

# Pollution Detection: Modeling an Underground Spill through Hydro-Chemical Analysis

James R. Garlick  
Savannah N. Crites  
Earlham College  
Richmond, IN 47374

Advisors: Mic Jackson and Tekla Lewin

## Summary

Data from ten monitoring wells in a region of suspected underground pollution are used to assess the source, time, and amount of pollutant released into the ground. The chemicals are sorted based on changes recorded in their concentrations over time to determine which were active pollutants during the data collection period and to account for discrepancy caused by an incomplete data set. Those chemicals found to be active during this time period change concentration simultaneously, indicating that each chemical is a component of a single leaking liquid involved in two major spills. The concentrations of selected active chemicals are combined to form a composite indicator whose concentration value is found at each well on each date. The composite indicator reveals that two spills occurred, the first between July 1991 and March 1993, and the second between January 1995 and April 1997, possibly continuing until the end of the data collection period. The primary chemical constituents of the leaking liquid are identified.

A Delaunay triangulation is used to interpolate a gradient of concentration for the composite indicator at each date between the monitoring wells. Given that the general flow of groundwater in this region is directed toward well 9, the time and location of the pollution source can be approximated based on changes in the concentration gradient over time. This spill is estimated to have originated in the region surrounding the point (8000, 4500). Following the initial triangulation, Voronoi polygons are used to construct a convex hull

---

*The UMAP Journal* 20 (3) (1999) 343–354. ©Copyright 1999 by COMAP, Inc. All rights reserved. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice. Abstracting with credit is permitted, but copyrights for components of this work owned by others than COMAP must be honored. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior permission from COMAP.

representing the total volume and position of the spill (the volume of the contaminated area). This polygon is comprised of smaller segments, each of a specific uniform concentration. The program Geomview is used to generate graphics of these polygons and convex hulls. A volume can be calculated at each concentration, and ultimately the total volume of polluting liquid can be found, if the concentration of the composite indicator in the original polluting liquid is known.

Finally, various testing and interpretation methods are explored and incorporated into a procedure for evaluating underground pollution. Each method is discussed in terms of its application to the scenario in Problem Two and uses information given in the data set to test the validity of the method.

## Introduction

Given the location and elevation of eight groundwater monitoring wells (two more wells exist at unknown locations), a complete chemical analysis taken periodically at each well between 1990 and 1997, and the general direction of groundwater flow, it is possible to accurately estimate the location, source, time of origin, and total volume of pollutants seeping underground. In the case of a suspected leak in a chemical storage facility built over homogeneous soil, cost and safety prohibited collection of analytical data directly below the suspected sight of the spill. Data from monitoring wells surrounding the periphery of but not necessarily directly in the suspected polluted region are used in a mathematical model to determine whether a leak has occurred, the time and location when the leak occurred, and the amount of liquid that has leaked during the data collection period.

## Assumptions

- All monitoring wells are located below ground and are contained within an aquifer (a geological unit capable of storing and transmitting substantial volumes of water). This aquifer has an unobstructed constant flow rate which is inversely proportional to the porosity of the soil medium. The monitoring wells are permanent, allow free flow through their measuring devices, have no effect on the chemical or geological composition of the region, and provide an accurate reflection of the surrounding area. This ensures that the wells themselves do not contaminate or pollute the region to be assessed [Soliman et al. 1997, 32].
- The volume of fluid is constant in each well, and all wells have the same volume. Assuming a consistent volume between wells allows direct ratios to be assessed comparing concentrations of solutes in each well.

- Different chemicals may travel through the aquifer at different rates. Chemical substances have a constant and specific ability to move in aqueous solutions depending on polarity of the molecules, hydrophobicity, and the initial concentration of each compound.
- Some chemicals found present in the data set occur naturally in the groundwater and are not products of pollution. Any chemical that exhibits no significant change in concentration at any monitoring well over the course of the data collection period can be removed from consideration in the data. In addition, certain naturally appearing chemical components of groundwater can be expected to fluctuate between standard levels.
- Concentrations of pollutants are highest near their source, and concentrations decrease as time and distance from their source increases.
- The given data set is incomplete. Some trends may be misrepresented or missed entirely due to lack of available data. Also, the values that are given must be appropriately evaluated so as not to treat the N/A values as zero.
- Discrepancies in the data can be attributed to variations in the equipment used or in sampling and analyzing techniques over the course of the study and should not always be interpreted as changes in the environment, especially those occurring on the same data in every sample tested.
- *Pollution* is defined as a contaminant that is harmful to an organism, while *contamination* refers to a greater concentration of a substance than would occur naturally without necessarily causing harm [Blatt 1997, 76]. In this problem, we assume that both terms refer to the artificial contamination of an underground region, regardless of the effect that the contaminants may have on organisms.

## Dealing with the Data

To use or interpret such a large and varied data set effectively, specific criteria must be employed to organize and sort the known information. We converted the data from its original spreadsheet form into a database so that we could set up queries and selectively access any portion of the information.

Several components of the data were not chemical concentrations but other factors necessary for a thorough chemical analysis, such as specific conductivity and total dissolved solids. These were separated and stored in another spreadsheet. Although some methods of modeling pollution use these measurements, our models do not, because we could not detect a significant pattern in these values to indicate the presence or absence of pollution.

Using line graphs mapping the concentration of a given chemical at all dates and at each well, we identified chemicals that exhibited a negligible change in concentration. These were removed and stored in a separate spreadsheet. This left 23 chemicals from an original set of 106 measurement categories.

## Determining the Presence of Pollution

From the rapid increases shown in the line graphs of chemical concentrations over time, it was apparent that new pollution had occurred in this region over the testing period. Those chemicals detected as new pollutants include: acetone, ammonia, arsenic, barium, bicarbonate, calcium, chloride, iron, lead, magnesium, manganese, nickel, nitrate/nitrite, potassium, sodium, TDS, sulfate, vanadium, and zinc.

The concentration of the majority of the chemicals in the active data set rise and fall together, indicating that each is a constituent of a single liquid involved in the spill. Although the concentrations of all active chemicals in the data set follow obvious trends, the changes in concentration are much more amplified for some than for others. We chose these amplified chemicals as indicator chemicals to track the movement of the spill. To further simplify spill detection, we added the concentrations of these indicator chemicals (chloride, sulfate, and nitrate/nitrite) together to form a composite indicator chemical, the concentration of which indicates the presence of pollution at each test site on a given date. We chose these chemicals also because they are common components of pollutants and are often used to monitor pollution [B.C. Ministry of Environment, Land, and Parks 1999].

In choosing chemicals to serve as indicators for a spill, it is essential to find chemicals that were measured consistently on the same dates and at all wells throughout the data collection period. Three chemicals in this data set that fit this criterion are chloride, sulfide, and nitrate/nitrite, and we used those in the composite indicator. Because the data set is not complete and the measurements were not taken consistently for all chemicals at all points or on all dates, it is important to ensure that the concentration of this composite indicator does not misrepresent trends in the movement of the spill due to a lack or abundance of data for a given well or on a given date. We went through the data set and eliminated dates that were recorded twice (taking an average of the concentrations listed at each well) and corrected other abnormalities in the data until each of the three chemicals had exactly one value at each test location on all dates needed. Exceptions to this include those wells for which values are not available at the beginning of the testing period; these are added as data from these wells became available.

## The Time of the Spill

A series of line graphs showing the concentration of the composite indicator at a given well over time can be used to estimate the time of the spill. When plotted together so that each line represents a monitoring well, these graphs of concentration over time show when concentrations first start to increase and at which well(s) this increase is first recorded. This record of which wells show the first rise in concentration provides a rough estimation of the location of the

source as well. [EDITOR'S NOTE: We cannot effectively reproduce the authors' graphs here in black and white.]

*Two spills probably occurred, the first between July 1991 and March 1993. During these times, the concentrations in wells believed to be closest to the spill increased dramatically, then receded back toward normal levels. The second probably began in January 1995 and continued at least until January 1997. At this time, concentrations were starting to descend, but this could result from a decrease in the rate of the spill and may not indicate that the leak stopped.*

## Locating the Source

The line graphs generated by queries from the database are extremely useful in determining the presence of a spill, the time at which it occurred, and the chemicals involved. However, finding the source of the spill is more effectively accomplished with a visual interpolation showing the concentration of the composite indicator at each well over time. This way, we can determine where the concentrations rose first and the general direction the spill moved in. Knowing the general direction of the spill, we can develop bounds within which the source of the spill must lie. This can be done in three dimensions by creating a Voronoi polygon. This method of interpolation organizes data points into a triangles with their natural neighbors and partitions areas around each known point into polygons such that an arbitrary point placed in the polygon is closer to that data point than any other. The triangulation of a map is unique and effectively weights the value of any point in the region as a function of its distance from three natural neighbors.

While the line graphs show approximate dates when a spill might have occurred and at which wells the changes in concentration were detected, the Voronoi polygon method interpolates between the known data points to show more precisely the location of the spill source. From a series of diagrams of the concentration of the composite indicator chemical at each well over a selection of dates, the progress of the spill is very apparent, and the location of the source can be found by following the flow patterns in the underground system backward from the point where the spill first occurred. [EDITOR'S NOTE: We do not reproduce the authors' maps.]

## A Procedure for Evaluating Underground Contamination

The problem of detecting the presence of underground liquids is an old one, and due to its applications in locating water sources, petroleum reserves, and mineral deposits, an abundance of information about techniques and methods is available. Drilling sampling or monitoring wells is clearly necessary at some

point to determine the exact properties of an underground region. However, such sampling and the analysis that follows is time consuming, dangerous, expensive, and has the potential to contaminate or destroy the flow of groundwater in the region. There are numerous surface or superficial measurements that help determine the most effective placement of such wells. In addition, data gained from existing wells can help determine the need for and placement of additional monitoring wells when properly applied. Several useful measurements can be gained from a surface geophysical survey before drilling a well, including gravitational, electrical, and magnetic conductance readings. These involve the passage of electrical current or magnetic fields through surface soil and measuring the drop in voltage or potential magnetism, as well as the density at a given location. By comparing the conductance of surface soils at various locations in the region, the presence of sand or gravel beds can often be detected below the surface [Walton 1970, 61]. This is useful because sand and gravel beds have a high porosity, or ability to contain free flowing fluids in the form of groundwater tables known as aquifers. These aquifers are the mechanism by which underground pollutants are contained and spread, so an understanding of the flow and direction of the aquifer is crucial to accurately predicting the location or the source of contaminants [Soliman et al. 1997, 32].

## Drilling and Monitoring A Well

Once the location of an initial well has been decided (surface measurements should indicate the presence of an aquifer), several types of wells are available. Because drilling the initial bore hole for the well is the most dangerous and expensive part of the process, permanent monitoring wells such as those used in collecting the data set for this problem are the most economical in the long term. Such a well should be capable of detecting the direction and rate of flow of fluids in the aquifer, determining the level of the water table, and providing core samples to be chemically analyzed in a laboratory. These wells must be permeable to water in the region and cannot disrupt flow in the aquifer or introduce new contaminants due to the drilling process or corrosion of the well itself over time.

## How Many Wells Are Needed?

The number of wells needed to determine the source, time, and volume of an underground chemical spill can vary widely based on the circumstances of the spill. For the models described here, a minimum of three wells is necessary. From the initial well, the direction and flow of the water system can be determined, along with the concentration values of chemicals dissolved in the groundwater. Additional wells should be drilled along the path of the water table, considering the general location of a chemical storage facility or other suspected source of contamination, if known. If contaminants are detected



by the initial well, others should be drilled “downstream” to the spill, and if no contaminants are detected, wells should be drilled “up stream,” or possibly along a different aquifer, depending on the geological constitution of the region.

When at least three wells are available to detect the flow and concentration of contaminants, the following models can be used to estimate the location of the source of any contaminants found to be present, as well as the time of a spill, and the total volume of liquid spilled. In each case, the more wells used for data collection, the more accurate predictions can be made about the spill.

## Model 1: A Graphical Approach

This model requires the assessment of chemical concentrations at a minimum of three different locations on at least three dates per location. The more wells or locations of data collection, the more accurate the model.

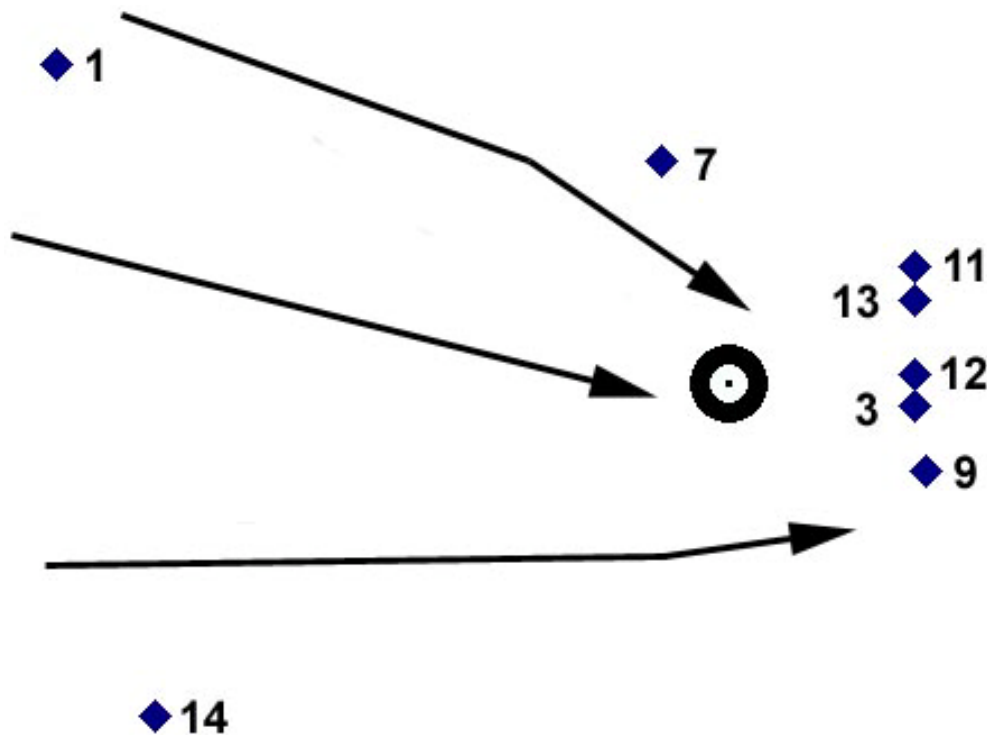
The first step is to enter the concentration values into a database so that they can be accessed by date, collection location, chemical, or concentration. Using the database, line graphs showing concentrations of all chemicals by date can be generated for each well or collection location. From these graphs, the presence of pollution can be determined, as well as the date of significant changes in concentration of measured chemicals. Dramatic increases in concentration indicate the introduction of a pollutant in this model. In many cases, as in this problem, many of the chemicals detected in a chemical analysis will rise and fall simultaneously, indicating that they are components of a common pollutant. It is possible that the chemicals would fall into two apparent groups, indicating that two liquids are leaking. In this case, one chemical, or preferably a group of detected chemicals, is consolidated to form an indicator chemical. This simplifies future graphs by allowing only one concentration value to be monitored.

In creating a consolidated indicator, it is imperative that the data be consistent. The same type of data must be available at all sites and on all dates, or adjustments must be incorporated to prevent anomalies in the data set from drastically misrepresenting the concentration of chemicals detected in the groundwater.

Having developed a consolidated indicator by adding the concentration values of representative chemicals at each site on a given date, we can generate new graphs to determine the time and source of a spill. Overlaying graphs showing concentrations over time at each well can effectively be used to determine the time of a spill. It is useful to collect enough data to develop a baseline concentration for the chemicals being measured. Tables published by the EPA, the British Columbia Department of the Environment, Land, and Parks, and other regulatory agencies list normal ranges of concentrations of various chemicals in groundwater and are also useful in distinguishing those chemicals present naturally in a system from those caused by pollution.

Interpreting these graphs is relatively simple. On a chart showing the location of each well or test location, identify those wells (if any) that never show a definite increase in concentration of the indicator. Then the well or wells reached first by high concentration values and the wells showing the highest concentration of chemical overall must be identified. Using this information, as well as the direction and rate of flow in the system (measured at each well or given in the problem statement), the contaminants can be traced back to an estimated source location.

The results of this process for each of the two spills indicated are included in **Figure 1**. Both spills reached wells 9 and 12 fastest, showing an increase at wells 3 and 11, and 7 later. This indicates that both spills began somewhere in the region surrounding the point (8000, 4500).



**Figure 1.** Location of the spills, indicated by a bull's eye, with directions of flow in the aquifer.

The delay of the spill reaching well 7 may be due to a pattern in the flow direction of the aquifer. It may at first seem strange that well 3 shows less concentration than well 7, which is clearly farther from the spill. This is probably due to the lower elevation of this well. The measurements taken at the top, bottom, and middle of the monitoring wells indicate that the spill is seeping in a downward direction, and that it never reaches the bottom of the wells. This also explains why well 13, directly in the path of the spill, never shows an increase in chemical concentration—its only samples were taken from the bottom of the well.



## Strengths and Weaknesses

The greatest strength of this method is its efficiency in interpreting a large and complex data set. Most computers are able to build and utilize such a database from a spreadsheet, and once the data are organized, the time needed to compute and interpret the results is minimal. Because it bases the placement of new wells on information gathered by existing ones and provides rough approximations with as few as three wells, the model is very efficient in terms of drilling and well maintenance. However, it provides only a very general approximation of the time and location of the source and has the potential to be greatly affected by irregularities in the positioning of the wells or uneven flow patterns in the groundwater system. In situations where samples are unavailable directly below a suspected pollution source, leaks cannot be detected until they have already penetrated into the groundwater supply, precluding attempts to stop the leak before it poses a problem to the surrounding community. This method provides no way to accurately determine the volume of polluting liquid spilled.

## Model 2: Interpolation with Triangulation in Three Dimensions

Using the same methods described above, this model requires the creation of a database and line graphs to determine the presence of a pollutant and the chemicals involved. It also uses changes in the concentration of a composite indicator of chemicals to monitor the flow of the pollutant.

Computational geometry describes a method known as *natural-neighbor interpolation* by which sets of highly irregular data can be organized and represented visually. Using Delaunay triangulation, a unique set of triangles can be arranged in an arbitrary set of points. The value of an arbitrary point is defined entirely locally based on the values of the three nearest known points, the vertices of the triangle in which the point lies.

Delaunay triangulation and Voronoi polygons are extremely useful for interpolation in this type of system for two primary reasons:

- They provide a linear system by which the value of any arbitrary point can be determined, and the original data points are exactly recovered if solved for using this system.
- The interpolation of every point is influenced only by its natural neighbors such that irregularities in the data set are reflected in the model but do not distort the accuracy of the model at other points [Sambridge et al. 1995, 3].

The computer program Geomview takes concentrations stored in the database and the location coordinates of the monitoring wells and generates a convex hull that represents of the spill as a whole. The convex hull is the outermost surface of a Voronoi polygon, comprised of smaller tetrahedrons, each called a

*datum*, representing the space between three “natural neighbors.” Each datum has a specific and constant concentration based on the known concentrations at the three points that define it [Watson 1992, 108].

The data entered into this program are divided such that only concentrations above an established baseline appear in the visual model representing the spill. The program weights each datum by volume at a specific concentration and, by adding each of these weighted concentrations, calculates the total volume of liquid in the spill. The convex hull can be used in this model to visually represent the location and volume of the entire contaminated area of the spill. We used Geomview to generate maps to show the spill defined by this data set at six dates during the data collection period and the volume of the contaminated area at that time. [EDITOR’S NOTE: We do not reproduce the authors’ maps.]

This model is also useful in determining the source of the chemical spill. It produces a new diagram on every date requested based on the concentrations entered from the data set for that date. At the start of the data collection period, no pollution is visible. As time progresses, the diagram clearly indicates which wells experienced higher than normal concentrations of the indicator chemical. The approximate source of contamination, as well as the direction of flow in the groundwater table, are apparent when several successive convex hulls are viewed together.

## Error Analysis

To test the error of this model and the linear model, a data set can be computed for which the source location of contaminant, the date on which the leak occurred, and the total volume of liquid spilled are known. The difference between predicted values and the actual location, time, and volume, divided by the actual values, determines the percentage error of the interpolation. The amount of error will depend on numerous factors specific to the individual test including how far the source is from the nearest monitoring well, the flow rates in the underground system, the size of the spill, the number of monitoring wells, and many more factors.

## Strengths and Weaknesses

This model is inherently stronger than the graphical method because it allows a three-dimensional visualization of the data and because it uses a unique and algorithmic interpolation to evaluate the presence of contaminants between known points. Delaunay triangulations and the Voronoi polygons and convex hulls that can be derived from them are extremely accurate when used to interpolate in highly irregular data sets, because of the natural-neighbor principle. The presence of irregular data points or wide variations in distribution of points is reflected in the resulting projection but does not result in a misrepresentation of the data or inaccurately skew the values of known points

[Sambridge et al. 1995]. The more data points available, the more accurate this interpolation is, because—like any interpolation—the model is most accurate nearest the known points.

Because the data set given with this problem is highly irregular and contains very few sampling locations, the predictions of this method may not be entirely accurate, but they are highly superior to most interpolation methods.

As with the previous model, this method cannot detect a spill until it has already entered the water table.

Also, the only obvious method of error analysis is to test more points. The values given are still approximations, and the process of sorting such a large data set is still rather tedious. If such a method were employed from the start of a project and data were collected in a specific and consistent manner, analysis using this model would be relatively simple using the database and the processes described here.

This model can determine whether or not a leak has occurred, approximate the time of the spill, and give an estimated location of the source of the spill. Its greatest advantage is its ability to determine the volume of contaminated area and to model its location underground. When the initial concentrations of chemicals in the leaking liquid are known, this model can also determine the volume of liquid that has leaked.

This model is extremely cost-effective, generating approximate boundaries of a spill based on known information. Successive wells should be drilled at these boundaries to ensure that the spill is in fact accurately predicted by the current information. If new wells detect additional areas of the spill, new boundaries will be generated and the process can be repeated. In either case, unnecessary wells are never drilled once a contaminated area is identified.

## Results and Recommendations

New pollution occurred in this region during the testing period between 1990 and 1997. Two separate spills of the same liquid occurred, the first beginning in July 1991 and ending approximately in March 1993, and the second beginning in January 1995 and tapering off about February 1997, although possibly continuing for the remainder of the collection period. These spills originated from the region marked in **Figure 1**.

The spill was composed of the following chemicals which were released into the ground as pollutants: acetone, ammonia, arsenic, barium, bicarbonate, calcium, chloride, iron, lead, magnesium, manganese, nickel, nitrate/nitrite, potassium, sodium, TDS, sulfate, vanadium, and zinc. The total volume of the polluted area by 1997 was approximately 32 million cubic feet. (When the initial concentration of the composite indicator in the liquid that spilled is available, it is possible to determine the total volume of liquid leaked by weighting each datum of a Voronoi polygon and finding the sum of the concentrations, multiplied by the total volume of contaminated area.)

Recommendations for future testing include:

- Identify surface properties to most effectively place the initial monitoring well.
- Test the same chemicals on the same dates at all wells to ensure complete and accurate data.
- Use the information generated by previous wells to predict the borders of the spill and place additional wells along this boundary to minimize the number of test locations needed to accurately determine the size of the spill.
- Determine the rate and direction of flow of the aquifer in which monitoring wells are located to accurately predict the time and location of the pollution source.

## References

- Blatt, Harvey. 1997. *Our Geologic Environment*. Upper Saddle River, NJ: Prentice Hall.
- British Columbia Ministry of Environment, Land, and Parks. 1999.  
<http://www.env.gov.bc.ca/wat/gws/gwbc/> .
- Flanagan, David. *Java in a Nutshell*. 1996. Sebastopol, CA: O'Reilly and Associates.
- Levy, Stewart, Tamara Munzer, and Mark Phillips. 1996. GeomView—1996.  
<http://www.geom.umn.edu/software/download/geomview.html> .
- Sambridge, Malcolm, Jean Braun, and Herbert McQueen. 1995. Geophysical parameterization and interpolation of irregular data using natural neighbours. *Geophysical Journal International* 122: 837–857.  
<http://rses.anu.edu.au/geodynamics/nn/SBM95/SBM.html> .
- Sedgewick, Robert. 1988. *Algorithms*. 2nd ed. Reading, MA: Addison-Wesley.
- Soliman, Mostafa M., Philip E. LaMoreaux, Bashir A. Memon, Fakhry A. As-saad, and James W. LaMoreaux. 1997. *Environmental Hydrogeology*. New York: Lewis Publishers.
- Levy, Stewart, , Tamara Munzer, and Mark Phillips. 1996. GeomView—1996.  
<http://www.geom.umn.edu/software/download/geomview.html> .
- Todd, David Keith. 1959. *Groundwater Hydrology*. 2nd ed. New York: John Wiley and Sons.
- Walton, William C. 1970. *Groundwater Resource Evaluation*. New York: McGraw Hill.
- Watson, David F. 1992. *Contouring: A Guide to the Analysis and Display of Spatial Data*. New York: Pergamon Press.