 25, 1982.

Vauclin, M., S. R. Vieira, R. Bernard, and J. L. Hatfield. 1982. Spatial variability of surface temperature along two transects of a bare soil. Water Resources Res. 18:1677-1686.

Vauclin, M., S. R. Vieira, G. Vachaud, and D. R. Nielsen. 1983. The use of cokriging with limited field soil observations. Soil Sci. Soc. Am. J. 47:175-184.

Vieira, S. R., D. R. Nielsen, and J. W. Biggar. 1981. Spatial variability of field-measured infiltration rate. Soil Sci. Sco. Am. J. 45:1040-1048.

Webster, R., and T. M. Burgess. 1980. Optimal interpolation and isarithmic mapping of soil properties. III. Changing drift and universal kriging. J. Soil Sci. 31:505-524.

VITA

Parichehr Hemyari

Candidate for the Degree of

Doctor of Philosophy

Thesis: A COMPARISON OF CLASSICAL STATISTICS AND GEOSTATISTICS FOR ESTIMATING SOIL SURFACE TEMPERATURE

Major Field: Soil Science

Biographical:

Personal Data: Born in Shiraz, Iran, December 31, 1955, the daughter of Mohamad Hadi Hemyari and Fatemeh Rahebi.

Education: Graduated from Mehrain High School at Shiraz in June, 1973; received Bachelor of Science degree in Soil Science from Pahlavi University in June 1977; received Master of Science degree in Soil Science from Oklahoma State University in May, 1980; completed requirements for the Doctor of Philosophy at Oklahoma State University in December, 1984.