United States
Department of
Agriculture

Forest
Service

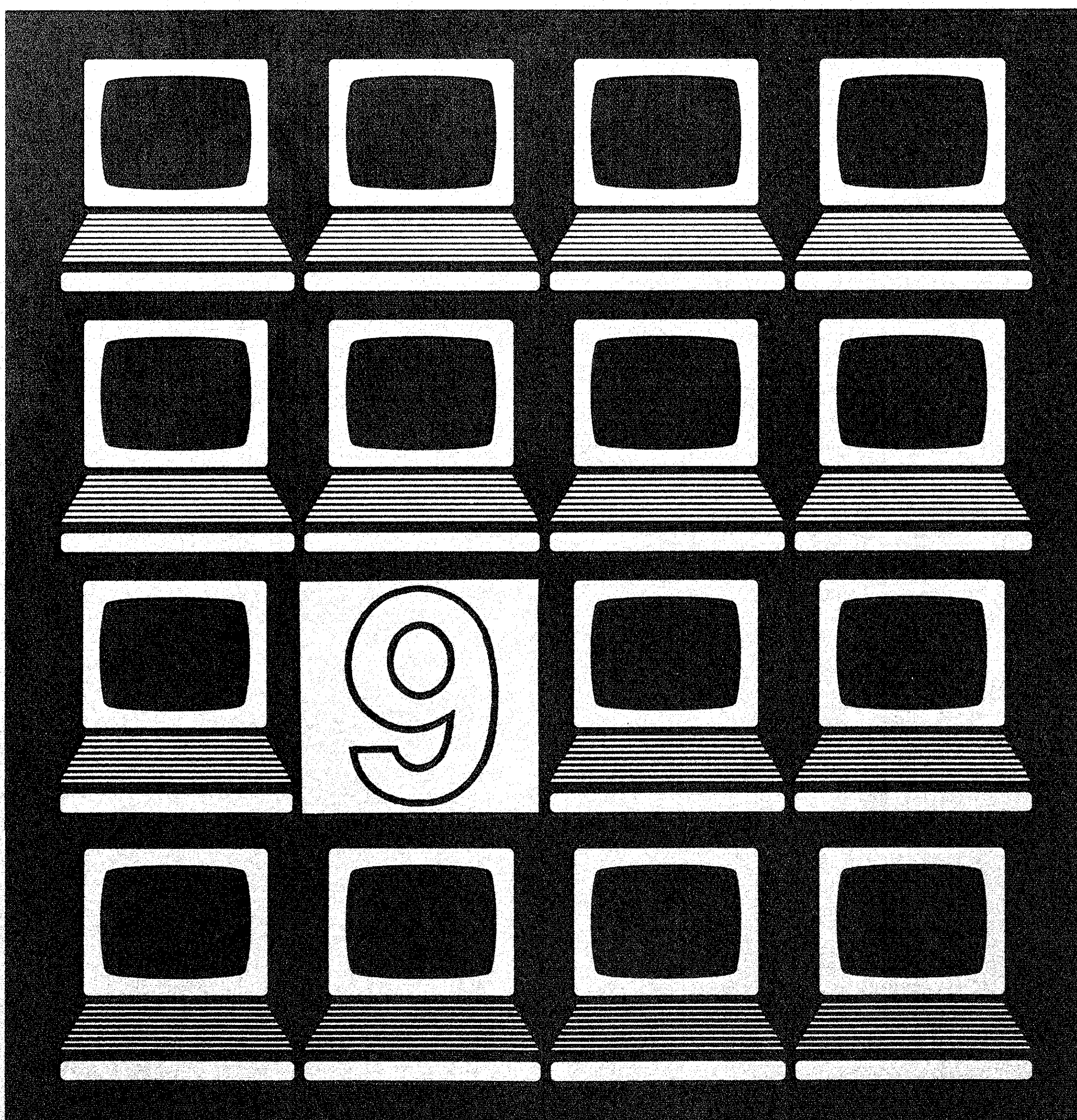
North Central
Forest Experiment
Station

General Technical
Report NC-167

The Microcomputer Scientific Software Series 9

User's Guide to Geo-CLM: Geostatistical Interpolation of the Historical Climatic Record in the Lake States.

Margaret R. Holdaway



Contents

North Central Forest Experiment Station
Forest Service—U.S. Department of Agriculture
1992 Folwell Avenue
St. Paul, Minnesota 55108
Manuscript approved for publication March 2, 1994
1994

Contents

North Central Forest Experiment Station
Forest Service—U.S. Department of Agriculture
1992 Folwell Avenue
St. Paul, Minnesota 55108
Manuscript approved for publication March 2, 1994
1994

The Microcomputer Scientific Software Series 9: User's Guide to Geo-CLM: Geostatistical Interpolation of the Historical Climatic Record in the Lake States

Margaret R. Holdaway

INTRODUCTION

Geostatistics (an emerging discipline combining statistics, mathematics, and the environmental disciplines) is the study of spatially or temporally correlated data. It is based on the simple notion that data closer together in space (and/or time) are more alike or correlated than those farther apart. This spatial variability or structure can be modeled via a mathematical function known as the variogram. A sophisticated statistical method known as kriging uses the information in the variogram to spatially interpolate observations at unknown locations using data from known observation points.

We now recognize that spatial variation is an important component of environmental phenomena and that methods are available to use our knowledge of the spatial dynamics. Therefore, it seems appropriate to transfer this expertise to the interpolation of climatic data and to facilitate its use by researchers.

THEORY

The elementary concepts used in geostatistics and kriging are briefly discussed in Appendix A.

OVERVIEW

Geo-CLM has been written primarily for researchers who want to interpolate the historical climatological record at their research plots. They face a number of problems in applying kriging methods to the interpolation of climatic data. The usual kriging approach is to interpolate a grid of values, which is then used to produce a contour map of the area. Scientists with research plots have a different set of needs: namely to estimate climate at a number of irregularly spaced plot locations. Also, environmental data used in kriging are normally entered in a (X,Y,Z) column format necessitating creation of separate climatic files for each time interval (repeating the X,Y station locations on each file). Estimating the long-term climatic record at a plot location becomes very tedious.

Furthermore, meteorological station locations are usually provided in latitude/longitudes, requiring a conversion to Cartesian units before using. A program is needed which will enter the station location data in latitude/longitude coordinates once at an initial stage, convert them to the appropriate Cartesian units (X,Y) in English or metric units, and then use one large file of condensed climatic (Z) data—with each row representing a slice in time of climate values across all stations and each column containing a station's time sequence of climatic values. With such a condensed climate format, it would be easy to find and edit values, tracking a climatic value down the appropriate column or across the appropriate row.

Another problem is that some "kriging" packages may appear to use kriging methods but actually use only the most simple underlying variogram model—linear through the origin. Therefore, they do not have the precision expected from using the correct model and parameters estimated from the data. Also these packages ignore potential directional influences.

With a program that performs multiple kriging runs, the accuracy and precision of the method over long-time intervals can be summarized by month and/or meteorological station (when performing a cross validation). When monthly temperature is interpolated,

---

Margaret Holdaway is a Mathematical Statistician with the Mathematical Models of Forest Dynamics of the North Central Region work unit at the North Central Forest Experiment Station, St. Paul, Minnesota.

results can also be summarized by growing degree days (GDD). Although the emphasis of Geo-CLM is on performing multiple climatic runs using pre-assembled compact data entry, researchers are also given the ability to execute single runs on any environmental data using the traditional data entry format. Finally, the program structure needs to be easily transportable to other regions with the user providing the appropriate regional files, climate data files, and parameter files.

Primary usage of Geo-CLM is for researchers who want to estimate the long-term historic time series of climate at their research plot locations. To summarize, the benefits of Geo-CLM are that it:

- addresses common concerns of researchers working with climatic data, including estimation of the climate at plot locations,
- makes long-term Lake States climatic data bases of monthly temperature and precipitation available to researchers in a ready-to-krige format,
- makes repeated regional kriging runs without messy data preparation,
- enables condensed climatic input data by placing meteorological station location values on a separate regional file,
- provides for conversion from latitude/longitude to Cartesian units,
- provides for data output in English or metric units,
- provides for accurate specification of variogram model and parameters, increasing the program's flexibility relative to some generic programs,
- allows temperature interpolations to be summarized in growing degree days,
- summarizes multiple run statistics over any time interval on cross validation runs,
- provides for a single run in a more interactive test mode, and
- implements methods that are transferable to other regions.

THE PROGRAM

The Geo-CLM program was developed on a Macintosh microcomputer and requires about 300 KB of storage for the application and peripheral files. The regional precipitation and temperature data files needed for time series runs take an additional 1,680 KB of storage. The application requires 500 KB to run. An analogous DOS version is available on request.

The core kriging algorithm uses source code from Geo-EAS 1.1 (Englund and Sparks 1988). It performs two-dimensional kriging using ordinary point kriging. Ordinary kriging is the most commonly used kriging method for environmental data. It is quite robust but has the most lenient assumptions. Point kriging is the estimation of a point as opposed to block kriging, often used in soil sciences, which averages the estimation over an area. Graphical plots of results have not been provided. Various output files from Geo-CLM can be reformatted as necessary for use in any graphical or statistical package for further analysis. Primary usage of Geo-CLM is for runs on the existing Lake States climatic data. However, when environmental data are entered by the user, a limit of 500 observations can be kriged.

INSTALLATION

To install the Mac version of Geo-CLM on your hard drive, drag the Geo-CLM folder on the first distribution diskette to your hard drive. Drag the individual climatic data files from the remaining distribution diskettes into the same folder. When complete, the Geo-CLM folder should contain the application, regional data files, climatic data files, parameter files, and example files. All Geo-CLM files (except example files) must be in the same folder for the program to execute properly.

To install the DOS version on your system, first create a subdirectory (for example, Geo-CLM). All files on the distribution diskettes should be copied into the subdirectory.

USING THE PROGRAM

Start Geo-CLM by double clicking on the application in the Geo-CLM folder. The Geo-CLM banner is displayed followed by keyboard prompts. The first two prompts determine the dominant features of the run: two run options and two kriging options are available. They are briefly described below along with other main features of the program. Terminate the program by holding down the Command key and typing a period (.). But this does not hold true when the file dialog box is displayed. Then, **[Cancel]** will allow for termination of the program or for alternative non-file actions.

MAIN FEATURES

Run Options

Multiple Climatic Time Series Run.—The program can perform a multiple climatic time series run on any of four Lake States regions for yearly time intervals varying from as early as 1900 to as late as 1992 (actually May 1992). Most critical decision prompts are entered from the keyboard, but for multiple runs the program expects all data to be entered from files. Regional information and names of corresponding climatic data files and parameter files (for precipitation and temperature) are entered from the regional data file. Once keyboard prompts have been answered and the region specified, the run commences. Starting time; 1-, 5-, or 10-year messages indicating progress; and finishing time are printed to the screen. Kriged data and some statistics are written either to user-defined output files or to the screen. These files can then be used as input into graphical or statistics packages for additional analysis.

Single Interactive Run.—For a single interactive run only one kriging run is performed, either on default precipitation or temperature data entered from one of the regional data files or on user entered data. Prompts are from the keyboard. Parameter data may be entered from the keyboard or from data files (on regional runs). The remaining data files are named from the keyboard. Single runs are interactive and flexible in nature and are designed primarily for testing. The effects of changing models, model parameters, anisotropy, angular direction, search radius, or data configuration can be demonstrated through analysis of the observation weights (when requested) or estimation accuracy (in cross validation runs). Kriged data and some statistics are written either to the screen or to user-defined output files.

Kriging Options

Both multiple time series or single interactive runs can be used to perform a *cross validation* or to *interpolate* new locations entered as a list of irregularly spaced plot locations.

Cross validation.—Cross validation is a means of statistically validating the accuracy and/or precision of the kriging results. In cross validation the temperature or precipitation value at each weather station in the data set is treated as missing, and its value is estimated by kriging using data from neighboring weather stations and the variogram. The kriged values are then compared to the original station observations.

When a time series run is executed, a set of cross validation summary statistics can be written to an output file, averaging various estimation results over the chosen time interval. These statistics include mean kriging errors, broken down by weather station and/or month; mean squared error (MSE); and other useful statistics. Taken together these statistics provide a picture of the accuracy and weaknesses of the kriging technique in general and of the variogram model and parameters used in particular.

Interpolation.—Kriging can also be used to interpolate climate at research plot locations (unknown observations) falling within a Lake States' region. The multiple run option provides an output interpolation data file of the long-term historical record for monthly precipitation, temperature, or annual growing degree days (GDD) over the selected time interval. The single interactive run option produces a one time kriging interpolation of a list of research plot locations for regional default data or for user-provided data when in the test mode (called Option 0). These four main features are summarized in figure 1.

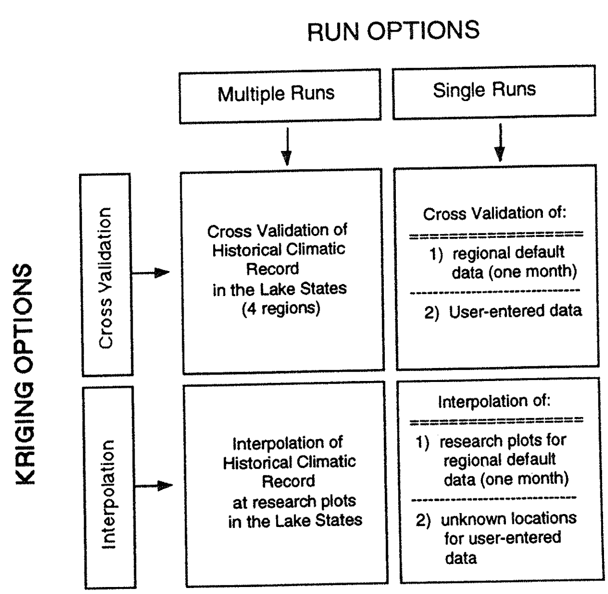


Figure 1.—*Main features of Geo-CLM.*

Climatic Elements

This program has been written to provide for the interpolation of monthly precipitation and temperature in four regions of the Lakes States. To execute a multiple time series run, a 93-year record of climatic data, starting in 1900, is provided in a condensed format. For single interactive runs, default regional precipitation and temperature data are entered automatically from the regional data file. Currently, all regions contain data for June 1980. These values can be altered on a regional file by the user.

Long-term monthly climatic records were obtained from the Historical Climatic Network of the National Climatic Data Center (NCDC) and augmented by secondary weather observer stations. Weather stations with moderately complete climatic records in the forested regions of Minnesota, Wisconsin, and Michigan were included. Monthly temperature and precipitation data were compiled from 1900 through May 1992. Details concerning the source and accuracy of the climatic data and editing procedures are given in Appendix B.

The climatic data files provide both spatial and temporal delineation of monthly precipitation and temperature. When requested, the program will also use the monthly temperature estimates to derive annual growing degree days at plot locations.

Climate Regions

The kriging algorithm is complex and extremely computationally intensive. Kriging may be impractical for more than 100 meteorological stations, especially with multiple time runs. One solution is to subdivide larger regions into smaller homogeneous regions. The Lake States has been subdivided into four smaller regions, which are kriged separately. Most regions overlap so that there are no obvious seams. For kriging locations very close to the seam, averaging the results of two runs, one from each region, may be desirable. Because of the Great Lakes influence on the topography of the region and the presence of the Mississippi and St. Croix Rivers, it is fairly easy to divide regions along State lines. In Wisconsin and Minnesota only the forested areas[1] were initially covered. However, a region in southeastern Wisconsin was later added. In Minnesota the prairie/farm areas were not included, except to incorporate an additional envelope of weather stations along the forest transition zone. The four final regions were Minnesota, Wisconsin, Michigan, and SE Wisconsin (fig. 2). Distinct symbols are used in the figure so that overlap stations are evident.

[1] *Weather stations located in counties designated as 16 percent or more forested as specified in the Forest Service Atlas of the Midwest (Merz 1979) were designated as "forested."*

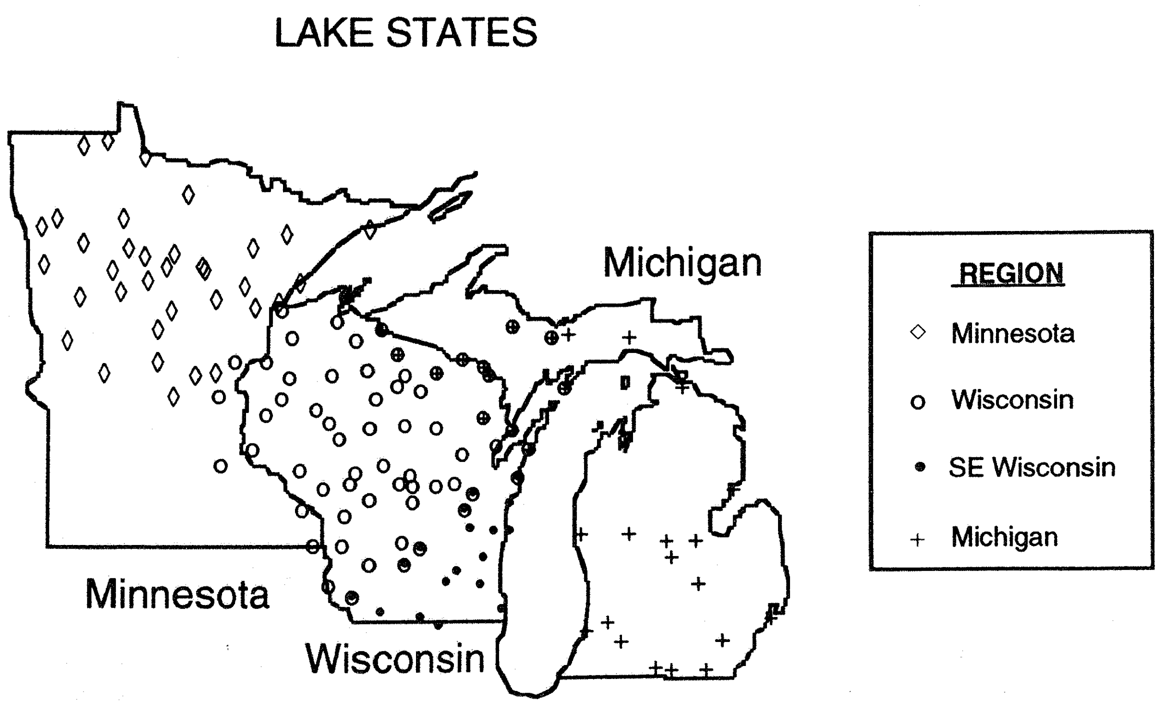


Figure 2.—*Location of meteorological stations in the Lake States by region.*

Generally, stations with 60 percent completeness in their records from 1900 through 1993 for both precipitation and temperature were selected. Many stations did not start recording observations until the 1930's. Inclusion of these stations provided for denser coverage in the latter part of the century. At times the allowable percent missing was increased to provide more coverage in sparsely forested areas or border areas.

This regional approach is general enough that users can transfer it to their own "regions" and climatic/environmental data by replacing the regional data file, climatic data files, and parameter files with their own data. Format and data conventions would still have to be honored. If attempted, the source code should be requested to alter regional prompts and measurement conversions.

Latitude/Longitude Conversion

A rectangular, or X-Y coordinate system is used to locate points on a flat, two-dimensional surface. In a rectangular coordinate system, an origin and two baselines (the x- and y-axes) are defined. In contrast, the latitude-longitude system is used to locate points on the Earth, reflecting the fact that the Earth is a sphere rather than a flat surface. This program uses a standard trigonometric formula to convert spherical coordinates to X-Y coordinates. To express latitude-longitude coordinates as X-Y coordinates, it is necessary to determine an origin, which then defines the two baselines (or lines of latitude and longitude). This origin should be south and west of all possible observation points to yield positive X,Y values falling in the first quadrant. This origin will be called the "SW corner point." A local SW corner point has been entered as the default value on each regional data file. At times, users may wish to use one common global SW corner point for all four regions. Such a value for the Lake States would be 41° N, 97° W. When using this value, X-Y coordinates for all four regions will be expressed in miles from this point and hence can be plotted and analyzed together.

Longitude and latitude are usually measured in degrees, minutes, and seconds, but they are represented by decimal degrees in computer calculations. This program offers three choices for entering this geographic location information—all of which are ultimately converted into X-Y coordinates. This location code is entered as a format-free numeric on the second record of each test data file or interpolation data file. The choices for entering geographic location are:

1 = latitude-longitude coordinates (in degrees and minutes),
2 = latitude-longitude coordinates (in decimal degrees), or
3 = Cartesian coordinates (X,Y).

If locations are entered as degrees and minutes, two fields (separated by a comma or space) are used in place of the one field used to enter decimal degrees or Cartesian data. When the geographic coordinates include seconds, they can be incorporated into the minutes value as a decimal, or into the decimal degrees value.

Certain rules apply to latitude-longitude coordinates. The longitude is the X-coordinate, and the latitude is the Y-coordinate. Latitude-longitude coordinates are read and written (longitude, latitude) so that they correspond to (X,Y). Within the program and as output, latitude and longitude values are always used and printed as decimal degrees. Normally, geographic locations in the United States have longitudes as negative and latitudes as positive. However, for simplicity, this program uses positive values for both.

DATA REQUIREMENTS

Multiple Time Series Run

Cross Validation

The input data files required for a multiple time series run are: *regional data file, climate data file,* and *parameter file.* Figure 3 clarifies how these components connect in a multiple time series run. Solid boxes depict input and output for a cross validation run. The geographic location information for meteorological stations is retrieved from a regional data file implicitly named from the regional keyboard prompt:

```
Specify Climatic Data Region:
    1. Minnesota
    2. Wisconsin
    3. Michigan
    4. Southeast Wisconsin
 Enter 1, 2, 3 or 4: 4
```

MULTIPLE TIME SERIES RUN

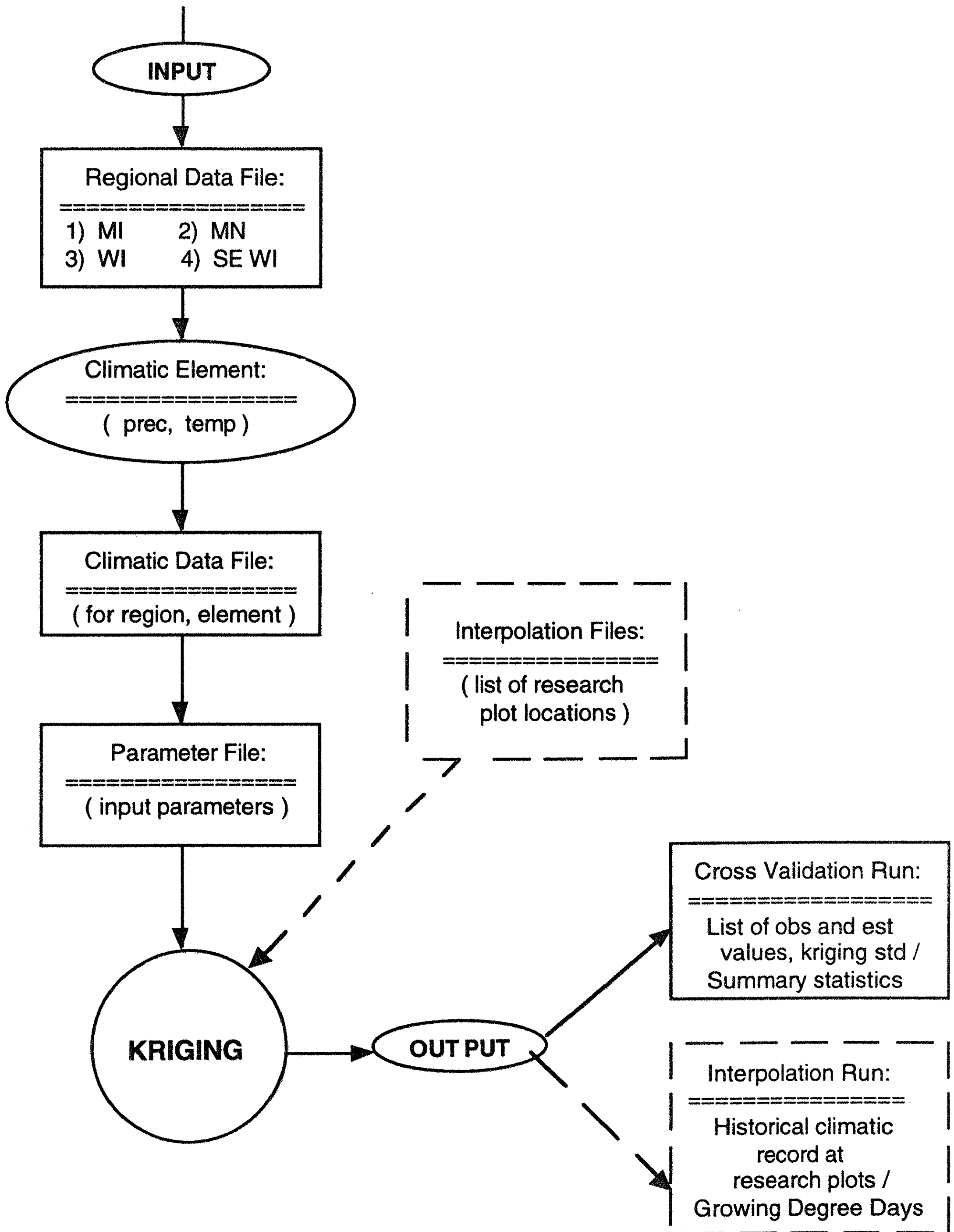


Figure 3.—*Flow chart of the primary aspects of a multiple time series run.*

The parameter files and climatic data files are identified from the regional data file and must be located in the Geo-CLM folder. All input data files are ASCII text files. All other input is interactive, using a question and answer format.

Regional data file.—The first action in a multiple run is to read a format-free regional data file in response to the identify region prompt. The regional data file contains general regional information followed by a listing of meteorological station information. The regional information includes: region name, names of parameter and climatic data files for both precipitation and temperature, local SW corner point, minimum and maximum longitudes and latitudes (in degrees and minutes), maximum time range of the climatic record and number of meteorological stations. The station information consists of a list of sequential numbers, longitude and latitude (in degrees and minutes)[2], default precipitation and temperature values, and station names in alphabetical order. Most of these data, along with the converted X,Y values in miles, are echoed out to the screen once the regional prompt is answered. For example, data printed to the screen for the SE Wisconsin region are:

[2]*Note that the geographic location code is 1 for all data entered from regional data files (for both multiple time series and single interactive runs).*

```
                    **** REGIONAL INFORMATION ****

Region: SE Wisconsin
SW Corner Point in long, lat   ( 91.0, 42.0)
Defined Boundaries  (min,max)
   Climatic Data Points
      Longitude  ( 87.0, 91.0)
      Latitude   ( 42.0, 45.0)
   Years   ( 1900, 1992)
Number of stations  (  18)
```

Number	Name	Long	Lat	Miles X	Miles Y	Default Z Prec(in)	Default Z Temp(F)
1	APPLETON	88.38	44.25	134.24	155.32	4.99	64.70
2	BELOIT	89.03	42.50	100.90	34.56	6.01	67.80
3	BRODHEAD	89.38	42.62	82.96	42.62	5.60	66.00
4	DARLINGTON	90.12	42.68	45.35	47.19	5.76	66.00
5	FOND DU LAC	88.45	43.80	130.82	124.26	5.13	64.30
6	KEWAUNEE	87.52	44.45	178.69	169.13	4.51	59.00
7	LAKE MILLS	88.92	43.07	106.89	73.65	3.87	66.00
8	LANCASTER	90.70	42.87	15.51	59.84	5.03	67.40
9	MANITOWOC	87.68	44.10	170.14	144.96	5.02	62.80
10	OSHKOSH	88.55	44.05	125.69	141.52	5.00	66.00
11	PLYMOUTH	87.98	43.75	154.76	120.82	7.17	64.70
12	PORTAGE	89.43	43.53	80.39	105.85	4.18	65.50
13	PRAIRIE DU SAC	89.73	43.32	65.00	90.90	2.80	68.00
14	RACINE	87.85	42.72	161.59	49.51	4.18	61.60
15	SHEBOYGAN	87.72	43.75	168.43	120.82	5.18	62.90
16	WATERTOWN	88.72	43.20	117.15	82.85	3.27	66.90
17	WAUKESHA	88.23	43.02	141.93	70.20	3.62	65.00
18	WEST BEND	88.18	43.40	144.50	96.66	3.32	63.20

Users may want to change some of the variables in the regional information section, such as the SW corner point, regional boundary points, or even the number of stations. For example, to enlarge the rectangular area of permissible interpolation locations, simply change the appropriate boundary value(s). Boundary points are further explained in the Secondary Features Section. Only in rare cases should the number of stations be altered. It can be decreased in value to remove the overlap stations. For example, in Wisconsin the last 11 stations are overlap stations in the Upper Peninsula of Michigan and an envelope of stations along eastern Minnesota. By altering the number of stations, meteorological stations in either Minnesota or Minnesota and Michigan can be removed from consideration, thus effectively decreasing the size of the data base and run time. The last seven stations in Michigan are Wisconsin overlap stations. For easy reference, all four regional data files are listed in Appendix C.

Climate data file.—Separate precipitation and temperature data files for each region (in English units) are available and accessed through the regional data file. Data are in a condensed format, and each data record contains one month's climate values for the region. Each file contains two header lines. Above each column of data is a consecutive number in the first line and an abbreviated station identifier in the second line. The climatic records that follow are in chronological temporal order and start with a time identification portion followed by a data portion. The ID portion identifies the year and month. The data portion contains meteorological observations for each station in the order listed on the corresponding regional data file. All data are written in integer format. The original precipitation and temperature observations were recorded in hundredths of inches and in tenths of degrees Fahrenheit, respectively. When the user specifies the climatic element type:

```
Specify Climatic Element:
   1. Monthly Precipitation
   2. Monthly Temperature
Enter 1 or 2: 2
```

the appropriate decimal conversion is made within the program. Missing values on the climatic data file are recorded as -999. The first 12 lines of SE Wisconsin's precipitation file are:

```
             1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
Year Mo App Bel Bro Dar FDL Kew LaM Lan Man Osh Ply Por PDS Rac She Wat Wau WeB
1900  1-999-999 145-999  60-999-999  91  88  55-999  75-999 195-999 152 143-999
1900  2-999-999 321-999 169-999-999 114 189 105-999 140-999 271-999 260 187-999
1900  3-999-999 230-999  94-999-999 198  98 105-999 113-999 120-999 179  91-999
1900  4-999-999 291-999 209-999-999 363 270 245-999 209-999  83-999 276 234-999
1900  5-999-999 108-999 119-999-999 272 112 230-999  86-999 175-999 199 121-999
1900  6-999-999 266-999 125-999-999 153 243 139-999 215-999 221-999 161 121-999
1900  7-999-999 714-999 563-999-999 942 459 944-999 828-999 666-999 508 703-999
1900  8-999-999 664-999 549-999-999 401 312 445-999 405-999 565-999 282 567-999
1900  9-999-999 240-999 306-999-999 477 327 151-999 290-999 182-999 274 202-999
1900 10-999-999 353-999 407-999-999 356 484 573-999 470-999 187-999 455 234-999
1900 11-999-999 189-999  94-999-999 181 246 120-999 195-999 419-999 230 186-999
1900 12-999-999  73-999  50-999-999  33 106  64-999  45-999  34-999  73  45-999
```

Should you desire to add stations to a region to provide denser coverage in an area of interest, station information (location and name) can be added onto the regional data file, the number of stations increased accordingly, and the appropriate columns of climatic values added to the climatic data files (for the precipitation and temperature files) with missing values entered as -999. Only 100 stations per line are allowed on the climatic data file. Any more than that must be added to a new line using the same column alignment. Normally, no more than 100 stations are recommended.

Parameter file.—The parameter file is an ASCII format-free text file (in English units) used to store a set of kriging input parameters by month. Separate precipitation and temperature parameter files for each region are provided. Each file contains necessary variogram model and parameter information, along with search information. The file contains 12 lines, one for each month. The form of each input line is: month, model, nugget, sill-nugget, maximum range, minimum range, angle, search radius, and maximum points. All are entered as numerical fields. "Model" is a numerical number (Spherical = 1, Gaussian = 2, Exponential = 3, Linear = 4); "nugget" is the nugget effect and must be greater or equal to zero; "sill-nugget" is the distance between the nugget effect and the sill and must be greater than zero; and "range" is the range of the spatial influence (a positive value). For a more detailed explanation of these parameters, see Appendix A.

In the case of *isotropy* (i.e., no directional influence), the "maximum" and "minimum range" are equal and represent, for the spherical model, the distance at which the curve reaches its sill. For the exponential and gaussian model, the effective range is used (i.e., the range at which the model asymptotically approaches the sill). For the linear model, the range and sill, taken together, define the slope. With isotropy, the angle is 0 degrees.

When a directional influence is present, Geo-EAS provides for *geometric anisotropy*, where the range changes with direction, while the sill remains constant. The "maximum range" is the longest range of the variogram, and the "minimum range" is the shortest range. The "angle" parameter is the direction of longest range. This direction is in degrees; 0 degrees is due east, increasing in a counterclockwise direction up to, but not including, 180 degrees. At a monthly temporal resolution, the climatic data show enough of an averaging effect of prevailing wind directions and weather patterns that strong anisotropy is uncommon. Weak anisotropy is handled very well by the isotropic model.

The "search radius" controls the neighborhood search used in kriging. Only data values within a circular search area are used. When entered from the parameter file, the search radius is in miles. For the station densities found in the Lake States, observation points beyond 80 miles usually have only a minimal influence in kriging. However, when weather station data are very irregularly spaced as in the northern forests, observations as far as 125 to 130 miles away may contribute needed information. Still, increasing the search radius also increases the kriging run time. Users can change the search radius on the parameter file to suit their requirements. Finally, "maximum points" is the maximum number of points used within the search radius. The maximum number of points allowable is 64. In a multiple time series run, the monthly values from the parameter file are echoed to the screen (fig. 4). Further information on choice of search radius and maximum points is given in the Search Radius Section of Secondary Features.

Output file.—On a cross validation run, the estimation output data are written to an output file named by the user. A run header is written to all output files before the actual data are printed. The output data from a cross validation time series run consist of year, month, station number, observed value, kriged estimate, error (estimate minus observed), kriging standard deviation, and number of neighbors found—displayed in columns. They are listed in the same time sequence as the initial climatic data, and within each time step, ordered by weather stations with non-missing data. See the example of an estimation output data file (fig. 5). If the file name prompt is [**Cancel**]ed, the output file is printed to the screen (without the header information). This is not recommended unless a very short time interval is used. (See Time Intervals under Secondary Features.)

Models: 1 - Spherical
2 - Gaussian
3 - Exponential
4 - Linear to range

Parameters used:
(English units)

Month	Model	Nugget	Sill-Nug	Max Range	Min Range	Angle	Radius	Max Pts
1	1	.060	.150	300.	300.	0.	150.	16
2	1	.040	.180	300.	300.	0.	150.	16
3	1	.100	.500	375.	375.	0.	150.	16
4	1	.100	1.000	285.	285.	0.	150.	16
5	3	.200	1.500	300.	300.	0.	150.	16
6	3	.500	2.800	420.	420.	0.	150.	16
7	1	.700	1.900	150.	150.	0.	150.	16
8	1	.700	1.900	150.	150.	0.	150.	16
9	1	.300	3.000	400.	400.	0.	150.	16
10	1	.100	.900	250.	250.	0.	150.	16
11	1	.060	.500	275.	275.	0.	150.	16
12	1	.060	.190	300.	300.	0.	150.	16

Figure 4.—*Echo of precipitation parameter file for Minnesota.*

Statistics output file.—If the user requests statistical output on a cross validation time series run, a statistics output file is also printed. Should the user **[Cancel]** the request to name a statistics output file, the statistical output will be printed to the screen. Then, appropriate sections can be saved by copying them from the screen before the run is terminated. The statistical output, pertinent concerns, and example output are discussed in detail in the Statistics Section of Secondary Features.

Interpolation.—The interpolation kriging option uses the same input data files as the cross validation option (see previous section for details) with the addition of an interpolation data file containing the geographic locations of research plots to be estimated. This situation is depicted in figure 3 by the dashed boxes for the additional input and alternative output.

```
DATE: 04/27/1993            TIME: 20:01:06
RUN TITLE         : Demonstration run on SE WI starting in 1900
KRIGING OPTION    : Cross Validation Run
REGION            : SE Wisconsin
CLIMATIC ELEMENT  : Precipitation
UNITS             : English
TIME INTERVAL     :   1900  -  1993
------------------------------------------------
  File Usage
     Input Files
        Regional Data File  : se_reg dat
        Climatic Data File  : SE_P
        Parameter File      : PAR Prec SE
     Output Files
        Estimation File     : Cross SE WI 1900-1993.out
------------------------------------------------

CROSS VALIDATION OUTPUT FILE:
```

Year	Mon	Stn	Observed	Estimated	Difference	K_Std_Dev	Nbrs
1900	1	3	1.450	1.248	-.202	.349	8
1900	1	5	.600	.962	.362	.307	7
1900	1	8	.910	1.157	.247	.412	7
1900	1	9	.880	.939	.059	.365	7
1900	1	10	.550	.863	.313	.322	7
1900	1	12	.750	1.031	.281	.335	8
1900	1	14	1.950	1.293	-.657	.353	7
1900	1	16	1.520	1.141	-.379	.309	8
1900	1	17	1.430	1.424	-.006	.313	7
1900	2	3	3.210	1.875	-1.335	.327	8
1900	2	5	1.690	1.622	-.068	.272	7
1900	2	8	1.140	2.348	1.208	.401	7
1900	2	9	1.890	1.605	-.285	.346	7
1900	2	10	1.050	1.756	.706	.289	7
1900	2	12	1.400	1.877	.477	.311	8
1900	2	14	2.710	2.225	-.485	.328	7
1900	2	16	2.600	1.985	-.615	.278	8
1900	2	17	1.870	2.500	.630	.279	7

Figure 5.—*Output estimation file for multiple time series cross validation run on SE Wisconsin data starting in 1900. Run header plus first 2 months printed.*

```
Interpolation data file for SE WIS
2 = dec degrees / dec long, dec lat, plot ID, comments
88.55,  42.51      1A   Near Lake Geneva
88.01,  43.74      2C   West of Shebogan
88.59,  43.68      8D   South tip of Lake Oshkosh
89.60   42.52     10B   On Illnois border near Monroe
```

Figure 6.—*Interpolation data file for four research plots in SE Wisconsin, entered in decimal degress (code = 2).*

Interpolation data file.—Data on the file are entered format-free; the first line contains an 80-column identifier. The second line contains the geographic location code, which can represent either latitude-longitude coordinates or Cartesian coordinates (in miles or kilometers). If plot location data are entered in miles, for example, they must correspond to the converted mile values calculated from the local SW corner point for the region used. The remaining rows contain format-free research plot locations—four fields for degrees and minutes observations; two fields for the rest. (There are no environmental observations on this file.) Any plot identifier or comments can be entered on the remainder of each row after a space or comma. Only 100 research plots can be interpolated at a time. If more plots are required, additional interpolation input files and runs will be necessary. Figure 6 shows an interpolation data file for the SE Wisconsin region with location values expressed in decimal degrees.

Output file.—The interpolation output file contains only the kriging estimates (there are no observed values and hence no errors). Following the run header, estimated values are listed in columns corresponding to the order of the research plots entered. The first row contains consecutive numbers; the second and third rows contain longitude and latitude (in decimal degrees) or X,Y values of the points in the order listed on the interpolation data file. The remaining rows contain year, month, and columns of estimated climate for each location, for up to 100 research plots. Figure 7 illustrates the first year's interpolation output

```
DATE: 04/27/1993          TIME: 22:37:41
RUN TITLE          : Demonstration run
2ND TITLE          : Interpolation data file for SE WIS
KRIGING OPTION     : Interpolation Run
REGION             : SE Wisconsin
CLIMATIC ELEMENT   : Temperature
UNITS              : English
TIME INTERVAL      :   1960  -  1960
------------------------------------------------
  File Usage
     Input Files
        Regional Data File   : se_reg dat
        Climatic Data File   : SE_T
        Parameter File       : PAR Temp SE
        Interpolation File   : SE (2).int
     Output Files
        Interpolation File   : Int SE WI 60-89.out
        GDD File             : GDD SE WI 60-89.out
------------------------------------------------

INTERPOLATION OUTPUT FILE:

             Consecutive point number
                 1      2      3      4
Longitude    88.55  88.01  88.59  89.60
Latitude     42.51  43.74  43.68  42.52
Year  Month
1960    1     23.9   23.5   23.0   22.7
1960    2     23.2   23.0   22.5   22.2
1960    3     22.5   22.9   22.2   21.5
1960    4     50.6   46.5   48.5   50.2
1960    5     56.4   53.5   56.0   56.3
1960    6     65.7   63.0   64.6   65.0
1960    7     70.7   68.4   69.6   69.8
1960    8     71.9   69.5   70.3   71.2
1960    9     66.6   63.5   63.8   65.7
1960   10     52.0   50.1   50.1   50.8
1960   11     40.3   39.0   38.9   39.7
1960   12     23.1   21.7   21.5   22.5
```

Figure 7.—*Interpolation output file for multiple time series temperature run on SE Wisconsin data for 1960.*

```
DATE: 04/27/1993          TIME: 23:55:14
RUN TITLE          : Demonstration run
2ND TITLE          : Interpolation data file for SE WIS
KRIGING OPTION     : Interpolation Run
REGION             : SE Wisconsin
CLIMATIC ELEMENT   : Growing Degree Days
UNITS              : English
TIME INTERVAL      :   1960  -  1969
-----------------------------------------------
  File Usage
     Input Files
        Regional Data File    : se_reg dat
        Climatic Data File    : SE_T
        Parameter File        : PAR Temp SE
        Interpolation File    : SE (2).int
     Output Files
        Interpolation File    : Int SE WI 60-69.out
        GDD File              : GDD SE WI 60-69.out
-----------------------------------------------

GROWING DEGREE DAYS OUTPUT FILE:
  Threshold value used =    35.0000
  Accumulated through month    7

             Consecutive point number
                 1      2      3      4
Longitude    88.55  88.01  88.59  89.60
Latitude     42.51  43.74  43.68  42.52
Year
1960          3156.  2797.  3019.  3096.
1961          3096.  2754.  2895.  2991.
1962          3332.  2907.  3095.  3219.
1963          3424.  2988.  3204.  3372.
1964          3520.  3173.  3392.  3487.
1965          3229.  2802.  3004.  3240.
1966          3147.  2853.  2990.  3153.
1967          3084.  2754.  2867.  3106.
1968          3397.  3032.  3166.  3377.
1969          3137.  2767.  2921.  3157.
```

Figure 8.—*Growing degree days (GDD) output file from the same run.*

starting in 1960, corresponding to the previous interpolation data file in figure 6. Again, **[Cancel]** file name prompt will print the interpolation output (minus header) to the screen.

GDD output file.—When growing degree days are also requested on a temperature run, an additional output file of annual growing degree days is printed to the specified file. Figure 8 lists the first few lines of such a GDD output file for the same interpolation data set used in figure 6. The general format is the same as above, but the monthly output is condensed to annual values. Growing degree days are discussed in more detail in the Secondary Features Section.

Single Interactive Run

Cross Validation

A single interactive run incorporates two main alternatives: the first uses climatic default values from one of the four regions and the second allows for any user-entered data. The general outline for a regional default run is similar to that used for a multiple time series run. A second feature has been added that allows for any data to be kriged (fig. 9). Input of data for a single run is specified in response to the regional prompt (which now includes an Option 0 test mode):

```
Enter 0, 1, 2, 3, or 4: 4
Enter your own climatic data or
  use regional default values?
    0. Test mode—enter your own
       X,Y,Z (or Long,Lat,Z) values
    1. Minnesota
    2. Wisconsin
    3. Michigan
    4. Southeast Wisconsin
```

Regional data file.—Responding with a region (1 - 4) provides for a single run similar to multiple regional runs, except that regional default values for precipitation or temperature are used. Current default values (for June 1980) can be changed on the regional data file. The condensed climatic data files are not used in single runs.

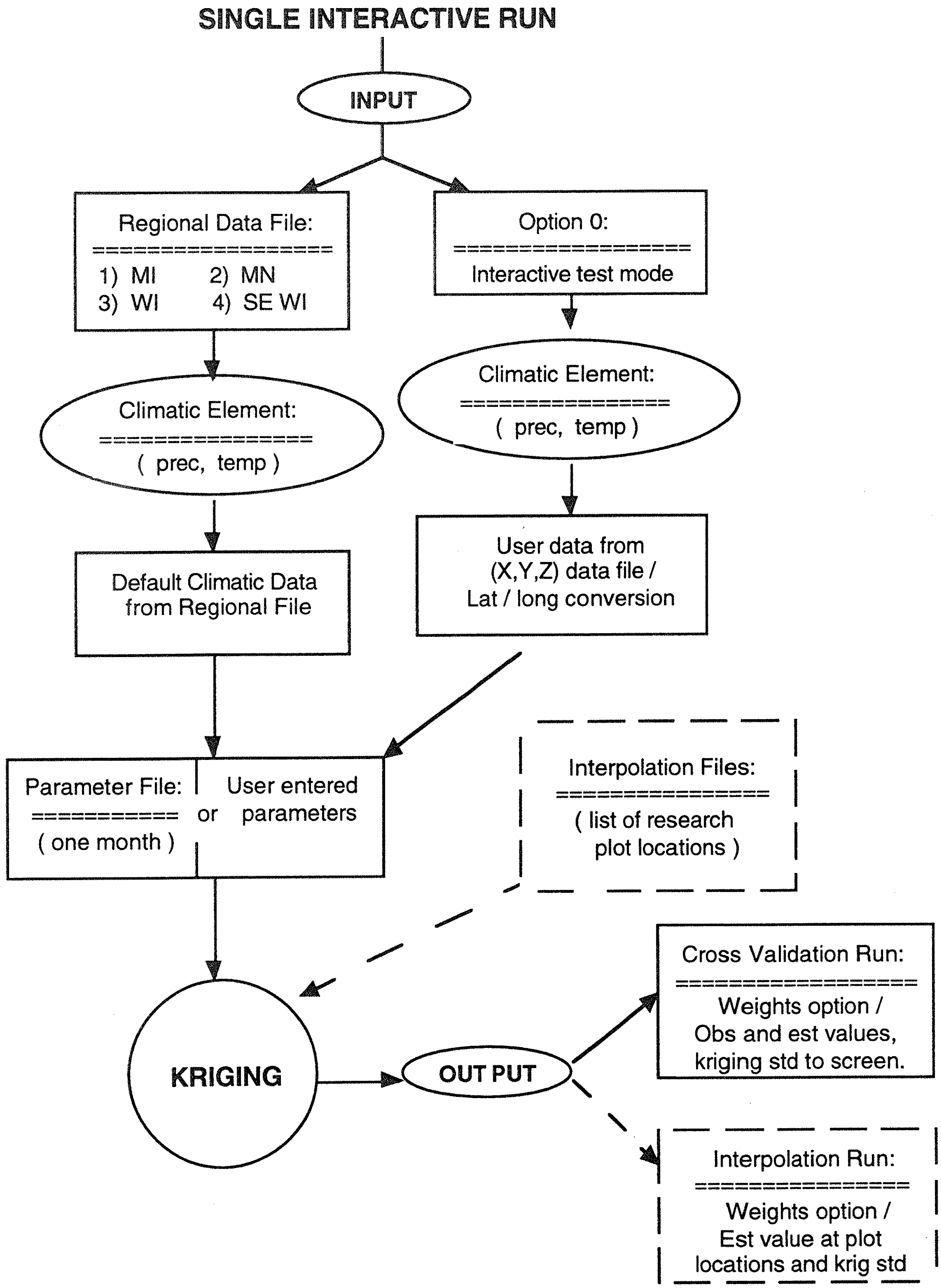


Figure 9.—*Flow chart of the primary aspects of a single interactive run.*

Test data file.—Responding with Option 0 (i.e., no region specified) to the regional prompt allows for entry of any environmental[3] data in the more traditional column format. Option 0 enables the program to mimic the usual kriging run but with the *additional* capability of entering the geographic locations in degrees and minutes (code 1), decimal degrees (code 2), or in Cartesian units (code 3).

Option 0 uses a previously generated data file. Obviously, entry of individualized geographic locations is necessary (information entered from the regional data file for all other run options). When latitude-longitude coordinates are used, the geographic location need not be related to any of the four Lakes States regions and consequently the SW corner point may need to be changed. A keyboard prompt furnishes the current SW corner point (the last values used from a regional or Option 0 run) and asks if new values are needed. If your geographic area is different, you must enter a new SW corner point:

```
Current values of SW corner point are ( LONG,
  LAT ) = (  97.00,   41.00)
Do you want to change these values? (Y/N): y

Enter new (LONG,LAT) values as decimal degrees:
  94, 44.3
New corner point values are (LONG,LAT):
  ( 94.0000,  44.3000)
```

Test data files for Option 0 can be created with any text editor using the traditional column format. The file starts with three header records. The first line is an 80-character descriptive title for the file; the second line contains the numeric value "1," "2," or "3" to identify the geographic location code; and the third line contains the name of the environmental variable. The remainder of the input data file holds the actual data values. The layout of the data is one line per location, and each data line contains the coordinate information followed by the environmental value. The data are entered in three or five fields corresponding to the code on the second record. The data must be in the following order: longitude, latitude, environmental value (or X-coordinate, Y-coordinate, environmental value). Longitude and latitude can be either expressed in degrees and minutes (code 1) or in decimal degrees (code 2). X and Y are expressed in any Cartesian units (code 3). For geographic locations expressed in latitude-longitude coordinates, the converted X and Y coordinates are in miles or kilometers depending on the measurements option. Each of the three data values (five in the case of degrees and minutes data) is entered in free-format and must be separated by at least one space or a single comma. As dimensioned, Geo-CLM will read and work with up to 500 data observations.

A further modification enables the user to add up to 10 environmental variables after the X,Y data. Names of environmental variables are entered on the third line in free-format. Either separate variable names by a space(s) or enclose names in quotes.

Figure 10 presents an example of a test data file entering geographic location in degrees and minutes of longitude and latitude and using an alternative climatic variable. Because the latitude-longitude coordinates are for a new region, the SW corner point is changed to 94° W, 44.3° N before entering the test data file (see screen prompts of previous example). English units are used.

After the data file has been read, the program echoes the first five observations to the screen with the appropriate converted X,Y values in the case of spherical coordinates. Cartesian X,Y data entered on the single interactive run can be in any geographic location units and environmental units. For more complex data sets and analysis, we suggest that the user obtain Geo-EAS[4].

Parameter data.—With a single interactive run using the default regional option, users

[3]*Obviously this includes climatic data, but can be used for any spatially distributed earth science variables.*

[4]*Government agencies and academic or research institutions can obtain a copy of Geo-EAS executable code with User's Guide at no charge by sending three preformatted high-density diskettes (5-1/4 inch or 3-1/2 inch) to:*

Evan J. Englund (Geo-EAS)
USEPA EMSL-LV, EAD
P.O. Box 93478
Las Vegas, NV 89193-3478
FAX: (702) 798-2454

ERRATA!!

Footnote 4 on page 14 contains wrong information for requesting copies of Geo-EAS. It is commercially available for MS_DOS machines from:

ACOGS
P.O. Box 44247
Tucson, AZ 85733-4247

```
Location of 25 prec obs stns in the Mpls/St.Paul metro area
1 = deg min / Z = Percent of days with prec.GE.0.10 inch May-Sept, 1964-1968
'Percent days P > 0.10 inch'
93 27    45 12    20.1    Anoka
93 19    44 49    22.7    Bloomington
93 34    44 54    23.8    Excelsior
93 11    44 38    25.1    Farmington
92 55    45 16    23.9    Forest Lake
93 03    45 04    21.4    Gem Lake
93 23    44 59    20.1    Golden Valley
92 52    44 46    22.6    Hasting
93 37    44 39    20.5    Jordan
93 10    45 08    21.8    Lino Lakes
93 14    44 59    23.4    Lower St.Anthony Falls
93 39    45 00    25.8    Maple Plain
93 13    44 53    22.6    Minneapolis - St. Paul Airport
93 13    45 05    21.3    Mounds  View
93 18    44 34    22.7    New Market
93 34    44 32    23.1    New Prague
93 09    44 28    25.2    Northfield
93 44    45 05    24.3    Rockford
93 18    44 56    25.6    St. Louis Park
93 04    44 59    23.0    St. Paul (1)
93 10    44 55    22.4    St. Paul (2)
93 02    44 52    26.0    So. St. Paul (1)
93 03    44 50    22.9    So. St. Paul (2)
92 47    45 02    23.0    Stillwater
93 30    44 59    26.8    Wayzata
```

Figure 10.—*Test data file of locations of 25 precipitation observing stations in the metropolitan area of Minneapolis/St. Paul, MN. The environmental variable is percent frequency that days with 0.10 inch or greater precipitation were observed from May through September, 1964 to 1968 (Baker and Kuehnast 1973).*

are given the choice of using one month's values from any generated parameter file or entering their own parameter values from screen prompts. If **[M]**etric units were previously specified, the appropriate values are converted to metric units before printing to the screen. For parameter file entry:

```
Use Parameter File? (Y/N): y

    Select a Parameter File.
Parameter File Name = PAR Prec SE
Which month's parameters?  (Enter
1,2...12): 7
```

If the user **[Cancel]**s the request for a parameter file name, the program allows for keyboard entry instead.

For Option 0 test data, parameter values must be entered from the keyboard. Variogram parameter values must be calculated in the same units as the test data. (This program does not provide for estimation of the variogram. Geo-EAS is one of many available programs that do provide it.) Example of keyboard entry of parameter values[5] for an isotropic model:

```
Use Parameter File? (Y/N): n

Enter MODEL
 (Sph=1, Gaus=2, Exp=3, Lin=4): 1
      Enter C0 (nugget effect): 1
              C1 (sill-nugget): 5
                   Major range: 50
                   Minor range: 50
                         Angle: 0
                 Search radius: 40
      Maximum number of points: 7
```

Whether parameter values are entered via a parameter file or from the keyboard, they are echoed to the screen in the form shown on the next page:

[5]*Models and parameters are discussed more fully in the Parameter file Section of Multiple Time Series Runs and Appendix A.*

```
PARAMETER OPTIONS

Run Name : Location of 25 prec obs stns in
           the Mpls/St.Paul metro area
Units:   English
Model:   Spherical
Nugget: 1.00          Z Label: Percent days
                               P > 0.10 inch
Sill-Nugget:   5.00   Search Radius:   40.00
Major Range:  50.00   Maximum Points:     7
Minor Range:  50.00
Angle:          .00
```

Permitting easy entry of one month's parameter values from an existing parameter file on regional runs or employing user-entered values allows great flexibility in analyzing the effects of various changes in parameter values.

Output data.—For single runs, the same basic output estimation data are saved to a user specified file or written to the screen (fig. 11). If the user wants to save the screen output from a run, go up to File: Save as ___ and specify a screen output name.

Interpolation

Again, the interpolation option on a single run uses the previously generated environmental data, along with an interpolation data file of research plot locations to be estimated. Dashed boxes (fig. 9) show the additional input necessary and the alternative output provided. Either a default regional run or an Option 0 run can be made.

Interpolation data file.—When either a default regional run or Option 0 run is performed, the form of the interpolation data file is identical to that discussed in the Multiple Time Series Section (see figure 6). When the regional default option is used, location code usage is the same as when making a multiple run. Plot location can be entered in all three codes, but if entered in miles or kilometers (code 3), they are measured from the regional SW corner point. Be careful to enter code 3 values accurately because no boundary check is made when Cartesian values are used.

When an Option 0 interpolation run is executed, data entry conventions are not quite as flexible. These runs require both a test data file (of locations with measured environmental data) and an interpolation data file (of research plot locations with unknown observations). Certain restrictions apply when matching location codes on the test data file to those on the corresponding interpolation data file. The interpolation file must follow the same coordinate type (i.e., lat/long or Cartesian) as the test data file with one exception—you can enter locations in lat/long coordinates on the test data file and in miles (or kilometers) on the interpolation data file. At times this may be a convenient approach.

To demonstrate, consider the following example. The metropolitan test data, used previously (fig. 10), had lat/long coordinates entered in minutes and degrees (code 1), but

Stn	X	Y	Z	Est	Diff	K_SD	Nbrs
1	134.239	155.318	4.990	5.028	.038	1.072	16
2	100.904	34.559	6.010	4.826	-1.184	1.120	16
3	82.962	42.617	5.600	5.134	-.466	1.070	16
4	45.352	47.185	5.760	4.877	-.883	1.153	16
5	130.819	124.261	5.130	4.878	-.252	1.006	16
6	178.686	169.129	4.510	5.006	.496	1.247	13
7	106.890	73.651	3.870	4.047	.177	1.010	16
8	15.509	59.843	5.030	4.980	-.050	1.323	12
9	170.141	144.964	5.020	5.203	.183	1.047	15
10	125.693	141.520	5.000	4.957	-.043	1.012	16
11	154.755	120.818	7.170	4.700	-2.470	.973	16
12	80.394	105.852	4.180	3.753	-.427	1.106	16
13	65.003	90.905	2.800	4.646	1.846	1.102	16
14	161.589	49.509	4.180	4.438	.258	1.253	16
15	168.428	120.818	5.180	5.473	.293	1.025	16
16	117.146	82.850	3.270	4.037	.767	.996	16
17	141.931	70.196	3.620	4.101	.481	1.047	16
18	144.496	96.657	3.320	4.747	1.427	1.029	16

Figure 11.—*Cross validation estimation results for SE Wisconsin, using default precipitation data (June 1980).*

```
Location of recreational/research areas in the Mpls/St.Paul metro area
3 = Miles / Estimate percent days with prec.GE.0.10 inch May-Sept, 1964-1968
44    48       St. Paul Agricultural Research Plots.  /  Miles from (94, 44.3)
49    37       Pigs Eye
25    32       Shakopee
35    59       Coon Rapids City Park
49    60       White Bear Yacht Harbor
46    53,      Rice Creek Regional Park
38    35.5     Minnesota River Wild Area
```

Figure 12.—*Interpolation data file for recreational/ research locations in the metropolitan area of Minneapolis/ St. Paul, MN, entered in miles (code = 3) from (94° W, 44.3° N).*

the interpolation file (fig. 12) contains metropolitan locations expressed in miles (code 3) calculated from the SW corner point (94, 44.3). The distance units (here miles) must match the measurement units specified with the test data file. Why would a user want to enter the interpolation data file in code 3? When it is awkward to determine lat/longs at local sites, it may be easier to find the SW corner point on a map and express the geographic locations in miles from that point.

Output data.— The interpolation output file on a single run can be saved to a file or is written to the screen and is similar to the estimation output file in figure 11, but without true values and errors. Figure 13 illustrates the interpolation output file for the current example.

SECONDARY FEATURES

Measurements

The measurements option allows you to specify how the various distance and climatic variables are displayed. This option is available when processing regional climatic data on single or multiple runs. In response to the measurements prompt, you can write the converted distance units in either English units (miles) or metric units (kilometers). All regional climatic data files and corresponding parameter files are in English units (°F or inches). When the user specifies [**M**]etric units on a regional run, the parameter values, kriged values, weights display and statistical summaries are all converted to metric units (°C or cm).

Boundaries

The boundary check is only used when making a regional interpolation run (for multiple or single runs). On each regional data file, default minimum and maximum latitudes and longitudes have been entered (in decimal degrees). These maximums and minimums form a rectangular envelope around the meteorological stations. When climate is interpolated at research plots, locations must fall within this envelope to be kriged or an error message will be given. These values provide only a crude boundary check. They can be changed by users to more accurately reflect their needs. The rectangle needs to be small enough to warn of potential outliers when entering plot locations for interpolation, and yet large enough to allow for extrapolation of locations

Plot	X	Y	Est	K_SD	Nbrs
1	44.000	48.000	22.678	1.321	7
2	49.000	37.000	23.787	1.324	7
3	25.000	32.000	22.379	1.617	7
4	35.000	59.000	21.008	1.510	7
5	49.000	60.000	22.534	1.532	7
6	46.000	53.000	21.925	1.301	7
7	38.000	35.500	23.085	1.441	7

Figure 13.—*Interpolation of percent frequency of days with precipitation greater than or equal to 0.10 inches for study locations in the metropolitan area of Minneapolis/ St. Paul, MN. X,Y in miles from (94° W, 44.3° N).*

outside the normal range of the station locations such as the "tip of the arrowhead" in northeastern Minnesota.

With an interpolation run using Option 0 (i.e., no region), no boundary checks are made. The user is responsible for exercising care in data entry. Should the location of an interpolation point be too remote, no near neighbors will be found and a missing value will be returned.

An additional computer check is made on all runs when spherical coordinates are used—all geographical locations are required to be in the first quadrant (i.e., north and east of the SW corner point).

Error Checking

Minimal error checking of model parameters is provided. Values on the supplied parameter files should be correct. However, errors may occur when the user interactively enters parameters in the test mode. When an error is encountered, an error message is written to the screen and corrected input is requested.

Missing Values

Missing values for environmental observations are represented by -999 for all forms of input. The "-999" was used because it fits into the climatic data field, but is not a possible monthly precipitation or temperature value after climatic values are converted to their decimal form. Missing values on the statistical output file are represented by *****'s and on the interpolation output file by -999. On the cross validation output file, missing observations are not printed.

Time Intervals

On multiple time series runs, the full time sequence of the climatic data file can be used or a shorter range can be specified, such as "1960,1989" to represent 30-year normals starting in 1960. The keyboard prompts are:

```
Do you want to specify starting year
 and last year of run (Y/N)?: y

Enter starting year and last year
 separated by comma or blank
 (Max Range 1900, 1992): 1960,1989
Range is   1960 - 1989
```

Note that both starting year and last year are entered simultaneously (separated by a comma or blank(s)). If only the starting year has been entered, the computer will wait until the last year is also entered.

Kriging starts in January of the first year and is completed in December of the second year listed. A single year can be run by specifying "19xx,19xx" for the starting and ending years. This is an easy method to evaluate any year's climate—with the output normally written to a file. This same option is available in single runs, except that only values for June 1980 are used, the output is generally written to the screen, and the weights display is available. Because the weights are not affected by the actual Z values, many aspects of the data can be explored using the default regional data in the single interactive mode. (The weights option is discussed later in this section.)

Geographic Location Code

The locations can be expressed in three ways—(1) latitude-longitude coordinates in degrees and minutes, (2) latitude-longitude coordinates in decimal degrees, or (3) Cartesian units. When a cross validation run is performed, location data on the regional data files are always in degrees and minutes. Option 0 runs allow for data entry in all three forms.

For interpolation runs, the interpolation data file can be in any form when regional data are used. For Option 0 runs, the user is not permitted to enter locations of environmental test data in Cartesian units and research plot locations in latitude-longitude coordinates. All other combinations are allowed.

Option 0 Test Mode

Option "0" means no region was specified to the regional prompt. This option allows users to enter any environmental data in the traditional X,Y,Z column format, and the location data can be entered in spherical coordinates. This option has already been discussed in detail in the Test Data File Section of Single Interactive Run.

Search Radius

Geo-CLM minimizes search strategies because many researchers using Geo-CLM are probably fairly unfamiliar with the intricacies of kriging. Also with multiple kriging runs, a fairly general approach, applicable across all runs simultaneously, is needed. The search strategy chosen is very simple: a circular search radius. Only those data falling within a search circle centered at the location being estimated are considered. An exhaustive search checks systematically for all data within the search radius and retains them.

In the Lake States, no strong anisotropy was found and therefore a circular search strategy is appropriate. Monthly climatic data represent an average condition, and strong anisotropy is unlikely to be found, even in other regions. The kriging algorithm will handle an anisotropic model, should it occur, but not the corresponding elliptic search strategy. This means that in the rare case of stronger anisotropy, the search strategy used is not as efficient as it could be.

Often the choice of search radius is based on data availability. Enough data must be found to provide a stable estimate without unnecessary smoothing. When a large search radius is used, kriging tends to smooth out the more local variation. To obtain greater local accuracy, use a smaller search radius that uses only the more local observations. However, because in the early 19th century, very few weather stations were operating, especially in the sparsely populated regions, the search radius needs to be large enough to pick up the more distant stations in these situations. One solution is to set the search radius to a larger value and restrict the number of observations within the search radius using the maximum points variable. This should enable more observations to be picked up in early years and yet provide an effective smaller search radius in later years. Trial and error may be necessary to find the best combination of search radius and maximum points.

To further complicate the situation, kriging is a numerically intensive procedure. Computation time for matrix inversion using ordinary kriging is proportional to the cube of the number of samples retained (Isaaks and Srivastava 1989). A doubling of the sample size increases the number of calculations eightfold. Execution, especially on multiple time series runs, may take many hours. Run times will depend on the number of meteorological stations, the search radius, maximum points, length of time interval, power of the computer's microprocessor, and in the case of interpolation runs, the number of research plots estimated. Preliminary runs using a shorter time interval can be made. Time is recorded at the start of the run, at appropriate 1-, 5-, or 10-year intervals, and at the run's conclusion. From these values, time for a full run can be estimated. If run time is too long, the search radius or maximum points may need to be reduced. However, if the search radius is too small, no near neighbors may be found, causing missing values to be returned.

Weights

The weights option is only available on single runs because such displays would provide a major disruption on multiple runs. The same affect, as if making a multiple run, is obtained by using the regional data on a single run. The weights option provides a means to view intermediate kriging results. It shows the kriging weights produced and several other intermediate results.

When cross validation runs are made using regional data, the first display is for the first listed weather station or observation point. The kriging weights display (fig. 14) provides information about the neighboring stations or observations found within the search radius. Header information includes location of the point to be kriged in the appropriate Cartesian units, datum number, and Z value. Stations are listed in order of increasing distances from the estimated point, followed by X,Y,Z values and the weights assigned to each point in the kriging algorithm. The location values are expressed in miles (or kilometers) from the SW

```
                         ---- WEIGHTS ----

        Point kriged is :(    134.239,    155.318)

     Datum #, Z value, Name :    1     4.990         APPLETON

Stn no      X(i)        Y(i)         Z(i)     Distance        W(i)        Name
---------------------------------------------------------------------------------

  10     125.693     141.520      5.000      16.230        .429       OSHKOSH
   5     130.819     124.261      5.130      31.245        .125       FOND DU LAC
   9     170.141     144.964      5.020      37.365        .141       MANITOWOC
  11     154.755     120.818      7.170      40.140        .062       PLYMOUTH
   6     178.686     169.129      4.510      46.544        .169       KEWAUNEE
  15     168.428     120.818      5.180      48.571        .019       SHEBOYGAN
  18     144.496      96.657      3.320      59.551       -.005       WEST BEND
  12      80.394     105.852      4.180      73.118        .050       PORTAGE
  16     117.146      82.850      3.270      74.456       -.002       WATERTOWN
  17     141.931      70.196      3.620      85.468       -.009       WAUKESHA
   7     106.890      73.651      3.870      86.124       -.004       LAKE MILLS
  13      65.003      90.905      2.800      94.565        .021       PRAIRIE DU SAC
  14     161.589      49.509      4.180     109.286       -.005       RACINE
   3      82.962      42.617      5.600     123.817       -.001       BRODHEAD
   2     100.904      34.559      6.010     125.275       0.000       BELOIT
   4      45.352      47.185      5.760     139.977        .012       DARLINGTON

# Neighbors:        16        Sum Weights:      1.00
   Estimate:      5.03         True Value:      4.99
    Std Dev:      1.07         Difference:    0.382E-01

 --------------------------------------------------------------------------------

 Print out next point? (Y/N):  N
```

Figure 14.—*First weights screen display on a single cross validation run for SE Wisconsin.*

corner point—when the latitude/longitude conversion is appropriate. Summary statistics for each point are given at the bottom of the display, including the number of near neighbors found, kriging estimate, true value, kriging standard deviation, and the difference (or error). For regional runs, the meteorological station names are added for user convenience. Output is similar for Option 0 except that the data for the cross validation run are the user's own test data.

As each weights display is printed to the screen, the user has the opportunity to cancel the remaining displays. Once the display is terminated, the program finishes kriging the remaining observation points and summarizes the results in a table consisting of weather station or observation location number, observed and estimated values, difference, kriging standard deviation, and number of neighbors found (see earlier example in figure 11).

The interpolation option provides information about weather stations or observations near the research plot being kriged. The weights display is slightly different because there is no known value of the kriged point, as there was with cross validation.

When performing a kriging study, it is a good practice to select a few representative locations and evaluate the kriging results, including the kriging weights and comparisons of estimates to nearby data. A single kriging run can be performed on selected data points using different variogram models and parameters. The goal is to gain an intuitive understanding of how the various variogram parameters (such as size of relative nugget effect) and data configurations (such as clustered or screened data) influence the estimation process. For example, it can be demonstrated that scaling the variogram by a multiple only affects the kriging variance but not the weights. Only the relative shape of the variogram model is needed to determine the weights and kriging estimate. These evaluations can be performed on either regional data or user-inputted data when in the single interactive mode.

Statistics

Answering **[Y]**es to the statistics prompt provides for summary statistics. These statistics, printed out to a user-specified file, are provided on multiple time series cross validation runs. (A shortened statistical summary is printed to the screen on single interactive runs.) Kriging errors are analyzed in three fundamental ways. First, the estimates calculated by ordinary kriging are compared to the observed values using the statistical measures of mean and standard deviation (fig. 15). These statistics are averaged by month. (Included also is the monthly correlation between the observed and estimated values.) Second, statistics are presented to evaluate the residuals from ordinary kriging (fig. 16). At a broad level, mean error (or bias), mean absolute error (MAE), and mean squared error (MSE) across all stations and time are given. At a more detailed level these statistics are broken down by month, weather station, and month-station averaged over the time interval. Third, the kriging standard deviation is given (fig. 17).

Analysis of the observed values and the kriging estimates (the first section given in figure 15) should exhibit the typical smoothing effect of kriging. Overestimation of low values and underestimation of high values will be observed, causing the distribution of the kriging estimates to have less spread than the distribution of the observed errors.

The statistical summary in the second section provides information on the distribution of the errors (fig. 16). These errors are defined so that a positive error is an overestimate and a negative error is an underestimate. Kriging estimates are theoretically unbiased with minimum variance. Therefore, the mean should be close to zero and ideally have minimal spread. Broad climatic trends, as indicated by mean errors, are usually reproduced quite well by kriging. However, spatial features of the errors may reveal limitations with the estimation procedure near isolated topographical features, along oceans or large bodies of water, or in mountainous regions with strong orographic effects. Because the errors presented are averaged over long time intervals, large errors may represent important problem areas, which unfortunately may be difficult to rectify without using additional detrending methods.

The third section provides overall average kriging standard deviations that are also broken down by month and meteorological station (fig. 17). One of the perceived benefits of kriging is that it provides a measure of the kriging variance. Unfortunately, the kriging variance is independent of the data values and cannot be used as a measure of estimation accuracy, but only as a ranking of the alternative data configurations. It should not be used to select a variogram model (Deutsch and Journel 1992).

Before a final time series interpolation run, it is a good practice to check the results of a cross validation statistics run using the same proposed conditions. Obvious problems can be identified with input parameters (variogram specifications, search radius etc.) or the data (e.g., anomalous values).

STATISTICS OUTPUT FILE:

**** STATISTICAL SUMMARY ****

-- ERROR --

------------ Month ------------

	1	2	3	4	5	6	7	8	9	10	11	12
Mean Z :	17.03	21.53	32.44	45.72	57.10	66.39	71.55	69.45	61.53	50.35	36.87	22.81
Mean Est :	17.03	21.53	32.44	45.76	57.18	66.47	71.60	69.46	61.52	50.34	36.87	22.81
Std Dev Z :	5.79	5.06	4.60	3.49	4.11	3.14	2.35	2.40	2.23	3.64	3.00	5.42
Std Dev Est:	5.64	4.87	4.48	3.28	3.83	2.80	2.08	2.20	1.97	3.50	2.78	5.28
Cor :	.98	.97	.97	.93	.94	.89	.84	.88	.84	.95	.94	.98
Num :	532	533	534	535	532	533	532	531	530	527	526	526

Figure 15.—*Statistical output file for a multiple time series run for temperature from 1960 to 1989 for SE Wisconsin (in English units). Comparison of estimated and observed values.*

------------ Month ------------

	ALL	1	2	3	4	5	6	7	8	9	10	11	12
Bias:	.022	0.00	0.00	0.00	.05	.08	.08	.05	.01	0.00	-.01	0.00	0.00
MAE :	.939	.96	.98	.84	.94	1.09	1.09	.99	.91	.93	.87	.78	.88
MSE :	1.539	1.54	1.72	1.28	1.56	2.12	2.00	1.65	1.34	1.50	1.37	1.09	1.29
Num :	6371	532	533	534	535	532	533	532	531	530	527	526	526

------------ Station ------------

	ALL	1	2	3	4	5	6	7	8	9	10
Bias:	.022	-.07	-1.87	1.55	.17	.08	1.08	-.23	-.23	.13	.37
MAE :	.939	.86	1.90	1.67	.60	.71	1.51	.65	.65	.56	.71
MSE :	1.539	1.13	4.57	3.69	.56	.84	3.55	.68	.66	.54	.86
Num :	6371	349	341	356	360	360	342	355	360	356	360

11	12	13	14	15	16	17	18
.11	.10	-.21	.05	-1.20	-.18	-.28	.98
.62	1.01	.62	1.09	1.48	.70	.61	1.02
.64	1.49	.62	2.12	2.92	.82	.72	1.50
358	355	359	349	360	360	334	357

------------ Month ------------

STN	1	2	3	4	5	6	7	8	9	10	11	12
1	1.13	.84	.42	-.61	-1.19	-1.19	-1.17	-.71	-.22	.08	.83	1.01
2	-1.48	-1.42	-1.76	-2.53	-2.53	-2.30	-1.83	-1.60	-1.92	-2.20	-1.50	-1.26
3	1.92	1.80	1.50	1.47	1.48	1.21	1.24	1.45	1.67	1.95	1.31	1.57
4	-.12	-.16	.04	.08	.25	.60	.39	.22	.31	.20	.26	-.01
5	.87	1.18	.45	-.25	-.61	-.59	-.56	-.34	-.23	-.17	.41	.86
6	-.53	.13	.08	1.81	2.79	3.05	2.53	1.73	1.34	1.18	-.23	-.58
7	.20	.13	-.08	-.30	-.48	-.12	-.35	-.40	-.57	-.66	-.24	.07
8	-.16	-.45	-.02	-.17	-.11	-.31	-.17	-.43	-.48	-.41	.10	-.13
9	-.25	-.04	.21	.11	.23	.09	.36	.48	.38	.32	-.05	-.26
10	.85	.91	.78	.33	-.13	-.22	-.11	-.04	.08	.12	.77	1.05
11	.31	.01	.20	-.24	-.55	-.50	-.05	.24	.49	.52	.47	.42
12	-.57	-.45	-.76	-.42	.34	.90	.98	1.00	.87	.25	-.39	-.49
13	-.19	.02	.07	.28	-.28	-.67	-.73	-.58	-.39	-.02	.09	-.10
14	-.80	-.85	.17	1.44	1.94	1.23	.74	-.30	-.67	-.74	-.91	-.98
15	-2.33	-2.55	-1.51	-.11	.67	.36	-.50	-1.28	-1.60	-1.59	-1.72	-2.20
16	.40	.47	-.24	-.48	-.50	-.66	-.42	-.33	-.25	-.26	-.16	.27
17	-.07	-.37	-.21	-.46	-.85	-.86	-.69	-.25	-.02	.20	.09	.10
18	.82	.70	.67	.88	1.13	1.39	1.41	1.33	1.09	.94	.71	.67

Figure 16.—*Distribution of kriging errors on statistical output file for a multiple time series run for temperature values from 1960 to 1989 for SE Wisconsin (in English units).*

-- KRIGING STANDARD DEVIATION --

------------ Month ------------

	ALL	1	2	3	4	5	6	7	8	9	10	11	12
Mean:	1.374	1.58	1.75	1.37	1.29	1.34	1.26	1.27	1.35	1.30	1.25	1.26	1.45
Num :	6371	532	533	534	535	532	533	532	531	530	527	526	526

------------ Station ------------

	ALL	1	2	3	4	5	6	7	8	9	10
Mean:	1.374	1.36	1.40	1.35	1.43	1.27	1.58	1.28	1.69	1.34	1.30
Num :	6371	349	341	356	360	360	342	355	360	356	360

	11	12	13	14	15	16	17	18
	1.25	1.37	1.37	1.58	1.30	1.26	1.32	1.29
	358	355	359	349	360	360	334	357

Figure 17.—*Kriging standard deviations on statistical output file for a multiple time series run for temperature values from 1960 to 1989 for SE Wisconsin (in English units).*

Lake Effect

The Great Lakes can have a marked effect upon local land temperatures in certain months. Lake Superior and Lake Michigan act as a heat reservoir in cool periods and as a cooling source in warm periods. As a result, temperatures along the lakes are relatively higher in cold months and cooler in warm months. The magnitude of this lake effect increases with increasing latitude. The farther north, the greater the temperature differential between water and land temperatures. However, along the Minnesota coast of Lake Superior the effect of the lake on the regional climate is minimized both because of prevailing westerly winds and because of the upland that rises abruptly from the lake shore (Baker and Strub 1965). Thus, the influence of Lake Superior in Minnesota is of importance only along the immediate lake shore.

Grand Marais is the most northerly meteorological station included in the climatic data base. It's temperatures are over 7° F warmer in December and January and over 6° F cooler in June than without the presence of Lake Superior. Residual kriging (Holdaway 1994) was used to detrend the Minnesota temperature data of these effects. With detrending, the program first subtracts off the lake effect from temperatures for the three stations on the north shore of Lake Superior. Then, kriging is used to interpolate the residuals (with lake effect trend removed). For a cross validation run, the lake effect is added back to the residuals before the estimation results and statistical summary are reported. However, for an interpolation run, the lake effect is subtracted from the residuals but not added back in. By removing the lake effect from lake stations (but not adding it back in), the algorithm does a better job of interpolating research plots in the general vicinity of Lake Superior but will not do as well on plots within 1 to 2 miles of the lake, where the lake effect is strongest. Handling this aspect properly would require location parameters specifying miles north along the lake and miles inland from the lake.

Currently, no accounting of the lake effect in Wisconsin and Michigan has been attempted. In northern Wisconsin and the Upper Peninsula of Michigan, elevation above the lake also becomes an important determinant. The lake effect in these regions is weaker than in Minnesota because of the lower latitude. Also, there are very few meteorological stations in the UP of Michigan. As a result, temperature interpolations in these areas will not be as accurate.

Growing Degree Days

Number of growing degree days is one of the most commonly used applications of monthly temperature data. Because so many biological and physiological processes are temperature regulated and because of the relative ease with which it can be calculated, number of growing degree days is an often used measure of the heat available for biological growth. It is based upon the concept that below a certain base temperature no growth occurs. By subtracting this minimum base temperature (known as the threshold temperature) from the daily average, the amount of heat available for species growth can be estimated.

In actual usage, growing degree days are often computed from monthly temperature values rather than daily temperatures. Using monthly values as opposed to daily values masks daily variations but does provide a good approximation. The formula used in Geo-CLM to calculate growing degree days (*GDD*) from monthly mean temperature is:

$$GDD = \sum_{i=1}^{12} (\bar{t}_i - T) * n_i$$

where, $\bar{t}_i$ = mean monthly temperature in i^{th} month;
T = threshold (or base) temperature;
and
n_i = number of days in i^{th} month.

The difference between mean t_i and T are to be summed for all positive values. Negative values are not counted. A historical record of annual growing degree days is computed from each year's monthly temperature values over the time interval specified by the user.

To accurately model forest stand/tree growth, researchers need to consider the environmental condition, including how climate affects growth and mortality of tree species. Accumulated growing degree days are often selected as a predictor when modeling tree growth. However, the application of growing degree days to tree growth may be more accurate when limited to the period from initiation of early spring growth until early or late July—a period over which soil moisture is usually not limiting to tree growth and when most height growth of the trees is completed. Thus, researchers may desire growing degree days to be summed only through June or July. This option is available through a keyboard prompt to specify the month when accumulation of growing degree days is terminated. The answer "12" will provide for an annual value.

Keyboard prompts for requesting growing degree days output, changing the threshold value, and indicating the last month through which growing degree days are accumulated are:

```
Calculate Growing Degree Days? (Y/N): y
Current threshold value (base tempera-
  ture) = 40.0  degrees F.

Do you want to change the threshold
  value? (Y/N): y
Enter new threshold value (32-70 de-
  grees F): 32

Accumulate GDD's through nth month
  (enter numerical value)
Enter 12 for untruncated GDD's: 7

      Select a GDD output file.
GDD File name = Out.GDD SE WI 60-89
```

PROGRAM AVAILABILITY

The compiled program can be obtained at no cost by sending three double density or two high density new, unformated 3.5-inch diskettes to:

Modeling (Geo-CLM program)
North Central Forest Experiment Station
1992 Folwell Avenue
St. Paul, MN 55108

Specify Mac or DOS version and include phone number with return address. Questions can be sent to Modeling at:

FAX: (612) 649-5285
Internet: holda001@staff.tc.umn.edu
DG: M.HOLDAWAY:S23A

APPENDIX A

GEOSTATISTICS

Estimating values at unsampled locations is an interpolation problem. Kriging, a statistical technique used in the spatio-temporal interpolation of geographic data, is one of the most refined methods available. The technique has been used frequently in soil science (Burgess and Webster 1980a,b; Webster and Burgess 1983) and hydrology (Delhomme 1978). Recently it has been used in analyzing ecological and environmental data (Fortin *et al.* 1989, Lefohn *et al.* 1988, Rossi *et al.* 1992, Seilkop and Finkelstein 1987). Kriging has been implemented in forestry (Samra *et al.* 1989) and climatology (Streit and Schwentker 1982, Bigg 1991, Hosang and Dettwiler 1991) including the study of orographic effects (Chua and Bras 1982, Dingman *et al.* 1988, and Phillips *et al.* 1992). For those readers unfamiliar with the basic concepts of kriging, a good fully developed introductory text can be found in Isaaks and Srivastava (1989).

Kriging deals with earth science variables distributed in space and/or time. Observations that are closer in space and/or time are more similar than more distant ones. The variation of an environmental variable has a structured component (a measure of the similarity between close observations) and a random component (Journel and Huijbregts 1978). The two components are represented in the variogram, a mathematical description of the spatial variability of the variable. The overall approach consists of two parts: modeling the variogram and then using kriging to spatially interpolate unknown data points based on weighted local averaging. The variogram is used to determine the weights applied to nearby known data points. Kriging provides estimates that are unbiased and that have minimum mean squared estimation error.

Variogram

The spatial dependence between observations is expressed by the semivariance, which is a measure of the average similarity between observations (station-to-station differences) at a given distance apart. The variogram is the function produced by plotting the semi-variance values against distance.

The experimental variogram may take many forms. Typically, it increases with distance between sample locations, rising to a more or less constant value (the sill) at a separation distance, called the range of influence. Samples separated by distances closer than the range are spatially related. Thus, the range defines the maximum radius from which neighboring samples are used for interpolation by the kriging program. The more or less rapid increase of the variogram represents the more or less rapid deterioration of the influence of a given sample point over more and more remote sample points. At the sill, no more influence remains. The "nugget effect" represents unexplained or "random" variance due to measurement error or spatial variability at a scale smaller than that of the shortest sampling distance. These three parameters are used to represent the variogram (fig. 18).

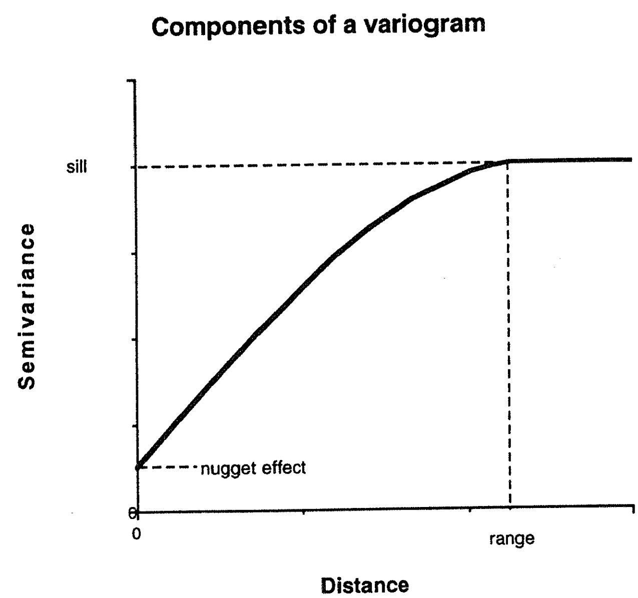


Figure 18.—*Variogram showing sill, range of influence, and nugget effect.*

The equations most commonly used to estimate parameters of the variogram are the linear, spherical, exponential, and gaussian model (fig. 19). Both the spherical and exponential models are nearly linear at the origin, but the spherical model stays linear longer and then makes a more abrupt change in slope as it approaches the sill. The exponential model is much slower in its approach to

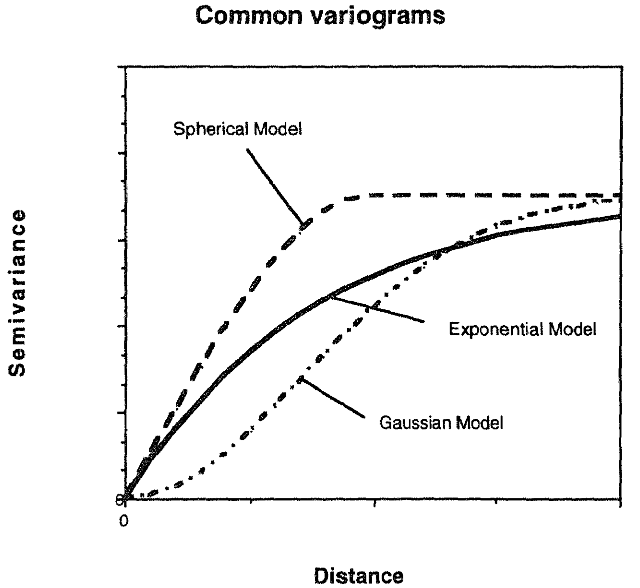


Figure 19.—*The three most commonly used nonlinear variogram models—the spherical, exponential, and gaussian models.*

the sill. The gaussian model has a parabolic behavior near the origin, which is indicative of a very continuous phenomenon so the estimation procedure makes much more use of the closest points (Isaaks and Srivastava 1989).

When the spatial structure is directionally oriented (called anisotropy), the estimation process becomes more complex, because the interpolation weights assigned to data points are functions not only of distance but also of direction.

Kriging

Once the spatial structure between observation points has been modeled by the variogram, the statistical procedure kriging uses the variogram to assign interpolation weights to data points within the neighborhood of the search radius. In kriging, the nearest few points are usually the most heavily weighted. Little is gained by including distant points unless there are few points close to the kriging location. This may occur with randomly sampled data.

Trends

Ordinary kriging assumes that the data points are stationary, i.e., they contain no significant trends over the spatial range (Journel and Huijbregts 1978). However, many environmental variables show local trends or even broad regional trends. These trends are usually not a problem because the stationarity assumption applies only to the search neighborhood (i.e., the radius around the point being estimated within which all data points are used in the kriging)—not to the entire data set. Data sets often are reasonably homogeneous within the smaller regions designated by the search neighborhood. Thus, ordinary kriging can be used even in situations of moderate drift without causing serious estimation bias (Journel and Huijbregts 1978).

To remedy the nonstationarity problem when the trend becomes too large, "universal" kriging was developed. Its simpler form, residual kriging, can also be used to "detrend" the data. Detrending is useful in dealing with stronger locally occurring conditions. One such local trend in the Lake States is the modifying influence of the Great Lakes on local climate, which can be quite large in certain months.

HISTORICAL CLIMATIC DATA BASES

"Secondary" Data Sources

Climatic observations are available from various sources such as the National Climatic Data Center (NCDC), public institutions, and individual researchers. Ideally, the data reflect the true state of the climate at a particular point in space and time. Realistically, different instruments, faulty instrument calibration, changes in measurement location or time of observation measurement, recording/processing/transcription errors, etc. can introduce various types and sizes of error (Guttman 1991). Because most historical climatic records were collected by others, current users do not know how inaccurate they are. These secondary data sources are often all that is available, but comprise an inherently weak data source.

Climatic data hold a unique position. There is an overwhelming need for it and yet long-term regional coverage can only be found by using NCDC data or NCDC data edited and adjusted by researchers. No other extensive data sources are available. Your choice is to spend vast amounts of your time trying to improve NCDC's original editing (which decreases with age of the data) or use the best available NCDC data source that has been cleaned up by others. For long-term climatic data, that source is the Historical Climatological Network.

Historical Climatological Network (HCN)

Climatic records should have few errors or biases, but this is seldom the case. Take for example the "Summary of Month" (SOM) data provided by the NCDC (NOAA 1986). These data have been edited for the more common and obvious recording/processing/transcription errors. However, the most stringent editing has been applied generally to the more recent data due to improved computerization. The amount of editing rapidly decreases as one regresses in time. Even so, the quality editing, when present in recent years, probably only accounted for the obvious human-computer generated errors. Beyond editing for these more gross errors, it is important that the climatic record be as free as possible of station discontinuities and inhomogeneities (due to station changes and changes in observation methods including time of observation bias). These errors are often much more subtle and hence very difficult to detect. Climate data sources from the NCDC have not been edited for these biases.

Researchers at the U.S. Department of Energy have developed climatological time series of monthly temperature and precipitation records adjusted for station inhomogeneities (Karl and Williams 1987) and time of observation bias (Karl *et al.* 1986). Their methods make use of station history information and correlations between candidate stations and neighboring stations to detect discontinuities. By these adjustment methods a Historical Climatological Network (HCN) of more than 1,200 stations across the United States was developed (Brower 1985). This network consisted largely of rural cooperative stations (90 percent have populations < 50,000) with a relatively long time series (≥ 80 years) of monthly temperature and precipitation data. Concentrating on stations from smaller communities and forest or agricultural locations attempted to remove the influence of urbanization. The adjustment technique tried to produce a climate time series as free as possible of discontinuous station inhomogeneities.

By minimizing the urbanization effect and the station/observation methods effect, a data source has been made available that is relatively free of non-climatic influences. This "cleaned-up" HCN data source with data through 1983 or 1984 was made available from the Minnesota State Climatology Office. There are 65 HCN stations in Minnesota, Wisconsin, and Michigan.

"Summary of Month" SOM Data

A second data source was used to supplement the above data. In Minnesota and Wisconsin, the HCN stations were augmented by 64 additional observer network stations from the SOM data tapes. The SOM data were also available from the Minnesota State Climatology Office, covering the entire century. They helped provide the necessary density to estimate spatial statistics at closer ranges or to provide data in the sparsely sampled regions. These data have undergone some editing but cannot be considered especially "clean."

Without undergoing an elaborate editing procedure, a simple method was used to identify gross outliers, such as those caused by misplacing the decimal point or the entry of a zero rather than a missing value code. All climate observations were run through the cross validation program, and each observation was estimated using only surrounding observations. The interpolated values were compared with the actual value, and observations with obvious gross errors were checked. Each identified problem observation was checked for validity and was either retained, replaced by hopefully an accurate though estimated value, or coded as missing if the value was obviously in error but for unknown reasons. This procedure helped identify those situations in which a very high precipitation observation, say 11.60 inches, was mistakenly recorded as the more realistic value of 1.16 inches.

The HCN data went through 1984 in Michigan, 1983 in Wisconsin, and 1988 in Minnesota. These data were augmented by the most recent SOM data provided by the Midwest Climate Center to complete records through May 1992. These data were also error checked by kriging. Therefore, the quality of the most recent data does not match that of the HCN data but should be fairly good because of the improved editing procedures instigated by NCDC in recent years.

In general, the quality of the supplemental minimally edited SOM data is unknown and decreases with age. These data were used to increase the spatial and temporal coverage. They comprised 50 percent of the stations in the data base, but represented only about 35 percent of the observations due to their shorter time records; most long-term stations are HCN data.

APPENDIX C

LISTING OF REGIONAL DATA FILES

1) Michigan

```
'Michigan'               Region name
'PAR Prec MI'            Parameter file for precipitation
'PAR Temp MI'            Parameter file for temperature
'MI_P'                   Climatic data file for precipitation
'MI_T'                   Climatic data file for temperature
91.00 41.00              Global SW corner point (LONG,LAT).  Global values = 97.00,41.00
   82    91              Minimum and maximum longitude
   41    48              Minimum and maximum latitude
 1900  1993              Default first and last year of run
   31                    Number of weather stations
  1   84  1   41 55   3.86   63.8  'ADRIAN               '  Id, longdeg, longmin,
  2   85 51   42 32   6.02   62.5  'ALLEGAN              '    latdeg, latmin,
  3   84 40   43 23   2.75   61.6  'ALMA                 '    defprec, deftemp, name
  4   83 44   42 17   3.77   64.9  'ANN HARBOR           '
  5   85 29   43 42   4.12   61.4  'BIG RAPIDS           '
  6   86 56   46 21   2.64   57.7  'CHATHAM              '
  7   84 28   45 39   2.80   58.5  'CHEBOYGAN            '
  8   85  0   41 57   6.59   64.4  'COLDWATER SP         '
  9   83 30   44 17   3.82   59.0  'EAST TAWAS           '
 10   86 42   45 41   3.44   56.1  'FAYETTE              '
 11   86 22   43 42   3.03   62.4  'HART                 '
 12   84 38   41 55   5.03   64.2  'HILLSDALE            '
 13   88  4   45 50   2.67   61.9  'IRON MT              '
 14   90 10   46 27   2.66   58.5  'IRONWOOD             '
 15   87 39   46 29   3.33   59.4  'ISHPEMING            '
 16   85 36   42 17   5.88   64.0  'KALAMAZOO            '
 17   84 13   43 37   3.34   63.8  'MIDLAND              '
 18   82 49   42 36   3.58   61.8  'MT CLEMENS           '
 19   84 46   43 35   3.17   62.4  'MT PLEASANT          '
 20   86 39   46 24   4.33   56.0  'MUNISING             '
 21   85 30   46 20   4.17   56.1  'NEWBERRY             '
 22   84 11   43  1   2.49   64.1  'OWOSSO               '
 23   86 17   42 24   7.11   62.6  'SOUTH HAVEN          '
 24   88 38   46  4   3.06   59.7  'STAMBAUGH            '
 25   90 49   46 53   3.69   61.3  'BAYFIELD, WI         '
 26   88 13   45 57   2.32   58.3  'BRULE IS, WI         '
 27   88 12   45 17   5.70   62.7  'CRIVITZ HIGH, WI     '
 28   89  8   45 54   3.86   59.4  'LONG LK DAM, WI      '
 29   87 38   45  6   5.68   63.1  'MARINETTE, WI        '
 30   89 53   46  8   2.88   60.4  'REST LAKE, WI        '
 31   87 20   44 52   4.69   61.9  'STURGEON BAY WI      '
```

2) Minnesota

```
'Minnesota'             Region name
'PAR Prec MN'           Parameter file for precipitation
'PAR Temp MN'           Parameter file for temperature
'MN_P'                  Climatic data file for precipitation
'MN_T'                  Climatic data file for temperature
97.00 45.00             Local SW corner point (LONG,LAT).  Global values = 97.00,41.00
   89    97             Minimum and maximum longitude
   45    49             Minimum and maximum latitude
 1900  1993             Default first and last year of run
   36                   Number of weather stations
  1   96 31   47 18   2.88   66.2  'ADA                   '   Id, longdeg, longmin,
  2   95 23   45 52   4.97   65.0  'ALEXANDREIA           '     latdeg, latmin,
  3   91 55   47 41   3.71   62.1  'BABBITT               '     defprec, deftemp, name
  4   94 37   48 43   2.47   63.2  'BAUDETTE              '
  5   94 56   47 30   2.51   63.0  'BEMIDJI               '
  6   93 48   48 12   3.48   62.0  'BIG FALLS             '
  7   94 38   47 23   2.81   62.3  'CASS LAKES            '
  8   92 31   46 42   2.46   61.9  'CLOQUET               '
  9   96 36   47 48   1.14   66.1  'CROOKSTON             '
 10   95 51   46 50   2.72   59.2  'DETROIT LAKES         '
 11   92  6   46 47   3.28   62.0  'DULUTH                '
 12   96  4   46 17   3.75   64.6  'FERGUS FALLS          '
 13   95 45   47 34   3.03   64.6  'FOSSTON               '
 14   90 21   47 44   2.32   53.2  'GRAND MARAIS          '
 15   93 30   47 14   3.40   63.6  'GRAND RAPIDS          '
 16   94 21   46 25   3.07   64.7  'GULL LAKE             '
 17   90 53   48 10   2.38   60.6  'GUNFLINT LAKE         '
 18   95 12   47 13   2.63   63.2  'ITASCA                '
 19   94 13   47 15   3.00   63.6  'LEECH LAKE DAM        '
 20   94 21   45 59   4.67   65.8  'LITTLE FALLS          '
 21   92 44   46 59   2.36   61.7  'MEADOWLANDS           '
 22   93 40   45 48   6.03   64.0  'MILACA                '
 23   93 18   45 53   4.37   63.1  'MORA                  '
 24   95  4   46 55   4.16   62.9  'PARK RAPIDS           '
 25   94  7   46 40   2.95   63.8  'PINE RIVER DAM        '
 26   93 35   47 15   4.32   63.1  'POKEGEMA DAM          '
 27   96 17   47 53   1.62   64.9  'REDLAKE FALLS         '
 28   95  2   47 52   3.45   62.7  'REDLAKE INDIAN AGNCY'
 29   95 46   48 51   1.75   61.4  'ROSEAU                '
 30   93 19   46 48   2.84   63.5  'SANDY LAKE DAM        '
 31   94  4   45 33   6.06   64.1  'ST CLOUD              '
 32   91 40   47  1   4.11   57.0  'TWO HARBORS           '
 33   92 33   47 30   5.83   61.1  'VIRGINIA              '
 34   94 35   47  4   3.01   62.9  'WALKER                '
 35   95 19   48 55   2.00   61.8  'WARROAD               '
 36   94  3   47 26   3.87   63.5  'WINNIBIGOSHISH DAM    '
```

3) Wisconsin

```
'Wisconsin'            Region name
'PAR Prec WI'          Parameter file for precipitation
'PAR Temp WI'          Parameter file for temperature
'WI_P'                 Climatic data file for precipitation
'WI_T'                 Climatic data file for temperature
94.00 42.00            Global SW corner point (LONG,LAT).  Global values = 97.00,41.00
   85    94            Minimum and maximum longitude
   42    47            Minimum and maximum latitude
 1900  1993            Default first and last year of run
   64                  Number of weather stations
  1  92 22  45 18   8.05  62.3 'AMERY                '  Id, longdeg, longmin,
  2  89  9  45  8   7.11  61.0 'ANTIGO               '    latdeg, latmin,
  3  88 23  44 15   4.99  64.7 'APPLETON             '    defprec, deftemp, name
  4  90 58  46 34   3.29  62.2 'ASHLAND EXP          '
  5  90 49  46 53   3.69  61.3 'BAYFIELD             '
  6  91 14  44 18   3.23  64.7 'BLAIR                '
  7  88 13  45 57   2.32  58.3 'BRULE ISLAND         '
  8  89 32  44 22   3.84  63.7 'CODDINGTON           '
  9  88 12  45 17   5.70  62.7 'CRIVITZ HI           '
 10  92  1  45 32   6.06  66.0 'CUMBERLAND           '
 11  92 22  46  1   3.76  62.9 'DANBURY              '
 12  89 32  44  7   5.04  66.0 'HANCOCK              '
 13  90 44  44 24   4.58  65.4 'HATFIELD             '
 14  91  8  45 13   8.10  65.7 'HOLCOMBE             '
 15  87 31  44 27   4.51  59.0 'KEWAUNEE             '
 16  90 42  42 52   5.03  67.4 'LANCASTER            '
 17  89  8  45 54   3.86  59.4 'LONG LAKE DAM        '
 18  87 38  45  6   5.68  63.1 'MARINETTE            '
 19  90  8  44 39   4.72  64.5 'MARSHFIELD           '
 20  90 23  44 11   5.94  63.5 'MATHER               '
 21  90 21  45  8  10.37  61.9 'MEDFORD              '
 22  90 39  46 19   2.54  60.4 'MELLEN               '
 23  89 44  45 11   7.13  61.1 'MERRILL              '
 24  89 44  45 53   2.85  59.1 'MINOCQUA DAM         '
 25  91 40  44 34   3.33  67.7 'MONDOVI              '
 26  90 38  44 32   5.63  64.2 'NEILLSVILLE          '
 27  88 44  44 23   5.07  65.9 'NEW LONDON           '
 28  87 57  44 54   6.37  61.1 'OCONTO               '
 29  88 33  44  3   5.00  66.0 'OSHKOSH              '
 30  90 27  45 56   5.02  61.8 'PARK FALLS           '
 31  89 26  43 32   4.18  65.5 'PORTAGE              '
 32  91  9  43  2   2.37  70.3 'PRAIRIE DU CHIEN     '
 33  89 44  43 19   2.80  68.0 'PRAIRIE DU SAC       '
 34  90 17  45 33   6.82  60.7 'PRENTICE             '
 35  89 53  46  8   2.88  60.4 'REST LAKE            '
 36  89 25  45 38   4.58  61.5 'RHINELANDER          '
 37  90 23  43 20   3.05  65.9 'RICHLAND             '
 38  92 37  44 52   6.23  67.0 'RIVER FALLS          '
 39  88 37  44 46   5.97  64.7 'SHAWANO              '
 40  91 49  46 21   2.99  63.9 'SOLON SPRGS          '
 41  90 48  43 57   3.66  65.8 'SPARTA               '
 42  91 53  45 49   2.91  63.7 'SPOONER              '
 43  90 56  44 58  12.25  65.0 'STANLEY              '
 44  89 34  44 30   3.55  64.9 'STEVENS PT           '
 45  87 20  44 52   4.69  61.9 'STURGEON BAY         '
 46  92  1  46 42   3.46  58.9 'SUPERIOR             '
 47  90 54  43 34   3.40  65.7 'VIROQUA              '
 48  89  4  44 21   5.02  64.9 'WAUPACA              '
 49  91 23  45 24   4.56  63.3 'WEYERHAUSER          '
 50  89 51  45 43   6.57  60.0 'WILLOW RES           '
```

```
 51    91  4    45 53    3.67    60.0    'WINTER                    '
 52    89 47    43 38    4.33    66.0    'WIS DELLS                 '
 53    89 48    44 23    3.42    64.7    'WIS RAPIDS                '
 54    86 56    46 21    2.64    57.7    'CHATHAM, MI               '
 55    86 42    45 41    3.44    56.1    'FAYETTE, MI               '
 56    88  4    45 50    2.67    61.9    'IRON MT, MI               '
 57    90 10    46 27    2.66    58.5    'IRONWOOD, MI              '
 58    87 39    46 29    3.33    59.4    'ISHPEMING, MI             '
 59    88 38    46  4    3.06    59.7    'STAMBAUGH, MI             '
 60    91 27    43 34    2.45    67.1    'CALADONIA, MN             '
 61    93 14    45 34    3.24    64.2    'CAMBRIDGE, MN             '
 62    93 11    44 40    4.76    67.4    'FARMINGTON, MN            '
 63    92 57    46  1    3.92    61.8    'HINKLEY, MN               '
 64    91 38    44  3    4.03    67.3    'WINONA, MN                '
```

4) SE Wisconsin

```
'SE Wisconsin'          Region name
'PAR Prec SE'           Parameter file for precipitation
'PAR Temp SE'           Parameter file for temperature
'SE_P'                  Climatic data file for precipitation
'SE_T'                  Climatic data file for temperature
91.00 42.00             Local SW corner point (LONG,LAT).  Global values = 97.00,41.00
   87    91             Minimum and maximum longitude
   42    45             Minimum and maximum latitude
 1900  1993             Default first and last year of run
   18                   Number of weather stations
  1    88 23    44 15    4.99    64.7    'APPLETON                  '    Id, longdeg, longmin,
  2    89  2    42 30    6.01    67.8    'BELOIT                    '      latdeg, latmin,
  3    89 23    42 37    5.60    66.0    'BRODHEAD                  '      defprec, deftemp, name
  4    90  7    42 41    5.76    66.0    'DARLINGTON                '
  5    88 27    43 48    5.13    64.3    'FOND DU LAC               '
  6    87 31    44 27    4.51    59.0    'KEWAUNEE                  '
  7    88 55    43  4    3.87    66.0    'LAKE MILLS                '
  8    90 42    42 52    5.03    67.4    'LANCASTER                 '
  9    87 41    44  6    5.02    62.8    'MANITOWOC                 '
 10    88 33    44  3    5.00    66.0    'OSHKOSH                   '
 11    87 59    43 45    7.17    64.7    'PLYMOUTH                  '
 12    89 26    43 32    4.18    65.5    'PORTAGE                   '
 13    89 44    43 19    2.80    68.0    'PRAIRIE DU SAC            '
 14    87 51    42 43    4.18    61.6    'RACINE                    '
 15    87 43    43 45    5.18    62.9    'SHEBOYGAN                 '
 16    88 43    43 12    3.27    66.9    'WATERTOWN                 '
 17    88 14    43  1    3.62    65.0    'WAUKESHA                  '
 18    88 11    43 24    3.32    63.2    'WEST BEND                 '
```

LITERATURE CITED

Baker, D.G.; Strub, J.H., Jr. 1965. **Climate of Minnesota. Part III. Temperature and its application**. Tech. Bull. 248. St. Paul, MN: University of Minnesota, Agricultural Experiment Station. 63 p.

Baker, D.G.; Kuehnast, E.L. 1973. **Climate of Minnesota. Part VII. Areal distribution and probabilities of precipitation in the Minneapolis-St. Paul metropolitan area**. Tech. Bull. 293. St. Paul, MN: University of Minnesota, Agricultural Experiment Station. 18 p.

Bigg, G.R. 1991. **Kriging and intraregional rainfall variability in England**. International Journal of Climatology. 11: 663-675.

Brower, W.A. 1985. **U.S. historical climatology network**. 3d conference on climate variations and symposium on contemporary climate. American Meteorological Society: 1850-2100.

Burgess, T.M.; Webster, R. 1980a. **Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging**. Journal of Soil Science. 31: 313-331.

Burgess, T.M.; Webster, R. 1980b. **Optimal interpolation and isarithmic mapping of soil properties. II. Block kriging**. Journal of Soil Science. 31: 333-341.

Chua, S.H.; Bras, R.L. 1982. **Optimal estimators of mean areal precipitation in regions of orographic influence**. Journal of Hydrology. 57: 23-48.

Delhomme, J.P. 1978. **Kriging in the hydro sciences**. Advance Water Resources. 1: 251-266.

Deutsch, C.V.; Journel, A.G. 1992. **GSLIB Geostatistical Software Library and user's guide**. New York: Oxford University Press. 340 p.

Dingman, S.L.; Seely-Reynolds, D.M.; Reynolds, R.C., III. 1988. **Application of kriging to estimating mean annual precipitation in a region of orographic influences**. Water Resources Bulletin. 24(2): 329-339.

Englund, E.; Sparks, A. 1988. **GEO-EAS (Geostatistical Environmental Assessment Software) user's guide**. Las Vegas, NV: U.S. Environmental Protection Agency. 183 p.

Fortin, M.; Drapeau, P.; Legendre, P. 1989. **Spatial auto correlation and sampling design in plant ecology**. Vegetatio. 83: 209-222.

Guttman, N.B. 1991. **Using secondary data sources**. Climatic Change. 18: 95-105.

Holdaway, M.R. **Spatial modeling and interpolation of monthly temperature using kriging**. (In prep.)

Hosang, J.; Dettwiler, K. 1991. **Evaluation of a water equivalent of snow cover map in a small catchment area using a geostatistical approach**. Hydrological Processes. 5: 283-290.

Isaaks, E.H.; Srivastava, R.M. 1989. **An introduction to applied geostatistics**. New York: Oxford University Press. 561 p.

Journel, A.G.; Huijbregts, C.J. 1978. **Mining geostatistics**. London: Academic Press. 600 p.

Karl, T.R.; Williams, C.N. 1987. **An approach to adjusting climatological time series for discontinuous inhomogeneities**. Journal of Climatology and Applied Meteorology. 26: 1744-1763.

Karl, T.R.; Williams, C.N.; Young, P.J. 1986. **A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States**. Journal of Climatology and Applied Meteorology. 22: 145-160.

Lefohn, A.S.; Knudsen, H.P.; McEvoy, L.R., Jr. 1988. **The use of kriging to estimate monthly ozone exposure parameters for the Southern United States**. Environmental Pollution. 53: 27-42.

Merz, R.W., ed. 1979. **Forest atlas of the Midwest.** St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Experimental Station. 48 p.

National Oceanic and Atmospheric Administration. 1986. **Environmental Data Service: summary of the month cooperative TD-3220**. Asheville, NC: National Climatic Data Center.

Phillips D.L.; Dolph, J.; Marks, D. 1992. **A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain**. Agricultural Forest Meteorology. 58: 119-141.

Rossi, R.E.; Mulla, D.J.; Journel, A.G.; Franz, E.H. 1992. **Geostatistical tools for modeling and interpreting ecological spatial dependence**. Ecological Monographs. 62(2): 277-314.

Samra, J.S.; Gill, H.S.; Bhatio, V.K. 1989. **Spatial stochastic modeling of growth and forest resource evaluation**. Forest Science. 35: 663-676.

Seilkop, S.K.; Finkelstein, P.L. 1987. **Acid precipitation patterns and trends in Eastern North America, 1980-84**. Journal of Climatology and Applied Meteorology. 26: 980-994.

Streit, U.; Schwentker, F. 1982. **Some problems of estimating areal rainfall means by kriging**. Proceedings of the symposium on hydrological research basins. Switzerland, Sonderh, Bern: 369-378.

Webster, R.; Burgess, T.M. 1983. **Spatial variation in soil and the role of kriging**. Agricultural Water Management. 6: 111-

Printed on recycled recyclable paper.

☆U.S. GOVERNMENT PRINTING OFFICE: 1994 - 557-249/80169

Holdaway, Margaret R.

1994. **The microcomputer scientific software series 9: user's guide to Geo-CLM: geostatistical interpolation of the historical climatic record in the Lake States**. Gen. Tech. Rep. NC-167. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station. 34 p.

Describes Geo-CLM, a computer application (for Mac or DOS) whose primary aim is to perform multiple kriging runs to interpolate the historic climatic record at research plots in the Lake States. It is an exploration and analysis tool. Addition capabilities include climatic data bases, a flexible test mode, cross validation, lat/long conversion, English/metric units, and growing degree days.

KEY WORDS: Kriging, computer application, temperature, precipitation, cross validation.