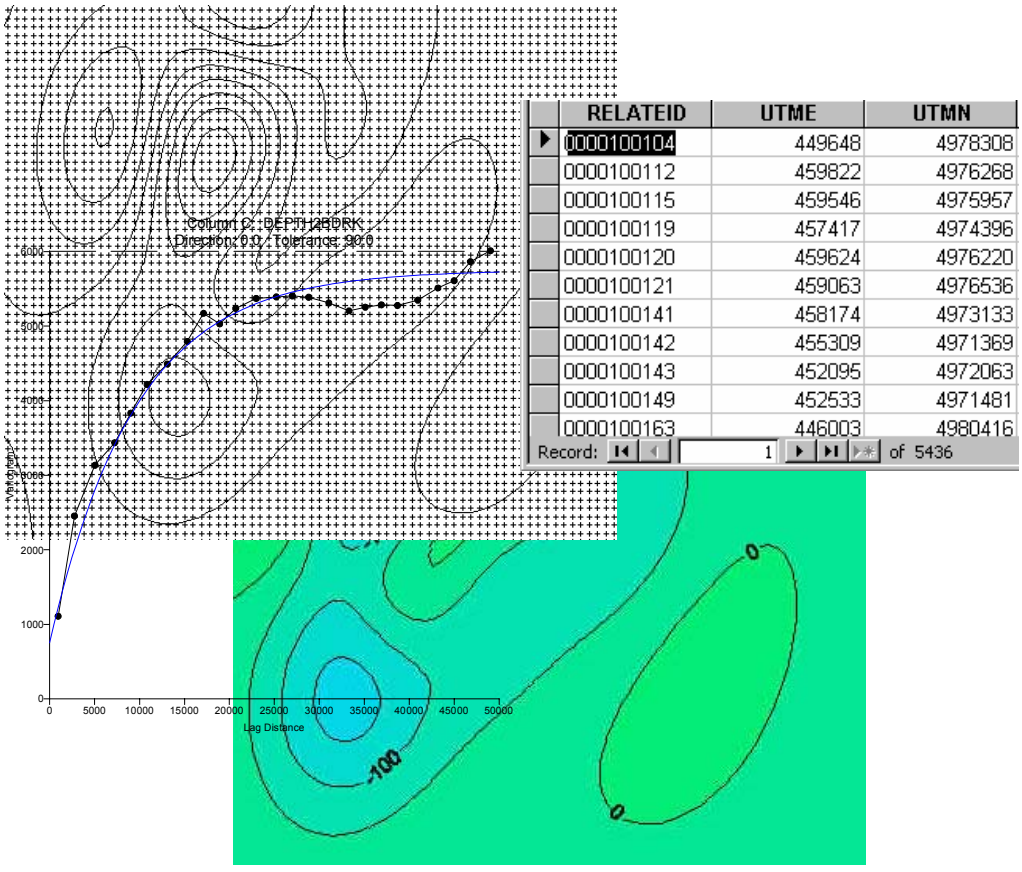
Metropolitan Area Groundwater Model Project

Preparation of Supporting Databases for the Metropolitan Area Groundwater Model

Version 1.00, August 2003

Andrew R. Streitz



RELATEID	UTME	UTMN
0000100104	449648	4978308
0000100112	459822	4976268
0000100115	459546	4975957
0000100119	457417	4974396
0000100120	459624	4976220
0000100121	459063	4976536
0000100141	458174	4973133
0000100142	455309	4971369
0000100143	452095	4972063
0000100149	452533	4971481
0000100163	446003	4980416

Record: 1 of 5436

Table of Contents

Figures

Figures (continued)

Metropolitan Area Groundwater Model Project

Preparation of Supporting Databases For the Metropolitan Area Groundwater Model

Version 1.00, July 2003

Andrew R. Streitz

Introduction

Objective

This report presents an overview of the many databases that support the Twin Cities Metropolitan Groundwater Model (Metro Model), a computer model that simulates regional groundwater flow in the seven-county Twin Cities metropolitan area (Figure 1). The modeled area includes Anoka, Carver, Dakota, Hennepin, Ramsey, Scott, and Washington Counties.

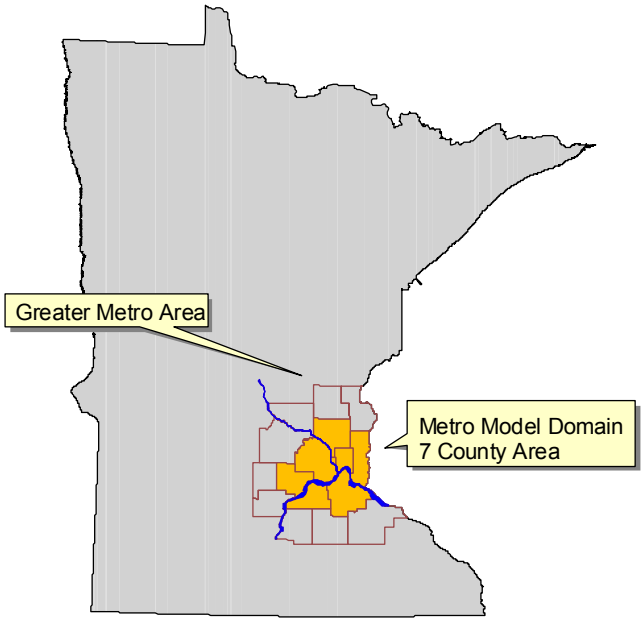


Figure 1. Database Coverage vs. Metro Model Domain

The Introduction presents background information regarding the Metro Model as well as the supporting data and information. It is followed by two main sections. The first presents descriptions of how the databases were created, and the second discusses the databases themselves.

A major goal of the Metro Model team during the creation of the groundwater model was to develop a solid data foundation, and to show how the resulting information was used to support the development of the Metro Model. The ambitious task of building of the Metro Model required an equally ambitious effort to build the elaborate network of supporting databases upon which the groundwater model stands. Some of the databases took the form of traditional geologic resources such as maps and cross-sections, but new data formats also were used, including geostatistically rendered three-dimensional displays, and cross-validated calibration datasets.

Summary of Metropolitan Area Groundwater Model

This report represents the first release version of the document, "Preparation of Supporting Databases for the Metropolitan Area Groundwater Model", available for widespread distribution, and referred to as the Data Report. Any future revisions to the Data Report will be reflected by incremental increases in the version number. However, due to the decision by the MPCA to terminate the Metro Model project it is unlikely that future revisions to this report will be made. Since this report addresses the development of the databases that support the Metro Model, it is appropriate to provide background information regarding the project in this section.

The following summary of the Metro Model was taken from the Metro Model Overview Report (Seaberg 2000).

> The Metro Model is a computer model developed by staff from the Minnesota Pollution Control Agency (MPCA) that simulates regional groundwater flow in the seven-county Twin Cities metropolitan area (Figure 1). The seven-county area is only the beginning however as the supporting databases were derived for a larger domain that extends out into the surrounding counties and comprise the greater metropolitan area. The computer model is based on the analytic element method and simulates multi-aquifer groundwater flow. It is available for use by groundwater scientists working in both the public and private sectors to aid in management decisions affecting groundwater. MPCA staff are applying it to problems of groundwater contamination, but it was developed with additional objectives in mind, to ensure that it has the broadest utility among all groundwater scientists. Since its development, the Metro Model and its supporting data have been used by a variety of planners, government agencies, and private consultants to help protect Minnesota groundwater resources.
>
> The Metro Model is a *regional* groundwater flow model that provides the regional context of groundwater flow in the metropolitan area. It can be an effective tool when modified to mindfully include local site-specific conditions. Application of the Metro Model necessarily requires that end-users insert local detail into the model to conduct site-specific modeling. By serving as the starting point for site-specific models, the Metro Model provides added value to a project because a local

model may be constructed with less time and money than would otherwise be required. Additionally, it may permit the user to spend more time developing the site-specific model since the context of regional flow is already provided. Moreover, the end product may be more technically robust, because the local detail is added against the regional backdrop that the Metro Model provides.

The Metro Model is actually comprised of four separate models that may be linked. Separate project summaries, prepared for each of these four models, as well as other information on the Metro Model can be obtained from the project website at:

http://www.pca.state.mn.us/water/groundwater/metromodel.html

Project Status and Contacts

The Metro Model was supported by the Legislative Commission on Minnesota Resources from 1996-99 with supplemental funding from the US Environmental Protection Agency and the Minnesota Pollution Control Agency (MPCA). It was then supported by the MPCA, through a three-year decline in staff funding. As a result of cuts in federal funds and the budget recommended by the Governor and passed by the Minnesota Legislature in 2001, the MPCA has lost about 10 percent of its former staffing level (over 70 positions). Therefore, the MPCA has had to reduce service in a number of programs and eliminate groundwater programs in order to carry on its highest environmental priorities. The Metro Model is one of the casualties of this action. Without a legislative appropriation expressly dedicated to the project, MPCA has no plans for future support of the Metro Model. Hard financial realities have necessitated cutbacks in many other program areas as well. The Metropolitan Council has contracted with the MPCA for one-quarter time support from one of the former Metro Model ground water modeling experts to provide technical assistance and training to the Council staff through the 2004 Fiscal Year. The goal is to enable continued updates of the Metro Model to develop applications that meet the Metropolitan Council's needs to the extent possible given their funding.

Outside the Metropolitan Council contract, MPCA staff have ceased all support for Metro Model activities. If end-users require assistance in the application of project resources, project staff recommend that they retain a qualified consultant experienced in hydrogeology, Geographic Information Systems, and groundwater modeling and engineering. Resources available through the Metro Model website will remain accessible until they become obsolete.

The following staff are still available to answer limited questions:

Andrew Streitz (218) 723.4929
andrew.streitz@pca.state.mn.us
Expertise: Hydrogeology, Geographic Information Systems (GIS), database management and manipulation, and geostatistics.

Doug Hansen (651) 296.9192
douglas.hansen@pca.state.mn.us
Expertise: Engineering applications, conceptual model, GIS, and model development, calibration, and application.

John Seaberg (651) 296.0550
john.seaberg@pca.state.mn.us
Expertise: Hydrogeology, conceptual model, GIS, and model development, calibration, and application.

Users' Advisory Workgroup

Throughout the course of the Metro Model project MPCA staff worked on a cooperative basis with interested governmental and private-sector parties from outside the MPCA, including government scientists, private consultants, and industrial representatives. As potential users of the model, these parties represented a Users' Advisory Workgroup for the model, providing valuable input into its development. This group consisted of approximately 30 professionals who met on a periodic basis to be apprised of progress and to give input on development of the Metro Model. This group was essential in providing critical technical review as well as guidance on the direction of model development and administration. Many of the ideas for innovative data analysis techniques came from meetings with these professionals. The Metro Model would not have reached its goals without the dedicated support of the members of the Advisory Workgroup.

Project staff are extremely grateful for the support, input, use, and critical feedback that groundwater professionals and end-users of the Users' Advisory Workgroup have provided over the years. Their peer-review and feedback have immeasurably improved the products and services that the project team was able to provide. Without that input, Metro Model resources would not be seeing the widespread application to the variety of groundwater management problems they do today.

Coordinate System and Units

The supporting databases of the Metro Model (as well as the Metro Model itself) are based on a Cartesian coordinate system (flat-plane x, y-coordinates). The coordinate system chosen for the Metro Model is Universal Transverse Mercator (UTM), Zone 15 North, using the NAD83 datum, which is also the standard for the MPCA and the State of Minnesota. The UTM coordinate system has units of meters and introduces a minimal amount of distortion to the projected Cartesian coordinate system. More information on this system and the reason it was selected for use in this project is available in the Overview Report.

Database Primer

Geographic Information Systems and Other Software Tools

Much of the work of the Metro Model project was handled in a Geographic Information Systems (GIS) environment. Specifically, the Metro Model team used ArcView, a proprietary GIS software package that is widely used among parties in Minnesota and the environmental field. Data and information that can be displayed within a GIS environment include bedrock geology, sand content of glacial drift, bedrock topography, thicknesses and surface elevations of the tops of selected bedrock layers, and model outputs including head calibration plots and piezometric surfaces. The GIS environment allows ready comparison of different location-oriented databases and coverages. For example, well and pumping test locations with hydraulic conductivity values may be superimposed on displays of geology or piezometric surfaces.

This software package was one of many used in the analysis of data for this report. ArcView can be used separately without interaction with other programs, but it was most often used to review datasets that were produced with some of the programs listed below. Most of the datasets presented in this report are the product of analysis by a full suite of proprietary programs used in sequence.

Proprietary software programs used for data handling, manipulation, and analysis are listed below, along with a brief description of how each was applied to the project. The mention of specific software packages in this report does not constitute an endorsement of a commercial product by the State of Minnesota or its employees.

Spreadsheet- EXCEL

Sorting and converting files between different formats, and graphing results. Able to produce a descriptive statistics analysis. Modifies datasets to facilitate transfer of data between analysis programs.

Database- ACCESS and ORACLE 8i

Dataset management, filtering, sorting, merging , and table generation. Able to produce specific filtered interpretations that can be exported directly into GIS.

Spatial Analysis- SURFER, VARIOWIN and GEOEAS

Variogram analysis, cross validation, dataset interpolation to a grid, contouring, and DXF graphic file generation. A quasi-GIS function, and a sophisticated spatial analysis set of tools for finding patterns in large datasets.

Three-dimensional Analysis- GMS and Savi3D

Dataset presentation and three-dimensional visualization.

The use of these software tools in developing the supporting databases will be described later in the report.

Central Role of Databases in Metro Model

Geologic and environmental databases played a critical role in the development of the Metro Model. A technically sound model can only be developed after a proper base of information is collected. One way to visualize this is to imagine a groundwater model as the apex of a pyramid made up of layers of supporting databases and assumptions (Figure 2). Each step up the pyramid is possible only because of the support provided from the step below, and each step up requires more sophisticated interpretations of the data.

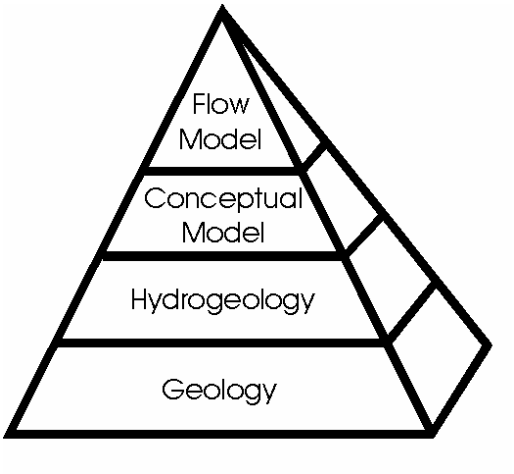


Figure 2. Groundwater Model Pyramid

Construction of a computer groundwater flow model, shown at the top of the pyramid, is based on a conceptual model of groundwater flow. A conceptual model may be described as a set of assumptions regarding groundwater flow expressed in words (Bear and Verruijt, 1990). As described in the Overview Report for the Metro Model (Seaberg, 2000):

> These assumptions include identification of the system's geometry, boundary conditions, type of flow, composition of the system, aquifer recharge and discharge zones, and hydraulic properties of the media. These assumptions represent a simplified perception of the hydrogeologic system intended to meet the objectives of the modeling effort by including only the features that are relevant to the questions being answered. Data and information used in the development of the conceptual model are approached from two fronts: 1) the hydrogeology is evaluated to identify features and processes likely to have a significant impact on the groundwater flow system; and 2) hydraulic head data are evaluated to ascertain indirectly the nature of interaction between hydrostratigraphic units and to help identify "hidden" hydrogeologic features that impact flow.

Simply stated, the conceptual model on the pyramid (Figure 2) is based on the hydrogeology of the system. And the hydrogeology of the system can only be understood once we understand the geology, which is depicted at the base of the pyramid. The geology provides some of the most basic information necessary to develop an idea of groundwater flow, since the geology constitutes the medium through which groundwater flows. The type of information collected at this level includes descriptions of all significant geologic units as well as their extents, thicknesses, and elevations.

External Databases and Maps

Listed here are databases that have been developed and compiled outside of the Metro Model project, and have been used as raw material for new geologic interpretations to benefit the Metro Model. They will be referred to as source databases.

Minnesota Geological Survey- County Well Index

The County Well Index (CWI) is a PC-based database system developed by the Minnesota Geological Survey (MGS) for the storage, retrieval, and editing of water well information. The database contains basic information on well records (e.g., location, depth, static water level) for wells drilled in Minnesota. The database also includes information on the well log, construction and water chemistry for many of the wells. CWI contains basic information for about 293,000 water-wells drilled in Minnesota. The data is derived from water-well contractors' logs of geologic materials encountered during drilling. Geologic well records are available for over 170,000 of the wells on the MGS website. Drillers have been required to submit information on new wells by the Minnesota water well construction code since 1974. After the development of each well the driller measures a water level elevation or head, which plays an important role in calibration of the Metro Model. CWI provided the raw data that was used by Metro Model staff on a regional basis to develop sand content maps of the glacial drift materials, calibration datasets, and regional piezometric surface maps for major aquifers.

More information and information on how to obtain a copy of CWI can be found at:

http://www.geo.umn.edu/mgs/cwi.html.

MGS Bedrock Coverages

In the fall of 1995 the Metro Model project contracted with the MGS to edge match and update the bedrock geology plates from the Twin Cities Metropolitan county Geologic Atlases over the entire seven-county metropolitan area. They delivered this map as a GIS coverage, along with coverages of the upper surface elevations of the St. Peter Sandstone, Prairie du Chien Group, Jordan Sandstone, and St. Lawrence Formation. The bedrock coverage was further updated in 1998. The following references support these metropolitan area bedrock geology coverages: Mossler and Tipping (2000), Mossler and Tipping (1996), Tipping and Mossler (1996).

The geologic map and upper surface elevation maps of selected hydrostratigraphic units were used by Metro Model staff in the development of the conceptual model. Additionally, the upper surface elevation maps were used to construct isopach maps of selected hydrostratigraphic units.

Department of Natural Resources, Water Resources Data

Obwell Network

Since 1944, the Department of Natural Resources (DNR) Division of Waters has managed a statewide network of water level observation wells, more commonly known as the Obwell Network. Data from these wells are used to assess groundwater resources, determine long term trends, interpret impacts of pumping and climate, plan for water conservation, evaluate water conflicts, and otherwise manage the water resource. Soil and Water Conservation Districts under contract with DNR Waters measure the wells monthly and report the readings to DNR Division of Waters. The U.S. Geological Survey also monitors some wells using continuous recorders, and readings are also obtained from volunteers at other locations. Currently, about 700 wells are being monitored in the observation well network. The DNR Obwell Network can be accessed at:

http://www.dnr.state.mn.us/waters/groundwater_section/obwell/waterleveldata.html

These high-quality data gave Metro Model staff a clear picture of piezometric conditions in certain areas, and helped them to look at past and current conditions of aquifer water levels.

Department of Natural Resources, State Water Use System

SWUDS

The State Water Use Data System (SWUDS) is a database of high volume wells that are permitted through the DNR Division of Waters. This database may be explored further at:

http://www.dnr.state.mn.us/waters/watermgmt_section/appropriations/index.html

Metro Model staff used the SWUDS database to develop model datasets of high-capacity pumping wells.

Data Analysis Techniques

Overview

This section describes the data analysis steps that were taken to convert the source databases into the Metro Model supporting databases. Some databases were used without additional processing, such as the Minnesota Geological Survey (MGS) bedrock maps. However, most of the source databases required modification because they were incomplete or because there was a need to identify biases and spatial patterns.

Biased and Incomplete Data

The CWI database is a large, spatially well-distributed database, but it is in many ways incomplete, at least for the purposes of building a groundwater model. This is because the database is based on well logs filled out and submitted by water well drillers, who are hired generally in response to new domestic and commercial construction. Though this ensures that hundreds of new logs are entered each month this source of information brings with it several limitations.

The first problem is location bias. The vast majority of wells in CWI are drilled for new residential homes, which in Minnesota means they are drilled within the rapidly expanding Twin Cities metropolitan area that stretches from Rochester to St. Cloud. Thus the spatial coverage of new data is denser in the metropolitan region than in Greater Minnesota. Even within this region development does not occur evenly, but is focused in specific townships and cities depending on growth-limiting factors such as highway expansions and changes in the Metropolitan Council's Metropolitan Urban Service Area (MUSA). Consequently, a large number of records might have been submitted from the west metropolitan area in 1978, while in 1999 a larger number might come from the east metro. For example, analyzing CWI by date of entry might provide a valuable review of residential development in the area, as first one county and then another receives hundreds of new homes. This time and location bias is not fatal to the use of the data, but is something that must be understood by potential users.

A second problem is location error. Wells are commonly reported with an incorrect location because the forms are submitted in the confusing noncartesian Township-Range-Section (TRS) system. As budgets allow, the MGS physically field locates individual wells, using Global Positioning System (GPS) technology to generate a UTM location coordinate. One principal advantage of the UTM system over the TRS is that the former is GIS ready (see previous section on coordinate systems). Because the Metro Model supporting databases are based on the field-located data and not the larger TRS dominated data, this location error does not apply. Outside of the Metro region however, the error associated with TRS-collected data would remain a problem. This of course remains a problem even if TRS coordinate locations are converted to the UTM coordinate system via a program such as SECTIC or MNCon.

A third problem is the collection of geologic data. Geologic units that are encountered during well construction may be described in colloquial vernacular, referring to water bearing sands in terms that may not be understood by everybody reading the log. Also, wells are drilled with the intent of finding water, and closer attention is paid to water-bearing strata, which generally receive a more detailed description than strata that cannot supply water. This bias is reflected in the notes recorded in the well log.

In conjunction with field locating wells, MGS geologists review the well logs, reinterpreting the descriptions and substituting standard geologic codes. Although they cannot compensate for a general lack of description on non-water bearing units, they are able to substitute standardized geologic codes for the more common colloquial vernacular. Of the hundreds of thousands of wells in the state-wide database, close to 100,000 have been field located and geologically interpreted, many of them in and around the Twin Cities metropolitan area.

The final location-based error has to do with the manner in which the ground surface elevation of the individual well is determined. Because the surface elevation for each well is taken from a topographic map on which the well location has been plotted, and because these maps have an elevation contour interval of 10 feet, the process yields an elevation error for each well of approximately plus or minus 5 feet. No practical solution exists to correct this error, since it would require a level survey for each well, a costly provision. For the purposes of the regional large-scale Metro Model project covering hundreds of square miles, an elevation error of a few feet is considered insignificant.

Data Trends

A simple definition of "trend" is a change in a parameter over time or space. Trends can be trivial or critical. To discover which you are dealing with, the data must be analyzed within the context for which it will be used, because different uses of the data will be sensitive to different trends. Trends that can affect groundwater include seasonal head fluctuation due to natural recharge, or increased domestic use in the summer, leading to increased pumping. Shorter period trends can be caused by storm events that can increase recharge to an aquifer via direct infiltration and by increasing surface water to groundwater flow from higher river and lake elevations. A seasonal trend can be tested for and, if found, can be accounted for using statistical methods such as yearly and rolling averages.

Spatial Analysis

Spatial statistical analysis (as used in this report) differs from regular descriptive statistics mainly in the inclusion of x, y- location to each data point. In spatial statistical analysis the statistical tool employed is not one of the traditional descriptive tools such as mean, median, and range. Instead it is the variogram which measures direction and strength of the spatial correlation. These techniques are especially powerful when the dataset has many data points and they are spatially well distributed, meaning that data is spread evenly throughout the area of interest.

The terms “many” and “well distributed” are not precise, and can only be determined in context. Generally, the sample number should be greater than 50 and the data points should fill the area of interest without big gaps. Finally, the points should be close enough together so as to produce a variogram (defined starting on p.13) with values that can be correlated. An example of a large, well-distributed dataset is displayed in Figure 3. Dakota County (outlined in red) shows good coverage for bedrock wells (almost all from wells in the Prairie du Chien-Jordan aquifer) in the display on the left, but inadequate data coverage for a countywide analysis of glacial drift wells in the display on the right. In the same way that a dataset mean gains power from a larger sample size, so too does an analysis based on the variogram gain from a spatially well distributed dataset.

As an illustration of the differences between statistical measures that use and don’t make use of location, consider the difference between determining the height of children in a school versus the height of plants in a field. For the children it is reasonable to disregard the location of individuals when calculating the average height of students in a classroom since the location of students within the classroom is certainly not fixed, and is therefore irrelevant to the measurement. But if you were analyzing the height of plants growing in a field then location could be critical to understanding the processes that led to different growth rates. (Even location can be brought to bear on the height of children. A correlation between height and home address that is random within a city may be meaningful across a country or between countries in the world.)

None of the shortcomings of the CWI database discussed previously render the data useless. On the contrary its size and relatively even spatial distribution over most of the state make it an ideal candidate for trend and spatial analyses. Earlier in this report specific examples were presented of different kinds of errors that exist in CWI. The same list will now be discussed in light of the use of data analysis techniques that can provide greater confidence in use of the data.

The first error was the location bias that stems from a spatially uneven collection of new data points. Specifically this could cause problems for analyses if there were significant areas within the region that were not sampled. Figure 3 shows that CWI data has accumulated in a spatially well distributed fashion over the last 30 years. Based on an analysis of CWI water elevations, trend analyses of the change in groundwater head over time for the unconsolidated Quaternary material (Layer 1) and the Prairie du Chien-Jordan aquifer (Layer 3) show no discernable patterns significant enough to affect the Metro Model. Within these aquifers there are local areas where groundwater levels are rising or falling.

If there is any area under-represented in the database it is in the highly urbanized center of the metropolitan area. This is an area that is both dependant on wells drilled before geologic logs were routinely recorded and incorporated into CWI, and areas served predominantly by public supply systems, not individual domestic wells. Estimates of parameters of interest, such as head measurements, can be made in this area, but where the data are sparse the confidence in the estimates is low.

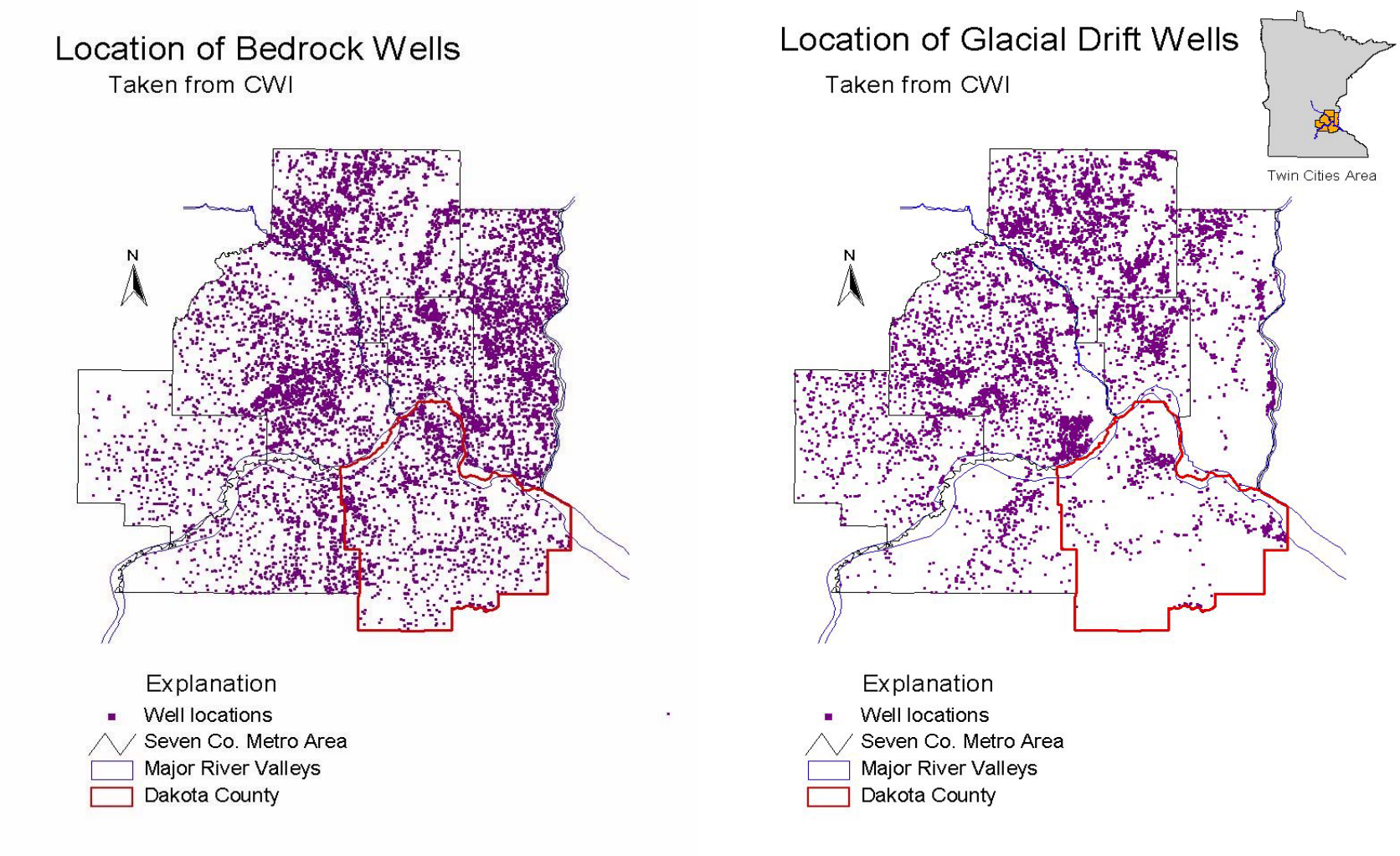


Figure 3. Location of Wells, Completed in the Bedrock and the Glacial Drift, Respectively

The second problem, location error, can also be improved with spatial analysis. Assuming that the error in field locating is random, meaning that the field staff do not produce locations that are consistently biased in one direction, then the combination of thousands of such wells can yield important information. Relatively small errors in many wells do not render those borings unusable. This applies to the original locations submitted at the time of drilling as well as the corrected locations; the use of GPS does not mean errors in field location disappear!

The third data concern regards characterization of the glacial drift materials. During well construction the driller is more attentive to the presence of water-bearing geologic units. This attention is carried over into the information written into the well log. Geologic units not suitable for providing water may tend to be given less attention. The result is the creation of a well log that can more accurately describe sand and gravel dominated units than those with higher percentages of clay, till, etc. This "sand or not-sand" approach lends itself to an analysis based on a binary system where the presence of sand and sand-like units are assigned a value of "1", and where non-sand-like units are assigned a value of "0". By simplifying the data in this way it is possible to do a pattern search with variogram analysis, based on the trait that was positively identified (ie. sand or sand-like materials). Incomplete descriptions of geology can be overcome by combining hundreds of neighboring well logs into a single interpretation. Imprecise or incomplete data can then be transformed through these statistical techniques into more powerful datasets.

An analysis based on a positively identified parameter is superior to a pattern search performed on a characteristic that is inferred (i.e. non-sand) because "non-sand" encompasses a wide range of possible geologic types, and these types will not correlate well as a group. Though well drillers may make mistakes about what is and isn't sand, this error is most likely randomly distributed. And this error can be greatly reduced in severity by grouping together the results of thousands of wells.

Up to this point, data analysis techniques have been described in only general terms, as methods for maximizing the utility of a dataset. Many of the concepts mentioned are commonly in use, including descriptive statistics and trend analysis. Before demonstrating how all these tools were used to analyze datasets for the Metro Model, it is only necessary to provide some background in the one analysis method that may not be widely known, spatial statistics.

Variograms

The concept of a variogram is briefly explained here, but a complete description may be found in Isaaks and Srivastava (1989). A variogram is a measure of the spatial correlation or continuity between location-based data points for a given parameter, such as water level measurements. It can provide insight into the presence of anisotropy, a measure of how a parameter varies preferentially with direction, by indicating directions of minimum and maximum correlation. Perhaps most significantly, a variogram forms the basis for interpolation between measured points

to obtain values on a regular grid used for contouring. The interpolation is done using a process known as kriging, which uses weighted linear combinations of data, minimizes the error variance, and reduces the mean residual to near zero. All this means that the interpolator used for estimating parameter values is guided by a spatial pattern that is manually constructed from actual field data collected from the area of interest, and reflects intrinsic variations in the parameter over that area. Most other interpolators (nearest neighbor, inverse distance, polynomial regression, etc.) are typically used as "black box" interpolators, automatically processing values without input from an operator, or taking advantage of an operator-defined measured variability of that parameter.

The variogram analysis can find patterns in a data population that are not apparent to the human eye or revealed with the use of descriptive statistics. A variogram is created by plotting the variance of data pairs against the distance between the paired values. The analysis starts with the selection of lag increments, the ranges of distances between all pairs of data points (Figure 4). Variance is plotted on the *y* axis, and lag intervals on the *x* axis.

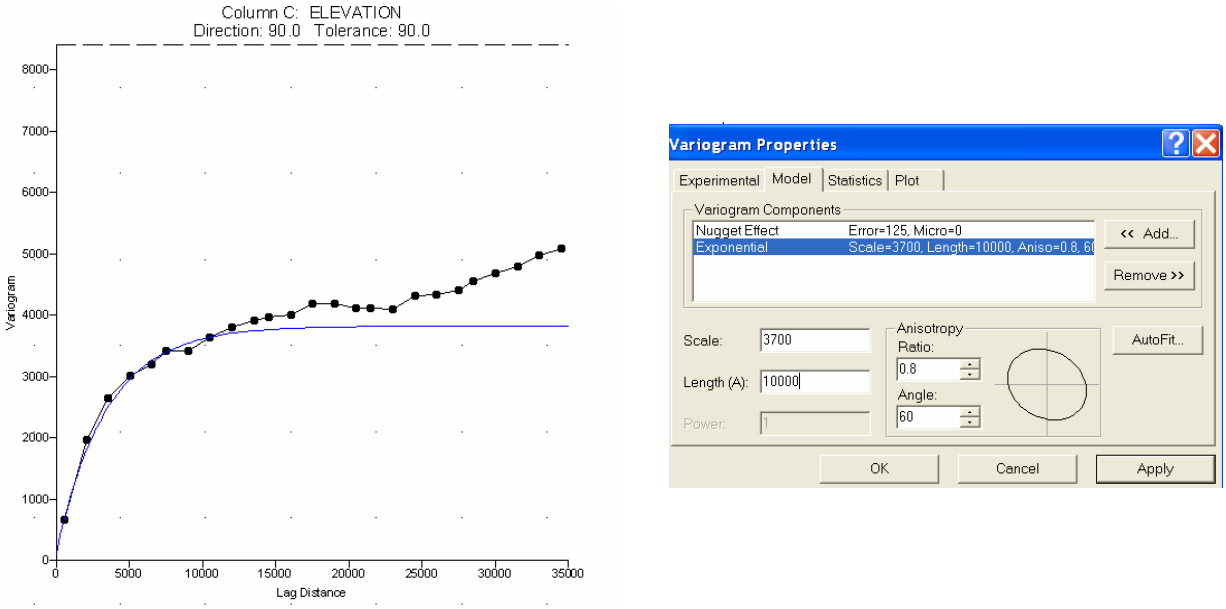


Figure 4. Determining a Variogram Model with SURFER 7

Data pairs from a population of data are then grouped according to lag interval. For example, if a lag interval of 100 meters (m) is selected then all data pairs lying within 100 m of each other are placed in the first lag (bin), all pairs of data points lying between 100 and 200 m of each other are placed in the second lag, etc. Differences in the values for each pair to be analyzed, such as groundwater head elevations, are then squared, summed together and finally divided by twice the number of pairs in the lag. This represents the variogram (or semi-variogram) value for that lag. The variogram is constructed by applying this procedure for each lag. The variogram value typically

increases as the distance between data pairs increases because values of closely spaced points tend to be more similar than for points lying farther apart. The distance on the X axis at which the variogram values reach a plateau as a function of increasing lag increments is an important measure of correlation of the value with distance and is called the Range. The value on the Y axis of the variogram at this point is referred to as the Sill, and increasing values on this axis are a measure of the variance of the dataset. The relationship between pairs of values is considered random at lag increments greater than the identified Range. The final step is to match a curve to the variogram lag points. The different measures of the variorgram are used in the calculation of estimates on a grid, and in the case of Figure 4 the best match comes from an Exponential model, Scale of 3,700, Anisotropy of 0.8, Nugget error of 125, Direction of 60 degrees, and a Range of 10,000. For more information on these terms see Isaaks and Srivastava.

A graphic presentation of the two-dimensional variogram's correlation pattern from Figure 4 is presented in Figure 5, and was generated by plotting the correlation distance (the Range) in 30 degree increments. The resulting plot is ellipse shaped, and when used with kriging to generate an interpolated value to a grid, is centered on a grid node. The value at the node was calculated from the weighted values of the data points located within the variogram ellipse, depicted as solid black points in Figure 5. The degree of anisotropy has been exaggerated for effect.

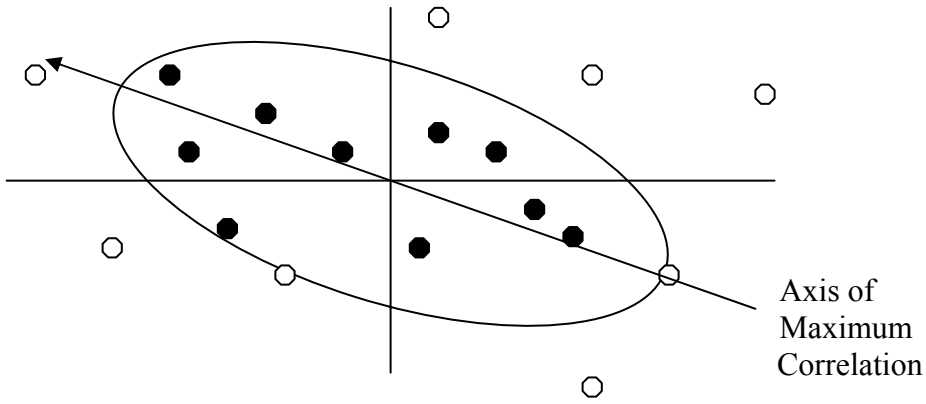


Figure 5. The Shape and Orientation of the Variogram Ellipse

Different programs can be used to develop variograms. Both SURFER and VARIOWIN were used in this work. VARIOWIN can handle a maximum of 1,200 wells in the variogram analysis. SURFER can handle a much larger number of data points in the variogram analysis through using a different methodology. SURFER's ability to use more wells in the calculation could be of greater use in certain areas because there is no need to randomly select a subset of wells to fit the 1,200 maximum allowed in VARIOWIN. In any case, the variograms calculated by both programs using the same datasets yielded virtually identical results. The variograms and data results displayed in this section were produced with SURFER.

This ends the overview of the tools used to prepare datasets for use in the Metro Model. The next section shows how they can all be used (and must be used) together in order to provide the best inputs to the groundwater model.

Data Analysis- from Simple to Complex

As an introduction into the use of analysis tools for the development of the Metro Model supporting databases, it might be helpful to demonstrate the strengths and weaknesses of some of the different tools. This will be accomplished by documenting progress toward an interpretation objective through the analysis of the CWI database. The simplest tools will be applied to the problem first, progressing to more sophisticated approaches, all the while showing how conclusions change with each tool used. The goal for this exercise is to describe the changes in groundwater elevations for the Prairie du Chien-Jordan Aquifer in Hennepin County.

Spreadsheet

The first and simplest tool is the spreadsheet. Any analysis performed with a spreadsheet is limited to the data handling capabilities of the spreadsheet, which sets a practical limit on the number of records the user can review, and the data that can be compiled. For instance, it is unlikely that someone would use a spreadsheet to sift through the thousands of well records available for the Prairie du Chien-Jordan Aquifer in Hennepin County. A workable strategy for a spreadsheet involves the selection (perhaps randomly) of a few well logs to represent the entire database. The data for a single well could be turned into a display such as Figure 6, where Ft AMSL stands for, “Feet above Mean Sea Level”.

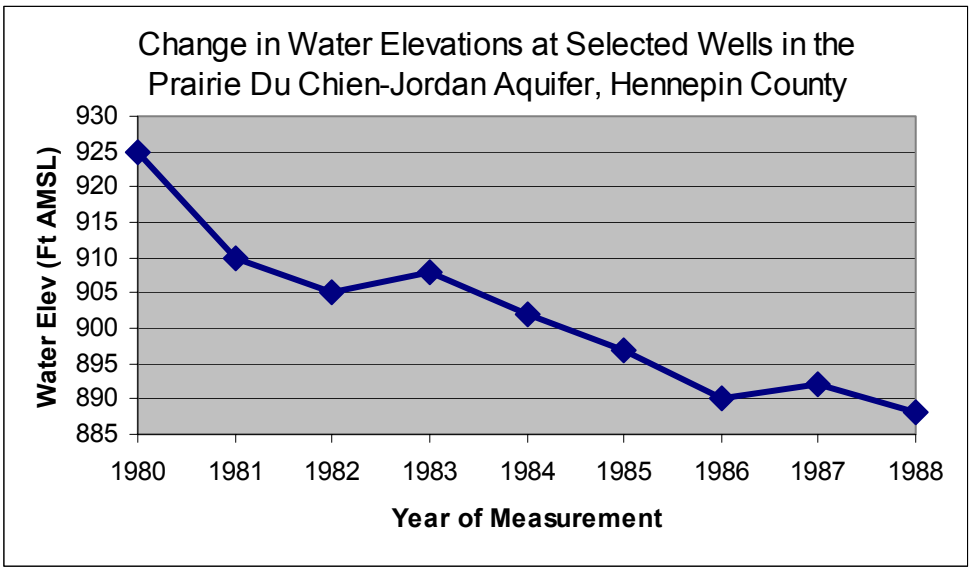


Figure 6. Change in Groundwater Elevations in Hennepin County

The display is simple and the conclusion is clear: groundwater elevations in the aquifer have been in decline almost continuously since 1980. One obvious objection is that a spreadsheet is unable to analyze the breadth of the database. Is enough known about the data to accept a single well as representative of what is happening in an aquifer across an entire county? Would similar conclusions result from considering more of the data?

Database

A more sophisticated tool that responds to these concerns is a relational database that is based on a software package such as ACCESS or ORACLE. Database software can analyze thousands of wells at a time, searching for specified characteristics, and this data is typically stored in tables which is the basis of the phrase "relational database". Instead of selecting individual wells that the user believes to be typical, large numbers of wells that match the requirements can be filtered with database tools to arrive at a more representative answer. This leads to a different conclusion from the simpler analysis above. Figure 7 shows two views of the data: the number of wells installed into the aquifer within a calendar year, and the average groundwater elevation of all wells completed in a given year (taken at the time of the well installation). The strength of this view of the data is that an average of all groundwater elevations from wells installed within a given year is more representative of the aquifer than is a single well. Also, by listing the number of wells per year the user can make an estimate of the confidence that can be placed in the calculation.

The clear trend of the previous display has now become more complicated. Is this as far as we need to proceed with our analysis, or is it possible that this analysis is also based on an incorrect assumption? Note the upward spike in the number of wells installed in 1987. This is a good reminder that CWI is not a groundwater network. New wells are installed in response to economic and societal pressures, not careful placement by groundwater

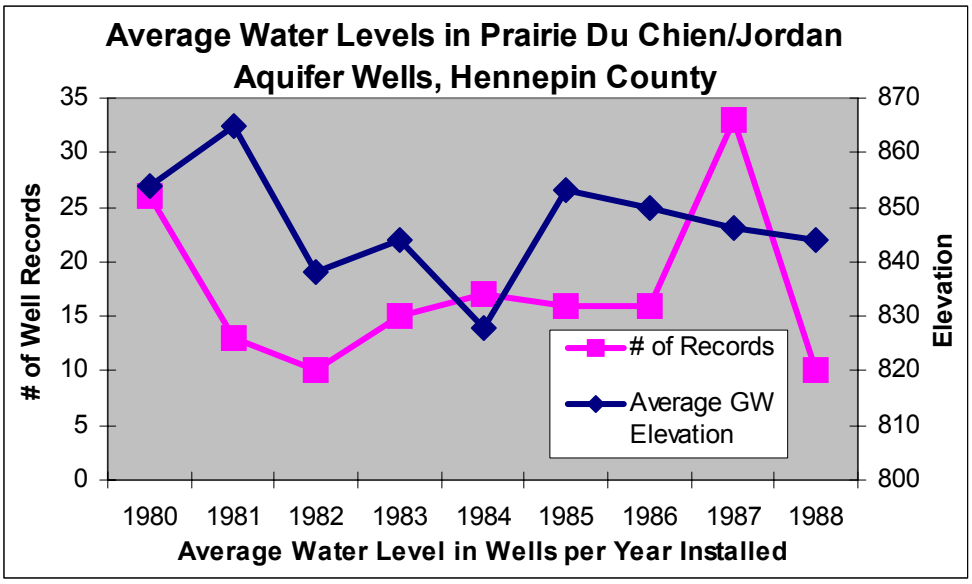


Figure 7. Average Annual Water Levels in Newly Constructed Wells

experts. Have all these 1987 wells been installed in the same area? Did a shift in the MUSA (Metropolitan Urban Service Area) line, the boundary that controls where development can and cannot occur, stimulate new drilling that is biasing the CWI database, and can we investigate this with database tools? These possibilities raise the issue of location, because a groundwater elevation really should not be represented by a single value across the entire range of the aquifer. How does the aquifer head elevation change across the area in question? A database is still the appropriate tool to filter the data, but the spreadsheet displays used in Figures 6 and 7 are limited in their ability to present geographic data effectively. The next step is to move the analysis to two dimensions, which is the province of GIS software.

Geographic Information System

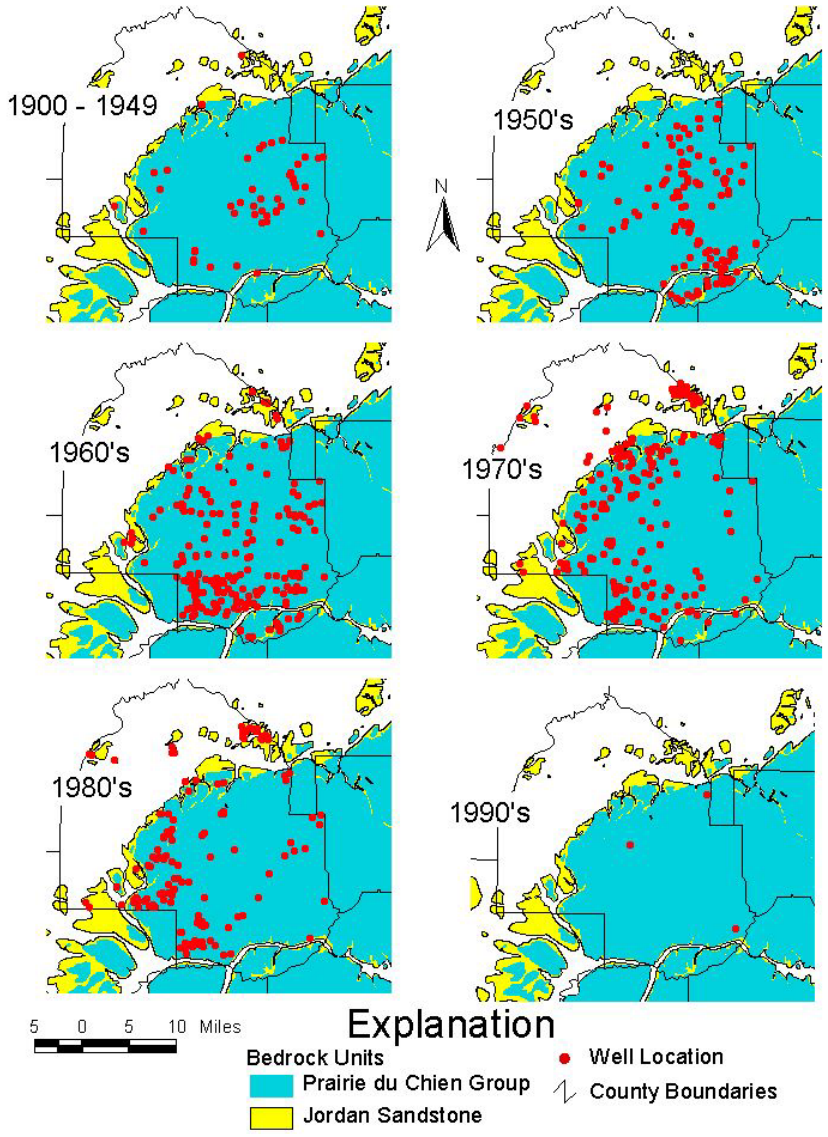


Figure 8. Location of New Wells by Decade and Selected Bedrock Units in Hennepin County

A proper analysis of this surface might start with the maps of Figure 8, showing the locations of wells, grouped by the decade they were installed in the Prairie du Chien-Jordan Aquifer in Hennepin County. The display shows a movement with time of new wells away from the urban core toward the northwest. This groundwater use trend matches the development of suburbs north and west of Minneapolis, and the growth of the community water supply system in the urban core during the last century that is predominately supplied by surface water.

This strong geographic trend means that yearly averages are not good predictors of water levels in this aquifer. The geographically balanced location of wells of the 1960's gives way in the following decades to the clustering of new wells along the margins of the aquifer's boundaries. To continue our example of Figure 7, consider Figure 9, showing just those wells installed in 1987. Of the 34 new wells installed that year, two-thirds are located in a small cluster in the southwest corner of the county (wells outlined in blue). This bias violates our assumption of representativeness of the data taken from CWI, a bias not apparent until the data were displayed in GIS.

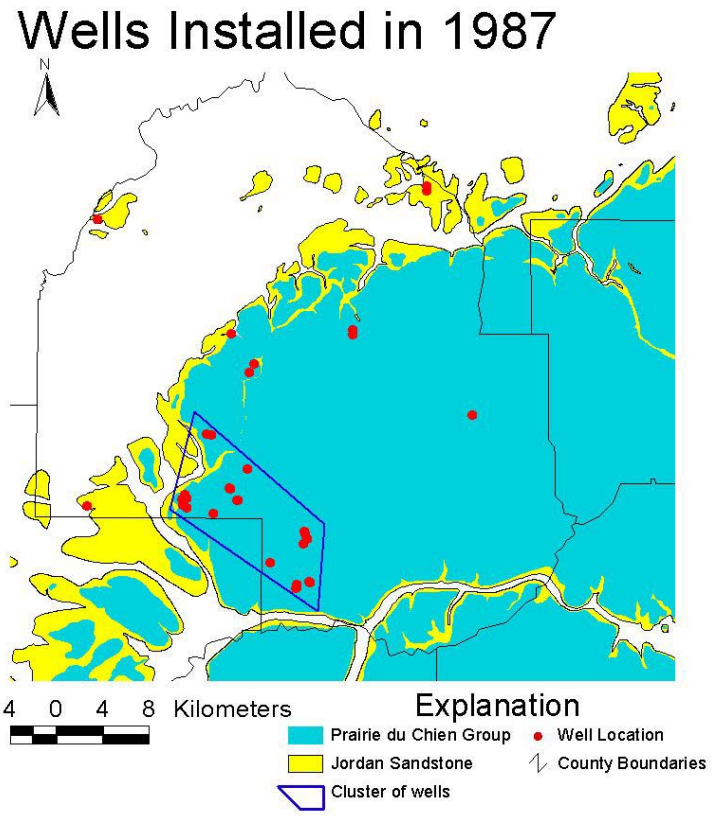


Figure 9. Location of Wells Installed in 1987,
Prairie du Chien-Jordan Aquifer, Hennepin County

How does this new understanding of well location change our interpretation of the spreadsheet in Figure 7? Given the temporal trend of new well locations, comparing yearly averages of groundwater elevations appears to be flawed analysis, at least as it relates to groundwater elevations across the entire aquifer. The problem was the assumption that a single value could represent a three-dimensional groundwater surface across its entire range. In hindsight this seems an unreasonable premise, but it is the kind of assumption that is driven by the use of available tools.

Three-Dimensional Analysis Using Spatial Statistics

Given that the groundwater surface is spatially distorted due to groundwater discharge and recharge, what display techniques are better suited for the task? Commonly used two-dimensional display methods include contours and shaded relief. Figure 10 shows a combination of shaded relief with a superimposed contour plot display of the data shown in Figure 9. Figure 10 shows the groundwater surface of the Prairie du Chien-Jordan Aquifer for 1987 in plan view. Figure 11 shows the location of this display within the Twin Cities metropolitan area. The scale is in UTMs, and the elevation in feet above mean sea level. Crosses mark well locations. Figure 12 shows shaded relief alone, but in an "edge-on" three-dimensional perspective.

From review of both displays it is apparent that the 1987 Prairie du Chien-Jordan Aquifer groundwater surface is not flat, and is strongly distorted in a few areas. Given the relatively few datapoints available for this dataset it is not possible to determine if the distortions are an artifact of the sampling. A larger, more spatially complete dataset for the Prairie du Chien-Jordan Aquifer in the same area displays a similar though more detailed view of this same groundwater surface (Figure 13).

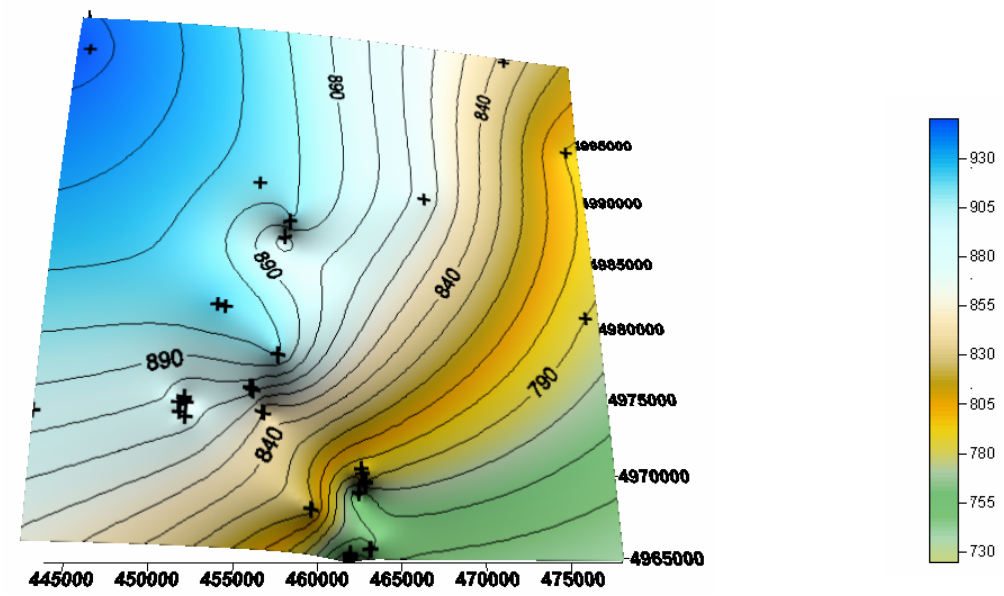


Figure 10. Groundwater Surface in 1987 for the Prairie du Chien- Jordan Aquifer

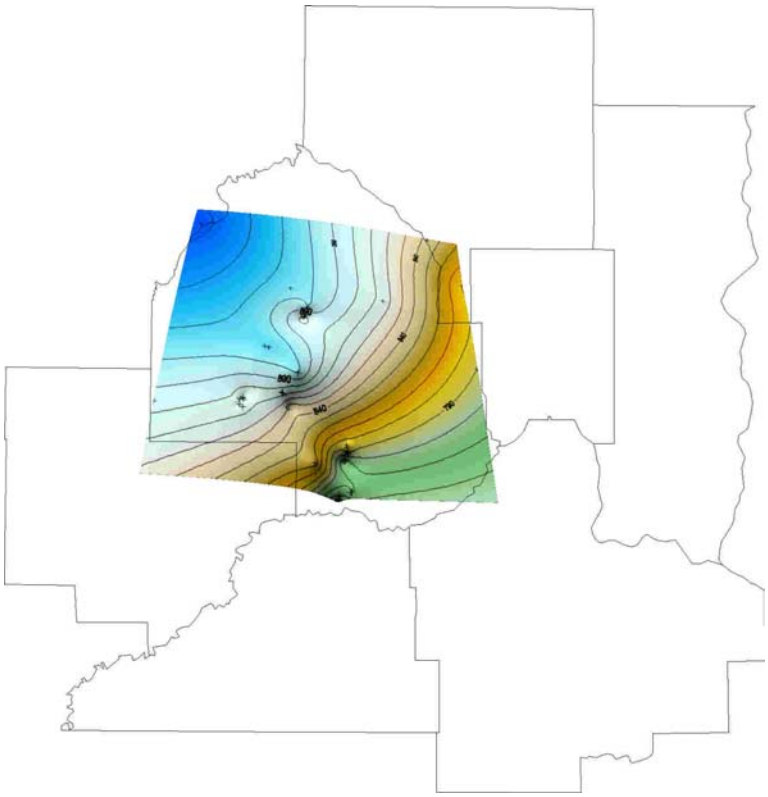


Figure 11. Location of Groundwater Site in Hennepin County

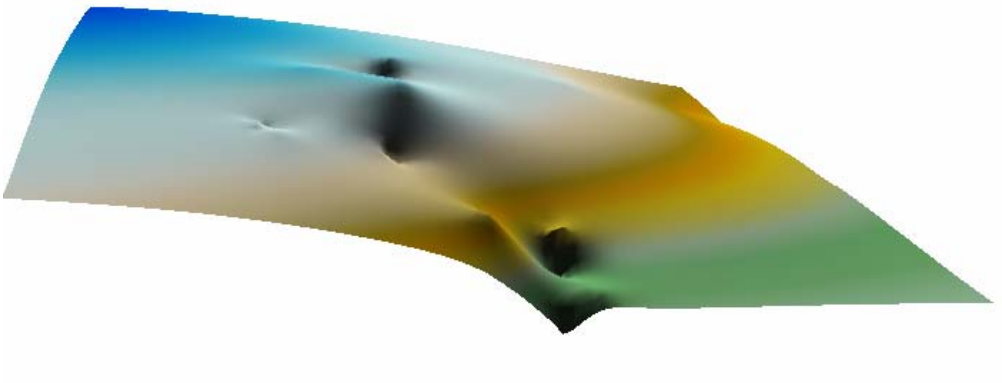


Figure 12. Groundwater Surface as Viewed from the South, Edge-on

Figure 13 finishes off this discussion by showcasing the two- and three-dimensional display techniques in combination with a spatially complete dataset. This presentation shows the groundwater surface of the Prairie du Chien-Jordan Aquifer from south of the Minnesota River looking northwest at Hennepin and Carver counties. The data is

geostatistically analyzed, and interpolated to a grid. The resulting contour plot is then superimposed over a surface plot, along with symbols marking a subset of the original data locations.

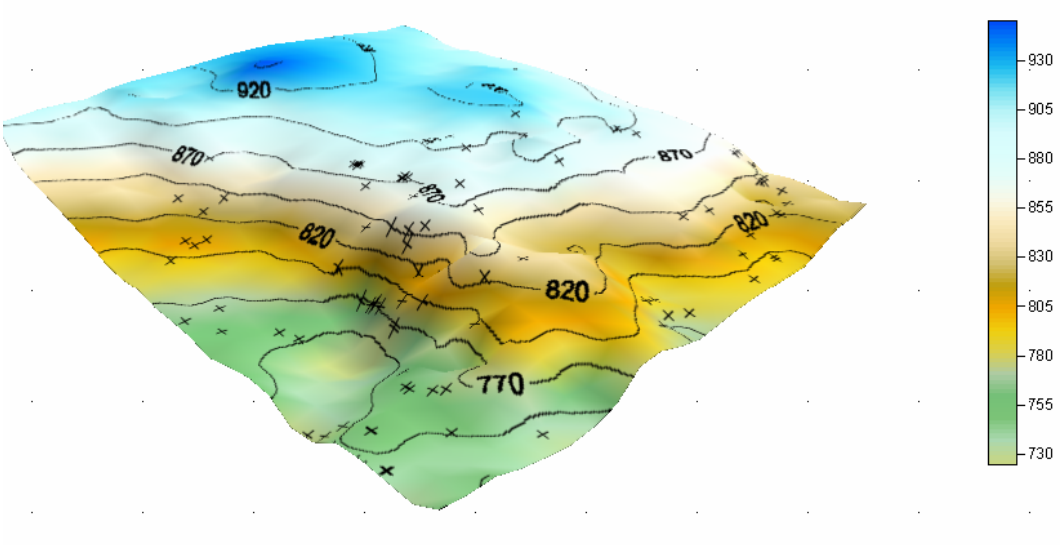


Figure 13. The Groundwater Surface of the Prairie du Chien-Jordan Aquifer in Three-Dimensions (elevations in feet Above Mean Sea Level)

At the start of any large project it is common to underestimate the amount of work necessary to complete tasks, especially when the project is breaking new ground. The Metro Model team was continually surprised at the amount of work that was needed to assemble the databases necessary for building the Model. The path just illustrated, that of analyzing data with increasingly sophisticated tools, is approximately the same route of discovery that the team went through as the project matured.

The larger lesson of this demonstration is that the strengths and weaknesses of a database must be understood fully before use, and the tools used to explore these characteristics must be sufficient to the task. The data should be tested with a variety of tools in order to confirm that the operator truly understands the dataset and what it is revealing about the groundwater system. It is not enough to produce an attractive display. And it is easy to assume that an answer is correct because it fits some preconceived notion, or because the result is straightforward. It is helpful to remember a quote by H.L. Mencken, “For every complex problem there's a solution that's clear, simple -- and wrong."

Analysis of Quaternary Deposits

Glacial Hydrogeology

It was very important to develop a conceptual hydrogeologic model of the unconsolidated Quaternary age glacial deposits, since they are represented in the Metro Model as the Layer 1 aquifer. This aquifer is critical to the model because it exerts great influence on the recharge rates for the underlying aquifers, and it is almost always the first aquifer impacted by groundwater contamination that originates at the ground surface. This section describes the development of the hydrogeologic conceptual model for the glacial drift materials using automated database and geostatistical techniques as described in the previous sections.

The geology of the Twin Cities area is described in some detail in the Metro Model Overview Report (Seaberg 2000), a portion of which, regarding the Quaternary geology, is repeated here:

> Unconsolidated Quaternary glacial deposits dominate the near-surface geology in the metropolitan area. Glacial sediments include relatively impermeable glacial tills and deposits of highly permeable outwash and ice-contact stratified drift, such as eskers, kames, and tunnel valley fans. Additionally, alluvium, generally confined to river valleys, is comprised of relatively permeable sands and gravels and relatively impermeable overbank type deposits. Much of the alluvium was deposited under glacial meltwater conditions. The glacial drift in the metropolitan area ranges from highly heterogeneous terrane, undifferentiated with no mappable units present, to zones showing significant continuous units that can be mapped over large areas.

Development of an effective hydrogeologic conceptual model evaded traditional techniques of characterizing glacial geology and stratigraphy. Early efforts focused on developing glacial geologic provinces based on glacial provenance, advance and retreat, and depositional environments. Although these techniques can yield effective models of glacial geology, the project team was not able to develop an effective hydrogeologic model for the deposits. This was largely attributed to the high variation typically seen in continental glacial deposits, compounded by the fact that the region has been subjected to multiple episodes of glaciation, originating from different areas. This meant that a new approach to characterizing the hydrogeology of the glacial deposits had to be developed.

The thick unconsolidated glacial drift deposits that are common throughout the Twin Cities metropolitan area are not, as a rule, homogeneous in the vertical direction, owing in part to the different depositional environments resulting from multiple glaciation events. However, the nature of glacial deposition results in deposits that typically display more uniformity in the horizontal direction than in the vertical. Consequently, the team decided to focus analysis on evaluating the horizontal (in the

x, y- directions) correlations that suggest continuity of stratigraphy using geospatial techniques, also known as geostatistics. To account for vertical variations in the drift, analyses of the lateral variability of the drift materials were conducted for four stacked discreet elevation intervals, resulting in a quasi three-dimensional representation.

An important issue to resolve early on in the analysis was to determine the scale and resolution of the analyses to develop the conceptual model. The convention chosen was to develop GIS displays and maps at a scale similar to maps of the Paleozoic-age aquifers used for the Metro Model's lower four layers. The scale of most maps used for the project is 1:100,000. This ratio is both a scale (e.g. 1 cm = 1 km) and an implicit expression of the resolution of the map. Detail that can be resolved by the human eye at this scale is approximately 160 meters, which is similar to the well and geologic boring location error of approximately 100 meters which is used to prepare maps of the glacial deposits.

Binary Analysis

The strategy for analyzing the Quaternary geology (Layer 1 of the Metro Model) was to take the information from thousands of well logs and develop a conceptual model that represents the presence and extent of permeable sands and gravels in the drift. By converting geologic descriptions into "1"s and "0"s as described earlier in the "Data Analysis Techniques, Overview" section, the analyses focused on the presence of sand and gravel. This approach is similar to the methods used to define mineral-bearing ore bodies.

The first step before well logs were selected was to identify the elevation at which glacial drift can be found. Figure 14 shows the distribution of glacial drift as a function of elevation, based on geologic material entries in CWI well logs. Each well log within the model domain was checked for the presence of Quaternary deposits for elevations ranging from 200 to 320 meters above Mean Sea Level (m MSL), taken in one-meter increments. The results are plotted as a histogram, plotting elevation against the number of wells containing Quaternary deposit information. The normally distributed dataset has a mode of 265 m MSL, with approximately 25,000 well borings containing Quaternary materials at this elevation. A cutoff was chosen to eliminate those elevations where the number of wells with information on Quaternary deposits dropped below 5,000. In this fashion the dataset was trimmed of the uppermost and lowermost intervals, leaving the interval range of 220 m MSL to 300 m MSL for the investigation of the Quaternary deposits. This 80-m interval contains 94 percent of all the available data on Quaternary deposits in the metropolitan area.

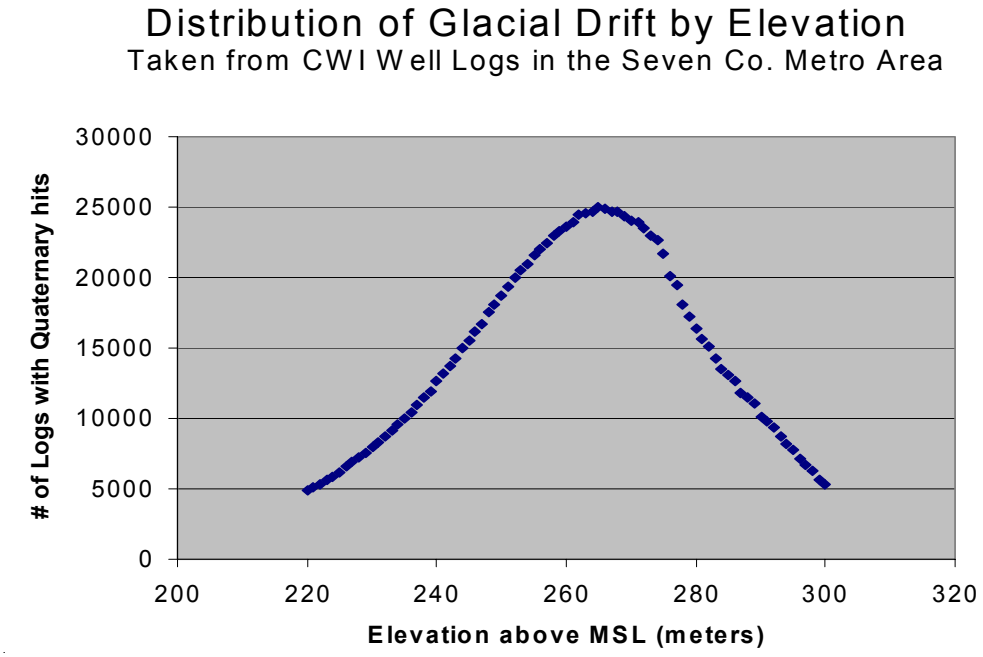


Figure 14. Number of Well Logs with Quaternary Entries by Elevation

After the selection of the elevation sequence, geologic logs were investigated at one-meter intervals for individual wells. A total of 81 horizontal planes were constructed at elevations designed to intersect the glacial drift at one-meter intervals, starting at 220 meters MSL. A binary coding system was used, with "sand" (highly permeable material) assigned the value of 1 and "not sand" (fine-grained materials with low permeability) a value of 0. Each log was therefore simplified into a column of 81 rows made up of codes for sand, non-sand, or no Quaternary information. As was discussed in a previous section (Data Analysis- from simple to complex), database software was used to manipulate the very large databases used for this analysis. These techniques are demonstrated later in the report in a section entitled, "Minnesota River Valley- a Demonstration of Data Analysis Techniques". The advantage of using a database to investigate geologic relationships instead of the development of the more traditional geologic cross-sections is that instead of relying on a relatively few, possibly representative boring logs, the interpretation instead rests on thousands of well logs. The database and geostatistical tools then distill the complexity of all those logs into an interpretation that is fairly easy to understand and is more representative of regional geologic conditions.

Sand Content Maps

One of the simplest methods to interpret these data is an averaged, two-dimensional representation of sand content. This has been done for the Northwest Province and the results displayed as Figure 15. For each of the 8,900 borings, sand content values

were averaged across all 81 intervals (where data exist at each well location). These data points were then spatially analyzed and interpolated to a grid measuring roughly 100 x 100 meters. Colors were assigned to signify the probability of finding sand at a grid node. The color code follows the visible light spectrum as produced by a prism. The color red represents a high probability of sand given as a percentage, blue is a low probability, with the other colors representing the continuum in between. Sand content is the term used in this report to describe the percentage of sand that could be found at a given location by averaging grain-size lithology types in the geologic column.

There are three broad zones of sand content visible on this map: low sand content (<20%) to the west, high sand content (>70%) to the east and southeast, and intermediate sand content between the two. This specific display was useful in gaining an overview of the sand content throughout the Quaternary system, but generalizes too much across a thickness of approximately 80 meters to be of use in developing specific elements in the groundwater model. After deciding that 80 meters was too thick, the next step was to split the sequence into finer layers.

Stacked Two-Dimensional Views

A modification of the technique just described was to break the 81 horizontal slices into four equal thickness intervals and then average the sand content at the well locations in each interval, and use kriging to produce a quasi three-dimensional view of the Quaternary geology (Figure 16). The interval divisions are: Quaternary Interval 1, extending from 280 to 300 m MSL; Interval 2, 260 to 279 m MSL; Interval 3, 240 to 259 m MSL; and Interval 4, 220 to 239 m MSL.

The resulting displays show a complex connection of sand units across and between layers. One important observation from a comparison of Figures 15 and 16, is the dominance of non-sand across most of Hennepin and Carver counties in the uppermost interval of Figure 16, extending from 280 – 300 m MSL. This has important implications for the treatment of infiltration into the top of Layer 1 in the Metro Model, and was incorporated into the Layer 1 recharge numbers. Similarly, due to the consistent sand content across the lower intervals, Intervals 3 and 4 were combined into the Model's Layer 1 aquifer.

The analysis was repeated using moving averages of horizontal planes in order to investigate the change in sand content in the vertical dimension. This was done by staggering the intervals five meters and reanalyzing. Contours of the number of actual values incorporated into each well location's sand content average were used to identify areas of sparse data (e.g. an elevation interval can intersect both bedrock and atmosphere). And contours could be developed of the number of switches between sand and 'not-sand' in each well log, which would highlight divisions between areas of homogeneous and heterogeneous sand deposits. Some additional analyses were performed for the project, but are not presented here because they merely confirmed

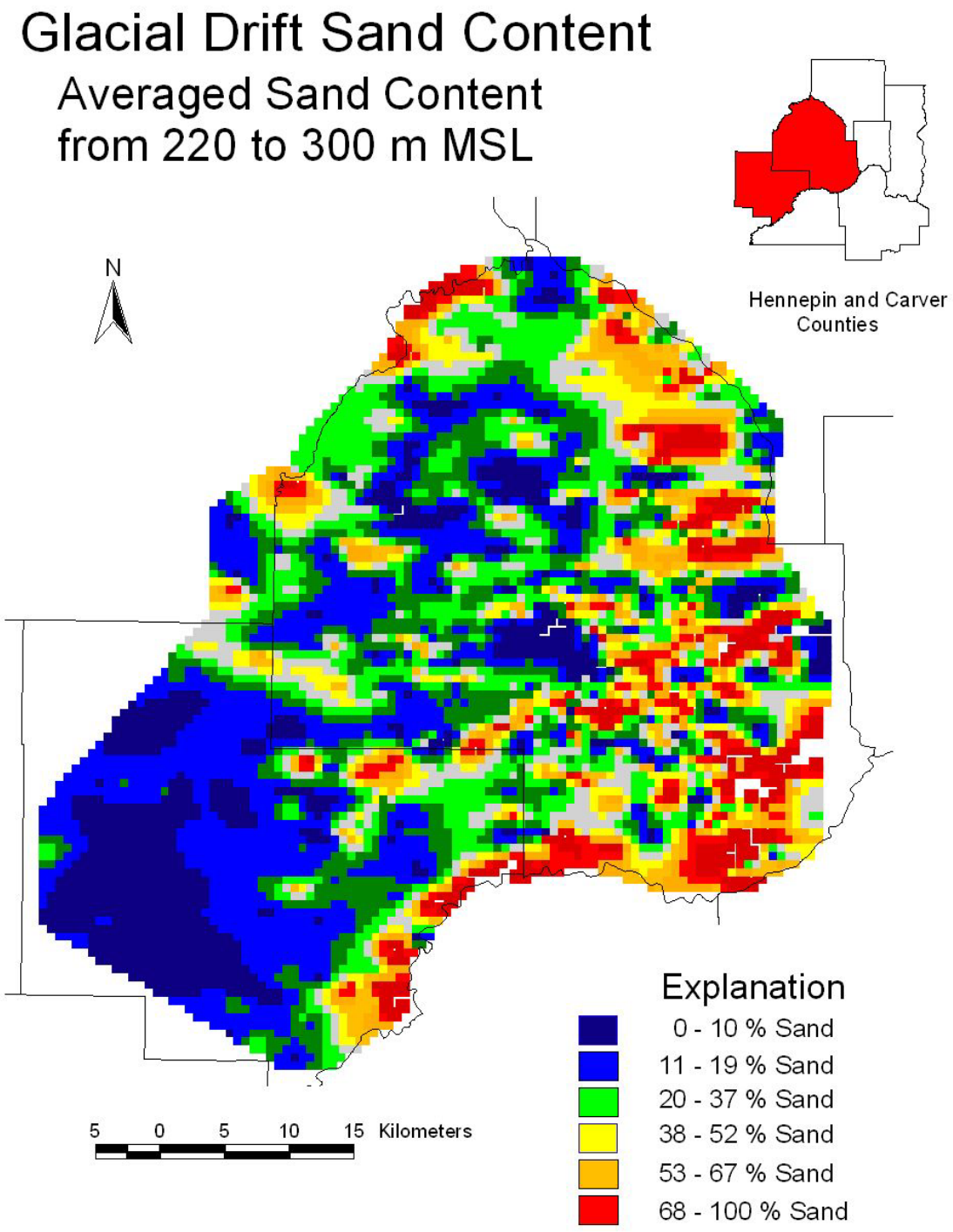


Figure 15. Averaged Sand Content for Northwest Province.

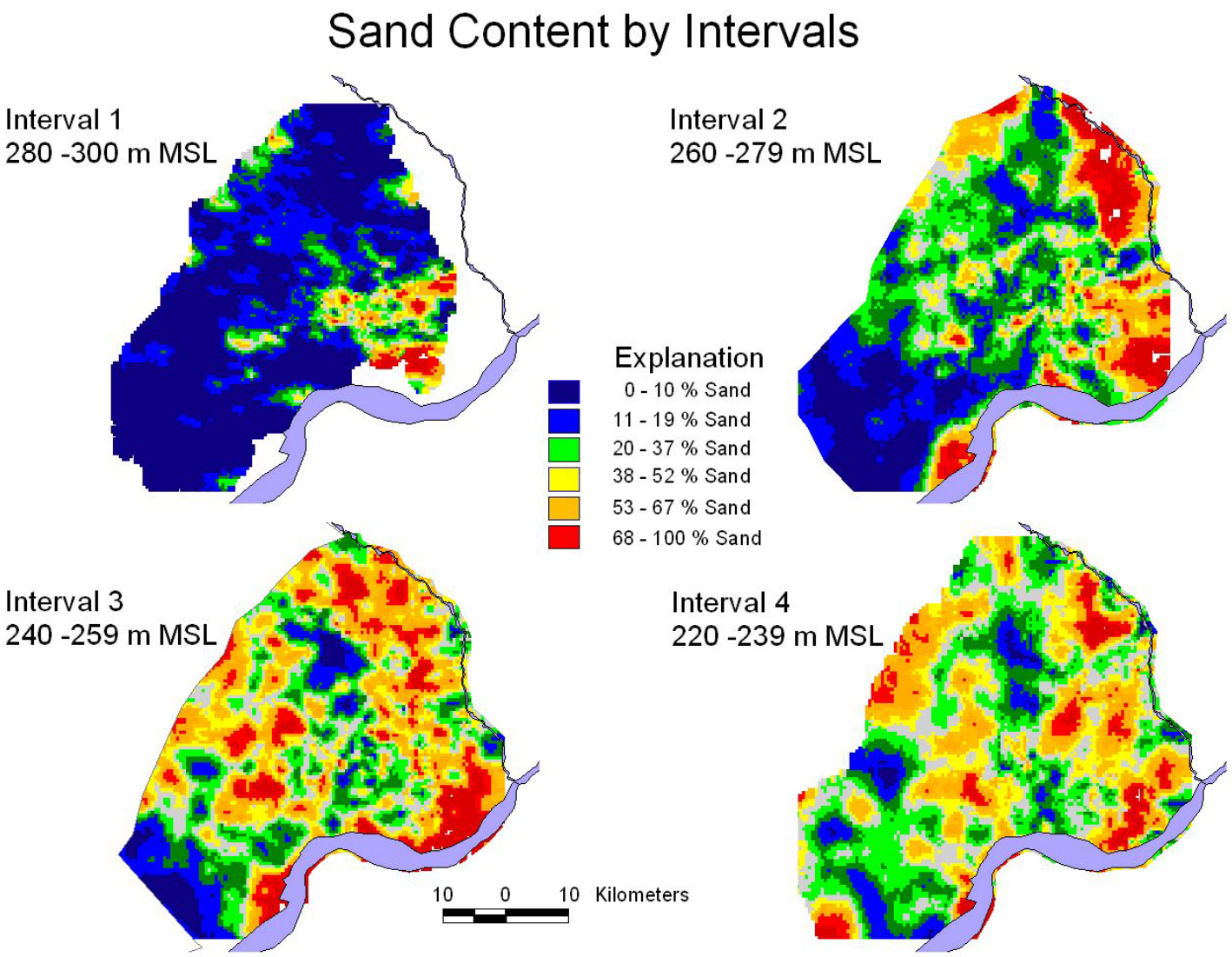


Figure 16. Averaged Sand Content for Northwest Province by Intervals

previous analyses. Finally, these results can be compared to existing geologic reports such as the MGS report "Geologic Atlas of Hennepin County, Minnesota" (MGS 1989). The pattern of high and low sand content visible in Intervals 1 and 2 of Figure 16 strongly resembles the geology of the Surficial Geology Plate (Plate 3 of 9).

Averaging Quaternary sand content data across intervals 20-meters thick helped to compensate for errors in all well elevations. The CWI database has an inherent well elevation error of up to approximately three meters due to the use of topographic maps to determine surface elevation. The topographic maps have a contour interval of 3 m (10 feet), which defines the maximum error. Because elevations of lithologic boundaries are calculated from the surface elevation, the error is propagated throughout the geologic log. Averaging across the thickness of the Quaternary deposits reduced the impact of this elevation error. Some analytic support for this level of error came from the error nugget of variogram analyses of a different media, groundwater heads for Layer 3. The nugget is an expression of the short scale variability and sampling error associated with sampling locational data. The larger the nugget, the greater the variability between closely spaced data points. The Layer 3 groundwater surface changed slowly with XY distance, and therefore the dataset of well heads was statistically very well behaved. (The dataset fits the Gaussian model of close agreement of heads among closely spaced wells and gradual change with greater distance. More information on this subject is to be found in the earlier section on variogram analyses.) The error nugget of the Layer 3 heads was approximately 9 m squared, yielding a root of 3 m, which corresponded to the maximum measured elevation error.

A comparison of different interpolating techniques indicated that variogram-guided kriging used in this analysis provided the most robust interpretation. Figure 17 presents a comparison of four different interpretations of the data found within Interval 3 of the Northwest Province, presented here in order of increasing sophistication: 1) the raw uninterpolated data, 2) natural neighbor interpolation, 3) kriging with a default variogram, and 4) kriging with a developed variogram. Options 2) and 3) are "black box" interpolators. Note the emerging pattern of sand and non-sand as the interpolators become more sophisticated. The kriging technique using the variogram developed from the actual data provided the clearest patterns, which follows from its reliance on measured correlations and weighted measurements of nearby logs. Results such as these led the team to use the kriging with a developed variogram for all spatial analyses. The kriged results yielded layouts of sand content that the team used to infer relative drift permeability that was effectively incorporated in the Layer 1 model. These relative permeabilities were also subsequently corroborated with other hydrogeologic evidence.

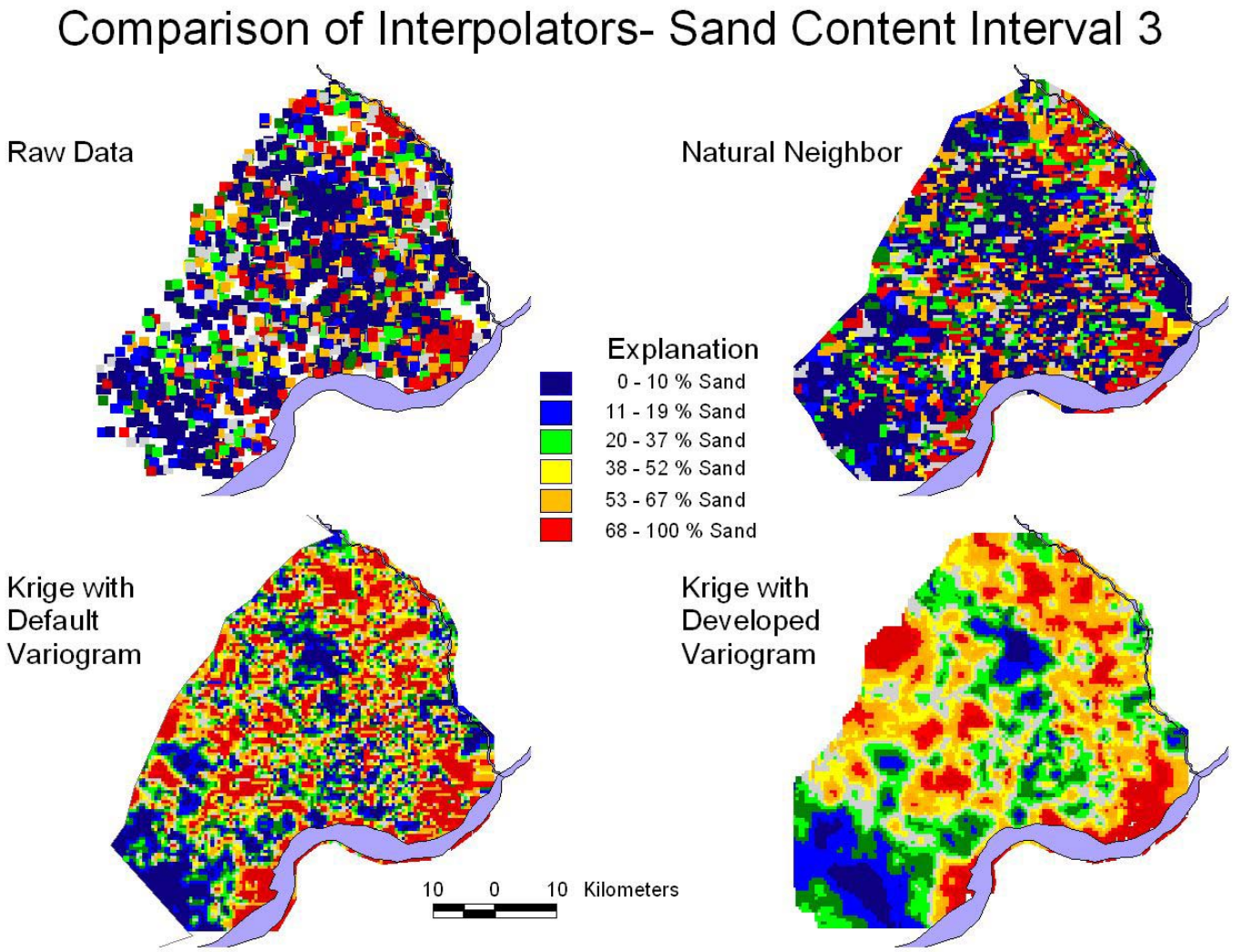


Figure 17. A Comparison of Interpolation Techniques; Northwest Province, Interval

Minnesota River Valley- A Demonstration of Data Analysis Techniques

Introduction

Up to this point the Data Report has been focused on the use of data analysis techniques to investigate the hydrogeology of the Twin Cities Metropolitan Area. However, these techniques are independent of groundwater models and properly stand on their own. To make clear that these tools can be used in geologic provinces different from the Twin Cities, as well as to provide geologic databases for an adjoining part of the State, we chose to investigate the Quaternary geology of the Minnesota River valley. The area chosen extends from the western boundary of the state into the southwest portion of the Twin Cities metropolitan area. Though this area lies outside the Twin Cities metropolitan area and therefore is not part of the Metro Model, the techniques used here are very similar to those used to develop the Metro Model supporting databases. It also demonstrates the applicability of the data analyses methods outside the metropolitan area.

The goal of this analysis was to characterize the hydraulic properties of the drift aquifer to better understand how the discharge of contaminated groundwater might affect surface water quality. This area is highlighted in Figure 18. Maps were prepared to represent surface elevation, thickness of the Quaternary cover over bedrock, and the average sand content of the top 100 feet of Quaternary material (the portion of the Quaternary that is assumed to be most affected by contamination).

Figure 18. Study area

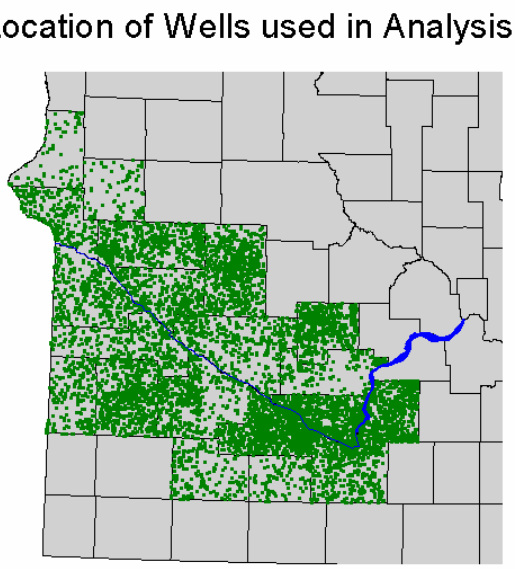


Figure 19. Well locations

Different analyses we re performed in this demonstration to provide graphical representations of surface elevation, depth to bedrock, and sand content.

Surface Elevation

The first analysis undertaken was to produce a surface elevation, which involved building a three-dimensional surface from the *x*, *y* and *z*- coordinates available in CWI, where *z* is the surface elevation at the well in feet above mean sea level. As of August 2002 there were over 330,000 wells in CWI, of which fewer than half have been field located and geologically interpreted. Of these, approximately 10,000 are located in the area of interest. It is these wells that were analyzed (Figure 19).

Database Queries

The best way to handle this large dataset was with database tools. CWI has well information grouped into ten different tables combined in ACCESS for filtering. Three of these tables were required to produce the elevation analysis with a single query, joined through the "RelateID" field, which is common to all CWI tables, and is the unique well number assigned by the MGS. The tables are:

Name	*General description*
C4ST	Geologic Unit information, including lithology types and elevations.
C4IX	Elevation information such as surface, depth to bedrock, bedrock, etc.
WWPT_N83	Coordinates in UTM NAD83

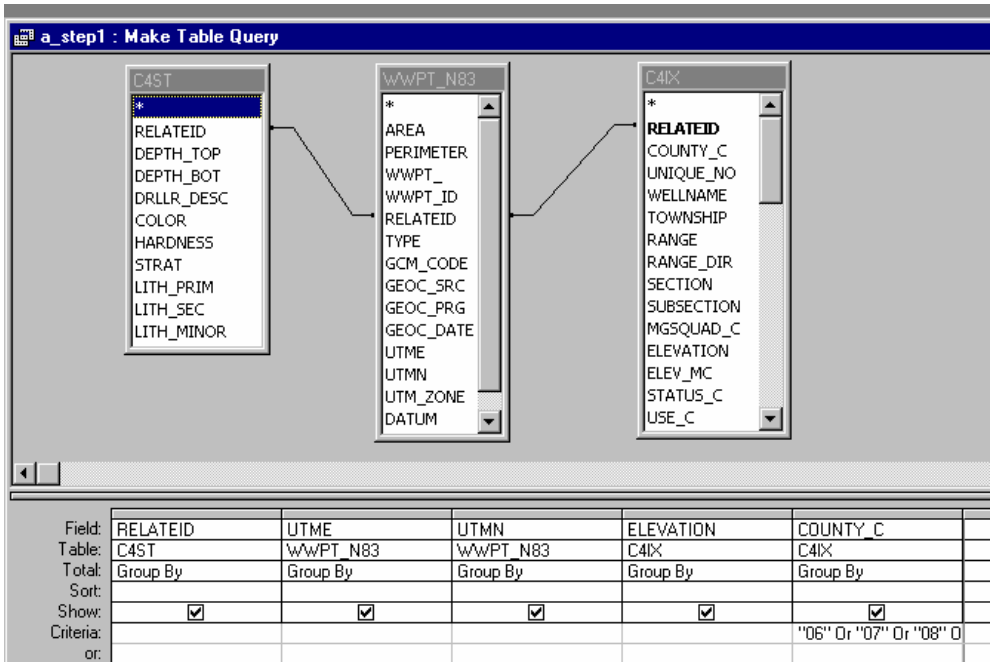


Figure 20. Query design in ACCESS

The act of tying these tables together with a common field limited the results of the query to only those entries (rows) that had the same RelateID codes in all three tables. This query is presented in Figure 20.

Refinements added to the query ensured that only one record is returned for each Unique number, and only wells falling within the target counties were selected. The query output is a table which when exported to GIS yielded the display seen in Figure 19.

Spatial Correlation- Variogram

The data obtained from the database query presented in Figure 20 was analyzed to determine spatial correlation, which is mathematically expressed as a variogram. Kriging was employed to apply the variogram to interpolate values on a grid. This technique accounts for both measured correlation and variability of the parameter—in this case, surface elevation—to be used along with weighting of nearby observed values to interpolate a value at each grid point node. Graphical depiction of gridded interpolated values (contours, for example) make it easier for the human eye to discern underlying data patterns. More information on variograms in particular and geostatistics in general can be found in the Data Analysis Techniques section.

In order to analyze the data spatially it was first exported to a program that can develop a variogram, such as SURFER. Figure 21 shows a variogram plot of the elevation

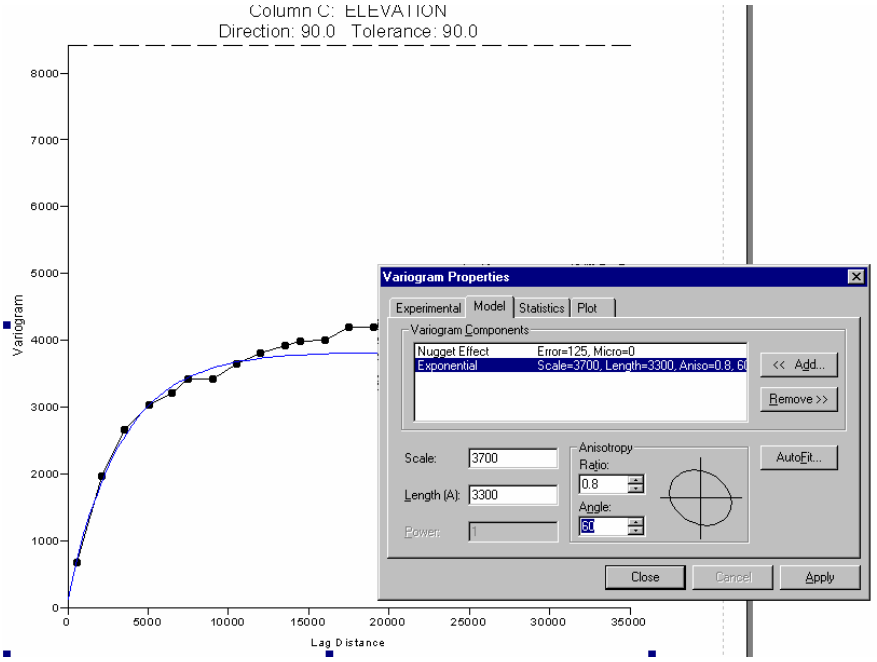


Figure 21. Variogram Analysis of Elevation Information

data collected in the database query. The variogram analyzes the elevations as a function of spatial location, and this particular variogram shows northwest-southeast orientation to the long axis of correlation, the longest direction over which surface elevations at each

well correlate. In this case the variogram (and the curve fit to the data) reveals that this correlation extends beyond 20,000 meters (20 km) along the major axis.

Following the kriging of values to a grid, the data may be exported to GIS for visual review. Figure 22 shows the results of the elevation analysis. Grid nodes, which on average are spaced 700 meters apart, are colored lighter or darker depending on the interpolated elevation. Higher elevations are darker colored. Visual inspection of Figure 22 reveals that the surface elevation of the study area forms a trough with the river running down the center. A second observation is that the elevation of the river valley drops as it nears the Twin Cities metropolitan area. Finally to the southwest can be seen the higher ground of the outer part of the Coteau des Prairies (Setterholm 1995). The Coteau des Prairies is comprised of glacial deposits several hundred feet that likely sit on top of a bedrock upland, presumably comprised of Cretaceous sedimentary rocks (Wright 1972). Though not directly related to an investigation of the Quaternary geology, this analysis provides information useful to understanding the geomorphology.

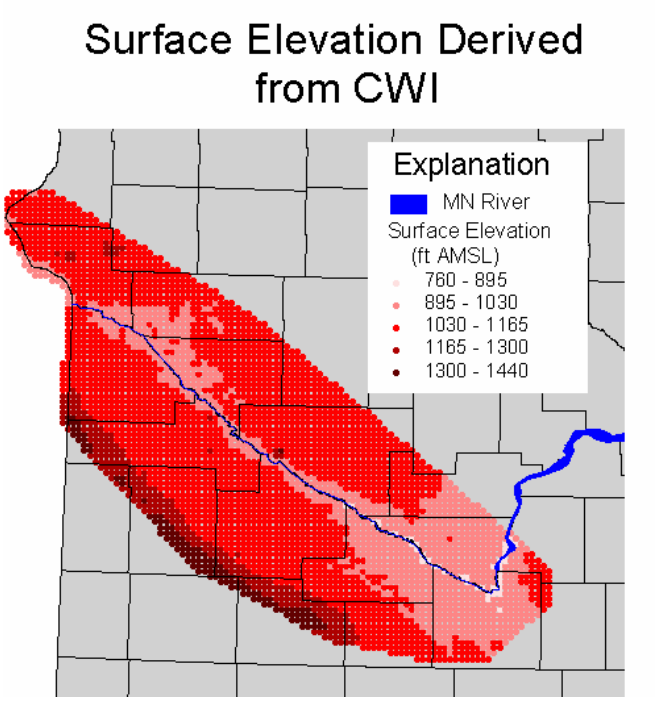


Figure 22. Surface Elevation

Other properties that can be defined with a numeric value can also be analyzed and displayed in this fashion. This same type of analysis was used to develop displays of thickness and sand content of the glacial drift materials, as presented in the following sections.

Quaternary Thickness

The preparation of a glacial drift thickness map (Figure 23) was only slightly more complicated than the surface elevation analysis. In the case of the thickness map the Z coordinate was the field labeled “depth-to-bedrock” in CWI, which is the thickness of the Quaternary materials in the special case where bedrock is encountered. Partially penetrating borings were not used in this analysis. The database queries resemble those used to produce surface elevation and will not be discussed further in this example.

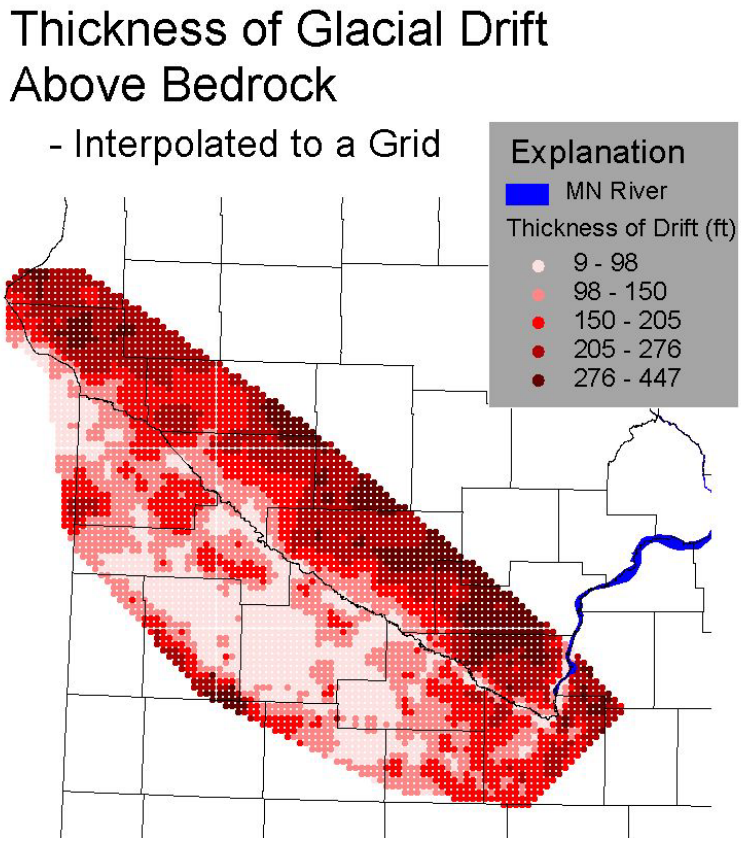


Figure 23. Thickness of Glacial Drift

The variogram analysis reveals the following about the thickness of glacial drift in the Minnesota River valley region:

- The long axis of of the correlation ellipse points northwest-southeast, which means it is the direction in which thickness of the glacial drift is the most constant.
- The correlation between wells in the northwest-southeast direction extends 5,000 meters (5 km). Past this distance the relationship between thickness of glacial drift between wells becomes random.

Finally, a review of Figure 23 shows the glacial drift is thicker north of the river, and thinner south of the river, though from inspection of other databases it is clear that the glacial drift thickens over the Coteau des Prairies.

Sand Content

As defined earlier, sand content is the term used in this report to describe the percentage of sand that could be found at a given location by averaging grain-size lithology types in the geologic column. This involves only the unconsolidated material above bedrock. It builds on the Quaternary thickness analysis, employing the same spatial statistical

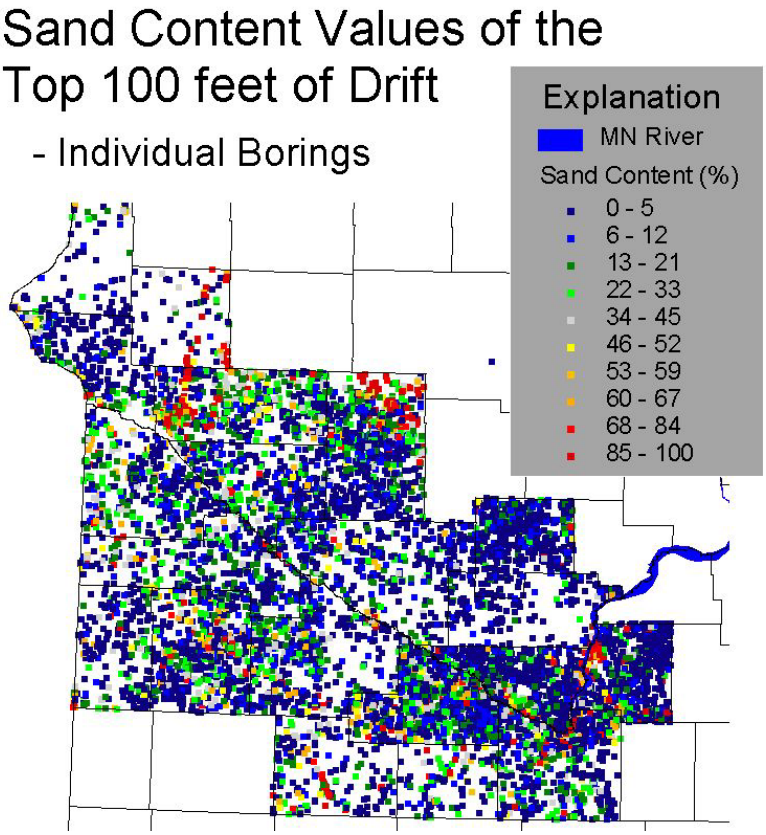


Figure 24. Sand Content Values- Individual Borings

techniques that, in effect, provide a spatial probability map of sand content. For this analysis the decision was made to include only the top 100 feet of geologic information from each boring. The database portion of this analysis was quite involved as it required a clear set of assumptions about what constitutes useable information, each of which generates one or more queries in the database. For example, unlike fields such as "Elevation" and "Depth2bedrock" which are unambiguous in providing a specific value, finding sand content requires that specific custom-built queries look for:

- Depth to bedrock,
- Maximum depth reached when bedrock is not encountered,

- The number of log entries recording alternating lithology sequences,
- Top and bottom elevations of each sequence, and
- Selection and translation of appropriate lithology codes into the sand content format.

Mean sand content values were calculated for over 10,000 wells, to a maximum depth of 100 feet (less than this where bedrock was within 100 feet of the surface). The sand content values for individual borings produced by the database ranged from 0 (no sand) at one extreme to 1 (all sand) at the other, and are displayed in Figure 24. The next steps included performing a variogram analysis on the sand content values, kriging values to a grid, and exporting the grid to GIS. The resulting analysis is presented in Figure 25.

Figure 25 shows clearly demarcated regions of low sand content that can be used for investigations ranging from infiltration and sensitivity studies, to a large scale groundwater model. The power of this statistical technique lies in its ability to take thousands of wells with complex well logs and produce an easily understood representation of a desired characteristic, in this case, sand content.

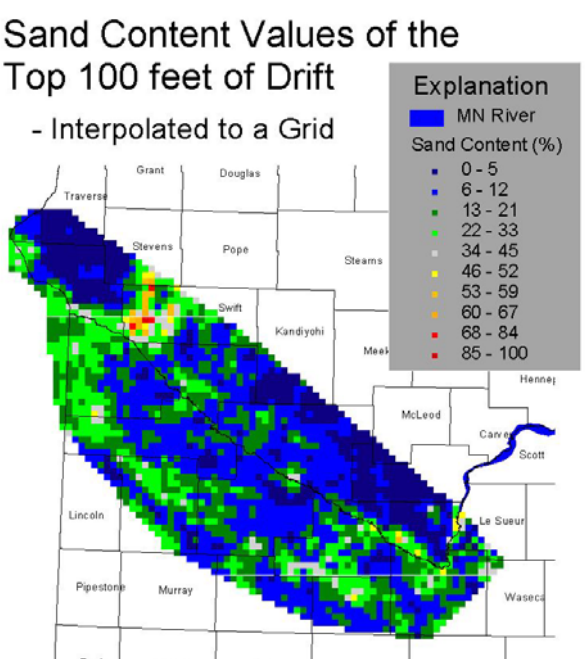


Figure 25. Sand Content Values-Interpolated to a Grid

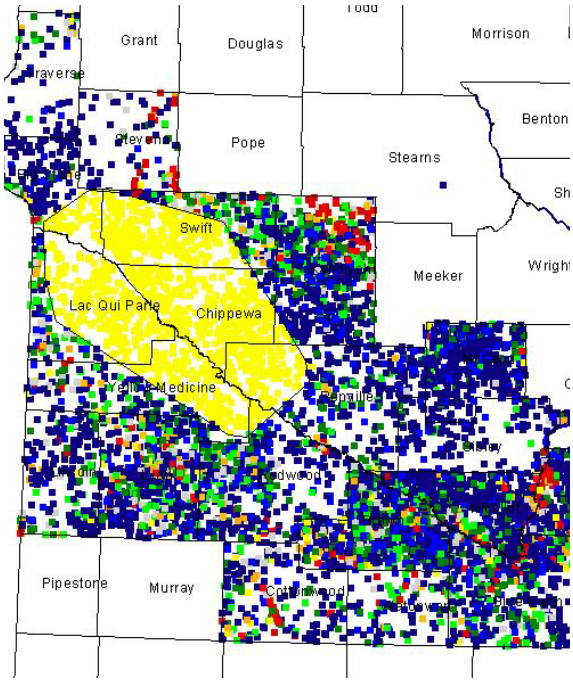


Figure 26. Focused Analysis Area

A refinement of this technique involves focusing the analysis to increase the resolution in an area of interest. This can be done either by interpolating new data points to a finer grid (Figure 25 consists of grid nodes placed approximately 3.2 kilometers apart), or by selecting a subset of the database output file and submitting these points to a new variogram analysis. The latter can yield superior results especially if the smaller area is part of consistent geologic terrane. An example would be an outwash plain where a high level of correlation between data points could be expected. If the area of the analysis encompasses too complex an area (for instance an outwash plain and a neighboring till

moraine) then the overall correlation between points declines because two different correlation patterns are superimposed onto each other. Figure 26 shows the selection of a subset of the larger dataset, filtering to include only borings from counties in the northwest portion of the Minnesota River valley, including parts of Big Stone, Chippewa, Kandiyohi, Lac Qui Parle, Redwood, Renville, Swift, and Yellow Medicine Counties. This reduces the dataset to be analyzed from 10,000 to 1,600 points.

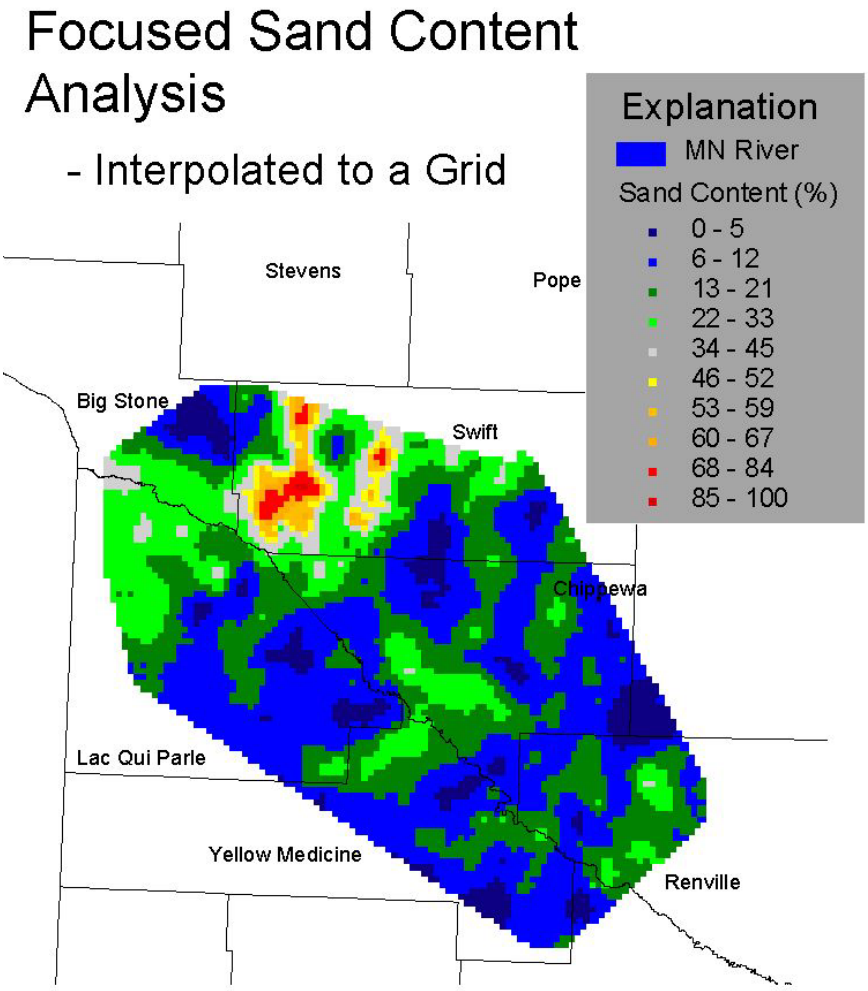


Figure 27. Results of Focused Analysis

Figure 27 shows a more finely resolved analysis of the Northwest corner of the Minnesota River valley, from borings selected in Figure 26. The grid spacing is now approximately one kilometer apart, and the variogram is marginally different due to the different datasets involved. A comparison of Figure 27 to Figure 25 shows the utility of this type of focused analysis.

Groundwater and Spatial Statistics

Introduction

Data analysis techniques described in this report up to this point have been focused on preparing databases to be used in the construction of a groundwater model. This section is concerned with describing the techniques used to build calibration datasets based on static water level data. These datasets are used to test the model and its assumptions, and potentially point to areas that need improvement. Because of the importance of this step the selection of inappropriate or unrepresentative head measurements from wells could have serious consequences for the model. Both the CWI database and spatial analysis have been discussed earlier, and it will now be shown how they can combine to produce filtered and representative datasets that simplify the task of building a groundwater model.

Ideally a calibration dataset would be based on a dedicated observation well network that had sufficient wells and adequate spatial distribution to meet basic data requirements for each aquifer modeled. Unfortunately no such network exists in the Metropolitan area. The DNR's Obwell Network offers high quality information, though the coverage is not ordinarily sufficient for calibration purposes. The only database that has sufficient coverage for calibrating the regional Metro Model is CWI. (The DNR network did provide wells for the Metro Model's Layer 5, Mt. Simon/Hinckley aquifer calibration dataset, but that owed more to the sparse number of that aquifer's wells in CWI than to anything else.) As was discussed earlier in this report, the short-comings of the CWI database can be minimized through the use of spatial statistics. In addition to the variogram there is another spatial tool that is useful in preparing a calibration dataset-- cross validation.

Cross validation

Cross validation is a process for comparing observed data against estimated values for the purpose of investigating just how representative the observed values are of the geographic neighborhood where they are found. In this case the value in question is groundwater elevation. For this project the first step in the cross validation process was the calculation of a variogram for the entire groundwater elevation dataset. An example of such a dataset would be the Layer 3 groundwater model, Northwest Province, where all the wells initially selected for the calibration dataset were located within the domain of this province. Typically this has meant that several thousand wells were analyzed to produce the variogram. Following the development of the variogram a subset of ~850 wells was then submitted to the cross validation process. This step calculates an estimated head for each well location using the variogram and all wells in the neighborhood except for the well in question. The weight of each neighboring well head elevation used in the estimate depends upon the variogram characteristics (see the previous section on variograms). In order to avoid problems associated with clustering (over-representation of an area in the database), an

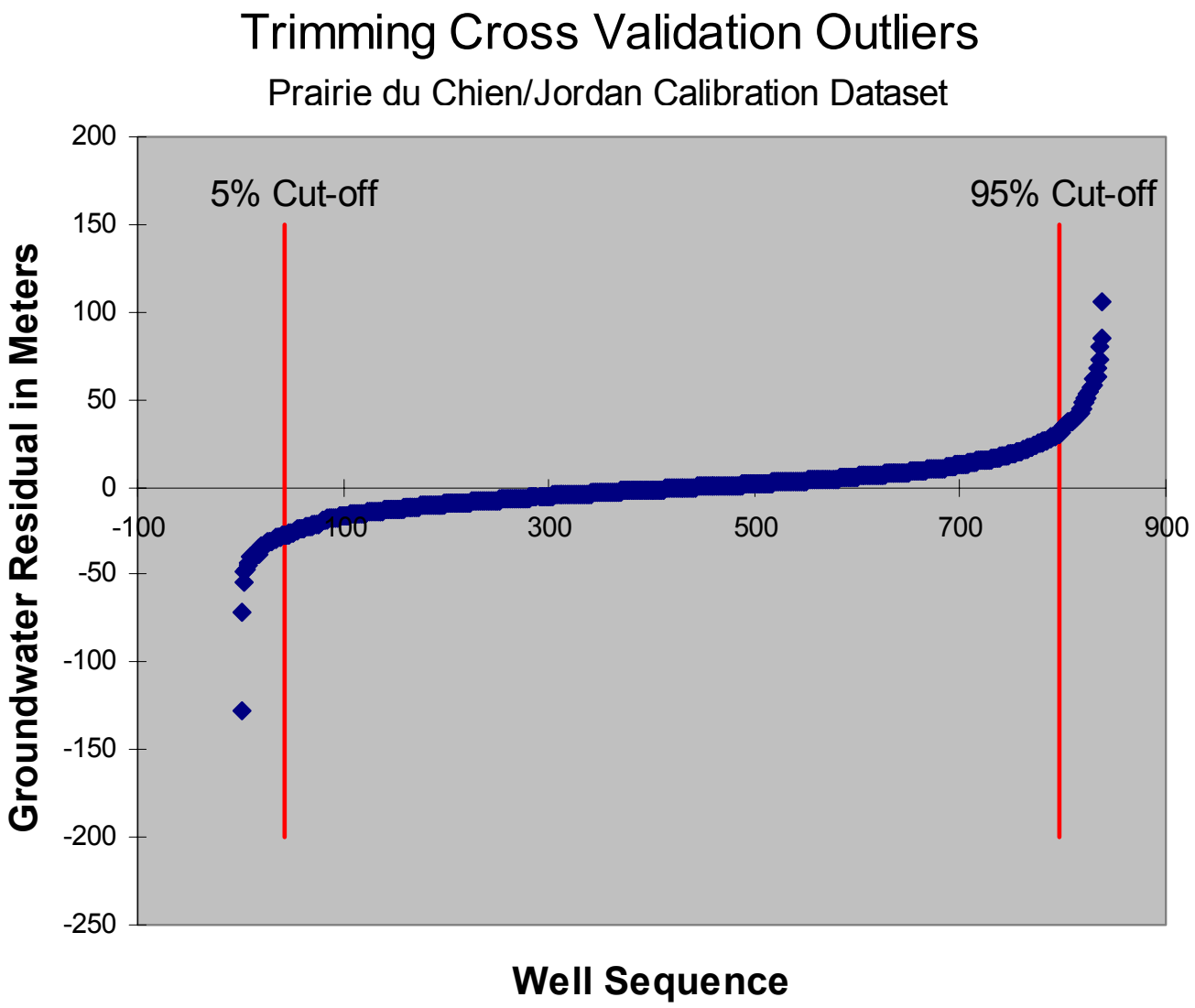


Figure 28. Ranked Residuals Outliers

exclusion zone of 100 m was used around each well in the cross validation procedure (see larger discussion of declustering later in this section). The cross validation process was repeated at every well location. The differences (residuals) between the observed and estimated elevations for all wells were then ranked and displayed (Figure 28). The Prairie du Chien/Jordan aquifer in the Northwest Province is used as an example.

Notice that the groundwater elevation residuals range from -130 to 110 meters. The decision made on this project was that a well with a relatively larger residual is a target for removal from the calibration database. This is based on the assumption that where a particular well head is not in good agreement with neighboring wells, this is either due to a measurement and/or recording error, or is caused by the presence of a local-scale hydrogeologic inhomogeneity in the vicinity of the well. Either case is undesirable for a regional-scale calibration dataset. Lastly, the possibility that a well's groundwater elevation is anomalous because of its proximity to an undocumented pumping well or other hydrologic feature, points out the need to investigate outliers before discarding data.

Upon completion of the cross validation 10 percent of the wells were discarded, the 5 percent of outliers (the wells with the largest absolute difference between observed and estimated values) at each extreme. This use of a cutoff is somewhat arbitrary, as a 1 or 20 percent cutoff could also be defended depending upon the shape of the outlier ranking curve. The advantage of using cross validation with CWI is that even after removing close to 100 wells from this dataset, there were still approximately 750 wells that were spatially well-distributed across the domain and representative of regional groundwater trends.

Bias in the selection of a cutoff criterion for outlier heads was investigated by graphically displaying the location of the outlier wells in the various hydrogeologic maps assembled for the project. One sign of bias would include the preferential grouping of outliers along bedrock margins, river boundaries or other hydrogeologic features. If wells labeled as "outliers" were found in large numbers in these spots this would suggest that the trimming process was incorrectly constructed, that it was identifying wells in areas of steep groundwater gradients. In the Metro Model calibration datasets, however, the well outliers were found to be evenly distributed geographically across the three hydrologic provinces.

The quality of the kriging estimate can also be expressed as the kriging standard deviation, which is a measure of the variation in the population of values located within the variogram search ellipse. This provides a basis to judge the difference in the observed and estimated values. If the kriging standard deviation is small and the difference between the observed and estimated parameter is large, one explanation is that the single groundwater elevation is anomalous, and that a human error is the cause. If instead the kriging standard deviation value is large, it may suggest the presence of a large anomaly that requires further investigation. The size of the

anomaly would be on the order of the variogram ellipse which defines the correlation distances.

Declustering

Clustered wells are useful for the variogram-based spatial investigation as they provide important clues on how the variable under study changes with small changes in *x*, *y*-distance. They can be a problem during cross validation however, requiring some special consideration. By adjusting the minimum distance input field to instruct the program to ignore wells within a certain radius when estimating a head value at each well, wells are compared to neighbors beyond the immediate cluster, where they would conceivably be in closer agreement. The intent is to move beyond inhomogeneities of a desired scale. For example if a well with a small residual from a small minimum distance instead has a large residual when the minimum is increased, that well may be a good candidate for trimming.

Head Calibration

Head calibration datasets were prepared for most of the individual models that make up the Metro Model using the spatial techniques described above. These datasets along with all other databases prepared for the Metro Model are available on the Agency website at:

http://www.pca.state.mn.us/water/groundwater/metromodel.html

Summary

Development of the Metro Model required a solid foundation of information and data regarding the hydrogeology in order to construct a defensible conceptual model. Some regional information and data derivations were already available for use, such as MGS' bedrock geology maps. However, other project information needs not met through existing resources required additional analysis by project staff to provide the regional information required for Metro Model development. Typically, these analyses involved a multi-step process, starting with raw data from CWI. The primary strength of the CWI dataset was that it contained tens of thousands of data points. However, there are two potential drawbacks to its use: 1) overall, it is not an high-quality dataset, and 2) the data were not necessarily collected for purposes that satisfy the objectives of the Metro Model. Yet, the application of geostatistical techniques to such a large population of data points produced robust interpretations on a scale appropriate to the regional scale used for the Metro Model development. These interpretations include sand content maps of glacial drift materials, piezometric surfaces, calibration datasets and other datasets listed in the report. The data analysis techniques developed for the Metro Model have utility beyond the narrow uses presented, and are accessible to the reader through popular technical software programs. Caution is urged to avoid "black box" manipulation, and to perform reasonable and defensible operations on the data.

Acknowledgments

The Metro Model was initially supported from 1995 through 1999 through a special allocation by the Minnesota Legislature as recommended by the Legislative Commission on Minnesota Resources, with additional support coming from the U.S. Environmental Protection Agency, and the MPCA. The program was essentially terminated in 2001, with only minor work continuing after that date, including the completion of this report.

The work products provided through this project would not have been possible without the help and support of many individuals who have contributed to the Metro Model project since it began in 1995. The dedication and efforts of the main project staff, John Seaberg, Doug Hansen, and Yuan-Ming Hsu have been instrumental in development and construction of the Metro Model and its supporting databases and conceptual model.

Special gratitude is extended to Don Jakes, an early supporter of the project within the MPCA and who provided supervisory oversight during the project's first five years. A special thanks to University of Minnesota Professor Randal Barnes, who provided the technical basis for much of the geostatistical treatment of data. And thanks to Bob Tipping at the MGS, a fellow student of Randal's, for his informed review and comment of many of the geostatistical approaches used for this project. And finally, thanks to the many members of the project Users' Advisory Workgroup, who were so helpful in their review of the many datasets created for the project . Their peer-review and feedback has immeasurably improved the work products of the Metro Model.

References Cited

Bear J. and A. Verruijt, 1990, Modeling Groundwater Flow and Pollution, D. Reidel Publishing Company, Dordrecht Holland, 414 p.

Isaaks E. and Srivastava R.M., 1989, Applied Geostatistics. Oxford University Press, New York, 561 p.

Mossler, J.H. and R.G. Tipping, R.G., 1996, Bedrock topographic map of the seven-county metropolitan area, Minnesota Geological Survey, unpublished manuscript map, scale 1:100,000, one digital file.

Mossler, J.H., and Tipping, R.G., 2000, Bedrock geology and structure of the seven-county Twin Cities metropolitan area, Minnesota, Miscellaneous Map series M-104, Minnesota Geological Survey.

Seaberg, J.K., 2000. Overview of the Twin Cities Metropolitan Groundwater Model, Ver. 1.00, Metropolitan Area Groundwater Model Project Summary, http://www.pca.state.mn.us/water/groundwater/mm-overview.pdf, 62 p.

Setterholm, D.R., 1995, Regional Hydrogeologic Assessment. Quaternary Geology – Southwestern Minnesota, Series RHA – 2, Part A, Plates 1 & 2, University of Minnesota.

Tipping, R.G. and Mossler, J.H., 1996, Digital elevation models for the tops of the St. Peter Sandstone, Prairie du Chien Group, Jordan Sandstone and St. Lawrence/St. Lawrence-Franconia Formations within the seven-county metropolitan area: Minnesota Geological Survey, unpublished manuscript maps, scale 1:100,000, four digital files.

Wright, H., 1972, Geology of Minnesota: A Centennial Volume, P.K. Sims and G.B. Morey (editors), Minnesota Geological Survey, University of Minnesota, St. Paul, Minnesota; 1972.