The Use of Principal Component Analysis for Interpreting Ground Water Hydrographs

by T.C. Winter, S.E. Mallory, T.R. Allen, and D.O. Rosenberry[a]

Abstract

Principal component analysis was used to define patterns in water table hydrographs at four small, lake-watershed research sites in the United States. The analysis provided insights into (1) characteristics of ground water recharge in different parts of the watersheds; (2) the effect of seepage from lakes on water table fluctuations; and (3) the effect of differences in geologic properties on water table fluctuations. At two sites where all of the water table wells were completed in permeable deposits, glacial outwash in Minnesota and dune sand in Nebraska, the patterns of water table fluctuation primarily reflected timing and magnitude of recharge. The water table had more frequent and wider ranges in fluctuations where it was shallow compared with where it was deep. At two sites where the water table wells were completed in sand or till, a glaciated mountain valley in New Hampshire and stagnation moraine in North Dakota, the patterns of water table fluctuations were strongly related to the type of geologic unit in which the wells are completed. Furthermore, at the New Hampshire site, the patterns of water table fluctuations were clearly different for wells completed in sand downgradient of a lake compared with those completed in sandy terraces on a mountainside. The study indicates that principal component analysis would be particularly useful for summarizing large data sets for the purpose of selecting index wells for long-term monitoring, which would greatly reduce the cost of monitoring programs.

Introduction

Measurements of water level in wells or piezometers are a routine part of ground water studies. The water level data can be used to plot hydrographs of hydraulic head versus time, which commonly are then used (1) to evaluate the response of the ground water system to either natural or human-induced stresses over the course of time at the location of that well; or (2) to select a point in time for construction of a potentiometric surface (such as a water table map) or a hydrologic section, which is a cross-sectional view of ground water flow. Hydrographs provide a view of hydraulic head over time at one point in space, and maps of a potentiometric surface or hydrologic sections provide a view of hydraulic head over space at one point in time. As a result, although most of the data are visually examined to do these types of analyses, the data generally are underused. Many measurements over time might be used once or twice to construct a water table map, or the data from only a few of the wells might be used to view the dynamics of the ground water system in specific parts of an area over time. To make additional use of data, statistical methods can quantitatively summarize all the data collected during the course of a project. One reason for doing so might be to better understand the areal distribution of various types of water level fluctuation patterns within a study area. Another reason might be to evaluate a data collection program with an eye toward reducing the number of wells that need to be measured.

The purpose of this paper is to discuss the use of principal components analysis, a multivariate statistical procedure, to interpret water level measurements in both time and space. The paper addresses the two reasons stated previously for evaluating the data in this manner: (1) to understand the areal distribution of various types of water level fluctuation patterns within a study area, and (2) to determine if fewer wells can be measured for long-term monitoring.

The data used in this study are from four field research sites in different hydrogeologic and climatic settings within the United States. The sites are near Mirror Lake in New Hampshire, Williams Lake in Minnesota, Cottonwood Lake in North Dakota, and Island Lake in Nebraska. The data consist of discrete measurements of the water level in wells over a period of 10 to 11 years. Measurements at all sites were made at least monthly during winter and weekly to biweekly during the remainder of the year. A minimum of 300 measurements of water level were made at each well during the study period. The number of wells used for this study was 50 for the Williams Lake area, 26 for the Island Lake area, 22 for the Mirror Lake area, and 26 for the Cottonwood Lake area. Although more wells are present at the sites, only those that had complete records for the period used could be included because the analytical method requires a full matrix of data. The requirement of a full matrix is a limitation of using this statistical method. However, small adjustments to the data set can sometimes be justified. For example, if a few wells are not measured on a given day, that date can be deleted from the matrix for all wells, or measurements from several dates can be grouped into one day for those times of the year when water level changes are minimal.

Site Descriptions

Mirror Lake Area, New Hampshire

The Mirror Lake drainage basin is located near the lower end of the Hubbard Brook valley in the White Mountains of central New

[a]Hydrologists, U.S. Geological Survey, Water Resources Division, National Research Program, Lakewood, Colorado 80225

Received December 1998, accepted October 1999.

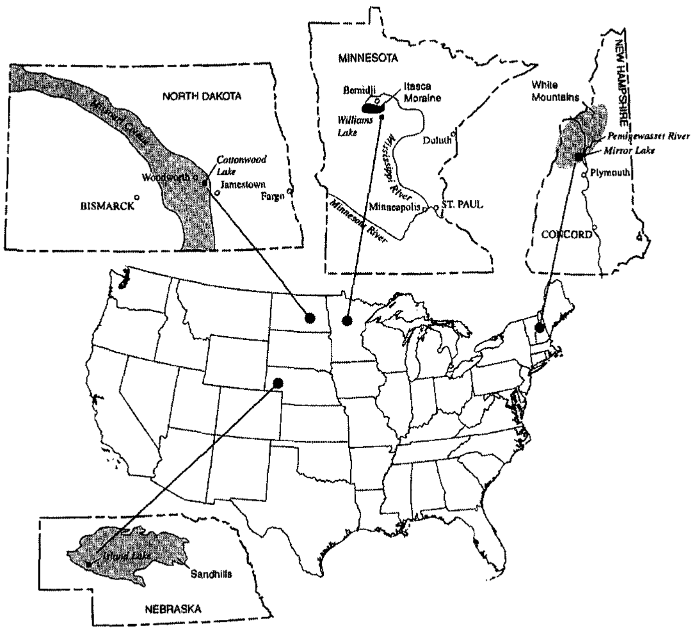


Figure 1. Location of Mirror Lake, New Hampshire; Williams Lake, Minnesota; Cottonwood Lake, North Dakota; and Island Lake, Nebraska.

Hampshire (Figure 1). The lake, which lies at an altitude of about 213 m, is bounded to the west and north by steep land slopes that reach a maximum altitude of 469 m. The drainage basin is underlain by fractured crystalline bedrock, which is overlain by glacial deposits. Thickness of the glacial deposits varies from zero to only a few meters in the higher part of the Mirror Lake drainage basin to as much as 30 m in the lower part (Winter 1984). Directly north of the lake, the glacial deposits reach a maximum thickness of about 53 m. Composition of the glacial deposits in most of the Mirror Lake area is silty, sandy till, containing numerous cobbles and boulders. Several terraces of limited areal extent and consisting partly or entirely of sand and gravel are present in the Mirror Lake drainage basin. Sand and gravel deposits are present in part of the area south of Mirror Lake, where they are as thick as 20 m in a buried bedrock valley that extends south from beneath Mirror Lake.

Williams Lake Area, Minnesota

The Williams Lake area is located on a small topographic ridge that extends south from a large east-west-trending glacial moraine, the Itasca Moraine, in north-central Minnesota (Figure 1). Physiographic features of the Williams Lake area are the result of glacial deposition, which is characterized by hummocky topography containing numerous isolated lakes and wetlands (Mooers and Norton 1997). Williams Lake lies at an intermediate altitude between Crystal Lake, which is about 4 m higher, and Mary Lake, which is about 5 m lower in altitude. Local relief in the Williams Lake area is about 25 m. The Williams Lake area is underlain by glacial deposits that are greater than 120 m thick. The deposits consist of thick alternating units of till and sand and gravel (Winter and Rosenberry 1997). The surficial unit consists of sand and gravel that is as thick as 22 m. The glacial deposits are underlain by Precambrian crystalline bedrock.

Cottonwood Lake Area, North Dakota

The Cottonwood Lake area is located on the Missouri Coteau, which is a large stagnation moraine that lies north and east of the Missouri River. The Cottonwood Lake area, situated on one of the higher parts of the eastern edge of the Missouri Coteau (Figure 1), lies about 120 m higher than the James River lowland to the east, and about 30 m higher than a small lowland within the Missouri Coteau about 3 km to the west. Local relief within the 80 ha that constitute the Cottonwood Lake area is about 33 m. The area is hummocky and has many closed depressions, most of which contain wetlands that have small surface-drainage basins. The glacial deposits in the Cottonwood Lake area are as thick as 140 m and consist predominantly of clayey, silty till (Winter and Carr 1980). The large clay content of the till causes it to crack upon drying, resulting in numerous fractures (Swanson 1990) that affect water movement through the deposits (Winter and Rosenberry 1995). Thin deposits of sand of limited areal extent are present in the Cottonwood Lake area. The glacial deposits are underlain by shale bedrock.

Island Lake Area, Nebraska

The Island Lake area is within the Crescent Lake National Wildlife Refuge, which lies within an extensive area of stabilized sand dunes in western Nebraska (Figure 1). Individual dune ridges, many of which have local relief as much as 30 m, extend for many kilometers. The dunes were formed on the Ogallala Formation, a sandy aquifer that extends throughout much of the Central Plains of the United States. The Ogallala Formation is as thick as 40 m in the study area. Shallow lakes and wetlands, which are common in the interdunal lowlands, are directly connected to this vast ground water resource (Winter 1986).

Methods

Field Procedures

The data used in this study are discrete measurements of the water level in water table wells and a few piezometers. The water table wells were constructed by augering a hole deep enough so the water table would not decline below the bottom of the well screen. For most wells in sandy terrain this depth was generally less than 2 m below the water table, but for the low permeability till in North Dakota the wells needed to be deeper. For wells completed in till, a sand pack was placed around the screen. For wells completed in sand, the sand pack generally was not needed because the sand would collapse around the screen to the level of the water table. For all wells, the annular space between the casing and drill-hole wall above the screen was backfilled with drill cuttings. For the Mirror Lake area, water level data for a few shallow piezometers were used because they were completed near the water table. The piezometers, designated as FSE-23, FSE-32, K2-21, and K3-22, were constructed by drilling a hole by the mud rotary method, inserting a screen, petal cement basket, and casing into the hole, and filling the annular space between the casing and drill-hole wall above the basket with cement. The only other piezometer constructed in a similar manner that was used in this study is well 12 in the Williams Lake area. Data for this piezometer were used inadvertently, but they were not deleted because the hydrograph pattern is similar to well 7, an adjacent water table well. Most wells have a diameter of 5 cm, but a few have a diameter of 3.2 cm. Three wells at the Island Lake site, wells 3, 4, and 5, are 10 cm in diameter. Most well screens are 45 cm long, but a few are 75 cm long. Measurements of water level were made using a steel tape.

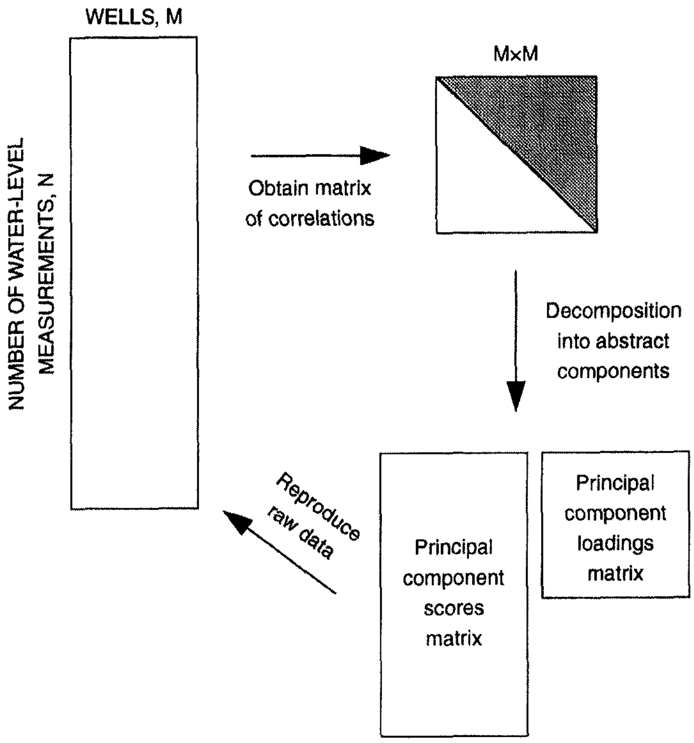


Figure 2. Schematic diagram showing the basic matrices in the principal components procedure (modified from Meglen 1991).

Analytical Procedure

The water level data were analyzed by principal component analysis, which is a multivariate statistical procedure that commonly is used to reveal patterns in large data sets. The goal when using principal component analysis is to determine a few linear combinations of the original variables that can be used to summarize the data set without losing much information. Principal component analysis quantifies the relationship between variables by computing the matrix of correlations for the entire data set. The matrix of correlations is decomposed into a scores matrix and a loadings matrix (Figure 2) by calculating and scaling eigenvectors and eigenvalues. These two matrices provide a means by which mutually independent axes that describe the data set can be derived (Meglen 1991). These axes, termed principal components, are linear combinations of the original variables that arise from the natural associations among the variables. This procedure does not require the analyst to make any assumptions about the data or variable structure, although the use of Pearson's product moment correlation coefficient to derive the correlation matrix assumes the data are multivariate normal. The procedure is described in numerous statistics books and is included in most statistical software packages. For this study, a statistical package developed by the SAS Institute Inc. (1989) was used to perform the analyses. Although the procedure is used commonly, the following is a brief description of the procedure for those unfamiliar with it. This description is taken largely verbatim from the concise description of the procedure provided by Lins (1985). The only changes to Lin's description made for this paper are the references to ground water levels instead of stream discharge and the use of the Minnesota site for the example.

For many data matrices, the range of values between variables can differ greatly; therefore, it is usually necessary to give each variable equal weight. This is done by calculating standard scores, which results in the water level variable having a mean of 0 and a standard deviation of 1. From the standardized field of about 300 water level measurements (N) and 50 well locations (M) at the Minnesota site, an M by N observation matrix, F, is formed wherein the nth column represents an M-component water level measurement vector, f_n. The identification of characteristic modes of water level variation occurs through determining which geometric form, represented by the vector e, has the greatest resemblance to all the water level vectors f simultaneously. Averaging across all f, this is accomplished by maximizing the quantity

Table 1
Percent of Variance Related to Principal Components

Principal Component	1	2	3	4	5
Williams Lake, Minnesota	81	12	2	1	1
Island Lake, Nebraska	80	11	4	2	1
Mirror Lake, New Hampshire	77	13	5	3	1
Cottonwood Lake, North Dakota	75	11	5	4	2

$$(e^TF)^2 \qquad (1)$$

subject to the condition

$$e^Te = 1 \qquad (2)$$

where e is an M-component vector representing the geometric form sought, and T denotes the transpose. Maximization of Equation 1 is equivalent to maximizing

$$e^TSe \qquad (3)$$

where S is the covariance or correlation matrix

$$S = F^TF/N \qquad (4)$$

Applying a Lagrange multiplier, λ, maximization of Equation 1 under the unit length constraint (Equation 2) corresponds to the unconditional maximization of

$$e^TSe - \lambda e^Te$$

which, when differentiated, yields

$$(S - \lambda I))e = 0 \qquad (5)$$

where I is the identity matrix equivalent in order to M, as the solution for vector e containing maximal resemblance to all water level measurements.

Solving for Equation 5 results in a set of eigenvalues, λ_k (k = 1, M), which can be placed as the elements in a diagonal matrix Λ, and a corresponding set of vectors, e_i (i = 1,M), which can be collected as columns into a matrix of principal components (E), for the covariance matrix (S). By arranging the λ_k in descending order, the corresponding principal components represent the geometric forms successively containing the highest resemblance to all water level observations, provided that each component is uncorrelated with all previously calculated components. Each eigenvalue represents the variance explained by its associated principal component. The variance explained by the first five principal components for each of the four sites is shown in Table 1.

The final step in the procedure is to identify systematic patterns of spatial and temporal variability in the original water level matrix.

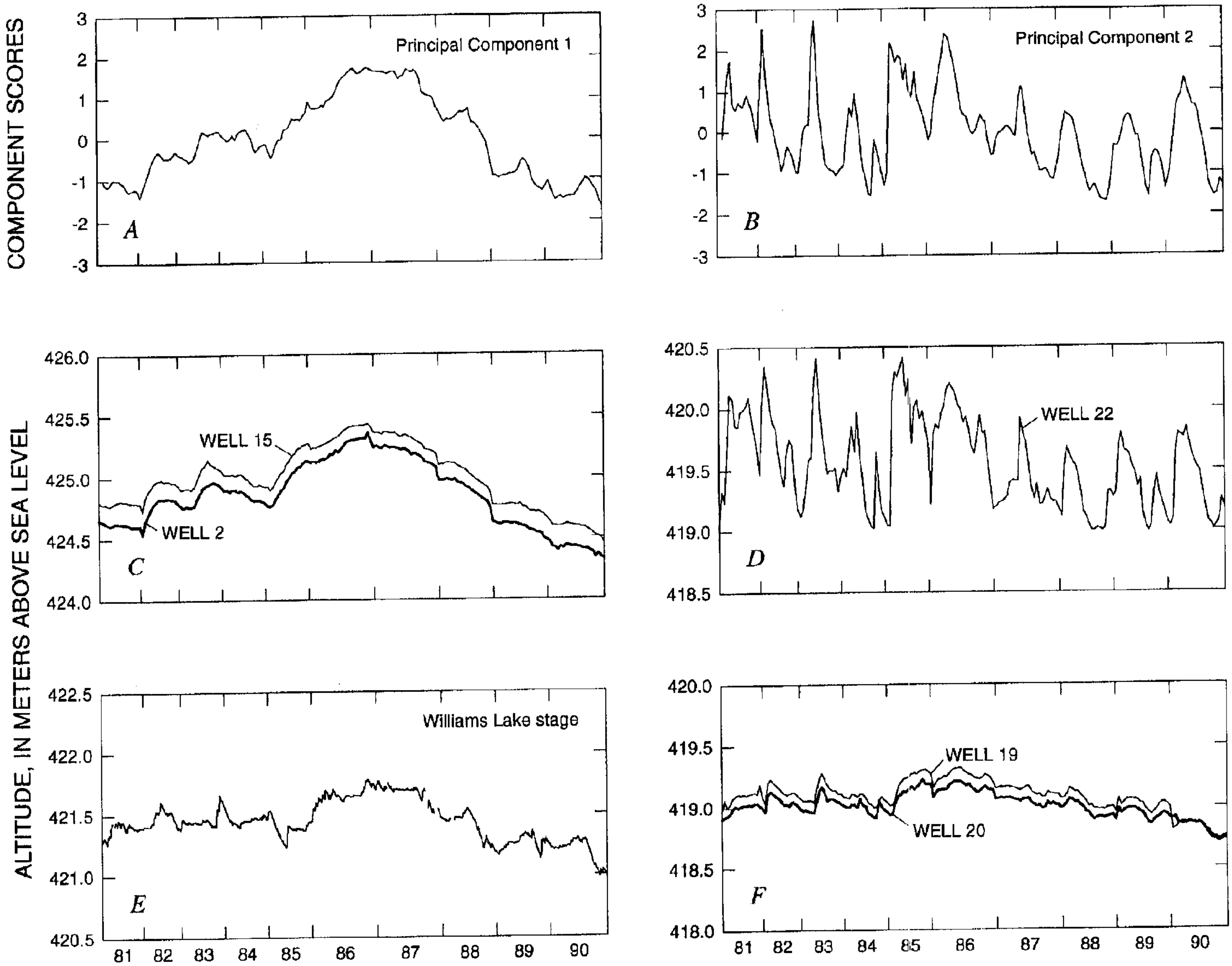


Figure 3. Hydrographs for the Williams Lake area, Minnesota: (a) component scores for principal component 1; (b) component scores for principal component 2; (c) water level in wells 2 and 15; (d) water level in well 22; (e) stage of Williams Lake; (f) water level in wells 19 and 20.

This is accomplished by calculating the component loadings and scores that reflect the underlying covariance or correlation structure of the data. That structure is inherent in the eigenvectors (e), calculated from either the covariance or correlation matrix, which are the basis of both loadings and scores.

Component loadings can be visualized as a measure of spatial similarity between the water level variables and each principal component. This similarity is expressed as a weighted relationship provided by the product of the matrices E and Λ such that

$$l_{ik} = e_{ik}\, \lambda_k^{1/2} \quad (6)$$

where l_{ik} is the loading of the kth principal component on the ith water level variable (i.e., the correlation coefficient between the kth component and the ith variable). For the complete set of variables and components, Equation 6 becomes

$$L = E\Lambda^{1/2} \quad (7)$$

and L is termed the matrix of component loadings.

Component scores are a measure of the temporal similarity between the observed pattern of water levels for a given date and each principal component. Component scores are computed as the inner product between a water level observation and a principal component:

$$c_{in} = e_i^T f_n \quad (8)$$

where c_{in} is the score of the nth observation on the ith principal component. For all observations and components, Equation 8 becomes

$$C = E^T F \quad (9)$$

and C is termed the matrix of component scores. The scores on any individual principal component will have a mean of zero, a standard deviation equal to the component's eigenvalue, and will be uncorrelated with the scores of all other components.

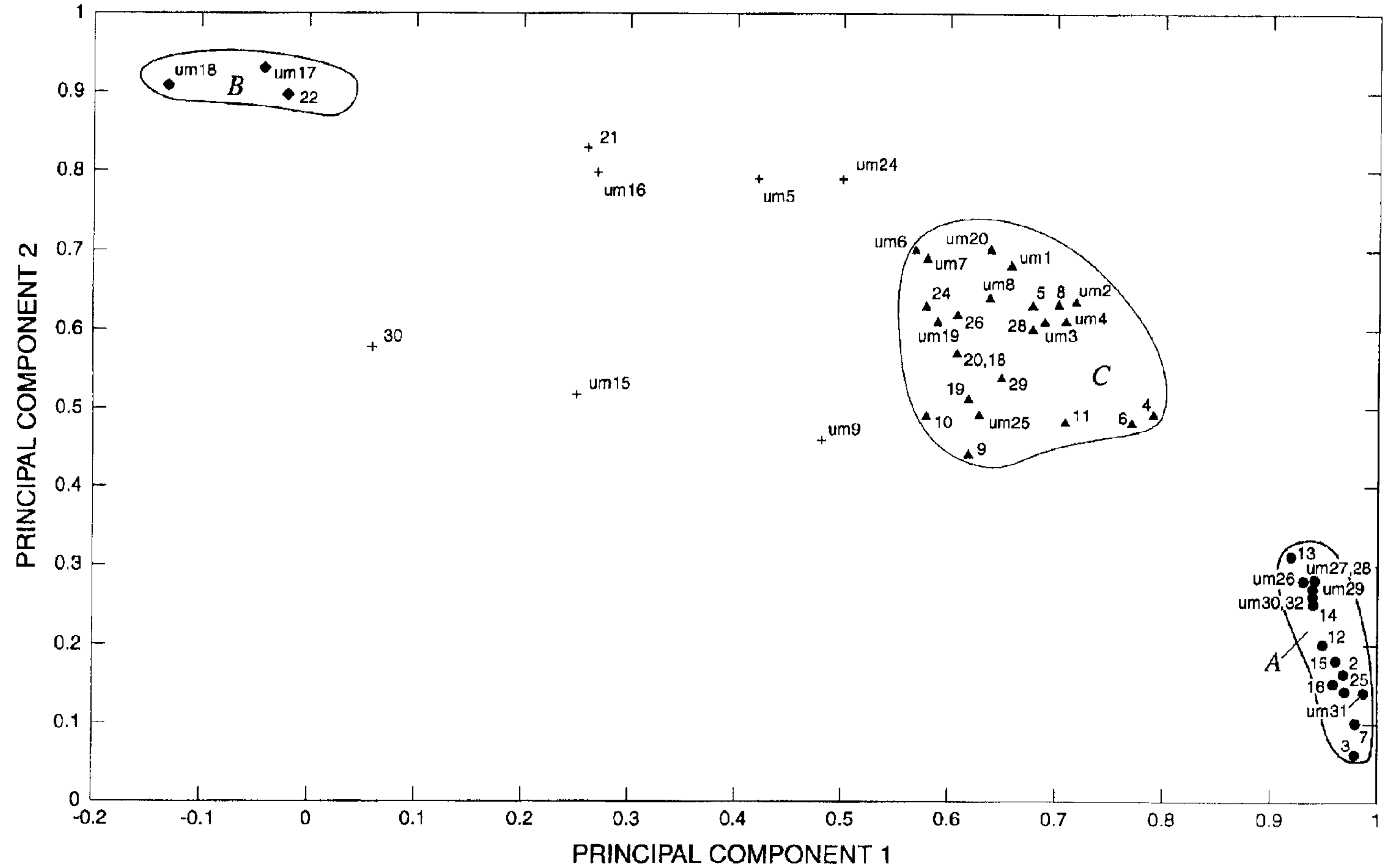


Figure 4. Scatter plot of component loadings for principal component 1 versus principal component 2 for water table levels in the Williams Lake area.

Results

Principal component analysis was selected for this study because an unbiased and efficient tool was sought that would facilitate analysis of the thousands of water level measurements that were made. The goal was to determine a few hydrograph patterns that would describe the general patterns of water level fluctuations over a 10- to 11-year period for each site. Furthermore, by determining the extent to which hydrographs at individual well locations relate to the statistically computed hydrographs, the areal distribution of hydrograph patterns could be mapped throughout the area of each of the four sites. It was anticipated that this type of information would be useful in understanding the response of the ground water sytem to natural processes, such as the magnitude and distribution of recharge, as well as indicating the relationship of water level fluctuations to contrasts in permeability of the geologic units. One benefit of this type of information is that it facilitates selection of a subset of only a few index wells that could be measured instead of the entire set with little loss of information.

All wells in the Williams Lake and Island Lake areas are completed in sand or silty sand. Wells in the Mirror Lake and Cottonwood Lake areas are completed in sand or in till. To minimize the effects of heterogeneous geology, the sites having somewhat uniform geology are presented first. The sites having more complex geologic conditions are then presented.

Williams Lake Area, Minnesota

For the Williams Lake area, the first principal component (WL-PC1) accounted for 81% of the variance in the water level data (Table 1). A hydrograph of component scores related to WL-PC1, which is a graphical representation of this variance in the data, is shown in Figure 3a. Hydrographs such as this are referred to herein as scores hydrographs. The second principal component (WL-PC2) accounted for 12% of the variance in the water level data (Table 1). A scores hydrograph for WL-PC2 is shown in Figure 3b

A plot of the component loadings for each well as they relate to WL-PC1 versus WL-PC2 (Figure 4) indicates that most of the wells fall into several groups. A large number of wells that have high loadings on WL-PC1 and low loadings on WL-PC2 on the lower right side of the diagram are designated as group A. Hydrographs of actual water levels for two of these wells (Figure 3c) indicate the close relationship of these actual hydrograph patterns to the scores hydrograph for WL-PC1 (Figure 3a). At the other extreme, a few wells that have high loadings on WL-PC2 and low loadings on WL-PC1 on the upper left side of the diagram (Figure 4) are designated as group B. A hydrograph of actual water levels in one of these wells (Figure 3d) indicates the close relationship of this hydrograph pattern to the scores hydrograph for WL-PC2 (Figure 3b). A third group consisting of many wells, group C, have relatively high loadings on WL-PC1 and moderately high loadings on WL-PC2. Hydrographs of actual water levels for two of these wells (Figure 3f) indicate some characteristics of the scores hydrographs for both WL-PC1 and WL-PC2, as might be expected from the moderate relationship of group C wells to both principal components. For ease of comparing hydrographs, the y-axis of the hydrographs for a given site has the same magnitude of scale. Furthermore, the variable spacing for each year on the x-axis reflects the number of measurements made for

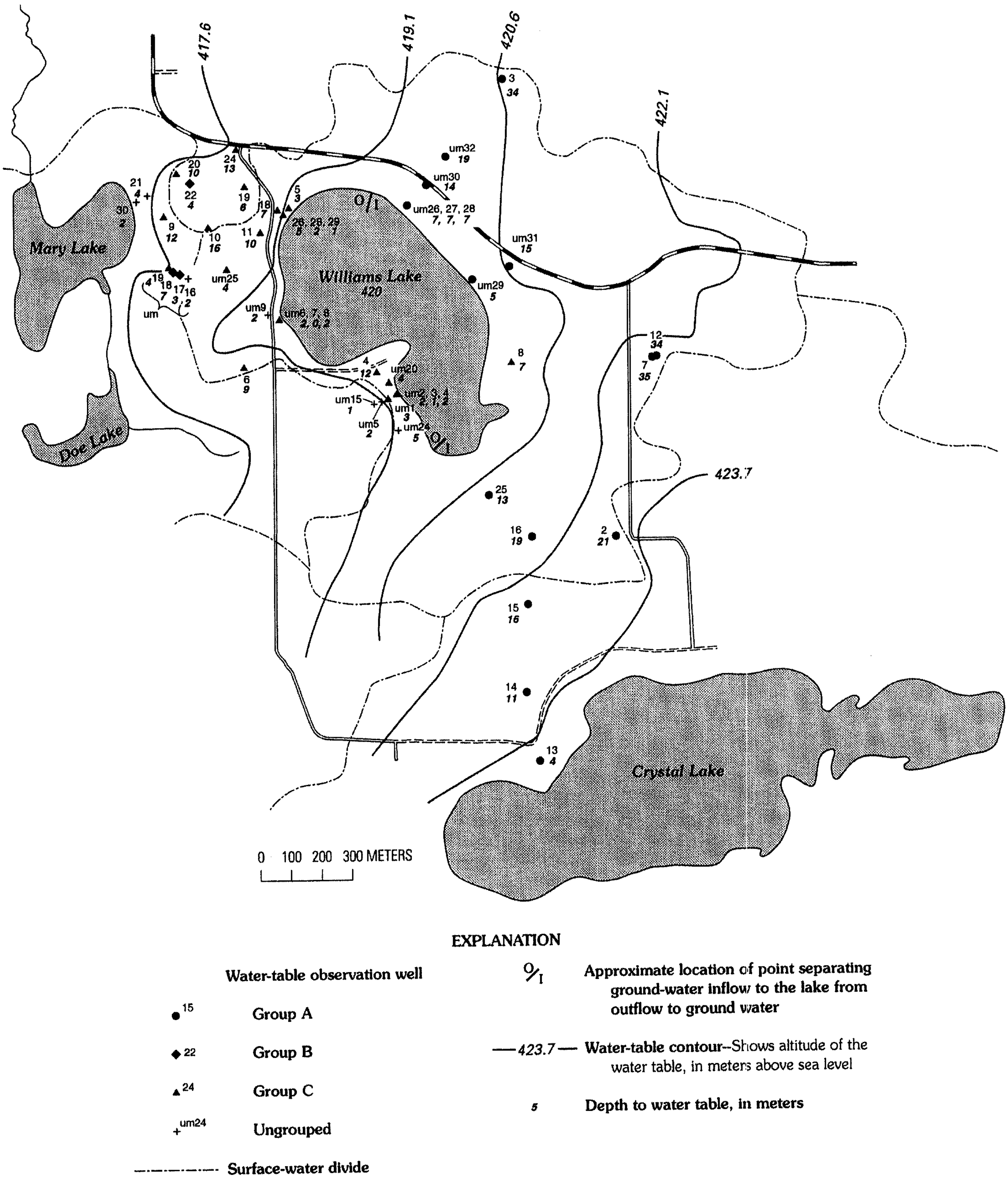


Figure 5. Configuration of the water table, depth to the water table, and areal distribution of well groups, as delineated in Figure 4, for the Williams Lake area.

a given year at each site.

A map of the areal distribution of the well groups identified in Figure 4 can be used to develop insight into the causes of the actual hydrograph patterns. For example, as shown in Figure 5, all group A wells are upgradient of Williams Lake, and the water table is relatively deep at these wells. Because of the deep water table in this area, the magnitude of individual water table fluctuations are relatively small, and they reflect more seasonal and longer-term recharge conditions. In contrast, the water table is shallow at the three group B wells. As a result, water table fluctuations in group B wells

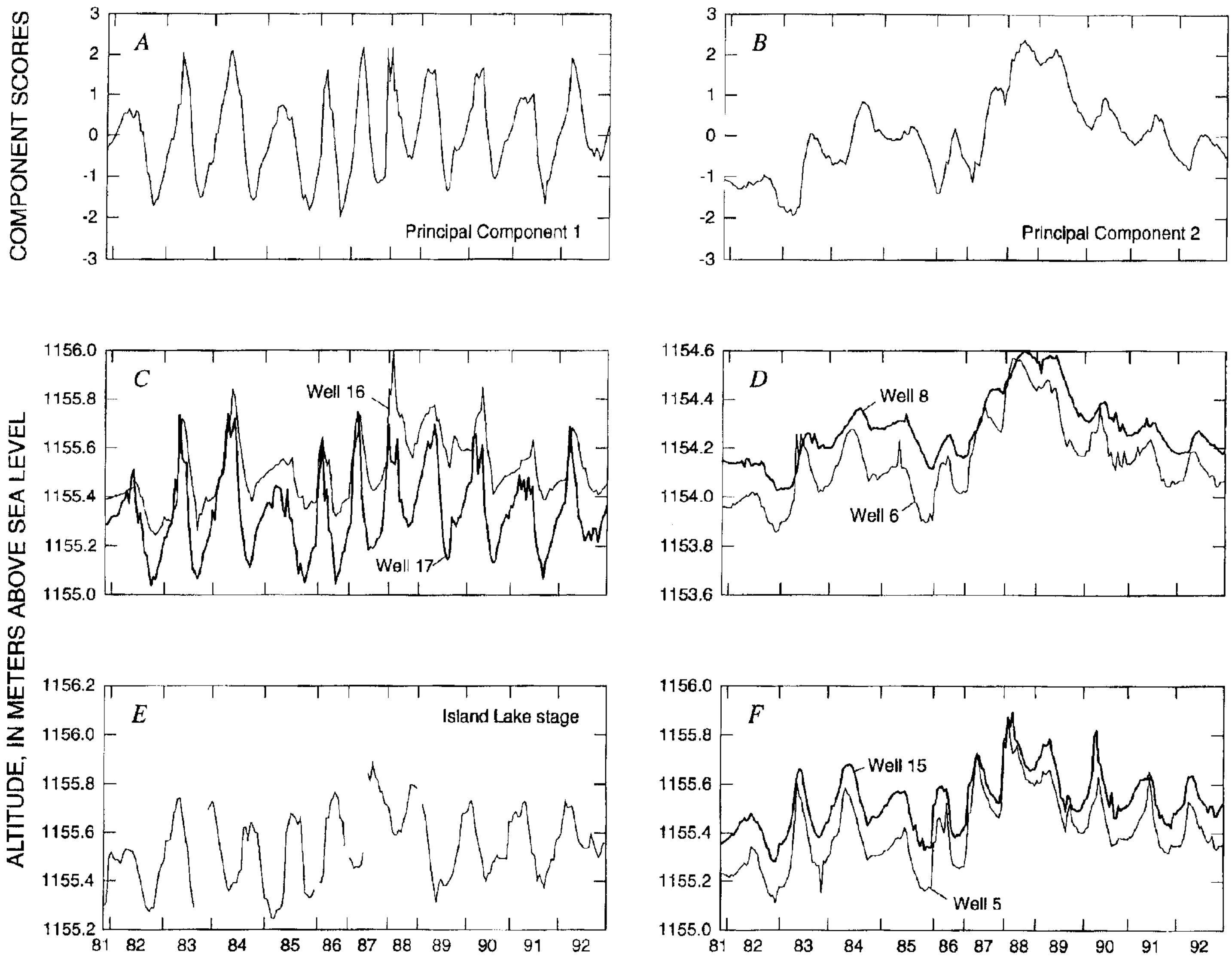


Figure 6. Hydrographs for the Island Lake area, Nebraska: (a) component scores for principal component 1; (b) component scores for principal component 2; (c) water level in wells 16 and 17; (d) water level in wells 6 and 8; (e) stage of Island Lake; (f) water level in wells 5 and 15.

are flashy; they respond quickly to recharge, and the magnitude of fluctuations is greater than those in group A wells. Group C wells, except for well 8, are all downgradient of Williams Lake, and the pattern of water table fluctuations in group C wells shows some similarity to the stage of Williams Lake (Figure 3e). Overall, the patterns of water table fluctuations in the Williams Lake area reflect the rate and magnitude of recharge, as related to (1) the depth to the water table, and (2) position of the wells with respect to being upgradient or downgradient of Williams Lake.

Island Lake Area, Nebraska

For the Island Lake area, the first principal component (IL-PC1) accounted for 80% of the variance in the water level data (Table 1). A hydrograph of component scores for IL-PC1 is shown in Figure 6a. The second principal component (IL-PC2) accounted for 11% of the variance in the water level data (Table 1). A hydrograph of component scores for IL-PC2 is shown in Figure 6b.

A plot of the component loadings for each well as they relate to IL-PC1 versus IL-PC2 (Figure 7) indicates that the wells plot along more of a continuum from the upper left to the lower right part of the graph; they do not fall into groups as distinctly as they did for Williams Lake. However, if wells 3, 4, 5, and 15 are ignored, the remainder of the data fall into two general groups of relatively widely scattered points. One group, A, has loadings greater than 0.6 on IL-PC1 and less than 0.6 on IL-PC2, and the other group, B, has loadings greater than 0.8 on IL-PC2 and less than 0.45 on IL-PC1. Several actual hydrographs of water levels in group A wells (Figure 6c) indicate the close relationship of these hydrographs to the scores hydrograph for IL-PC1. Two hydrographs of group B wells (Figure 6d) show the close relationship of water level fluctuations in these wells to the scores hydrograph of IL-PC2 (Figure 6b). For both groups, wells were selected for plotting to indicate the similarity of the hydrographs even though they plot some distance apart within the group on Figure 7. Wells 5 and 15 plot between groups A and B on Figure 7. Hydrographs for these wells (Figure 6f) indicate the intermediate pattern of fluctuations between groups A and B. Furthermore, they have a strong similarity to the hydrograph of the stage of Island Lake (Figure 6e).

A map of the areal distribution of the well groups identified in Figure 7 for the Island Lake area is shown in Figure 8. At all group A wells, the water table is relatively shallow, generally about

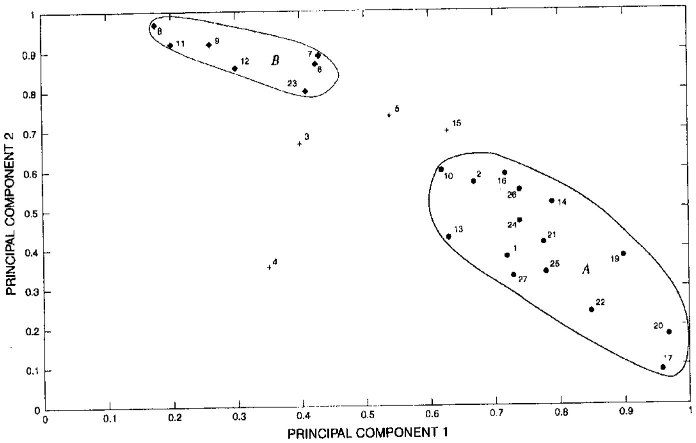


Figure 7. Scatter plot of component loadings for principal component 1 versus principal component 2 for water table levels in the Island Lake area.

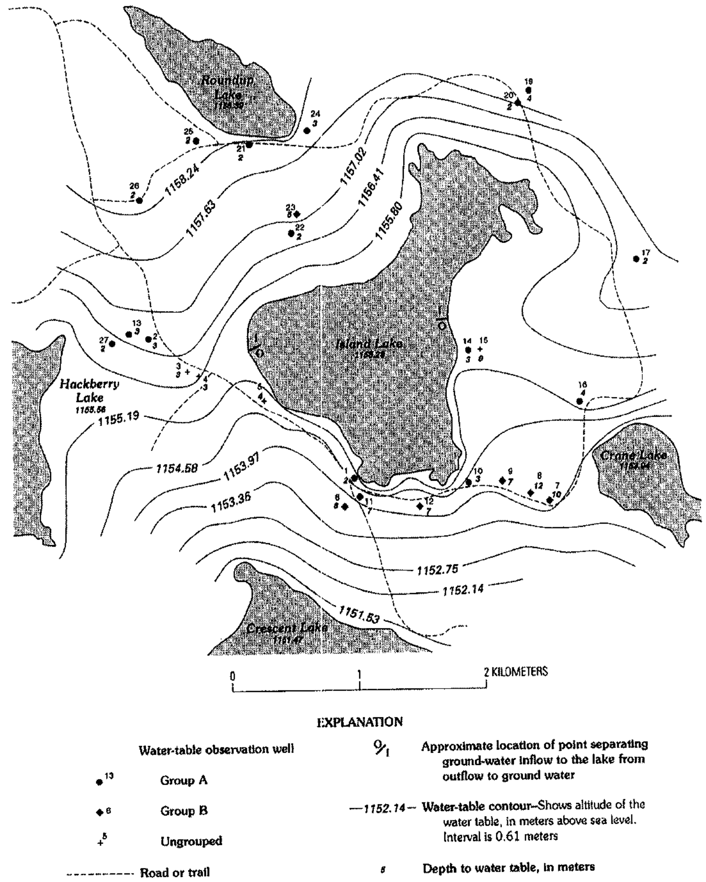


Figure 8. Configuration of the water table, depth to the water table, and areal distribution of well groups, as delineated on Figure 7, for the Island Lake area.

3 m deep or less. In contrast, the water table is relatively deep at the group B wells. Therefore, the explanation for the hydrograph patterns in the Island Lake area is similar to that for the Williams Lake area. Where the water table is shallow (group A), water table fluctuations are flashy; they respond quickly to recharge and the magnitude of fluctuations are significantly greater than those in wells where the water table is deep (group B). At the group B wells, the magnitude of individual water table fluctuations are relatively small, and they reflect more seasonal and longer-term recharge conditions. The intermediate position of wells 5 and 15 in Figure 7 may be a result of the relatively deeper water table at these wells compared with other nearby wells. In fact, well 15 could possibly have been included in group A. It is interesting to note that wells 3, 4, and 5 are the 10 cm diameter wells. In contrast to the Williams Lake area, Island Lake does not seem to affect the hydrograph patterns of the water table downgradient of it. At Island Lake, the depth to the water table seems to be the most important factor in determining water table fluctuation patterns.

Mirror Lake Area, New Hampshire

the Mirror Lake area, the first principal component (ML-PC1) accounted for 77% of the variance in the water level data (Table 1). A hydrograph of component scores related to ML-PC1 is shown in Figure 9a. The second principal component (ML-PC2) accounted for 13% of the variance in the water level data (Table 1). A scores hydrograph for ML-PC2 is shown in Figure 9b.

A plot of the component loadings for each well as they relate to ML-PC1 versus ML-PC2 (Figure 10) indicates that most of the wells fall into two distinct groups: Group A has loadings greater than 0.7 on ML-PC1 and less than 0.4 on ML-PC2; group B has loadings greater than 0.55 on ML-PC2 and less than 0.5 on ML-PC1. An actual hydrograph of water level in a group A well (Figure 9c) indicates the close relationship of this hydrograph pattern to the scores hydrograph for ML-PC1. It also shows some similarity to the hydrograph of the stage of Mirror Lake (Figure 9e). This similarity is most evident for 1986, when the lake level, the scores hydrograph for ML-PC1, and the hydrograph for group A wells declined little during the summer. The hydrograph of Mirror Lake stage has a different appearance from the other lakes in this study because it is the only one of the four lakes that has a surface outlet; therefore, high stages are cut off when lake water flows over the outlet structure. An actual hydrograph of a group B well (Figure 9d) shows the close relationship of water level fluctuations in this well to the scores hydrograph for ML-PC2 (Figure 9b).

A map of the areal distribution of the well groups for the Mirror Lake area is shown in Figure 11. A water table map was not prepared for the Mirror Lake site because the mountainous terrain limits the avalability of sites to drill observation wells. However, the data available indicate that the configuration of the water table is somewhat similar to the configuration of land surface in this steep mountainous terrain. All group A wells are located downgradient of Mirror Lake, they all are completed in sand or sandy till, and the depth to the water table is less than 5.5 m. The outlier near group A on Figure 10, well K2-21, also is downgradient of Mirror Lake, but it did not fall tightly into group A. The reason for this may be that it has the greatest depth to the water table of all the wells downgradient of the lake. Although this is not evident on the map because of rounding, the water level is about half a meter deeper at well K2-21 than it is at well 10. All group B wells are located on the hillside upgradient of Mirror Lake. The group B wells can be separated into two subgroups (Figure 10): group Ba wells, which have loadings of at least 0.8 on ML-PC2, are completed in sand on hillside terraces; group Bb wells are completed in sandy till. The greater magnitude of water table fluctuations in till (well 17, Figure 9f) compared with the fluctuations in sand (well 11, Figure 9d) is evident. The distinct outlier, well 15, is completed at the contact between till and crystalline bedrock.

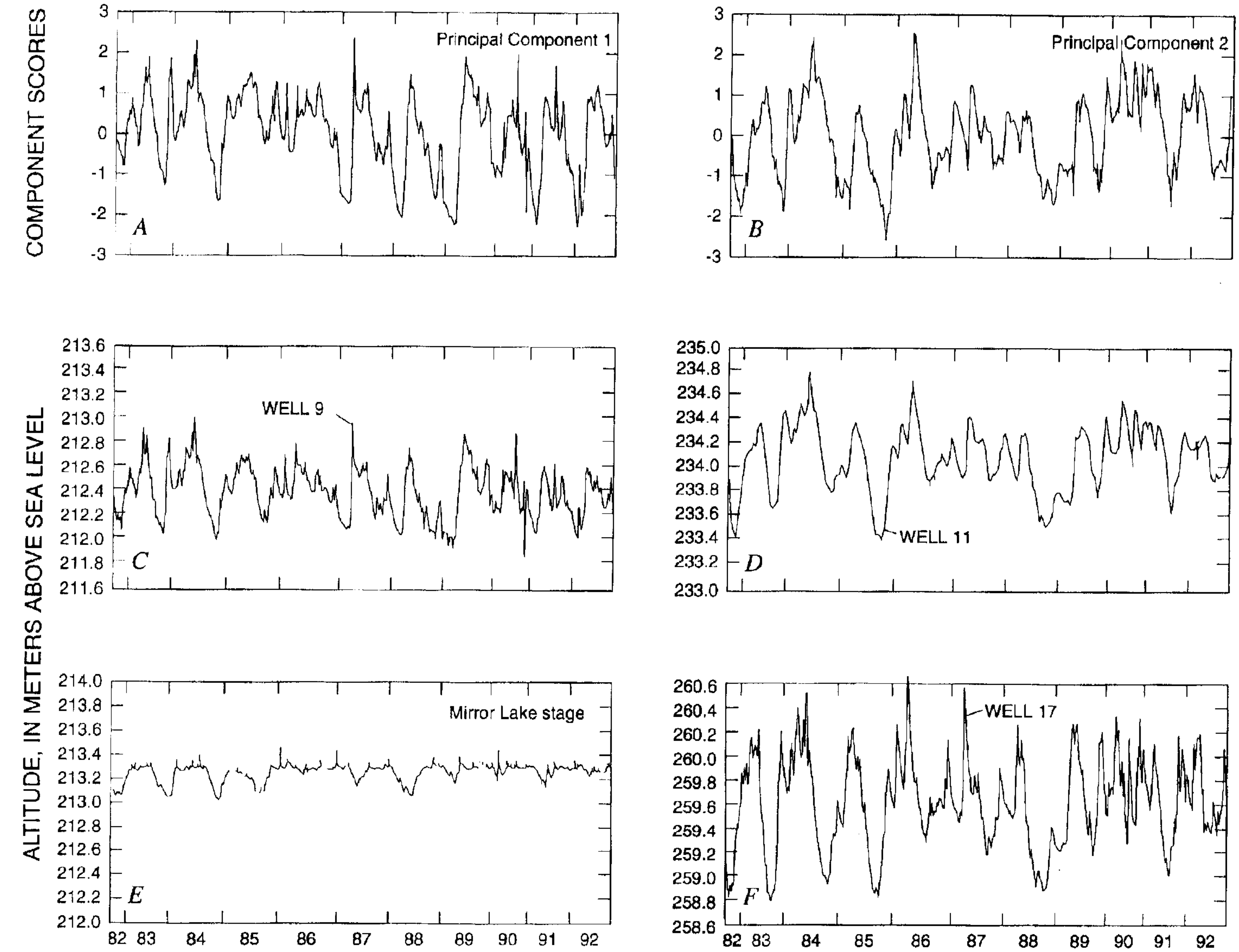


Figure 9. Hydrographs for the Mirror Lake area, New Hampshire: (a) component scores for principal component 1; (b) component scores for principal component 2; (c) water level in well 9; (d) water level in well 11; (e) stage of Mirror Lake; (f) water level in well 17.

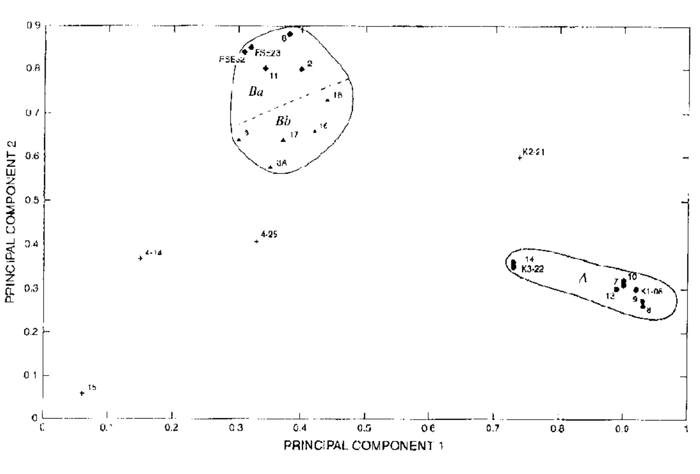


Figure 10. Scatter plot of component loadings for principal component 1 versus principal component 2 for water table levels in the Mirror Lake area.

Cottonwood Lake Area, North Dakota

For the Cottonwood Lake area, the first principal component (CL-PC1) accounted for 75% of the variance in the water level data (Table 1). A hydrograph of component scores for CL-PC1 is shown in Figure 12a. The second principal component (CL-PC2) accounted for 11% of the variance in the water level data (Table 1). A scores hydrograph for CL-PC2 is shown in Figure 12b. The amount of variance accounted for by the third principal component is the highest, along with that for the Mirror Lake area, compared with the third principal component for the four sites. The amount of variance accounted for by the fourth principal component is highest for the Cottonwood Lake area compared with the other sites. The percent of variance explained by the third and fourth principal components at the Cottonwood Lake site do not appear to be meaningful in a statistical sense, but actual hydrograph patterns of wells completed in sand (discussed later) have some similarity to the scores hydrograph for CL-PC4 (Figure 12d).

A plot of the component loadings for each well as they relate to CL-PC1 versus CL-PC2 (Figure 13) shows the greatest scatter of points compared with the other three sites. However, even with

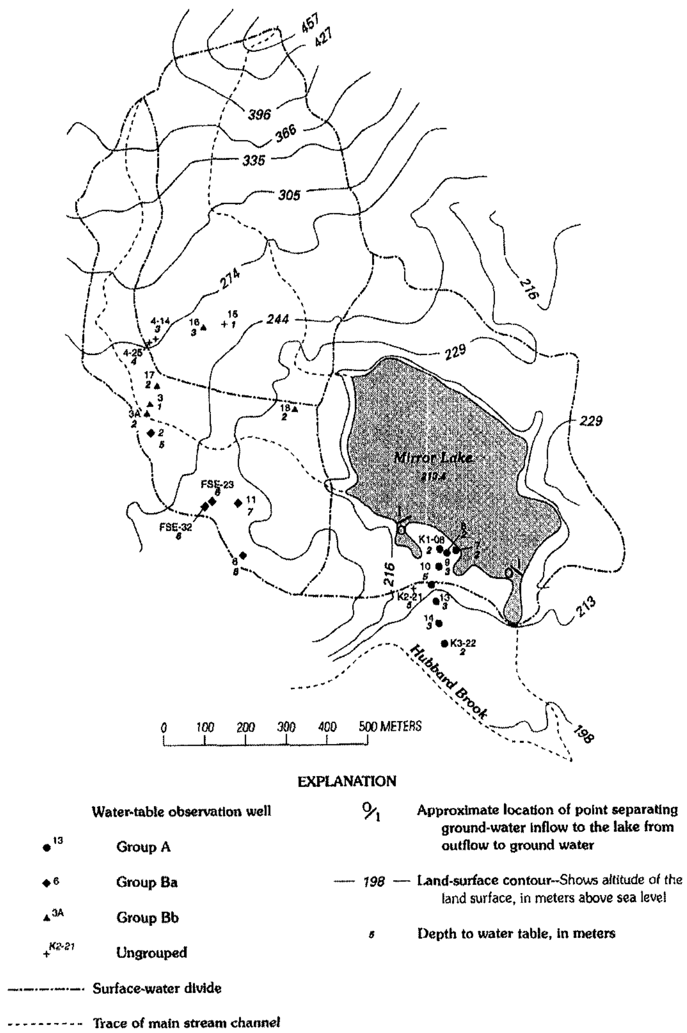


Figure 11. Configuration of the land surface, depth to the water table, and areal distribution of well groups, as delineated on Figure 10, for the Mirror Lake area. Ground water moves into Mirror Lake throughout most of its littoral zone except for an area along its south side, as indicated on the map.

the scatter, it is possible to determine some relationships of the hydrograph patterns to hydrogeologic conditions at the site. For example, a group of wells that have loadings greater than 0.6 on CL-PC1 and less than 0.5 on CL PC2, group A, have actual hydrograph patterns (Figure 12c) similar to the scores hydrograph for CL-PC1. However, a group of wells that have loadings greater than 0.6 on CL-PC2 and between 0.3 and 0.7 on CL-PC1, group B, also have actual hydrograph patterns (Figure 12e) similar to the scores hydrograph for CL-PC1. The major difference between the hydrograph patterns for group A wells and group B wells is that the magnitude of fluctuation is greater for the group B wells. This general pattern also is similar to the pattern of fluctuations in the stage of Wetland P1 (Figure 12g). The scores hydrograph for CL-PC2 appears to have little similarity to the actual hydrographs for any of the wells in the Cottonwood Lake area. Interestingly, all of the wells that have loadings less than 0.4 on CL-PC1 and less than 0.8 on CL-PC2 on Figure 13 (group C) have hydrograph patterns somewhat similar to the scores hydrograph for CL-PC4 (Figure 12d). Several examples of the hydrograph patterns for these wells are shown on Figure 12f.

A map of the areal distribution of the well groups identified for the Cottonwood Lake area is shown in Figure 14. Most group A wells are located at intermediate positions within the ground water flow system, which moves generally southeast to northwest through the area. Group A wells also generally are upgradient of wetlands that receive inflow from ground water. Most group B wells are near wetlands that recharge ground water. The greater magnitude of fluctuation in group B wells (Figure 12e) compared with group A wells (Figure 12c) probably is related to their position within the ground water flow system. By being high in the flow system and having a small ground water contributing area upgradient of the well, the water table declines to greater depths between recharge events compared with the water table lower in the flow system.

In contrast to all wells in groups A and B, which are completed in till, all wells in group C are completed in buried sand. The hydrograph patterns for wells completed in buried sand (Figure 12f) are distinctly different from those completed in till (Figures 12c and 12e) and resemble the scores hydrograph for CL-PC4 (Figure 12d). All of the wells completed in buried sand, except well 29, are completed in a single, fairly extensive sand unit, and they all have hydrographs similar to wells 14 and 22 (Figure 12f). The hydrograph for well 29, which is completed in a small isolated sand, is also plotted on this graph. The greater magnitude of fluctuations of the water table at well 29 compared with the magnitude of fluctuations at wells 14 and 22 is a reflection of the poorly permeable till. The slow movement of water in till has a greater effect on small isolated aquifers than it does on more extensive aquifers.

Discussion

At four sites in different hydrogeologic and climatic terrains, the statistical procedure of principal component analysis proved useful in defining water table hydrograph patterns that could be interpreted in terms of (1) ground water recharge characteristics; (2) position of an observation well relative to surface water bodies; and (3) differences in geologic conditions. For example, at two sites where all of the water table wells were completed in permeable deposits, glacial outwash in Minnesota and dune sand in Nebraska, the water table fluctuation patterns primarily reflected timing and magnitude of recharge. That is, where the water table is shallow, water table hydrographs have more frequent and wider ranges in fluctuations compared with where the water table is deep. At the Williams Lake area in Minnesota, the water table fluctuation patterns downgradient of the lake indicate some effects of seepage from the lake. However, this does not seem to be the case for the Island Lake area in Nebraska.

At the two sites where the water table wells were completed in sand or till, New Hampshire and North Dakota, the patterns of water table fluctuations are strongly related to the type of geologic unit in which the wells are completed. Furthermore, at the Mirror Lake area in New Hampshire, the patterns of water table fluctuations are clearly different for wells completed in sand and sandy till downgradient of Mirror Lake compared with those completed in terraces on the hillside upgradient of the lake. In the hillside terraces, fluctuation patterns are different between wells completed in sand compared with those completed in till.

Principal component analysis of water level measurements in wells did not lead to any profound new insights into what might be determined from conventional analysis of ground water hydrographs. However, the method made analysis of the hydrographs effi-

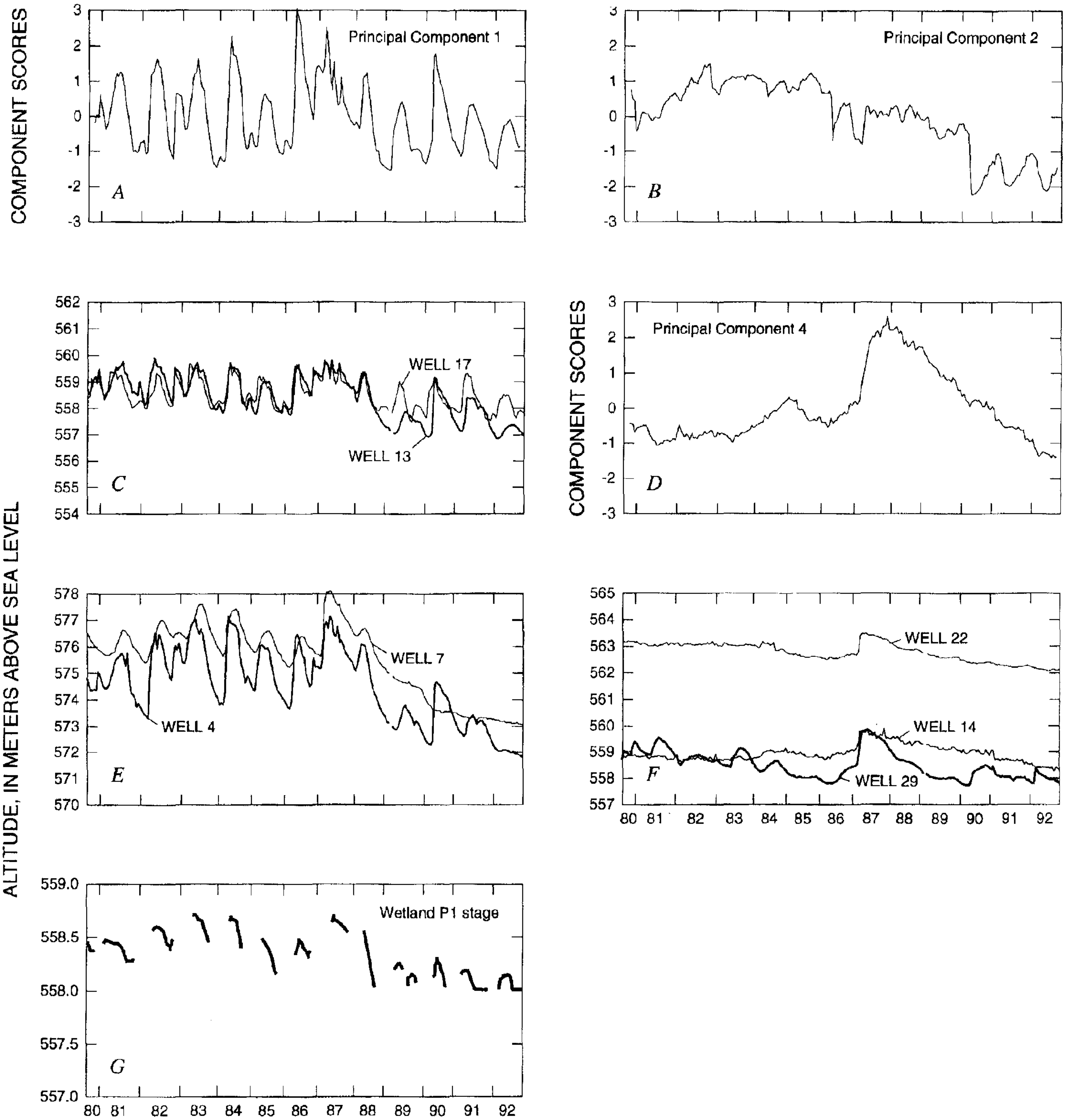


Figure 12. Hydrographs for the Cottonwood Lake area, North Dakota: (a) component scores for principal component 1; (b) component scores for principal component 2; (c) water level in wells 13 and 17; (d) component scores for principal component 4; (e) water level in wells 4 and 7; (f) water level in wells 14, 22, and 29; and (g) stage of Wetland P1.

cient, objective, and reproducible. Because of the speed and objectivity of the procedure, this type of analysis can be useful in ground water resource studies, for evaluating ground water monitoring programs, and selecting index wells for long-term monitoring.

For ground water resources studies, the areal distribution of different hydrograph patterns can be used to determine the areal characteristics of recharge in a study area. Principal component analysis also is a powerful tool to summarize information in large data sets of ground water monitoring networks. For example, maps could be prepared for regions, watersheds, or political units that show different hydrograph patterns in different parts of those areas. As a follow-up to areal mapping of hydrograph patterns, the procedure would be

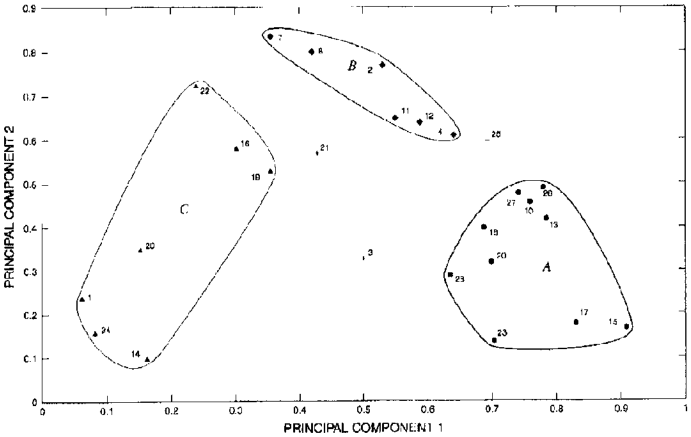
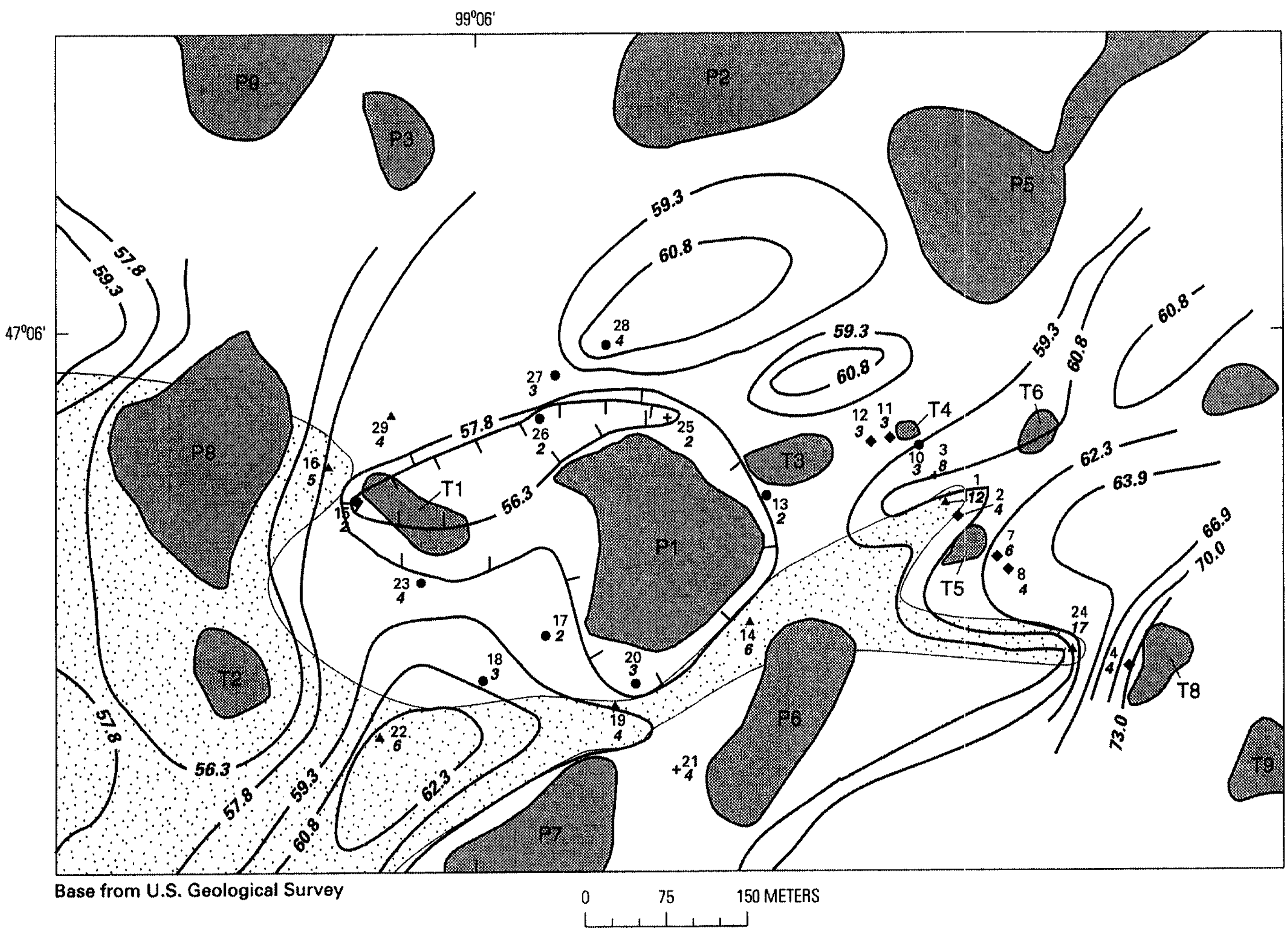


Figure 13. Scatter plot of component loadings for principal component 1 versus principal component 2 for water table levels in the Cottonwood Lake area.

especially useful for selecting index wells from a large number of candidate wells for long-term monitoring of hydrologic processes in a watershed or aquifer. The Williams Lake area is used as an example. Because of the close grouping of wells in group A (Figure 4), the hydrograph pattern for well 15 (Figure 3c) could be considered representative of the part of the watershed that encompasses the group A wells. Therefore, well 15 could be used as the index well for long-term monitoring for group A, and measurement of all other wells in that group could be discontinued. Similarly, well 22 could be an index well for the area represented by group B wells, and well 19 could be an index well for the area represented by group C wells. Additional wells could be selected for index wells depending on the monitoring needs, but one well from each of the three groups would be the recommended minimum.



EXPLANATION

- Buried sand deposit—At surface near wetlands T2 and P8
- Semipermanent wetland
- Seasonal wetland
- — 59.3 — Water table contour—Number is in meters greater than 500 meters above sea level. Interval is variable
- *4* Depth to water table, in meters

Water-table observation well

- ● 13 Group A
- ◆ 17 Group B
- ▲ 24 Group C
- + 21 Ungrouped

Figure 14. Configuration of the water table, depth to the water table, and areal distribution of well groups, as delineated in Figure 13, for the Cottonwood Lake area.

Conclusion

The results of this study indicate that principal component analysis can be a valuable tool for interpreting large data sets consisting only of water level measurements, and for selecting index wells representative of hydrologic conditions in various parts of aquifers or watersheds.

Acknowledgments

We are indebted to Dr. Robert R. Meglen, formerly of the University of Colorado at Denver, for making us aware of the propriety of principal component analysis for evaluating large data sets consisting only of water level measurements.

References

Lins, H.F. 1985. Interannual streamflow variability in the United States based on principal components. *Water Resources Research* 21, no. 5: 691–701.

Meglen, R.R. 1991. Examining large databases: A chemometric approach using principal component analysis. *Journal of Chemometrics* 5, 163–179.

Mooers, H.D., and A.R. Norton. 1997. Glacial landscape evolution of the Itasca/St. Croix moraine interlobate area including the Shingobee River headwaters area. U.S. Geological Survey Water-Resources Investigations Report 96-4215, 3–10.

SAS Institute Inc. 1989. *SAS/STAT User's Guide*. Version 6, 4th ed., vol. 1. Cary, North Carolina.

Swanson, K.D. 1990. Chemical evolution of ground water in clay till in a prairie wetlands setting in the Cottonwood Lake area, Stutsman County, North Dakota. M.S. thesis, Department of Geology, University of Wisconsin–Madison.

Winter, T.C. 1984. Geohydrologic setting of Mirror Lake, West Thornton, New Hampshire. U.S. Geological Survey Water-Resources Investigations Report 84–4266.

Winter, T.C. 1986. Effect of ground-water recharge on configuration of the water table beneath sand dunes and on seepage in lakes in the sandhills of Nebraska. *Journal of Hydrology* 86, 221-237.

Winter, T.C., and M.R. Carr. 1980. Hydrologic setting of wetlands in the Cottonwood Lake area, Stutsman County, North Dakota, U.S. Geological Survey Water-Resources Investigations Report 80–99.

Winter, T.C., and D.O. Rosenberry. 1995. The interaction of ground water with prairie pothole wetlands in the Cottonwood Lake area, east-central North Dakota, 1979-1990. *Wetlands* 15, no. 3: 193-211.

Winter, T.C., and D.O. Rosenberry. 1997. Physiographic and geologic characteristics of the Shingobee River headwaters area, Minnesota. U.S. Geological Survey Water-Resources Investigations Report 96-4215, 11–17.