2000

Determination of Semivariogram Models to Krige Hourly and Daily Solar Irradiance in Western Nebraska*

G.G. Merino

D. Jones

David Stooksbury

Kenneth Hubbard

Determination of Semivariogram Models to Krige Hourly and Daily Solar Irradiance in Western Nebraska*

G. G. MERINO

Department of Agricultural Engineering, University of Concepcion, Chillan, Chile, and Department of Biological System Engineering, University of Nebraska, Lincoln, Nebraska

D. JONES

Department of Biological Systems Engineering, University of Nebraska, Lincoln, Nebraska

D. E. STOOKSBURY

Department of Biological and Agricultural Engineering, University of Georgia, Athens, Georgia

K. G. HUBBARD

High Plains Climate Center, University of Nebraska, Lincoln, Nebraska

(Manuscript received 6 June 2000, in final form 28 September 2000)

ABSTRACT

In this paper, linear and spherical semivariogram models were determined for use in kriging hourly and daily solar irradiation for every season of the year. The data used to generate the models were from 18 weather stations in western Nebraska. The models generated were tested using cross validation. The performance of the spherical and linear semivariogram models were compared with each other and also with the semivariogram models based on the best fit to the sample semivariogram of a particular day or hour. There were no significant differences in the performance of the three models. This result and the comparable errors produced by the models in kriging indicated that the linear and spherical models could be used to perform kriging at any hour and day of the year without deriving an individual semivariogram model for that day or hour.

The seasonal mean absolute errors associated with kriging, within the network, when using the spherical or the linear semivariograms models were between 10% and 13% of the mean irradiation for daily irradiation and between 12% and 20% for hourly irradiation. These errors represent an improvement of 1%–2% when compared with replacing data at a given site with the data of the nearest weather station.

1. Introduction

During the last four decades, many scientists have addressed the problem of determining the spatial and temporal variability of solar radiation. The reasons for this attention are the great number of applications for this climate variable and the limitations in the historical density of radiation measurement networks imposed by economic constraints.

In solar energy engineering, solar radiation data are used for the design and performance simulation of systems such as thermal solar collectors, photovoltaic systems, and passive systems such as space heating and solar buildings. Engineers use monthly mean values of daily solar irradiation for preliminary sizing and productivity estimation of solar energy systems (Markvart 1994). However, the use of monthly means of daily values limits the accuracy of estimates, because most solar energy systems exhibit a nonlinear dependence upon weather variables. Thus, computer simulations are generally used to accommodate the complexity involved with design optimization of solar energy systems (Fiksel et al. 1995). These simulations require hourly climatic and load variables as input data.

Quite often, the engineer or the researcher faces the problem of not having hourly or even daily solar radiation data available at the site under study. In such situations, one alternative is to extrapolate solar radiation data measured at other locations. In such a case, it becomes necessary to assess the spatial variability of

* This paper has been approved as Journal No. 12746 of the Nebraska Agricultural Experiment Station.

Corresponding author address: Gabriel Merino, Dept. of Biological Systems Engineering, 212 L. W. Chase Hall, Lincoln, NE 68583-0726.
E-mail: gmerino@unlserve.unl.edu

solar radiation not only in terms of the radiation regime itself but also with regard to the economic consequences.

Many authors have studied the extrapolation and interpolation of daily and hourly solar radiation. Generally speaking, they determined the spatial variability between data recorded at different weather stations as a function of the station separation distances. To quantify the spatial variability, most of the authors have used the correlation coefficient (Kerr et al. 1968; Long and Ackerman 1995; Barnett et al. 1998), the coefficient of determination (Kerr et al. 1968; Hubbard 1994), the coefficient of variability (Suckling 1983), or the standard deviation of the differences of measured solar irradiation (Suckling 1985; Long and Ackerman 1995). These statistics have been determined as functions of separation distance between station pairs. In almost every study, the statistics were calculated from sets of aggregated data. From the calculated variability, conclusions have been made about the errors involved when using the data of a particular weather station as being representative of a nearby site.

In relation to the approaches described above, there are limitations that should be highlighted. 1) The correlation coefficient r may be low when the range of values of solar radiation is small, even under high spatial coherence (Hay and Hanson 1985). 2) As demonstrated by Gunst (1995), when calculating the above statistics for station pairs using aggregated data, diurnal and seasonal trends inherent to solar radiation force the calculated statistics to account not only for spatial coherence but also for temporal correlation. 3) Most of the studies have used aggregated data for a period of 1 yr or less to estimate spatial correlation; however, as determined by Hubbard (1994), when using aggregated daily data, it is necessary to consider at least 5 yr of data to get stable values of the determination coefficient. 4) When spatial correlation is estimated for daily and hourly solar radiation from aggregated data, the estimated values do not represent the coherence of the radiation field in individual days or hours of the period from which the data were obtained (Long and Ackerman 1995). This fact makes statistics calculated from aggregate data useless for inferring extrapolation errors at any particular day or hour.

The above limitations have prompted the need to address the problem of extrapolation of hourly and daily solar irradiation from the perspective taken by Bland and Clayton (1994) and by Bland (1996), that is, using kriging as the extrapolation method. Bland and Clayton (1994) analyzed the spatial structure of solar radiation in Wisconsin by using satellite-derived measurements of daily radiation to investigate the effects of network density on errors of prediction when only sparse data are available. Bland and Clayton (1994) included 44 days of 1984 in their study. On individual days, all locations in the complete database were estimated by kriging, using the semivariogram from that day and semivariograms from other days. Semivariograms of solar radiation calculated on individual days were essentially linear. The researchers concluded that, when semivariograms from the day with the greatest variation are used, errors less than 4.6 MJ m^{-2} day^{-1} can be expected for 80% of the area within the measurement grid on 90% of the days, given a 270-km spacing between stations. Increasing network density fourfold typically halved absolute errors of estimation. The errors of prediction by kriging were calculated as the absolute difference between the true value for the point from the complete dataset and the kriged value. Bland (1996) studied the adequacy of the current array of ground-based pyranometers in the Midwest using kriging over a network covering 655 000 km^2 (four states). Bland (1996) concluded that the array of 51 pyranometers provided sufficient data to interpolate daily irradiation with a median error of 2 MJ m^{-2} day^{-1} for partially cloudy days at locations 40–80 km away from the nearest measurement. These conclusions were made from the analysis of 44 days of 1992 taken within the period of June to November.

In this study, we examine the capability of a network of weather stations in western Nebraska to estimate (by kriging) hourly and daily solar radiation at any point within the region covered by the network. This capability was judged according to the extrapolation errors obtained when the data at a particular site are estimated using the data of all the stations in the network simultaneously.

One of the principal kriging problems is to select one from among the unlimited number of semivariogram models that ensure unique solutions to the kriging equations. Because of the high variability of solar radiation with time, finding semivariogram models that can be applied to long time series of daily and hourly solar irradiation is a difficult task and at the same time a mandatory one in order to make kriging an efficient process.

Instead of using an aggregate set of daily or hourly solar radiation data to estimate semivariogram models, all days and a set of 600 h of 1996 were analyzed individually. In that way, the problem of seasonal and diurnal trends affecting the spatial variability was avoided. From the semivariograms obtained, semivariogram models were developed that can be applied to any day and hour of the year.

Accordingly, the objectives of this work were: 1) to determine semivariogram models for kriging hourly solar irradiation, 2) to determine semivariogram models applicable for kriging daily solar irradiation, and 3) to quantify the errors associated with estimates derived from hourly and daily solar irradiation data with the models obtained.

2. Materials and methods

a. Data

Hourly and daily integrated values of global solar irradiation on a horizontal surface were analyzed in this

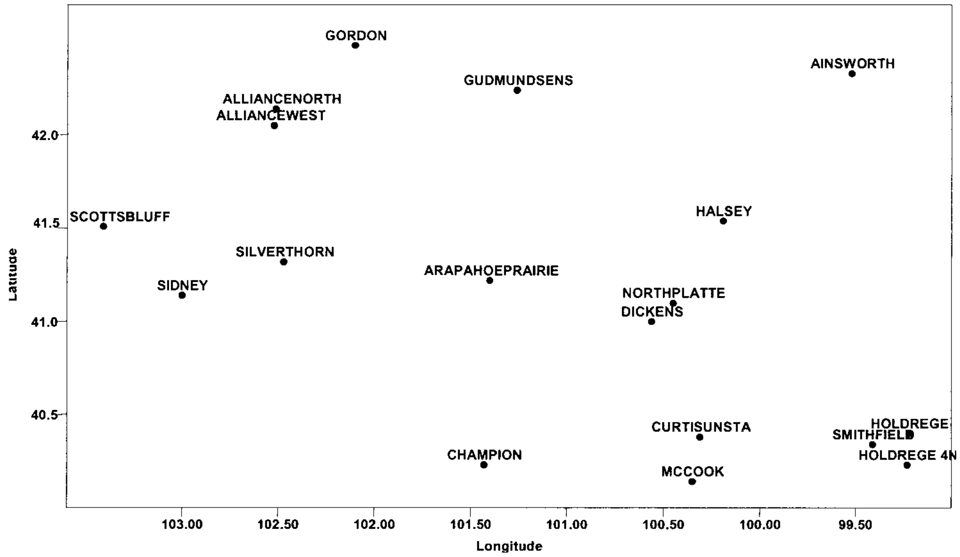

FIG. 1. Map of climate stations in western Nebraska.

study. The data used are from 1996 and come from 18 weather stations in western Nebraska. These stations form part of the High Plains Regional Climate Center's (HPRCC) Automated Weather Data Network.

At these stations, a silicon photodiode detector, the Li-Cor, Inc., 200,[1] is used to measure solar radiation. The sensors are calibrated periodically with an Eppley precision spectral pyranometer. This calibration gives an absolute accuracy of about 5% (Barnett et al. 1998). Maintenance is performed by the HPRCC. Data are downloaded, quality controlled, and archived by the HPRCC.

A map of the weather stations used is shown in Fig. 1. The latitude and longitude differences between stations are not greater than 2.5° and 4.5°, respectively, and the distances fluctuate between 10 and 375 km. With these small ranges in latitude and longitude, possible problems of anisotropy were avoided (Gunst 1995).

b. *Ordinary kriging*

1) THE KRIGING SYSTEM

In this study, ordinary kriging was evaluated as a procedure for extrapolation of hourly and daily solar irradiation data. Ordinary kriging is a linear estimator by which an estimated value of solar irradiation in a particular site can be calculated, from solar radiation measured at other sites, according to the linear combination

$$Z^*(x) = \sum_{j=1}^{k} \omega_j Z(x_j), \tag{1}$$

where Z is solar irradiation, $Z^*(x)$ is the estimated value of Z at point x, $Z(x_j)$ is the measured value of Z at point x_j, ω_j is the weight given to observed value $Z(x_j)$ (these weights are allowed to change as estimates are computed at different locations), x is the coordinates [Universal Transverse Mercator (UTM)] of an estimated point, x_j is the coordinates (UTM) of a measured value, and k is the number of measured values used in the estimation. The weights ω_j are calculated using the ordinary kriging system of equations (Isaaks and Srivastava 1989).

A common approach when solving the kriging system of equations is to employ what is called a semivariogram function $\gamma(h)$. An estimate of the semivariogram can be calculated from the equation

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [Z(x_i) - Z(x_{i+h})]^2. \tag{2}$$

In this equation, $Z(x_i)$ and $Z(x_{i+h})$ are values of the variable Z measured at points x_i and x_{i+h}, respectively, separated by a distance h. Data measurements sites are rarely spaced exactly at h distances apart; therefore, to use Eq. (2) to estimate semivariograms, data must be grouped into pairs with similar separation distances of about h (distance lags). Each lag contains $n(h)$ number of pairs.

Once the semivariogram function has been computed from the sampled values of Z at different locations, the next step is to fit a parametric semivariogram model to the sample semivariogram $\gamma(h)$. There are two reasons for not using the sample semivariogram directly in the ordinary kriging system. First, the kriging system of equations may need semivariogram values for distances that are not present in the sample data; this requirement will depend on the location at which Z is being estimated. Second, the use of the sample semivariogram

[1] Mention of a vendor does not imply endorsement over other vendors.

does not guarantee the existence and uniqueness of the solution to the ordinary kriging system of equations. To accomplish such requirements, a semivariogram function that is negative definite must be provided (Armstrong and Diamond 1984). A common method used to fit parametric semivariogram models to the sample semivariogram is least squares curve fitting (Cressie 1985). That approach was taken in this work using the proprietary software GS+ (Robertson 1998).

The most common semivariogram models used to fit sample semivariograms and therefore used to describe the spatial variability of the variable under study are Gaussian, spherical, exponential, and linear. The parameters that characterize these models are the nugget, sill, and range and are calculated through the fitting process. As is reported later in this paper, the semivariograms that fit the sample semivariograms obtained when analyzing daily and hourly solar irradiation were in almost every case the linear or spherical models.

The parameters that determine the semivariogram models to fit the sample semivariograms were calculated using GS+ (Robertson 1998). First, this software calculated the sample semivariogram from measured values of solar irradiation and the spatial coordinates corresponding to these values by using Eq. (2). Then, curves corresponding to spherical, linear, Gaussian, and exponential were derived using least squares curve fitting. As a result of this process, the parameters for each model and the regression coefficient R^2 of the fitting procedure were determined. The semivariogram model with the higher value of R^2 was selected as the appropriate model to represent the sample semivariogram.

2) Cross validation

This technique was performed using the public-domain software package GEO-EAS (Geostatistical Environmental Assessment Software; Englund and Sparks 1988). Cross validation was used to compare kriged estimates obtained with different semivariogram models and true measured values. This process allowed a comparison between different semivariogram models and a quantification of the errors associated with the kriging process. The errors were calculated as the difference between the estimated (kriged) solar irradiation at the stations and the actual value of solar irradiation measured at these stations. The estimated value at a particular station was obtained by discarding the actual value temporarily from the dataset and then estimating (kriging) that value using the remaining data from the rest of the stations. This procedure was repeated for all the stations. Thus, a list of true values and estimated values was obtained for the set of stations, and the distribution of errors was analyzed.

The distribution of errors was analyzed using three summary statistics: the mean error (ME), the mean absolute error (MAE), and the root-mean-square error (rmse):

$$\mathrm{ME} = \frac{\sum_{i=1}^{k} [Z^*(x_i) - Z(x_i)]}{k}, \tag{3}$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{k} |Z^*(x_i) - Z(x_i)|}{k}, \tag{4}$$

$$\mathrm{rmse} = \left\{\frac{\sum_{i=1}^{k} [Z^*(x_i) - Z(x_i)]^2}{k}\right\}^{1/2}, \tag{5}$$

where k is the number of measured values in the dataset $Z^*(x_i)$ is the estimated value of variable Z at point x_i, and $Z(x_i)$ is the actual value of Z at point x_i.

The ME quantified the overall bias and detected if the semivariogram model was producing overestimation or underestimation in the kriging process. The MAE quantified the net bias generated, and the rmse incorporated both the bias and the spread of the error distribution (Isaaks and Srivastava 1989).

3. Data analysis and results

a. Determination of generic semivariogram models for hourly and daily solar irradiation

For each day of 1996, a sample semivariogram was calculated using Eq. (2) and daily solar irradiation data from the 18 weather stations. The next step was to fit a semivariogram model to each sample semivariogram using the least squares curve-fitting procedure. The model with the greatest value of R^2 was taken as most representative of the sample semivariogram. The outputs of this process were the semivariogram model parameters and a value of the corresponding regression coefficient, R^2. After semivariogram models were determined for every day of 1996, they were grouped according to the season of the year to which they belonged. Because it was considered visually impossible to associate any of the semivariogram models with the sample semivariogram when the R^2value was less than 0.3, only semivariogram models that were determined to have a regression coefficient greater than 0.3 were kept for further analysis. To determine if particular days (with low or high irradiation) were arbitrarily left out of the analysis when dropping semivariogram models with low R^2, graphs were made of average daily irradiation (over the 18 stations) versus R^2 for each season. The result of this is shown in Fig. 2 for the summer season. No significant correlation was found in this figure between the average daily irradiation and R^2, and similar results were found for the other seasons.

In the curve-fitting process, the exponential model always represented a very low occurrence frequency (less than 6%). Therefore, only spherical and linear models were considered as feasible candidates for kriging daily solar irradiation.

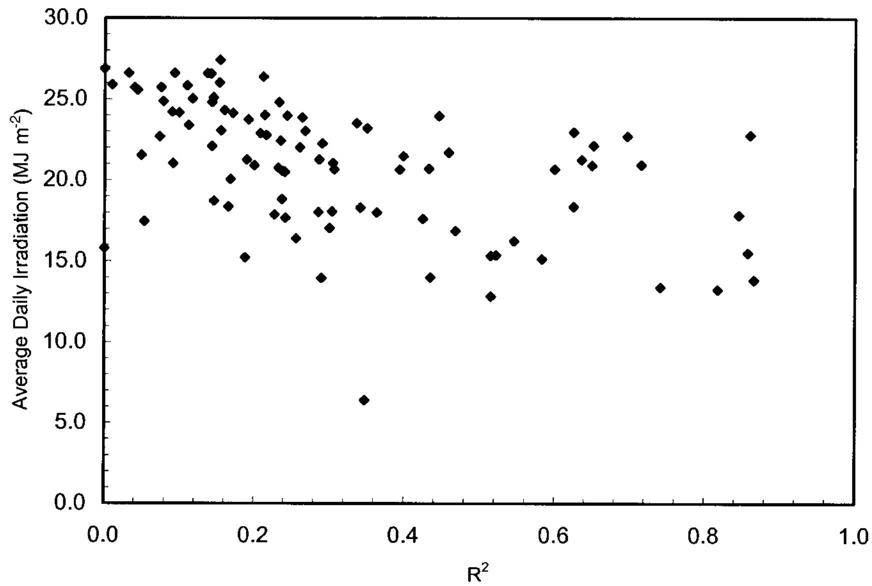

FIG. 2. Average daily irradiation of 18 climate stations vs regression coefficient R^2 of the best-fit model to the sample semivariograms for all days in summer.

The next problem to solve was to find only one spherical model and one linear model for each season. The approach taken was to average the semivariogram parameters for a given model over individual days (Gunst 1995). In summary, using this method for each season, eight semivariogram models, two for each season (linear and spherical), were produced.

The same procedure to obtain semivariogram models for daily irradiation was applied to obtain the models for hourly irradiation. The only difference in the analysis was that not all hours of 1996 were analyzed. Initially, 30 days were randomly selected from each season, and then from each day, 5 h were again randomly selected. This procedure gave 600 h to be analyzed. The same procedure described above for daily data was applied to these hours. An important problem found at this point was the high variability between the semivariogram model parameters calculated for different hours. Thus, to obtain semivariogram parameters that were really representative of each season, more random days and therefore more random hours were added to the analysis until the model parameters calculated from the averages became stable.

To determine if particular hours (with low or high irradiation) were left out of the analysis when dropping semivariogram models with low R^2, graphs were made of average hourly irradiation (over the 18 stations) versus R^2 for each season. The result is shown in Fig. 3 for the summer season. No significant correlation was found in this figure between the average hourly irradiation and R^2, and similar results were found for the other seasons.

The semivariogram models obtained for daily and hourly solar irradiation data are presented in Table 1 and Table 2, respectively. In these tables, R^2 is the average regression coefficient of the models from which the parameters were derived. The frequency column accounts for the fraction of the total semivariograms calculated for each season that corresponded to a particular type of model. For example, for daily irradiation, 15.2% of the sample semivariograms calculated for summer were best fit with linear models and 21.7% with spher-

TABLE 1. Linear and spherical semivariogram model parameters by season, average regression coefficient R, and frequency of occurrence for daily solar irradiation.

Season	Model	Nugget (MJ m^{-2} day^{-1})^2	Sill (MJ m^{-2} day^{-1})^2	Range (km)	R^2	Frequency (%)
Summer	Linear	1.07	46.55	343.21	0.63	15.2
	Spherical	0.07	28.61	187.56	0.46	21.7
Autumn	Linear	0.33	14.80	343.21	0.65	35.1
	Spherical	0.15	9.17	233.80	0.48	26.3
Winter	Linear	0.11	4.96	343.21	0.69	49.4
	Spherical	0.22	3.40	248.59	0.51	27.4
Spring	Linear	0.48	29.65	343.21	0.71	35.8
	Spherical	0.50	21.14	222.97	0.54	33.7

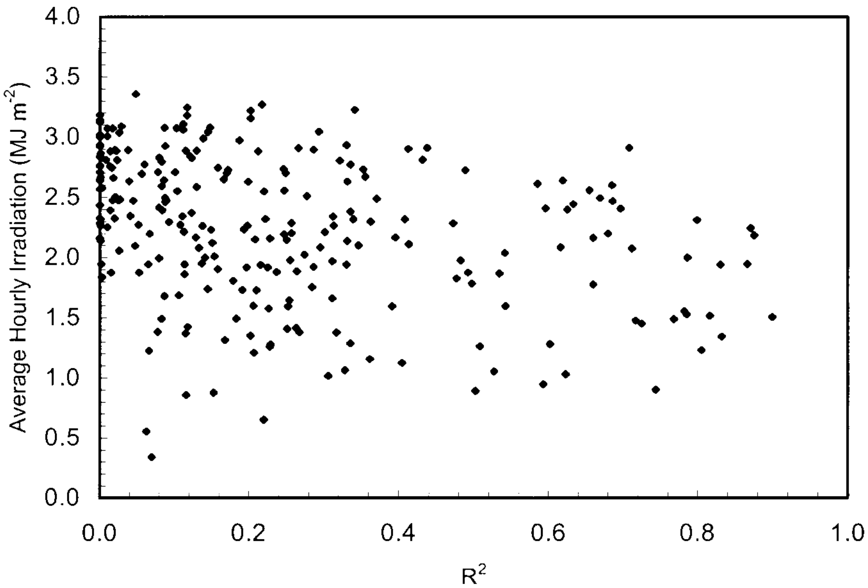

FIG. 3. Average hourly irradiation of 18 climate stations vs regression coefficient R^2 of the best-fit model to the sample semivariogram for 250 h in summer.

ical models (see Table 1). The information in Table 1 indicates that linear models fit the sample semivariogram with more frequency than the spherical semivariogram and with a higher average regression coefficient, which agrees with the results reported by Bland and Clayton (1994). Also, the nugget values for the linear model found here agree with the values reported by Bland (1996).

The seasonal variability of daily and hourly solar irradiation was reflected in the considerable seasonal variations of the sill parameter for the linear model as well as for the spherical model. This result agrees with that observed by Suckling (1985). The same issue is evident from the variations with season observed in the range parameters for the spherical model. The maximum range obtained in this model was about 250 km for daily and for hourly irradiation, which is over 100 km less than the largest distance between stations in the network. This supports the hypothesis that the network chosen was big enough to describe daily and hourly spatial variability. The range values reported for the linear models do not change, because they correspond to the maximum lag distances used when the sample semivariograms were calculated. Actually, a linear model does not have a range.

b. *Cross validation using the generic semivariogram models generated*

The goal of this process was to compare the performance (in kriging) of the semivariogram models generated for every season with the semivariogram models that best fit the sample semivariograms obtained for any particular day or hour.

To perform cross validation with semivariogram models for daily and hourly solar irradiation, 30 days and 30 h were randomly selected within every season of 1996. In each of these 120 days and 120 h, three cross validations were performed using the daily (hourly) solar irradiation measured at the 18 stations. The first cross validation was made using the linear semivariogram model determined for the corresponding season from

TABLE 2. Linear and spherical semivariogram model parameters by season, average regression coefficient *R*, and frequency of occurrence for hourly solar irradiation.

Season	Model	Nugget (MJ m^{-2} h^{-1})^2	Sill (MJ m^{-2} h^{-1})^2	Range (km)	R^2	Frequency (%)
Summer	Linear	0.050	0.90	343.2	0.54	20.8
	Spherical	0.026	0.54	240.0	0.46	12.8
Autumn	Linear	0.024	0.70	343.2	0.62	15.6
	Spherical	0.009	0.24	200.0	0.46	25.2
Winter	Linear	0.010	0.19	343.2	0.62	20.4
	Spherical	0.005	0.10	210.0	0.33	31.8
Spring	Linear	0.022	0.58	343.2	0.73	28.2
	Spherical	0.016	0.40	250.0	0.48	21.4

TABLE 3. Cross validation results of daily solar irradiation by season for 120 days in 1996 using spherical model, best-fit models, and the NN procedure. Statistics are defined in text.

Season	Model	r^2	ME (MJ m^{-2} day^{-1})	MAE (MJ m^{-2} day^{-1})	Rmse (MJ m^{-2} day^{-1})
Summer	Spherical	0.54	0.12 (0.6%)*	2.80 (13.1%)*	3.89
(21.31)**	Best-fit	0.53	0.14 (0.7%)	2.76 (13.0%)	3.94
	NN	0.50	−0.75 (−3.5%)	2.80 (13.1%)	4.25
Spring	Spherical	0.80	0.23 (1.3%)	1.96 (11.3%)	2.87
(17.30)	Best-fit	0.79	0.12 (0.7%)	1.95 (11.3%)	2.95
	NN	0.71	−0.88 (−5.1%)	2.14 (12.4%)	3.64
Autumn	Spherical	0.89	0.09 (0.9%)	1.16 (11.5%)	1.70
(10.08)	Best-fit	0.90	−0.06 (0.6%)	1.12 (11.1%)	1.64
	NN	0.86	−0.30 (−3.0%)	1.30 (12.9%)	1.89
Winter	Spherical	0.89	0.08 (0.9%)	0.89 (10.4%)	1.20
(8.54)	Best-fit	0.90	0.08 (0.9%)	0.87 (10.2%)	1.18
	NN	0.82	−0.23 (−2.7%)	1.04 (12.2%)	1.46

* Percentage of the mean irradiation.
** Mean irradiation (MJ m^{-2} day^{-1}).

which the particular day (hour) studied belonged. The second cross validation was made with the spherical semivariogram model determined for the same season. Last, a third cross validation was made using the semivariogram model that best fit the sample semivariogram calculated with the data corresponding to the day (hour) studied. This third model, determined through a fitting procedure, varied from day (hour) to day (hour) and could be any of the semivariogram models presented before (linear, spherical, Gaussian, or exponential).

The output of the process described above was three sets of estimated and measured solar radiation for each of the 120 days (hours) randomly selected. Each set consisted of 18 measured and 18 estimated values, since 18 stations were considered. On the other hand, each set corresponds to the particular model used to generate the estimated data (linear, spherical, or the model fit to the sample semivariogram). To evaluate quantitatively the performance of each model, the statistics MAE, rmse, and ME were applied to these sets after they were grouped together according to the season considered. Thus, for example, for summer and spherical model, the statistics were applied to the complete set of actual and estimated values at the 18 stations in the 30 days (hours) selected.

No significant differences were found between the statistics calculated with the actual and estimated solar radiation using the linear model and the spherical model; therefore, only the results with the spherical model are reported in this paper. The resultant statistics are presented in Table 3 and Table 4 for daily and hourly irradiation, respectively. In these tables, the coefficient of determination r^2 of the correlation made between the measured and estimated values was also included. The ME and the MAE statistics are given as actual values and as a percentage of the average measured irradiation (mean irradiation) over the 18 stations and over the days (hours) considered in the cross validation.

For the worst case (summer), graphical representations of correlation between actual and estimated values with the spherical model are reported in Fig. 4 and Fig. 5 for daily and hourly irradiation, respectively. As can

TABLE 4. Cross validation results of hourly solar irradiation by season for 120 h in 1996 using spherical model, best-fit models, and NN procedure. Statistics are defined in the text.

Season	Model	r^2	ME (MJ m^{-2} m^{-1})	MAE (MJ m^{-2} m^{-1})	rmse (MJ m^{-2} m^{-1})
Summer	Spherical	0.68	−0.05 (−2.3%)*	0.35 (16.1%)*	0.48
(2.18)**	Best-fit	0.70	−0.04 (−1.8%)	0.33 (15.1%)	0.46
	NN	0.63	0.01 (0.5%)	0.38 (17.4%)	0.51
Spring	Spherical	0.81	0.01 (0.5%)	0.23 (12.6%)	0.33
(1.83)	Best-fit	0.82	0.003 (0.2%)	0.23 (12.6%)	0.33
	NN	0.75	−0.027 (−1.5%)	0.27 (14.8%)	0.39
Autumn	Spherical	0.82	0.01 (0.9%)	0.21 (19.6%)	0.31
(1.07)	Best-fit	0.84	0.01 (1.9%)	0.20 (18.7%)	0.29
	NN	0.80	−0.02 (−1.9%)	0.22 (20.6%)	0.33
Winter	Spherical	0.87	0.01 (1.0%)	0.15 (15.6%)	0.22
(0.96)	Best-fit	0.88	0.01 (1.0%)	0.15 (15.6%)	0.21
	NN	0.83	−0.03 (−3.1%)	0.17 (17.7%)	0.26

* Percentage of the mean irradiation.
** Mean irradiation (MJ m^{-2} h^{-1}).

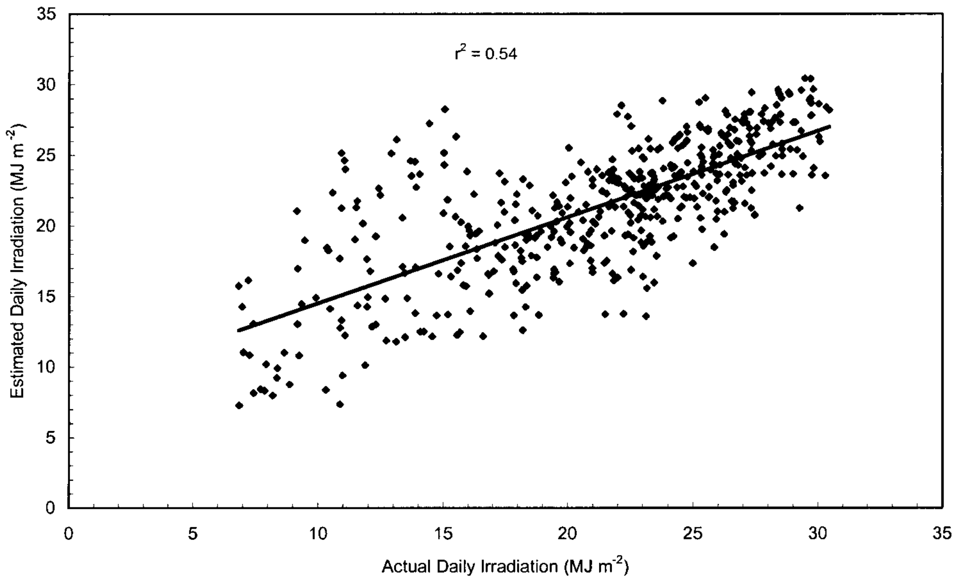

FIG. 4. Estimated vs actual daily solar irradiation using the spherical semivariogram model for daily irradiation in summer.

be seen in Fig. 4, the spread in estimated daily values is higher for intermediate values of solar irradiation than for low or high solar radiation. This result indicates that kriging, like any other extrapolation method (Long and Ackerman 1995), performs better for overcast or clear sky days than for partly cloudy days. The same phenomenon was observed for hourly solar radiation, with a lower degree of differentiation due to the fact that the errors obtained with kriging are higher for hourly data than for daily data (in percentage).

To compare the performance of kriging with a simpler procedure to estimate missing solar radiation data, cross validation was also performed using the nearest neighbor (NN) technique. In this procedure, each estimated data value for each station was obtained by substituting for each actual data value with the actual data from the nearest station. This analysis was performed for daily and hourly solar radiation using the data from the same hours and days that were used to test the semivariogram models.

The results reported in Table 3 and Table 4 indicated no significant differences in the statistics r^2, ME, MAE, and rmse when the generated models (linear and spherical) and the best-fit models (for individual days and hours) were used in cross validation. The same conclusion can be made for hourly data (Table 4). The meaning

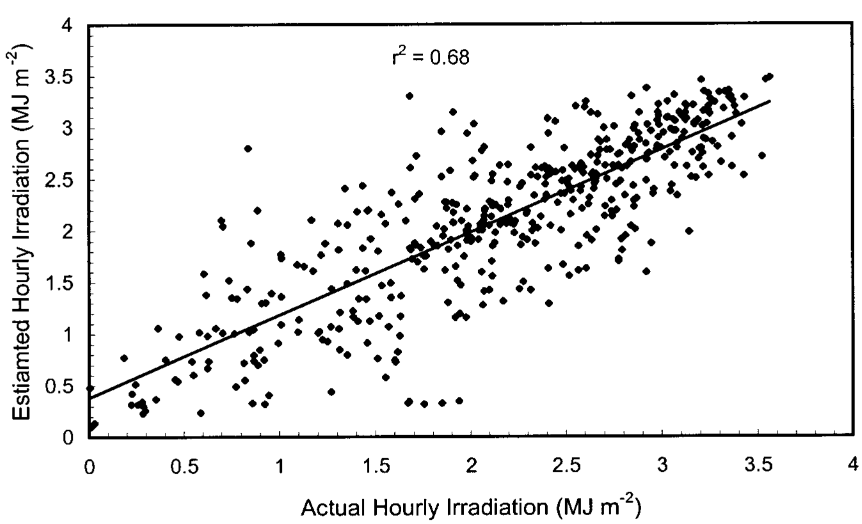

FIG. 5. Estimated vs actual hourly solar irradiation using the spherical semivariogram model for daily irradiation in summer.

of this is that the spherical or the generic linear models generated (Tables 1 and 2) could be used to perform kriging on any day (hour) without developing a specific semivariogram for that particular day or hour.

The best result obtained with these semivariogram models in cross validation for daily radiation was obtained in winter, when the MAEs were less than 1 MJ m^{-2} day^{-1}, which corresponded to about 10% of the daily measured irradiation. The determination coefficient r^2 between actual and estimated daily irradiation was about 0.9 in this case. On the other hand, the worst results were obtained in summer, with an MAE of about 2.8 MJ m^{-2} day^{-1}, which corresponded to about 13% of the mean measured irradiation. In this case, r^2 was about 0.5. These results represent a considerable improvement when compared with the errors predicted by Suckling (1985) and Kerr et al. (1968).

For hourly irradiation, the best result in cross validation was also obtained in winter, when the MAEs were about 0.15 MJ m^{-2} h^{-1}, which corresponded to about 15% of the hourly measured irradiation. The determination coefficient r^2 between actual and estimated hourly irradiation was about 0.9 in this case. On the other hand, the worst results were obtained in summer, with an MAE of about 0.35 MJ m^{-2} h^{-1}, which corresponded to about 16% of the mean measured irradiation. In this case, r^2 was about 0.7. Thus, we have larger percentage errors when kriging hourly data using the generated and best-fit models than when kriging daily data. However, considering the errors obtained and the distances between stations in the network, a considerable improvement is obtained when compared with the results reported by Hay and Hanson (1985) and Barnett et al. (1998).

When using the generic spherical (or linear) semivariogram models or the best-fit models, the results based on cross validation were better than the results obtained with the NN procedure (see Tables 3 and 4). Thus, for the MAE an improvement between 1% and 2% (Tables 3 and 4) of the mean irradiation can be obtained using kriging versus the NN procedure.

4. Conclusions

In this paper, spatial variability of daily and hourly solar irradiation was analyzed from a different point of view. Ordinary kriging was evaluated as a method for extrapolation of irradiation within a network of weather stations in western Nebraska.

To determine semivariogram models to perform kriging of solar irradiation, days and hours were analyzed individually to avoid the influence of seasonal and daily trends in the spatial structure. From this analysis, seasonal semivariogram models (linear and spherical) were developed and tested against the semivariogram models that best fit a particular day or hour. Because the results indicated no significant differences in kriging when using the linear, spherical, or best-fit semivariogram models, it can be concluded that the seasonal linear models and the seasonal spherical models will work equally well for kriging solar radiation. This finding is considered to be of major importance, because it allows development of computational programs for kriging irradiation in the region without choosing semivariograms on a daily or hourly basis.

All kriging approaches gave the poorest performance in the summer season, which was linked to intermediate solar radiation values that apparently exhibit higher variability due to the complexity of summer convective cloud systems. The seasonal mean absolute errors obtained in kriging, within the network, when using the spherical or the linear semivariogram models were between 10% and 13% of the mean irradiation for daily irradiation and between 12% and 20% for hourly solar irradiation. These errors are not considered to be excessive, given that the accuracy of measurements has been estimated to be 5%. These errors represent an improvement of 1% to 2% when compared with those obtained by replacing data at a given site with data from the nearest weather station. In any case, these errors have to be evaluated in terms of the specific application of interest. Such analysis will be addressed in a future work.

REFERENCES

Armstrong, M., and P. Diamond, 1984: Testing variograms for positive-definiteness. *Math. Geol.,* **16,** 407–421.

Barnett, T. P., J. Ritchie, and G. Stokes, 1998: On the space–time scales of the surface solar radiation field. *J. Climate,* **11,** 88–96.

Bland, W. L., 1996: Uncertainty of daily insolation estimates from a mesoscale pyranometer network. *J. Atmos. Oceanic Technol.,* **13,** 255–261.

——, and M. K. Clayton, 1994: Spatial structure of solar radiation in Wisconsin. *Agric. For. Meteor.,* **69,** 75–84.

Cressie, N. A. C., 1985: Fitting variogram models by weighted least squares. *Math. Geol.* **17,** 563–586.

Englund, E., and A. Sparks, 1988: GEO-EAS 1.1 user's guide. U.S. EPA Environmental Monitoring Systems Laboratory, Office of Research and Development, Las Vegas, NV 89193-3478.

Fiksel, A., J. W. Thornton, S. A. Klein, and W. A. Beckman, 1995: Developments to the TRNSYS simulation program. *J. Sol. Energy Eng.,* **117,** 123–127.

Gunst, R. F., 1995: Estimating spatial correlations from spatial–temporal meteorological data. *J. Climate,* **8,** 2454–2470.

Hay, J. E., and K. J. Hanson, 1985: Evaluating the solar resource: A review of problems resulting from temporal, spatial and angular variations. *Sol. Energy,* **34,** 151–161.

Hubbard, K. G., 1994: Spatial variability of daily weather variables in the High Plains of the USA. *Agric. For. Meteor.,* **68,** 29–41.

Isaaks, E. H., and R. M. Srivastava, 1989: *An Introduction to Applied Geostatistics.* Oxford University Press, 561 pp.

Kerr, J. P., G. W. Thurtell, and C. B. Tanner, 1968: Mesoscale sampling of global radiation analysis of data from Wisconsin. *Mon. Wea. Rev.,* **96,** 237–241.

Long C. N., and T. P. Ackerman, 1995: Surface measurements of solar irradiance: A study of the spatial correlation between simultaneous measurements at separated sites. *J. Appl. Meteor.,* **34,** 1039–1046.

Markvart, T., 1994: *Solar Electricity*. John Wiley and Sons, 228 pp.

Robertson, G. P., 1998: *GS+: Geostatistics for the Environmental Sciences*. Gamma Design Software, Plainwell, Michigan, 152 pp.

Suckling, P. W., 1983: Extrapolation of solar radiation measurements: Mesoscale analyses from Arizona and Tennessee Valley Authority regions. *J. Appl. Meteor.*, **22,** 448–494.

——, 1985: Estimating daily solar radiation values in selected mid-latitude regions by extrapolating measurements from nearby stations. *Sol. Energy*, **35,** 491–495.