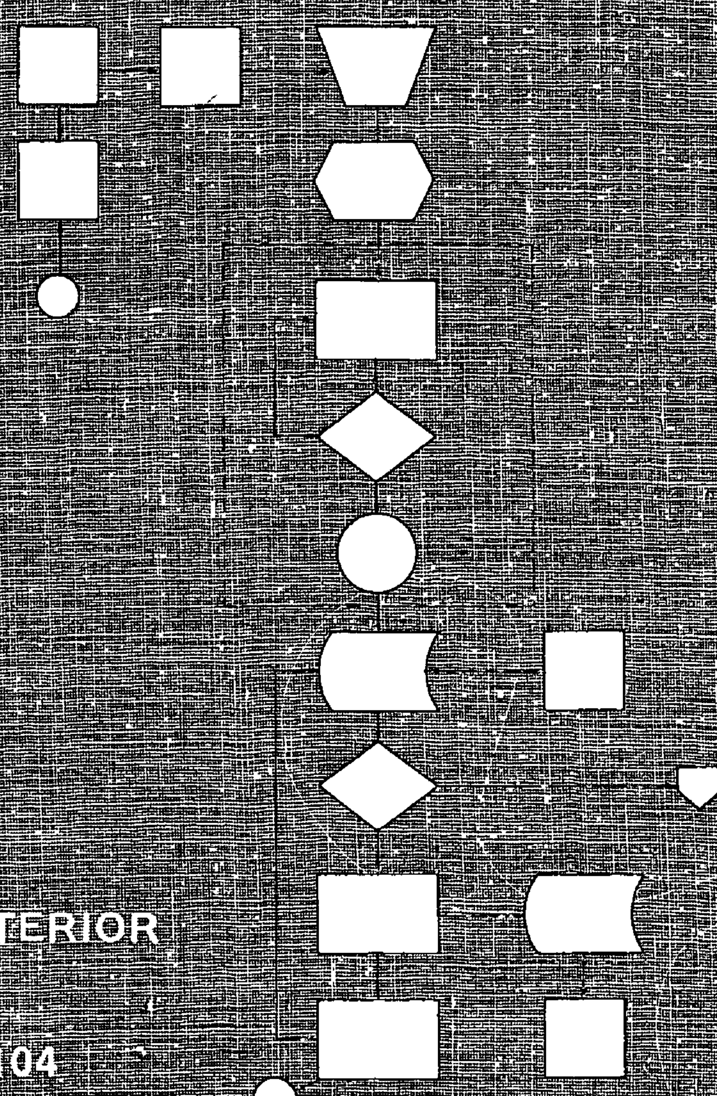
DATA-BASE SYSTEM FOR NORTHERN MIDWEST REGIONAL AQUIFER-SYSTEM ANALYSIS



PREPARED BY
UNITED STATES DEPARTMENT OF THE INTERIOR
GEOLOGICAL SURVEY

WATER-RESOURCES INVESTIGATIONS 80-104

50272-101

REPORT DOCUMENTATION PAGE

REPORT DOCUMENTATION PAGE	1. REPORT NO.	2.	3. Recipient's Accession No.
4. Title and Subtitle DATA-BASE SYSTEM FOR NORTHERN MIDWEST REGIONAL AQUIFER-SYSTEM ANALYSIS			5. Report Date January, 1981 6.
7. Author(s) A.L. Kontis and R.J. Mandle			8. Performing Organization Rept. No. USGS/WRI-80-104
9. Performing Organization Name and Address U.S. Geological Survey, Water Resources Division 1815 University Avenue Madison, Wisconsin 53706			10. Project/Task/Work Unit No. 11. Contract(C) or Grant(G) No. (C) (G)
12. Sponsoring Organization Name and Address U.S. Geological Survey, Water Resources Division 1815 University Avenue Madison, Wisconsin 53706			13. Type of Report & Period Covered Final 14.

15. Supplementary Notes

16. Abstract (Limit 200 words)

A computerized data-base system was developed to facilitate collection and use of large quantities of data for a model study of the Cambrian-Ordovician aquifer system in the Northern Midwest. Input to the data-base system consists of either point values or digitized contour maps of hydrogeologic data required by the model. Uniformly spaced model-node values are then computed from the discrete data by two-dimensional interpolation. Nonuniformly-spaced modal values may be obtained by fitting two-dimensional polynomials (bicubic splines) to surfaces formed by uniformly spaced points and solving for the surface values at the node locations.

The data-base system has the following attributes: 1) manual handling of data is minimized and machine handling of data is maximized, 2) given unequally spaced point data over the extent of the study area, model input arrays for any reasonable uniform node spacing can be rapidly computed, 3) accuracy of computed node values are generally compatible with accuracy and spatial distribution of point data, 4) a relatively large class of nonuniformly spaced node configurations can be computed, 5) data within files can be easily accessed and edited, and 6) the occurrence of data processing errors at various stages of data-base generation is monitored by machine contouring the computed grids.

17. Document Analysis a. Descriptors

*Data storage and retrieval, *Regional analysis, * Hydrogeology, Computer models, Mapping, Automation

b. Identifiers/Open-Ended Terms

*Data-base system, Digitization, Interpolation, Coordinate transformation, Contour mapping

c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
No restriction on distribution	UNCLASSIFIED	27
	20. Security Class (This Page) UNCLASSIFIED	22. Price

DATA-BASE SYSTEM FOR NORTHERN MIDWEST REGIONAL AQUIFER-SYSTEM ANALYSIS

A. L. KONTIS AND R. J. MANDLE

U. S. GEOLOGICAL SURVEY

Water-Resources Investigations 80-104

December 1980

UNITED STATES DEPARTMENT OF THE INTERIOR

CECIL D. ANDRUS, *SECRETARY*

GEOLOGICAL SURVEY

H. William Menard, *Director*

For additional information write to:

U. S. Geological Survey
1815 University Avenue
Madison, Wisconsin 53706

CONTENTS

ILLUSTRATIONS

CONVERSION TABLE

For the reader who may prefer to use metric units rather than inch-pound units, the conversion factors for the terms used in this report are listed below:

Multiply inch-pound units	By	To obtain metric unit
foot (ft)	3.048×10^{-1}	meter (m)
foot per day (ft/d)	3.048×10^{-1}	meter per day (m/d)
square foot per day (ft^2/d)	9.290×10^{-2}	square meter per day (m^2/d)
mile (mi)	1.609	kilometer (km)

DATA-BASE SYSTEM FOR NORTHERN MIDWEST REGIONAL AQUIFER-SYSTEM ANALYSIS

A. L. KONTIS AND R. J. MANDLE

ABSTRACT

The U.S. Geological Survey is conducting a study of the Cambrian and Ordovician aquifer system of the northern Midwest as part of a national series of Regional Aquifer-Systems Analysis (RASA). An integral part of this study will be a simulation of the ground-water flow regime using the Geological Survey's three-dimensional finite-difference model. The first step in the modeling effort is the design and development of a systematic set of processes to facilitate the collection, evaluation, manipulation, and use of large quantities of information. A computerized data-base system to accomplish these goals has been completed for the northern Midwest RASA.

The input to the data-base system consists of either point values or contour maps of the hydrogeologic data required by the model. Digital samples of contoured surfaces are obtained by machine digitization. Uniformly spaced model-node values are then computed from the discrete data by two-dimensional interpolation. The interpolator uses the regional (long wavelength) characteristics of the data to compute a surface with minimum total curvature. Local (short wavelength) data are merged with the regional surface to produce the model-node values at the desired spacing. Nonuniformly spaced nodal values may be obtained by fitting two-dimensional polynomials (bicubic splines) to surfaces formed by uniformly spaced points and solving for the surface values at the node locations. The uniformly or nonuniformly spaced node values constitute the output of the data-base system and in turn are the input arrays of data values for the ground-water flow model.

Management of the data base is facilitated by forming a data file for each model input parameter and for other hydrogeologic data used in the study. The data files are subdivided into elements consisting of individual model-layer arrays. In this form the data base can be readily accessed and edited. In addition to the interpolation programs, software components of the system include programs: (1) to transform geographic coordinates on a Lambert conformal conic projection to cartesian coordinates and vice versa, (2) to machine contour gridded data, (3) to expedite manipulation and editing of data, and (4) to perform various interarray computations.

The data-base system has the following attributes: (1) manual handling of data is minimized and machine handling of data is maximized, (2) given unequally spaced point data over the extent of a study area, model input arrays for any reasonable uniform node spacing can be rapidly computed, (3) accuracy of computed node values are generally compatible with accuracy and spatial distribution of point data, (4) a relatively large class of nonuniformly spaced node configurations can be computed, (5) data within data-base files can be easily accessed and readily edited, and (6) the occurrence of data processing errors at various stages of data-base generation is monitored by machine contouring the computed grids.

Functioning of the data-base system is illustrated by the sequence of steps required to produce a model-input array of transmissivity, given raw geologic and hydraulic conductivity data.

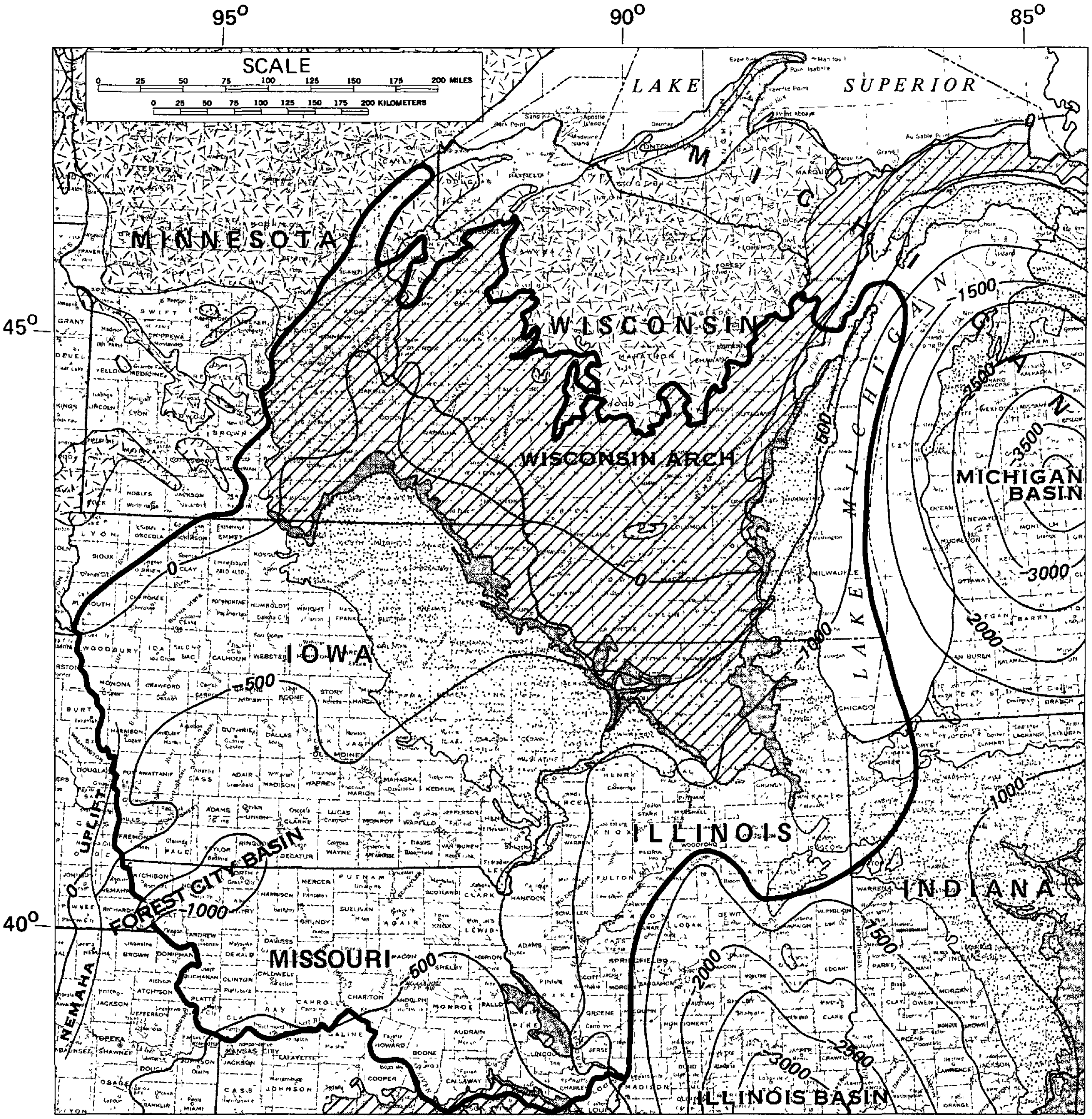
INTRODUCTION

The U.S. Geological Survey has begun a series of hydrogeologic investigations of major ground-water provinces within the United States. These investigations are part of the Geological Survey's Regional Aquifer-System Analysis (RASA) Program (Bennett, 1979). The primary goal of this program is to gain an understanding of the regional hydrogeology of each aquifer system, including the nature of the hydrogeologic units, ground-water flow system, and water-quality distribution. This information will form the basis for determining effects of future development on regional aquifer systems and thereby provide to State, local, and regional officials the information needed for efficient management of their ground-water resources.

In October 1978, a 4-year RASA study of the Cambrian and Ordovician aquifer system in the northern Midwest was begun (Steinhilber and Young, 1979). The study area (fig. 1) covers approximately 161,000 mi^2 including most of Wisconsin and Iowa, northern Illinois, northwestern Indiana, southeastern Minnesota, and northern Missouri.

A major part of the northern Midwest RASA effort will be the generation of several ground-water flow models. The three-dimensional finite-difference ground-water flow model of Trescott (1975), and modifications of it (Trescott and Larson, 1976), will be used to simulate the regional flow system and to estimate aquifer response to future development. To describe the regional flow system, a multilayer three-dimensional model with uniform node spacing is being developed for the entire study area (fig. 1). In addition, a relatively detailed model of the Chicago-Milwaukee area will also be developed.

An important part of the model-development effort is gathering the hydrogeologic data required by the model. These data, constitute the model data base. Because of the large areal extent of the study area and the multilayer nature of the model, large quantities of information are involved. To minimize time and effort expended on development and handling of the data, a data-base system has been developed for the northern Midwest RASA. The data-base system will provide a means by which data, in various forms, can be readily compiled, accessed, manipulated, and edited. In addition to data needed for model development, any hydrogeologic data gathered during this study, such as water-quality or geologic data, can be stored, accessed, edited and displayed.



Base from U. S. Geological Survey
U. S. base map, 1:2,500,000

Geology modified from P. B. King and H. M. Beikman (1974). Structure from P. B. King (1969)

EXPLANATION

- Mississippian and younger formations
- Silurian and Devonian formations *Mostly dolomite*
- Ordovician Maquoketa Shale
- Ordovician and Cambrian formations *Mostly sandstone and dolomite*
- Precambrian sandstone
- Precambrian crystalline rock
- STUDY AREA BOUNDARY
- —500— STRUCTURE CONTOUR *Shows altitude of top of Precambrian crystalline basement. Contour interval 500 meters. National geodetic vertical datum of 1929*

Figure 1. Location and general bedrock geology of the northern Midwest regional aquifer study area.

The authors wish to acknowledge the assistance of Mr. James J. Frawley, formerly of the U.S. Naval Oceanographic Office and presently with the Phoenix Corp., who wrote the Lambert transformation computer programs, and Mr. William E. Rankin, of the U.S. Naval Oceanographic Office, who wrote the contouring program.

DATA-BASE SYSTEM

Definition of system

The data-base system is defined as a set of processes which are applied to hydrogeologic data, in a relatively raw form, to produce refined data which can be directly inserted into the ground-water flow model. The raw data consist of irregularly spaced point values over the areal and vertical extent of the study area. The computer-model input consists of two-dimensional arrays the elements of which represent values of the geohydrologic properties of model finite-difference blocks. With the exception of pumping nodes, the raw data for a particular model parameter and layer may be considered as discrete samples from a continuous surface, which ideally are of sufficient quantity, quality, and distribution to define the essential characteristics of the surface. The nucleus of the data-base system is application of a suitable two-dimensional interpolation method to the raw point data to produce values of the surface at any appropriate node spacing. Other aspects of the system facilitate this process and allow for efficient management of the data base.

The data-base system will be described in terms of system flow diagrams (figs. 2 and 3) and illustrated by presenting the various steps involved in computing a transmissivity model-input array of the Mount Simon Sandstone of Cambrian age, the lowermost unit of the Cambrian and Ordovician sedimentary sequence. This prototype array is developed for an assumed uniform node spacing of 8 mi. The maps presented in this report are based on preliminary data and are not to be taken as being definitive. These maps are presented only to facilitate explanation of the various steps in the data-base system.

Acquisition of data

The creation of the data base is carried out by a central RASA staff located in Madison, Wis., working with project chiefs from each of the Geological Survey's district offices located in the six states included in the study area. The first and most important step in the process is undertaken at the state level, wherein all available hydrogeologic data in each state are assembled and analyzed. The rawest form of the data for a particular parameter is usually a nonuniformly spaced set of point values. Given the large area covered by the regional model there will generally be high variability in data density including large gaps in data coverage. Commonly, such data gaps are filled-in manually by interpretive contouring if sufficient data exists. In this approach, the map compiler's knowledge of the local hydrogeology together with the observed data can, in many cases, result in a realistic representation of the parameter surface. Alternatively, computerized methods are also available to fill in data gaps. These methods include, two-dimensional least squares, two-dimensional polynomial interpolation, and various statistical estimation techniques. The applicability of these methods will depend on the mathematical or statistical

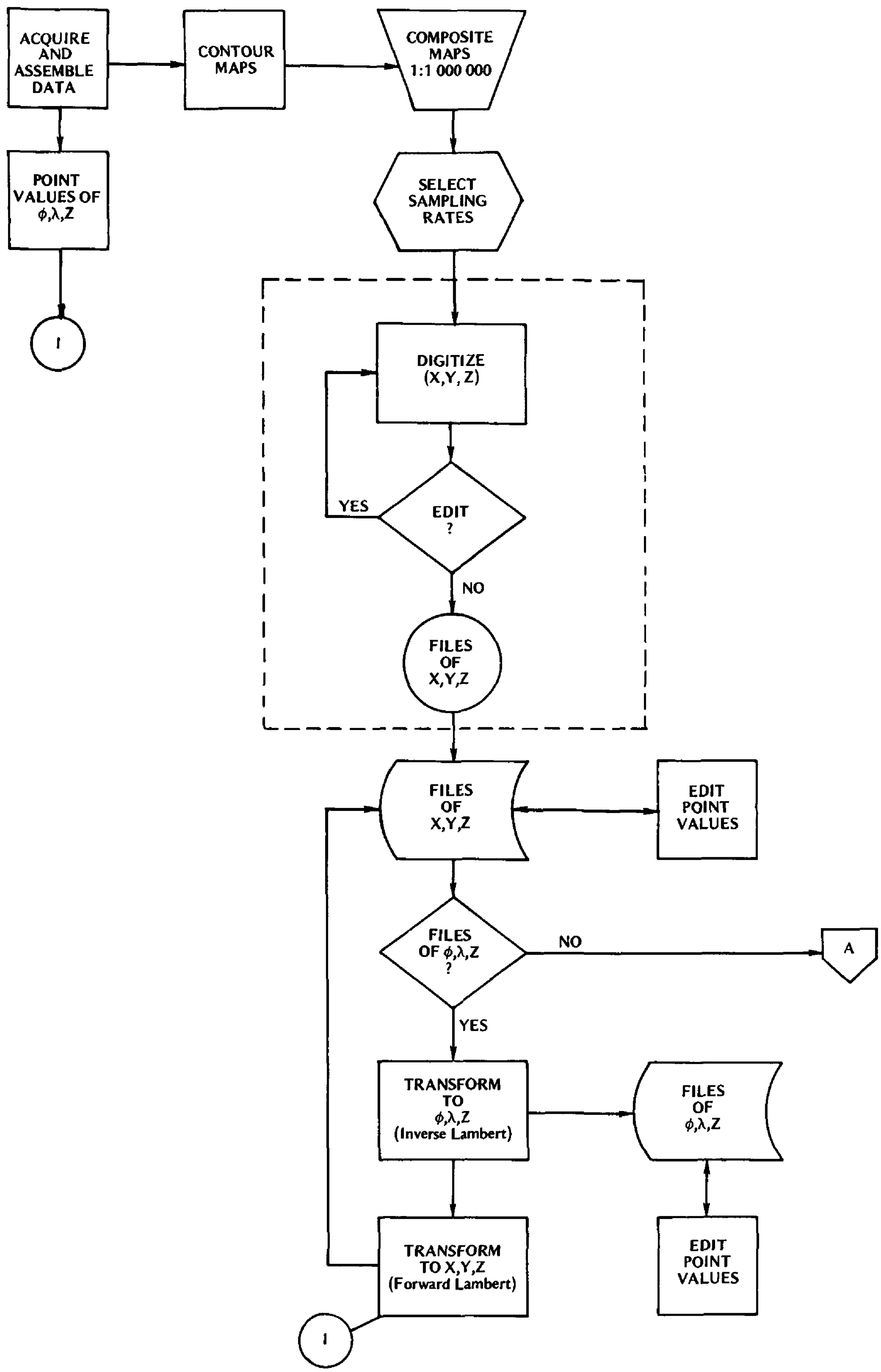


Figure 2. Flow diagram of the data-base system, - Part 1.

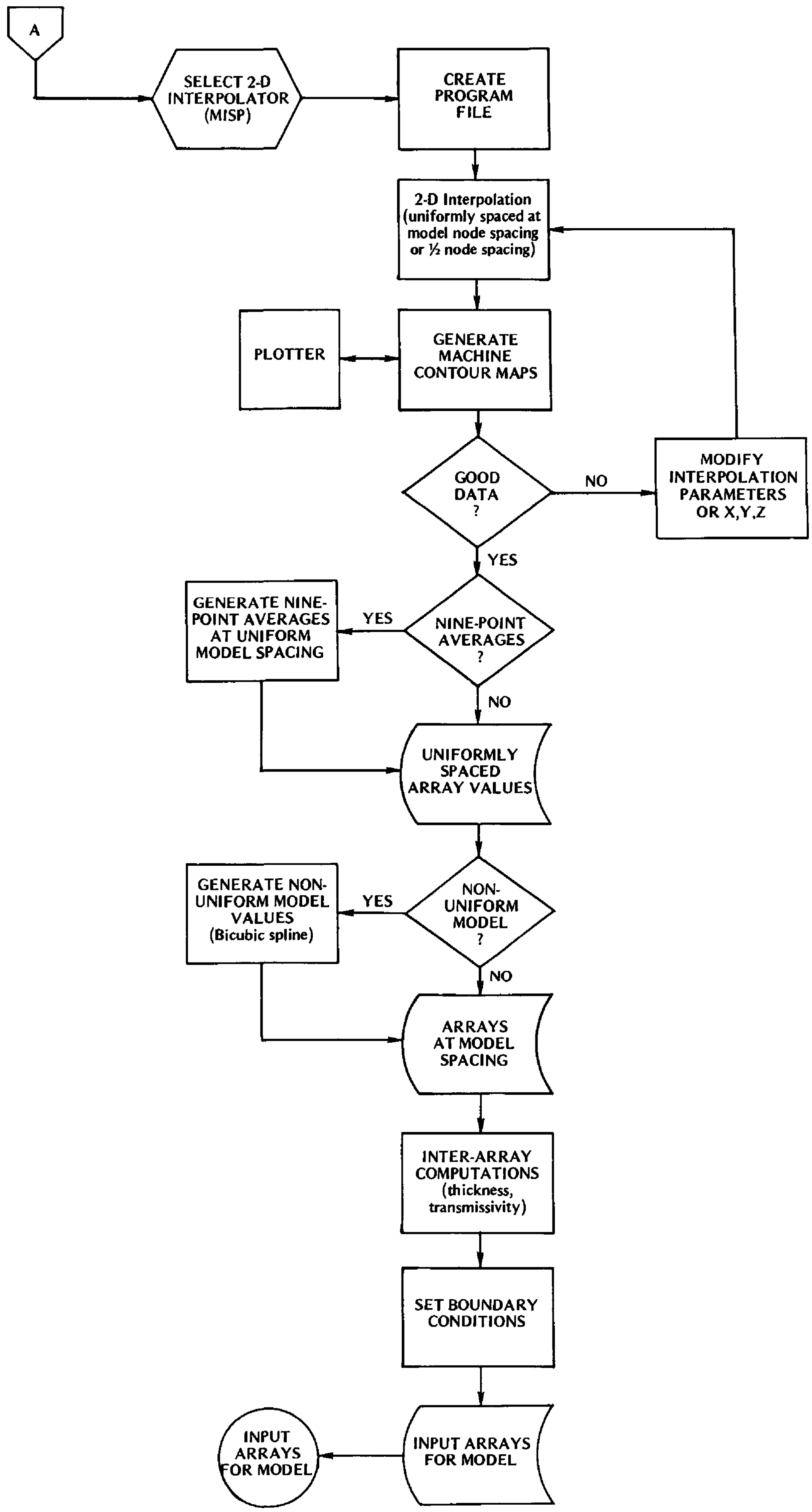


Figure 3. Flow diagram of the data-base system, - Part 2.

assumptions of the methods, the amount and distribution of the observed data and the required accuracy of the data. The data received by the central RASA staff from each state is either in the form of manually derived or computer generated contour maps of a given parameter or as sets of point values of the parameter in that state.

Composite maps

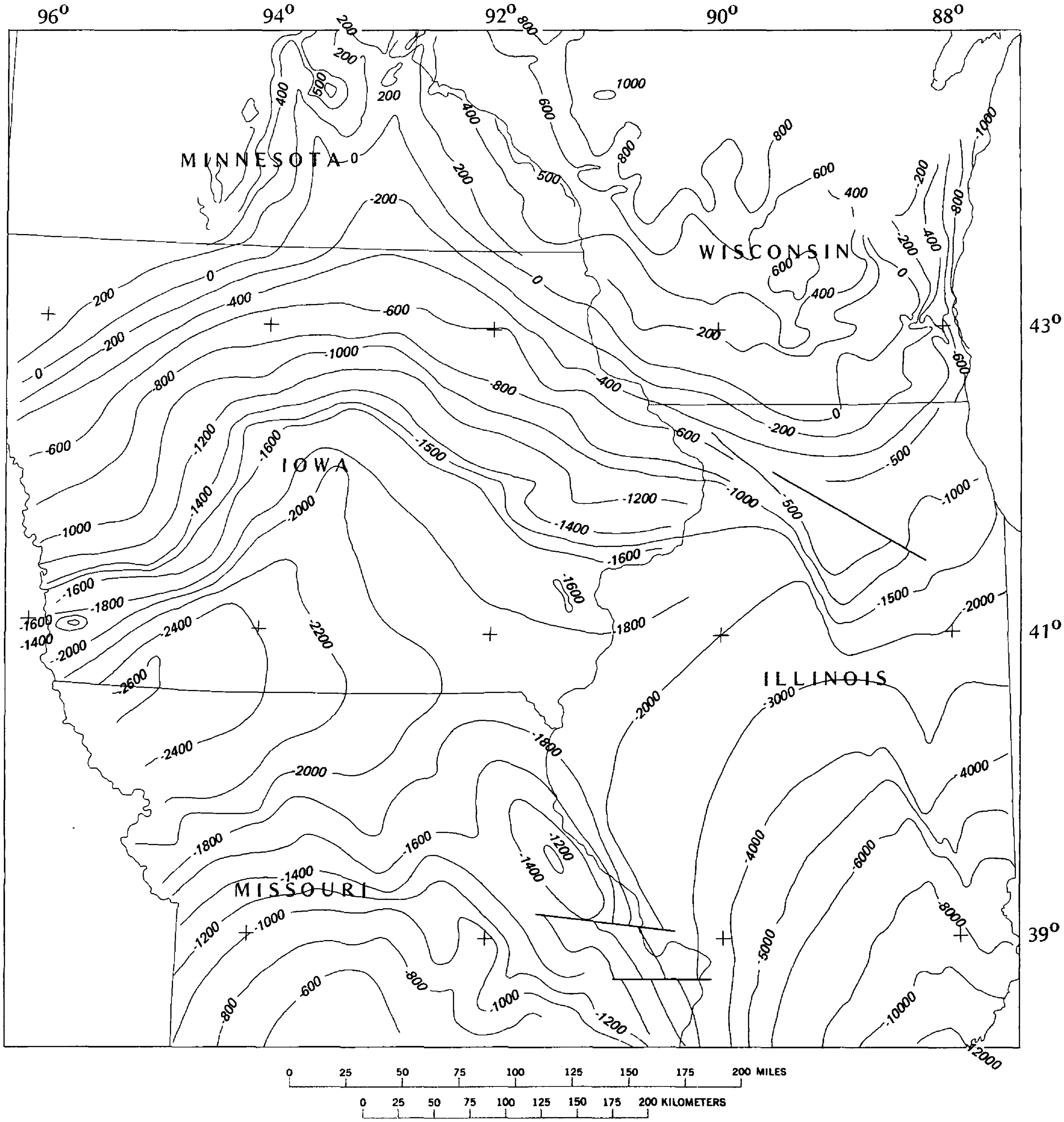
Assuming receipt of contour maps of a particular type of data from each state, the next step in the process is to compile a composite map of the entire region by meshing individual state maps (fig. 2). The primary purpose of the composite map is to provide data in a form that can be readily digitized. The digitized data in turn represent input data to a mathematical interpolation process. In generating the composite map, it is, in some cases, necessary to adjust contour lines in the vicinity of state boundaries to ensure data compatibility across the boundaries. The composite maps for the regional data base are at a scale of 1:1,000,000, on a Lambert conformal conic projection. An example is a map of the structure of the top of the Mount Simon Sandstone (fig. 4). For the sake of clarity only part of the structure map for the entire study area has been shown. Note on figure 4, the contour lines that are truncated either signify faulting or represent areas where there is enough information (contour lines) where the surface configuration can be adequately determined by mathematical interpolation.

Selection of sampling rates

The composite maps are digitized along contour lines. Before digitization a determination is made as to which contours are to be sampled as well as the sampling rate along each contour line (fig. 2). The determination of sampling rates is basically an attempt to sample a surface with enough points to sufficiently represent computationally relevant features while at the same time minimizing the total number of points, thereby reducing the costs of digitizing and subsequent data processing.

The factors which govern sampling rate sufficiency and relevancy of features include the type of interpolator being applied and the model node spacing. In general, most interpolation methods require only a few suitably distributed points to define accurately features with uniform gradients. A larger number of points are required to represent accurately features with a higher degree of curvature. The model node spacing in turn determines which features of the composite map are of computational significance in the model. Considering the surface map to be composed of the superposition of individual features of different wavelengths (as in a Fourier analysis), those wavelengths shorter than twice the node spacing are effectively removed by the interpolation process. For example, if a model consists of a uniform node spacing of 8 mi, then only continuous surface features with wavelengths greater than 16 mi will be discernible. Likewise, for a finer-spaced model, shorter wavelengths will be of significance.

With these considerations in mind, figure 4 shows the contour lines selected for digitization of the altitude of the top of the Mount Simon Sandstone. The sampling interval along each contour line is 4 mi (0.25 in. at the map scale of 1:1,000,000) except in areas containing contour closures, where the sampling rate is increased to 1.6 mi (0.1 in. at map scale).



EXPLANATION

——— 200 ———

Structure contour

Shows altitude of top of Precambrian basement. Contour interval variable. National geodetic vertical datum of 1929

———————

Fault

Figure 4. Altitude of top of Mount Simon Sandstone (manually derived).

Digitization

Early in the planning stages of the data-base system an analysis was made of the advantages and disadvantages of purchasing or leasing digitizing equipment, purchasing or developing software, and "contracting out" the entire digitizing operation. On the basis of savings in cost and time, it was decided that the digitizing would be contracted to the University of Wisconsin Cartographic Laboratory which has in place all of the hardware and software required to digitize efficiently and accurately. Probably the largest benefit that can result by contracting the work is that the RASA staff is able to devote their time and effort to other aspects of the modeling process.

The Cartographic Laboratory digitizing system consists of a Bendix Data Grid-1 digitizer linked to a PDP-11/34 minicomputer.[1] Initially, digitized data are placed on a disk drive and interactively manipulated and edited with software developed by the Cartographic Laboratory. To check the generated data, a crude plot of each digitized contour line by straight line connection of the digitized points is made. Comparison of the crude line plot with the original map reveals the existence of errors, which can readily be corrected. After the errors are corrected the digitized data are placed on magnetic tape as a set of X,Y,Z values (fig. 2), where X and Y are horizontal and vertical distances in inches from an arbitrary map origin and Z is the contour value. For the Mount Simon structure (fig. 4), the selected sampling rates resulted in a total of 2,830 digitized X,Y,Z values.

Data-base files

On the basis of accessibility and efficiency, all of the data-base system processing is done on the University of Wisconsin UNIVAC 1100/82 computer. The form of the data base for computational purposes consists of a data file for each type of data, for example, geologic structure. The data for each model layer is in turn a data element within that data file. The relationship of the data elements to the data files is shown in figure 5. The structure top of each formation or groups of formations is stored as a separate element in the geologic-structure data file. The data files are part of the computer mass storage (magnetic drum) and can be easily accessed and operated on by use of standard editing processors so that additions, deletions, and modifications of the data-base elements can be readily effected. Thus, the digitized Mount Simon Sandstone structure is placed in the data base and resides as a data element of the geologic structure data file (fig. 2).

Coordinate transformations

The next step in the system processing is to decide whether to maintain data files in both X,Y,Z form and parallel data files in latitude (ϕ), longitude (λ), Z form. One reason for maintaining parallel files is that it may be desirable to have the data referenced to a nonarbitrary coordinate system that will be useful for future applications, that is, in latitude and longitude. In addition, throughout the RASA study there will be an ongoing data-collection effort. When new or updated data in ϕ,λ,Z form become available, it may be useful to have files in which the data can be directly inserted.

---

[1] The use of trade names in this report is for identification only and does not imply endorsement by the U.S. Geological Survey.

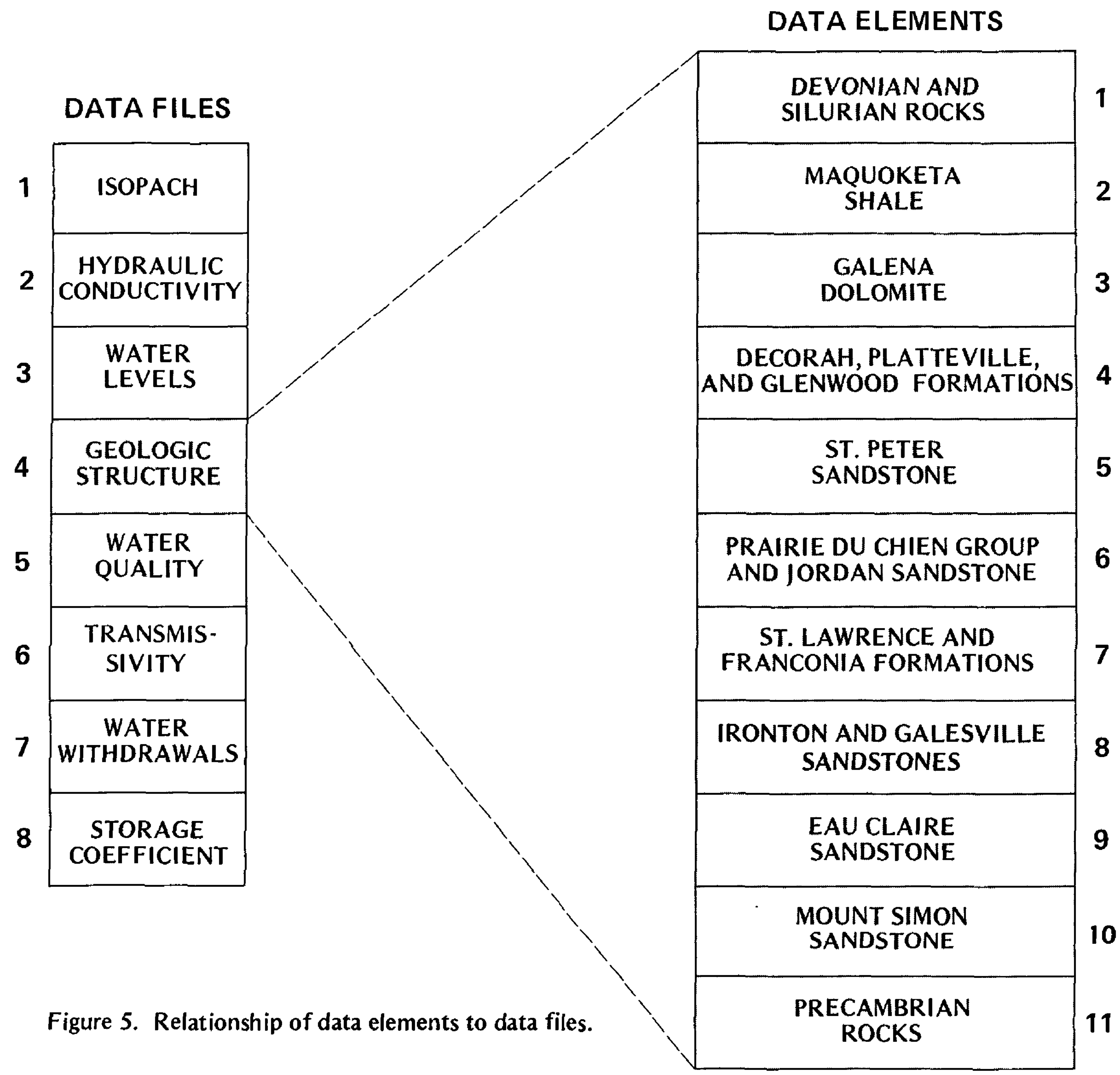


Figure 5. Relationship of data elements to data files.

To convert the data files in X,Y,Z to ϕ,λ,Z a program termed the Inverse-Lambert transformation (J. J. Frawley, written commun., 1979) is applied (fig. 2). This program uses X,Y distances from the map origin to three geographic reference points located on the standard parallels of the Lambert projection together with the eccentricity and equatorial radius of the reference spheroid to compute ϕ and λ for any other points. The transformation is computed from Lambert mapping equations developed by Adams (1918) and Thomas (1952).

On the other hand, if the original data are point values in ϕ,λ,Z, rather than a contour map, then it is necessary that these data be converted to X,Y,Z because the interpolation process requires data in distance form. For example, the areal distribution of the point values of hydraulic conductivity of the Mount Simon Sandstone is insufficiently known at present to produce a contour map manually. Consequently, the conductivity data is in ϕ,λ,Z point value form at the beginning of the data-base flow (fig. 2). In preparation for gridding and subsequent computation of the transmissivity of the Mount Simon Sandstone, the conductivity values must be transformed from ϕ,λ,Z to X,Y,Z. Another application of this transformation is in the determination of the model-node locations of discharge or recharge wells. The ϕ,λ well coordinates may be converted to X,Y form, thereby locating the corresponding model-node location.

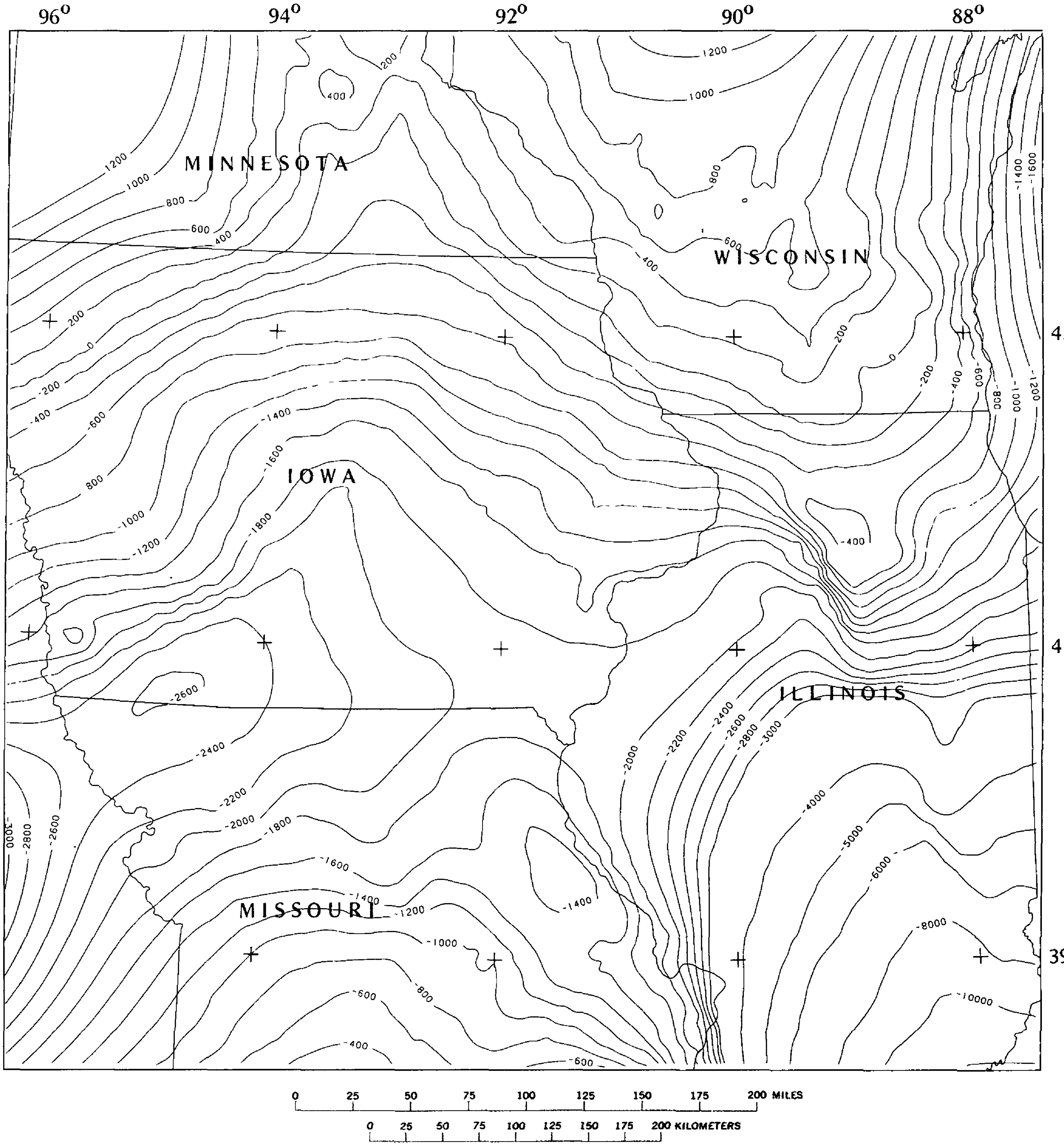
The change from geographic to cartesian coordinates is implemented by a program termed the Forward-Lambert transformation (J.J. Frawley, written commun., 1979) (fig. 2). Here the geographic coordinate of the desired map origin, Lambert mapping equations, (Adams, 1918; and Thomas, 1952) and reference spheroid parameters are used to calculate the transformation.

Both programs were tested for accuracy by first applying the Inverse-Lambert transformation to the digitized Mount Simon Sandstone structure X,Y values to produce values in ϕ,λ. The Forward-Lambert process was then used to transform back to X,Y. Comparison of the original and computed sets resulted in X,Y differences in the order of hundredths of an inch (at the scale of 1:1,000,000), which are sufficiently accurate for the purposes of this study.

Selection of two-dimensional interpolation method

As previously mentioned, various techniques exist with which uniformly spaced, two-dimensional grids can be generated from unequally spaced data. In general there is no single technique that will be appropriate for all distributions of input data and for all applications. Consequently, an analysis should be made of available methods with regard to the nature of the input data and the required accuracy of the gridded output data (fig. 3). At present, the technique selected for the northern Midwest RASA is a modification of a method developed by Briggs (1974) and implemented by Swain (1976). The method termed MInimum curvature SPline (MISP) is based on the premise that where there are large data gaps on a surface, regional or long wavelength characteristics of the given data should be used to fill in the gaps. By this approach the probability of introducing large magnitude interpolation errors is reduced.

In the MISP method, the regional characteristics of the input data are derived by a series of averaging processes. Suppose that the desired output spacing of the interpolation process is 4 mi. For a particular grid cell, input data within a 4x4-mi area are weighted as to distance from the grid point and averaged. This average value is assigned to the grid point. A coarser grid spacing, usually three times that of the desired output spacing, is now specified. Over the 12x12-mi extent of the coarse grid, values are developed by averaging the above averages, thereby smoothing the short wavelength components of the input data. An algorithm based on the minimum total curvature criteria is then used to compute a regional surface at the 12-mi grid spacing. The minimum total curvature property ensures that this computed surface is the smoothest surface that can be fit to the given data. To obtain regional values at the desired 4-mi output spacing, a one-dimensional spline interpolation is used on the coarse regional surface points first in the X direction, then in the Y direction. The program then merges this relatively smooth surface with the



EXPLANATION

-4000

Structure contour

Shows altitude of top of Mount Simon Sandstone. Contour interval 200 feet except southern Illinois with variable contour interval. National geodetic vertical datum of 1929

Figure 6. Computer-generated map showing altitude of top of Mount Simon Sandstone at 4-mile node spacing (interpolated from digitized data).

original data. For cells in which input data exist, the regional surface values are modified to reflect the original data. Where no input data exist, the regional surface values are used. The final result is a surface which reflects the shorter wavelength information inherent in the input data, but which is relatively smooth in regions lacking data.

Creation of program file

To facilitate program modifications and maintenance, all software which is used to implement the data-base system is placed in the computer as elements of a program file (fig. 3). Included in the file are the two-dimensional interpolator (MISP), contour program, interpolation program to compute values at unequally spaced node positions (bicubic spline), and the coordinate transformation programs (Inverse and Forward Lambert). In addition, the file contains an assortment of programs to manipulate and edit arrays, assign model-node locations to pumping centers, compute model-unit thickness from geologic structure maps, and compute transmissivity from thickness and hydraulic conductivity maps.

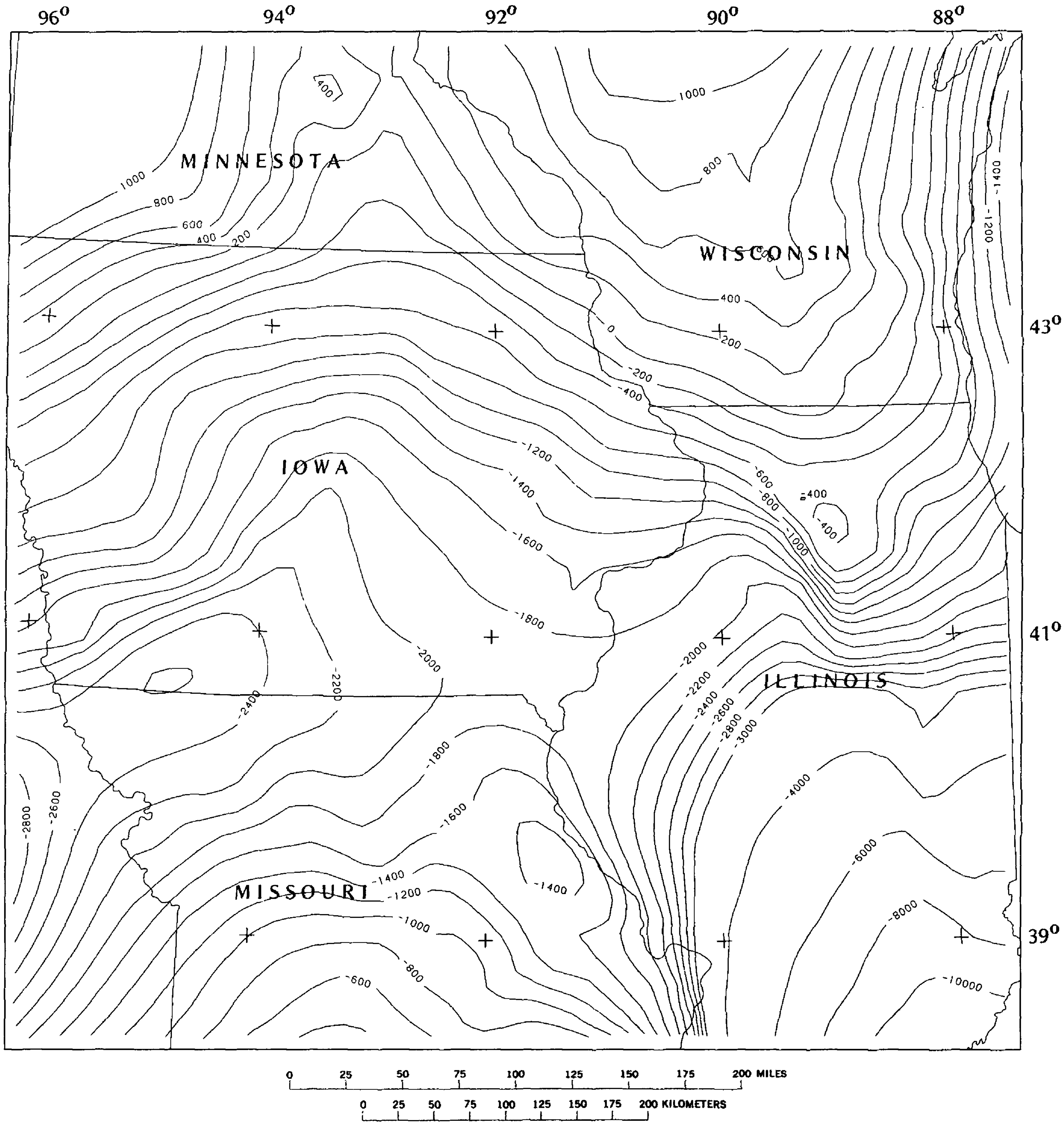
Two-dimensional interpolation

To illustrate application of the MISP interpolation method, the digitized Mount Simon Sandstone structure data was interpolated at a spacing of 4 mi and machine contoured (fig. 6) using a program by W.E. Rankin (written commun., 1980). Comparison of this contour map with the original contour map (fig. 4) shows that the interpolated surface replicates the regional characteristics of the original surface, whereas local features in the order of several miles are greatly attenuated in amplitude. The smoothness attribute of the MISP technique is evident in those areas where there are no data, that is, in the northwest, northeast, and southwest corners of the map.

By comparing machine-generated contour maps with the manually derived composite maps, it can be determined whether the essential features of the original surface appear in the gridded data (fig. 3). If not, appropriate modifications are made to the X,Y,Z data file and the interpolation process is repeated until a satisfactory gridded data set is achieved. Alternatively, the gridded data may be directly edited to reflect the desired changes.

Uniformly spaced grids

In the three-dimensional finite-difference model, a specific node value of a given parameter is assigned on the assumption that the nodal block has uniform hydrogeologic properties. Consequently, the parameter value of the node should be an averaged value over the extent of the block. To generate the model arrays for a desired node spacing for the case where there is significant spatial variability over the dimensions of the blocks, the MISP interpolation process may initially be applied to produce an equally spaced grid at one-half the desired node spacing. The nodal values are then obtained by successive 9-point averages of the gridded data (fig. 3). For the Mount Simon Sandstone structure the data were interpolated at a 4-mi spacing (fig. 6) so that the averaging process produces a model array with node spacing of 8 mi (fig. 7). The results of a similar process applied to the Precambrian basement surface is given in figure 8.



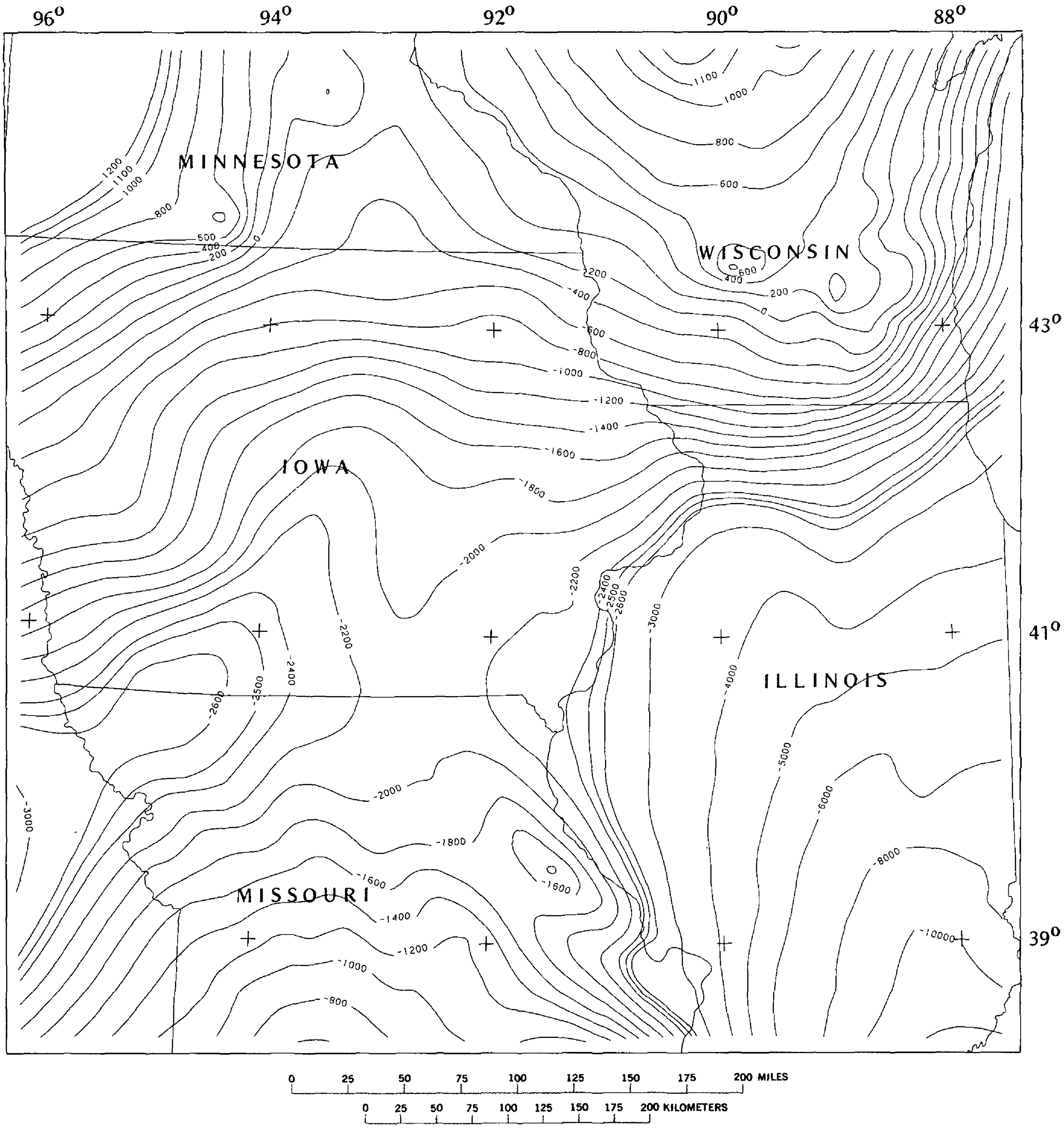
EXPLANATION

—— -200 ——

Structure contour

Shows altitude on top of Mount Simon Sandstone. Contour interval variable. National geodetic vertical datum of 1929

Figure 7. Computer-generated map showing altitude of top of Mount Simon Sandstone at 8-mile node spacing (9-point average of data gridded at 4 miles).



EXPLANATION

-2200

Structure contour

Shows altitude of top of Precambrian basement. Contour interval variable. National geodetic vertical datum of 1929

Figure 8. Computer-generated map showing altitude of top of Precambrian basement at 8-mile node spacing (9-point average of data gridded at 4-miles).

In cases where the data are known only at widely spaced points or where there is little spatial variability, the above process is probably unnecessary. The model arrays may then be obtained directly by setting the interpolation interval at the model node spacing. An example is the hydraulic conductivity of the Mount Simon Sandstone where at present only a few discrete points are available. These data were interpolated directly with the MISP technique to produce gridded values at an 8-mi spacing (fig. 9).

Nonuniformly spaced grids

If a nonuniform model grid is desired, then an additional process may be applied to the equally spaced points to produce values of the surface at nonuniform node locations (fig. 3). The technique selected to implement this process is bicubic-spline interpolation (de Boor, 1962) and is a modification of a program by Davis and Kontis (1970). Initially a two-dimensional cubic polynomial, f(x,y) (fig. 10), is determined for each cell as defined by sets of four points of the equally spaced grid. In the expression of f(x,y) (fig. 10), the $C_{m,n}^{i,j}$ are a set of 16 bicubic coefficients that describe the surface for the i,jth cell. Given the bicubic coefficients, the value of the surface can then be solved for any point (x,y) within the cell. The programing required to compute values of the surface at any arbitrary spacing can be tedious. If, however, a certain class of unequally spaced nodes is considered, the programing is relatively easy.

In particular, consider a set of cells with dimensions of $\Delta X \cdot \Delta Y$ where $\Delta X = \Delta Y$ and specify that each individual cell be subdivided forming blocks of size $(\Delta X/M \cdot \Delta Y/N)$ where M and N are integers. For example in figure 10, the initial equidimensioned cells are subdivided in the X direction by the sequence $M = \{1,1,1,2,4,3\}$ and in the Y direction by $N = \{4,3,2,1,1,1,1\}$ to produce the given set of unequally spaced model blocks. Solving the bicubic polynomials at the center of each block produces the node values. Alternatively, if there is significant spatial variability in the parameter, such that the interpolated point value would differ significantly from an average value, the bicubic polynomial may be formally integrated over the extent of each new block to produce the average node values.

A disadvantage of computing values for nonuniform nodes with spacing less than the equally spaced points is that, as previously discussed, the equally spaced gridding process filters out shorter wavelengths that may exist in the original data so that the unequally spaced node values are generated from a relatively smooth surface. If the shorter wavelength information is required, the initial equally spaced gridding may be set at some finer spacing and then an appropriate algorithm may be applied to obtain the nonuniform node values.

Interarray computations

Final interarray computations when necessary are performed (fig. 3) as a concluding aspect of the data-base system. In particular, given the structure tops and hydraulic conductivity of each model layer, model layer thickness and transmissivity may be determined. In the prototype Mount Simon Sandstone model-layer computation, the Mount Simon and Precambrian structural tops have been gridded (figs. 7 and 8) along with the hydraulic conductivity in the Mount Simon

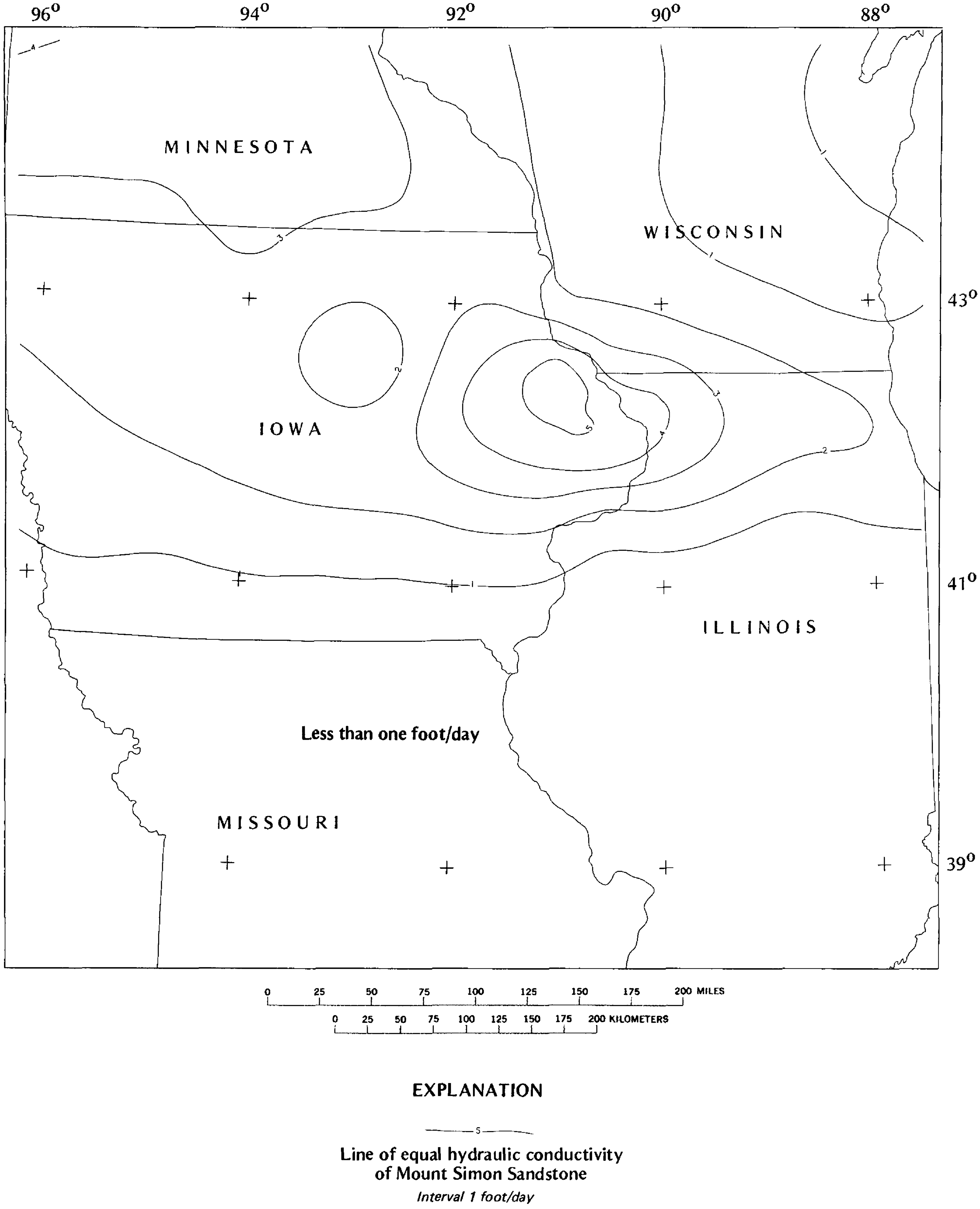


Figure 9. Computer-generated map showing hydraulic conductivity of Mount Simon Sandstone at 8-mile node spacing (interpolated from point data).

$$f(x,y)_{j,i} = \sum_{m=o}^{3} \sum_{n=o}^{3} C_{m,n}^{j,i} (x-x_j)^m (y-y_i)^n$$

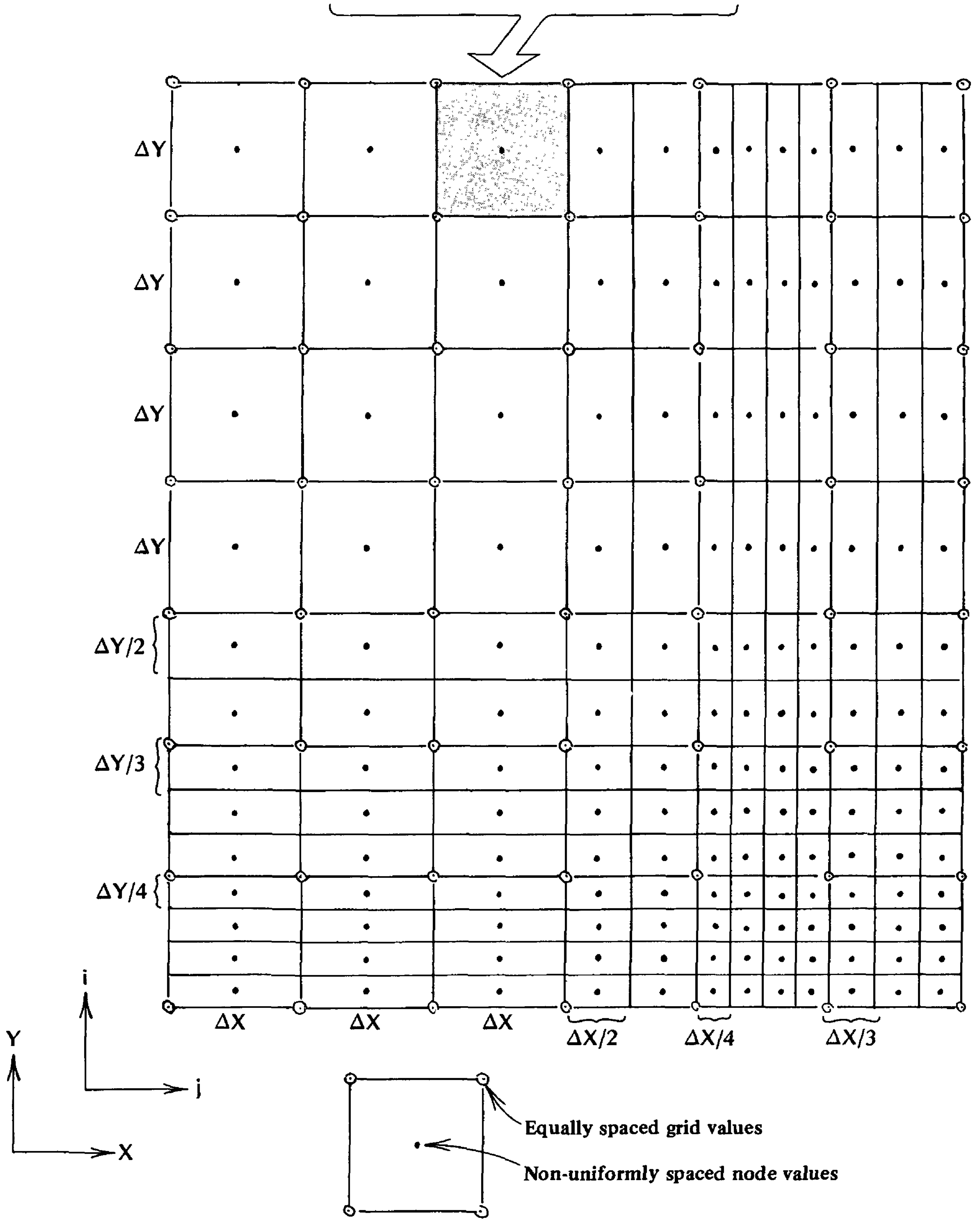


Figure 10. Bicubic-spline method for generating non-uniformly spaced model node values.

(fig. 9). From these data, the Mount Simon Sandstone thickness (fig. 11) and transmissivity (fig. 12) are then readily computed. The hydraulic conductivity (fig. 9) and subsequent transmissivity map shown on figure 12 are based on a few preliminary values and are given here only to illustrate the functioning of the data-base system.

Boundary conditions

A final step in preparing data for input to the model is implementation of model-boundary conditions. In general, the areal model boundaries of a region will not coincide with the rectangular interpolation boundaries. The interpolation process generates a value at the specified node spacing over the extent of the rectangular grid boundary. Consequently, nodes outside the model boundaries are set to appropriate boundary condition values (fig. 3).

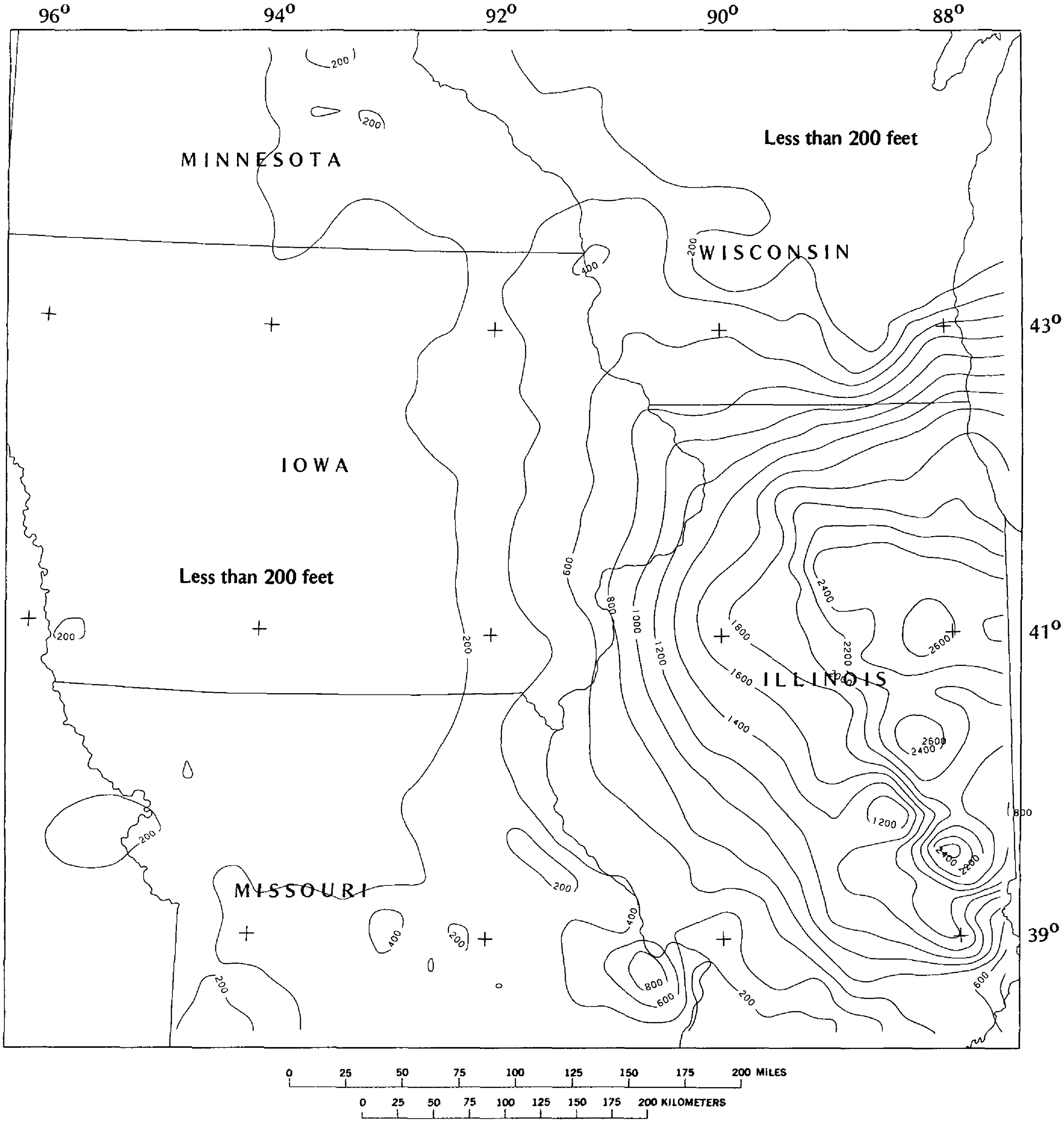
Model input arrays

Application of the data-base system to all required model parameters (fig. 5) results in array sets that are in the model program input formats. Parameter arrays that are spatially continuous may be obtained by implementing those parts of the system that were used to create the prototype Mt. Simon Sandstone arrays. The water withdrawal input is formed by determining the array positions of individual wells by coordinate transformation and summation of discharge rates of wells within each nodal block. When a model run is to be made, a temporary mass storage data file is produced by inserting appropriate elements from the data-base files (fig. 5). Upon execution, the model input data is read from the temporary file.

The various phases of the data-base system have been presented as a sequence of steps. To expedite generation of the data base, several of these steps may be incorporated into a single process. For example, two-dimensional interpolation (fig. 6), 9-point averaging (fig. 7), and setting nodes outside the model boundaries to suitable values may be treated as a single step. In addition, by applying the system directly to isopachs rather than structure contours, model node thickness values may be obtained more efficiently.

SUMMARY AND CONCLUSIONS

The northern Midwest RASA data-base system has been designed on the basic concept that, given a sufficient amount of data, model-node values can be generated by a suitable computer-based interpolation process. In particular, model-node values can be determined with an accuracy commensurate with the accuracy and spatial distribution of the original data and accuracy and node spacing of the model. Starting with point data or digitized data from contour maps, equally spaced grid points are computed by two-dimensional interpolation. For data with high spatial variability the grid spacing is half that of the model-node spacing. Model-node values are then obtained by successive 9-point averages of the computed data. For data with little spatial variability, uniform node values are computed directly at the model-node spacing. Machine contouring of gridded data serves as a quality control. If significant errors occur, they can be readily detected and corrected. For nonuniformly spaced models, a method based on fitting bicubic splines to equally spaced data is used to compute values for a large class of node configurations. Point data coordinates exist in terms of latitude and longitude or in distance units from a map origin. Conversion from one system to the other is implemented by coordinate transformations based on the Lambert conformal conic projection. By construct-



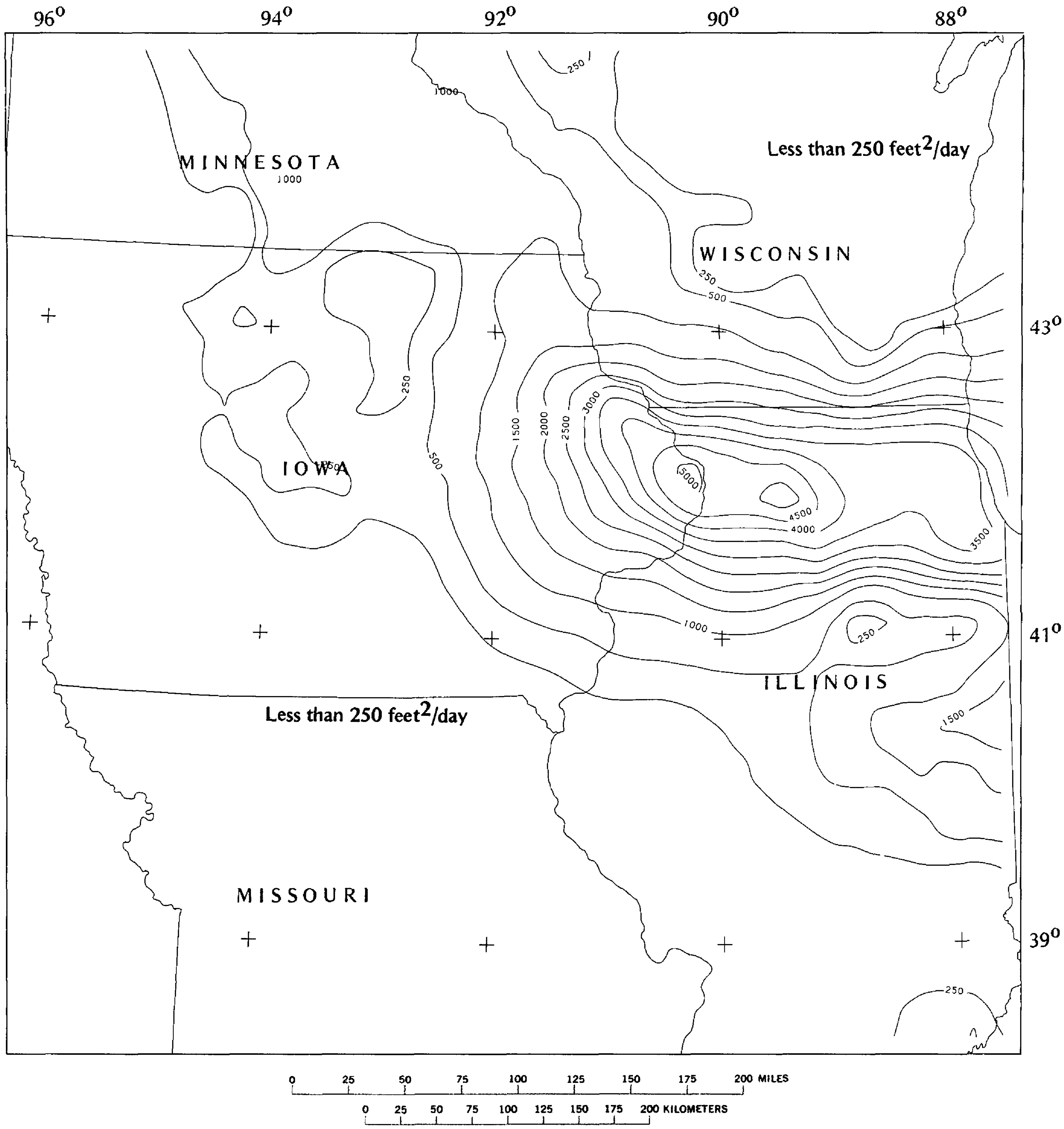
EXPLANATION

600

Line of equal thickness of
Mount Simon Sandstone

Interval 200 feet

Figure 11. Computer generated map showing isopach of Mount Simon Sandstone at 8-mile node spacing.



EXPLANATION

—— 1000 ——

Line of equal transmissivity of Mount Simon Sandstone

Interval 500 feet²/day with supplemental 250 feet²/day

Figure 12. Computer generated map showing transmissivity of Mount Simon Sandstone at 8-mile node spacing.

ing a data file for each model parameter as part of the computer's mass storage, elements of the data files can be rapidly accessed and efficiently managed.

Although the discussion of the data-base system concerned itself with model data input, other data, such as water-quality data may be readily processed using this data-base system.

The basic software used in the system including the MISP and bicubic-spline interpolators, coordinate transformation, and contouring programs were in existence before the development of this data-base system. Nevertheless, the programs required differing degrees of modification and extension, and considerable time and effort were required to design, implement, and test the entire system. Each finite-difference ground-water model is unique with regard to areal extent, number of model layers, and node configurations. The problem of rapidly developing model-input arrays from discrete point data is, however, common to all modeling. It is hoped that the conceptual approach and software elements presented here may reduce time and effort expended on data-base system development of future projects. Although the system was developed for use on a UNIVAC-1100/82, with minor variations, it can be implemented on any mainframe computer.

REFERENCES

Adams, O.A., 1918, General theory of the Lambert Conformal conic projection: U.S. Dept. of Commerce, Coast and Geodetic Survey Special publication, No. 53, 38p.

Bennett, G. D., 1979, Regional ground-water systems analyses: Water Spectrum, V.11, No.4.

Briggs, I. C., 1974, Machine contouring using minimum curvature: Geophysics, v. 39, no. 1, p. 39-48.

Davis, T. M., and Kontis, A. L., 1970, Spline interpolation algorithms for track-type survey data with application to the computation of mean gravity anomalies: U.S. Naval Oceanographic Office Technical Report 226, 50 p.

de Boor, C., 1962, Bicubic spline interpolation: Journal of Math and Physics, v. 41, no. 1, p. 212-218.

King, P.B., compiler, 1969, Tectonic maps of North America: U.S. Geological Survey map, 1.5,000,000.

King, P.B., and Beikman, H.M., compilers, 1974, Geologic maps of the United States: U.S. Geological Survey map, 1:2,500,000.

Steinhilber, W. L., and Young, H. L., 1979, Plan of study for the northern Midwest Regional Aquifer System Analysis: U.S. Geological Survey Water-Resources Investigations 79-44, 20 p.

Swain, C. J., 1976, A FORTRAN IV program for interpolating irregularly spaced data using the difference equations for minimum curvature: Computers and Geosciences, v. 1, no. 4, p. 231-240.

Thomas, P.D., 1952, Conformal projections in Geodesy and Cartography: U.S. Dept. of Commerce, Coast and Geodetic Survey Special Publication, No. 25,

Trescott, P. C., 1925, Documentation of finite-difference model for simulation of three-dimensional ground-water flow: U.S. Geological Survey Open-File Report 75-438.

Trescott, P. C., and Larson, S. P., 1976, Documentation of finite-difference model for simulation of three-dimensional ground-water flow: Reston, Va., U.S. Geological Survey Open-File Report 76-591. (Supplement to Open-File Report 75-438)