Evaluation of Ground Water Monitoring Network by Principal Component Analysis

by Subhrendu Gangopadhyay[1], Ashim Das Gupta[2,4], and M.H. Nachabe[3]

Abstract

Principal component analysis is a data reduction technique used to identify the important components or factors that explain most of the variance of a system. This technique was extended to evaluating a ground water monitoring network where the variables are monitoring wells. The objective was to identify monitoring wells that are important in predicting the dynamic variation in potentiometric head at a location. The technique is demonstrated through an application to the monitoring network of the Bangkok area. Principal component analysis was carried out for all the monitoring wells of the aquifer, and a ranking scheme based on the frequency of occurrence of a particular well as principal well was developed. The decision maker with budget constraints can now opt to monitor principal wells which can adequately capture the potentiometric head variation in the aquifer. This was evaluated by comparing the observed potentiometric head distribution using data from all available wells and wells selected using the ranking scheme as a guideline.

Introduction

The design of a ground water monitoring network is a multi-objective design problem. For any monitoring network design, however, the question would be either (1) to measure the ground water level; (2) to measure the solute concentration in ground water; or (3) both. Processing the information on the ground water level and the solute concentration can provide an answer to all issues pertaining to ground water management. The key points in designing a monitoring network are, therefore, (1) to define the objectives; (2) to select the spatial location of monitoring wells; and (3) to ascertain the sampling frequency. The objectives of the monitoring program can be ambient monitoring, detection monitoring, compliance monitoring, research monitoring, or a combination of these objectives (Todd et al. 1976). Substantial literature has been generated through the years to identify spatio-temporal attributes of wells that satisfy these objectives. However, the focus has been on the design of a ground water quality monitoring network. Loaiciga et al. (1992) suggested that the considerations for ground water quality monitoring network design are the spatial and temporal coverage of sampling sites, the competing objectives of the monitoring program, and the uncertainty on the geologic, hydrologic, and environmental conditions. According to these authors, the two general approaches to network design are the hydrogeologic and statistical approaches. The hydrogeologic approach is based on judgment of quantitative and qualitative hydrogeologic information, without the use of advanced statistical techniques. The statistical approach has been further classified into simulation, variance-based, and probability-based techniques. The simulation approach is based on the synthetic generation of random fields of aquifer characteristics using geostatistical models (Journel and Huijbregts 1978). In ground water quality management, the simulation approach was demonstrated by Massman and Freeze (1987a, 1987b) and Wagner and Gorelick (1987, 1989). Application of neural networks and genetic algorithm optimization integrated with flow and transport simulation models are found in Rogers and Dowla (1994) and McKinney and Lin (1994). The variance-based approaches can be further classified into global method (Olea 1984), variance reduction analysis (Rouhani 1985), and optimization. Of these methods, the optimization approach has found considerable application. This approach considers the design problem as a mathematical programming problem consisting of an objective function, typically minimizing an estimation error, subject to constraints. Integer or mixed integer programming algorithms like branch and bound algorithm are commonly used to evaluate the presence or absence of a well at a particular location (Carrera et al. 1984; Loaiciga 1989; Hudak et al. 1995; Wagner 1995). The variance-based approaches maximize the information by minimizing the estimation variance. The probability-based approaches consider the probability of exceeding a certain level of the field variable as the criterion to be controlled in the network design problem (Rouhani and Hall 1988).

The philosophy behind the design of a ground water level observation network is essentially the same as that for a ground water quality monitoring network. However, the objectives in designing a ground water level observation network can be to determine: (1) the effect of withdrawals on recharge and natural discharge conditions, (2) the hydraulic characteristics of ground water systems, and (3) the extent and degree of confinement of aquifers (Heath 1976). The literature on design of a ground water level observation network is limited. Jawad and Hussien (1988) developed a method

[1]Division of Water Sciences, UNESCO, Paris, France
[2]Water Engineering and Management Program, School of Civil Engineering, Asian Institute of Technology, Bangkok, Thailand
[3]Department of Civil and Environmental Engineering, University of South Florida, Tampa, Florida
[4]Corresponding author
Received July 1999, accepted August 2000.

based on the correlation measure, RV coefficient (Robert and Escoufier 1976). They applied this method to evaluate an existing network to monitor aquifers of the Erbil hydrogeological basin in Iraq. The RV coefficient expresses the correlation between the two vectors, X (the population) and Y (the sample) and has a value between 0 and 1. A maximum RV value usually occurs when the dimension of Y is increased to approach the dimension of X.

The objective of this research is to develop and test a new method for evaluating a network of ground water level observations. Through the years, ground water monitoring networks have expanded tremendously, and many networks today consist of dozens, if not hundreds, of sampling wells. At a certain stage, municipalities have to rationalize their ground water monitoring networks and ask questions such as how sampling from a particular well can help explain the dynamic variation of potentiometric head in the aquifer and for a municipality facing budget constraints, what subset of observation wells should be selected to continue monitoring in the near future.

At the core of the new proposed method is the multivariate technique of principal component analysis. Principal component analysis was used in surface water hydrology to identify the important geomorphological parameters that contribute to runoff from a catchment (Haan 1977). To evaluate a ground water level monitoring network, the principal component analysis is used to discriminate against the value of information collected from monitoring wells. Thus facing budget constraints in the near future, a manager for a municipality can prioritize sampling from the monitoring network of the aquifer. The management authority can choose to continue monitoring the wells that capture most of the dynamic variation in the aquifer.

The proposed method is tested on the monitoring network for the Bangkok Aquifer. For the Bangkok Aquifer, the objectives of the ground water level observation network are to model ground water flow, and to evaluate land subsidence in Bangkok. The described approach deals with the problem of evaluating an existing ground water monitoring system.

Principal Component Analysis

When ground water level data are collected from p adjacent monitoring wells, these data are often correlated. This correlation reflects the complexity of the aquifer hydrogeology and indicates that some of the information (annual water level in this case) collected from one well is also contained in the remaining p-1 wells. Thus the objective of principal component analysis is to evaluate this correlation to save on the number of independent variables that describe the potentiometric head variation in the aquifer. This reduces the number of variables needed to be measured or observed. A review of the particular aspects of principal component analysis technique important for the present application has been outlined in this section.

Let $\underline{\underline{\mathbf{X}}}$ be the observation matrix of deviations from the mean of order n x p, where n is the number of observations on p variables. Here n is the number of years for which water level measurements have been considered for p adjacent monitoring wells. The mean water level at a well is assumed to be the average of the n annual water level values for the well. So in $\underline{\underline{\mathbf{X}}}$, water levels for a well represent the deviation from their respective mean water level. It should also be noted that for each of the p wells, data may not be available for all the n years. The missing observations in this case have been filled using a ground water flow model, and the implications have been discussed in the following section. This ensure $\underline{\underline{\mathbf{X}}}$ has n values for the p wells being analyzed. It is also considered that this original p-variate set of observations in $\underline{\underline{\mathbf{X}}}$ contain a correlation. This correlation can be analyzed using principal components to identify the relative importance of any well in representing variations of potentiometric head among the p wells. This is done by characterizing the variance of $\underline{\underline{\mathbf{X}}}$ with q (q < p) principal (orthogonal) components and evaluating the correlation between the ith original variable and the jth principal component. The jth principal component, $\underline{z_j}$ (n x 1 column vector), is the linear function

$$\underline{z_j} = \underline{\underline{\mathbf{X}}}\,\underline{\mathbf{a}_j} \tag{1}$$

where $\underline{\mathbf{a}_j}$ is a p x 1 linear transformation coefficient vector corresponding to $\underline{\mathbf{z}_j}$. From the theory of principal component analysis (Haan 1977) it can be shown that $\mathbf{a}_j$ is the characteristic vector associated with the characteristic root λ_j of $\underline{\underline{\mathbf{S}}}$. The characteristic roots λ_j (roots of the scalar λ) are obtained by solving the equation defined by

$$|\underline{\underline{\mathbf{S}}} - \lambda \underline{\mathbf{I}}| = \underline{0} \tag{2}$$

where $\underline{\mathbf{I}}$ and $\underline{0}$ are the p x1 unit and zero vectors, respectively; $\underline{\underline{\mathbf{S}}}$ (order, p) is the estimate of the variance-covariance matrix of $\underline{\underline{\mathbf{X}}}$, given by

$$\underline{\underline{\mathbf{S}}} = \underline{\underline{\mathbf{X}}}^{\mathrm{T}}\,\underline{\underline{\mathbf{X}}} / (\mathbf{n}-\mathbf{1}) \tag{3}$$

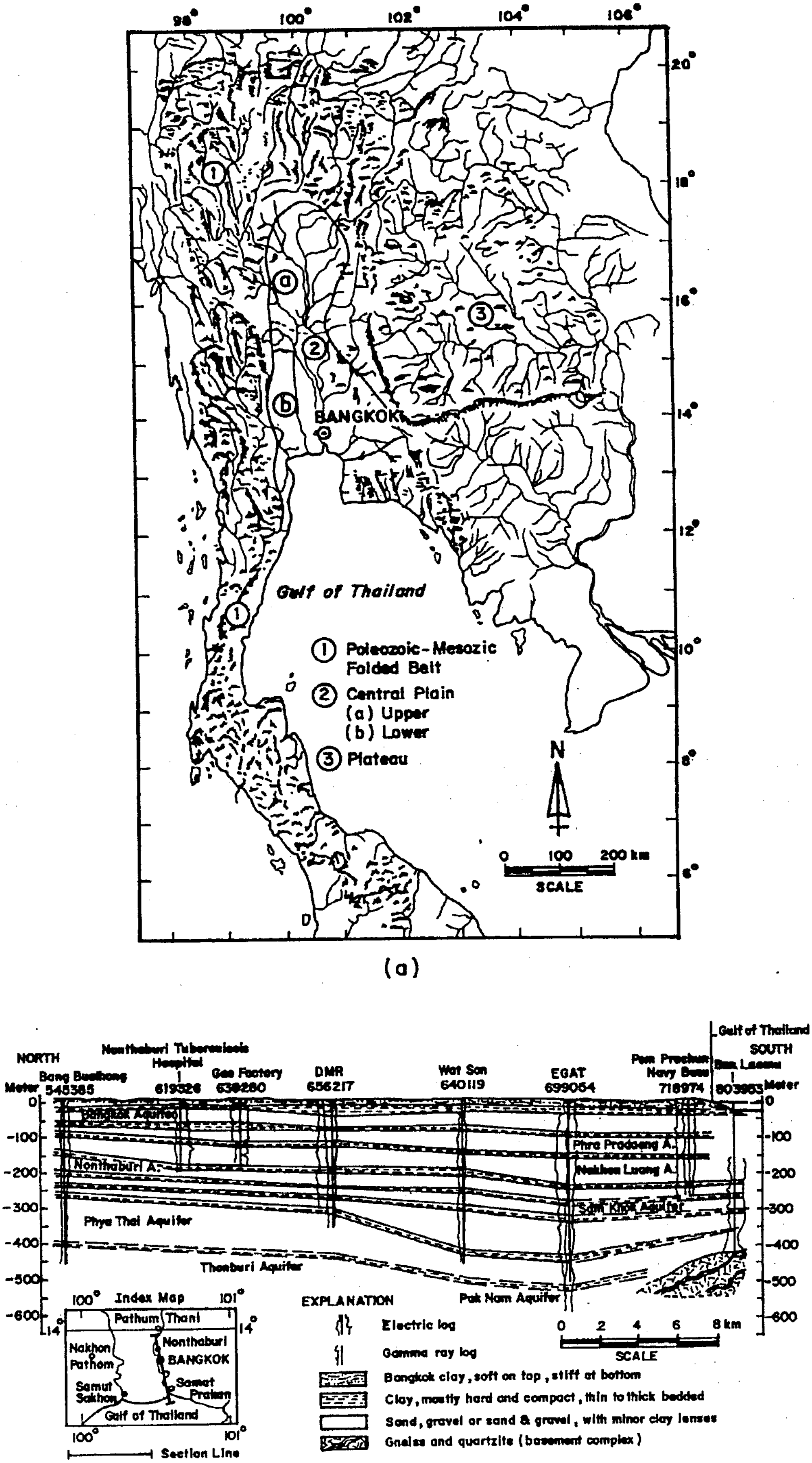
and superscript T denotes the matrix-transpose operator. The solution of Equation 2 is subject to, normalizing constraint: $\underline{\mathbf{a}_j^{\mathrm{T}}}\,\underline{\mathbf{a}_j} = 1$; and orthogonality constraint: $\underline{\mathbf{a}_i^{\mathrm{T}}}\,\underline{\mathbf{a}_j} = \underline{\mathbf{a}_j^{\mathrm{T}}}\,\mathbf{a}_i = \mathbf{0}$. These constraints ensure a unique solution and uncorrelated principal components. Thus, corresponding to $\lambda_1, \lambda_2, ..., \lambda_p$ we have the n x 1 column vectors $a_1, a_2, ..., a_p$ respectively. Then Equation 1 can be extended to $\underline{\mathbf{a}_i^{\mathrm{T}}}$

$$\underline{\underline{\mathbf{Z}}} = \underline{\underline{\mathbf{X}}}\,\underline{\underline{\mathbf{A}}} \tag{4}$$

where $\underline{\underline{\mathbf{Z}}} = (\underline{z_1}, \underline{z_2}, ..., \underline{z_3})$ is the transformed n x p matrix of n values for each of p components, and $\underline{\underline{\mathbf{A}}} = (\underline{\mathbf{a}_1}, \underline{\mathbf{a}_2}, ..., \underline{\mathbf{a}_p})$ is the pxp linear transformation coefficient matrix. In $\underline{\underline{\mathbf{Z}}}$, the first q components explain most of the variation contained in $\underline{\underline{\mathbf{X}}}$, and the remaining p-q provide only a small contribution to the total variation and they may be neglected. The advantage now in working with $\underline{\underline{\mathbf{Z}}}$ is that these new variables (principal components) are mutually uncorrelated and the dimensionality of the original problem is reduced. In this analysis however, the objective is to use the correlation between $\underline{\mathbf{x}_i}$ and $\underline{\mathbf{z}_j}$ to determine the importance of well i in representing the water level fluctuation among the group of p wells. Combining Equations 1, 3, and 4, it can be shown that the correlation between the ith standardized observed variable and the jth computed component can be expressed as (Haan 1977)

$$\mathbf{Cor}(\mathbf{x}_i, \mathbf{z}_j) = \lambda_j^{1/2}\,\mathbf{a}_{i,j} \tag{5}$$

where $\mathbf{a}_{i,j}$ (i = row, j = column) are the elements of the coefficient matrix $\underline{\underline{\mathbf{A}}}$. Equation 5 can thus be used to calculate the elements of the correlation matrix that depicts the degree of correlation between the original observed variable and the derived principal component.



;ure 1. (a) Physiographic map of Thailand showing the Lower Central Plain. (b) Hydrogeologic north-south section of the Lower Chao Phraya lta showing principal aquifers of the Bangkok metropolis.

This correlation matrix is commonly referred to as the factor loading matrix, and the elements given by Equation 5 are called factor loadings. To aid in the interpretation of the factor loading matrix an orthogonal transformation such as varimax rotation is often performed (Haan 1977). Setting a cutoff value for factor loading (measure of significant correlation), wells that have loadings below this significance level for all the extracted components may not be observed. In other words, only q (q < p) wells that have significant correlation with the extracted components can adequately explain potentiometric level variation in the group of p wells.

Study Area and Steps in Analysis

The Bangkok Metropolitan area is situated on the flood plain and delta of the Chao Phraya River which traverses the Lower Central Plain of Thailand. The plain, also known as the Lower Chao Phraya Basin, extends about 200 km from north to south and about 175 km from east to west. It is bounded on both east and west by mountain ranges and on the south by the Gulf of Thailand. To the north, the plain is bordered by a series of small hills dividing it from the Upper Central Plain, with the Chao Phraya River as the interconnection (Figure 1a). The Lower Central Plain was formed on a geologic fault/flexure depression filled with clastic sediments. The subsurface strata overlying the basement were described as fluvial and deltaic sediments aged from Recent to Oligocene. The topmost stratum is called Bangkok Clay. It is about 20 m thick in the Bangkok area and consists of about 12 m of soft marine clay stratum overlying 8 m of desiccated stiff clay stratum. Based on well log information, the subsurface strata are interpreted to have eight aquifers existing to a depth of 660 m (Piancharoen and Chuamthaisong 1976), as shown in Figure 1b. The ground water development for the public water supply in the Bangkok area began in 1954 with an abstraction of 8360 m^3/d, and by 1982 the pumpage had increased to 1.4 million m^3/d. With this rapid increase in ground water abstraction, the potentiometric level dropped from near to the ground surface to a level of 46 m below ground surface for the second aquifer and 53 m below ground surface for the third aquifer. The report by Cox (1968) provided an early indication of the problem of land subsidence in the Bangkok area resulting from ground water pumping. Quantitative assessment of land subsidence was not conducted until early in 1978. From mid-1978 to 1982, the Royal Thai Survey Department conducted seven runs of first order leveling at half-year intervals to monitor benchmarks. These results, together with measurements undertaken by the Asian Institute of Technology (AIT) at 27 observation stations in 1978 and at four additional stations in 1981 onward, provided useful information regarding land subsidence in the Bangkok area. A maximum subsidence of 54 cm was observed from 1978 to 1982 and as much as 1.14 m of settlement occurred between 1940 and 1980 (AIT 1981).

The theory of principal components is applied to the monitoring network of the aquifer in Bangkok. A description of data and a step-by-step procedure for network evaluation follows.

Ground Water Monitoring Network in Bangkok

The ground water monitoring network in Bangkok was first established in 1978 under the comprehensive study program on ground water and land subsidence. The network was aimed at monitoring potentiometric levels and water quality in the three main aquifers of Phra Pradaeng (PD), Nakhon Luang (NL), and Nonthaburi (NB). At that time, the number and locations of monitoring wells were selected based on the rates of decline of potentiometric level, land subsidence, and the extent of the critical zone with regard to the depletion of ground water level and land subsidence. After the initiation of the project on Mitigation of Ground Water Crisis and Land Subsidence in Bangkok Metropolis in 1985, additional monitoring locations were identified, and 258 monitoring wells at 103 monitoring stations were designed to be installed over a study area of 5600 km^2. In the majority of these wells, water levels are measured on a monthly basis (DMR 1994).

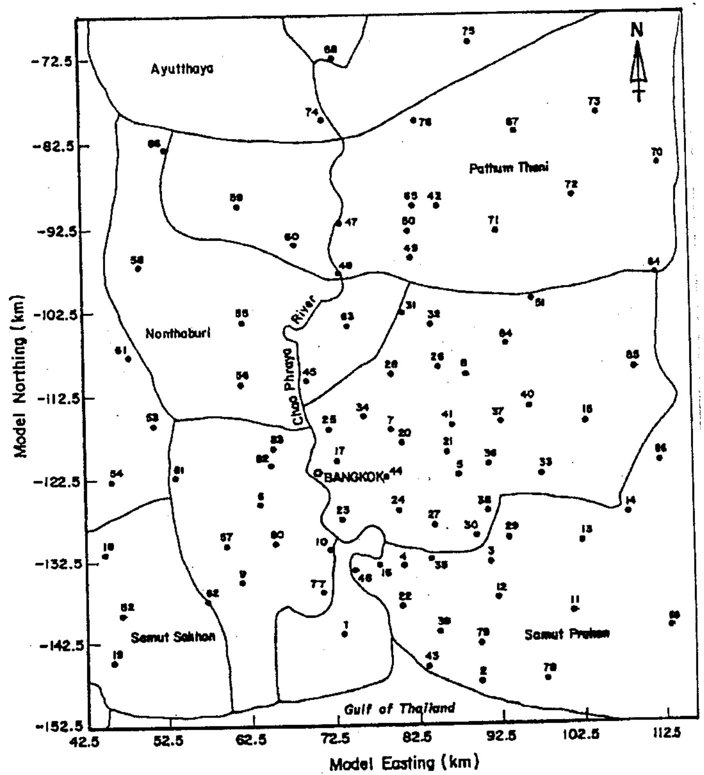


Figure 2. Location of the ground water monitoring wells in the Phra Pradaeng (PD) Aquifer.

The 258 monitoring wells are distributed as 86 PD wells, 87 NL wells, and 85 NB wells. The location of the 86 PD wells are shown in Figure 2. Observations in the monitoring wells were started in 1978 and they were few in number. Because the construction of the present number of wells took place gradually, missing data were filled up for the newly constructed wells using a calibrated simulation model. Thus, for this study period, the well database consists of a series made of 13 rows, representing years 1978 to 1990, and 86 columns, representing the 86 PD monitoring wells.

From 1978 to 1983 in about 40% of the wells missing data were filled using the simulation model, and from 1984 to 1990 the number of wells for which filled data has been used is within 20%. By filling up missing data using a calibrated flow model, an artificial correlation may be added. This, however, is an intrinsic constraint of the present approach. Even with this shortcoming of the present technique, this extension is the natural approach for real world problems and basically may be the only way applicable. With this approach it is possible to proceed and provide more insight into the problem. The flow model used in this case is based on the MODFLOW code (McDonald and Harbaugh 1988). Its setup and calibration have been extensively described in Gangopadhyay (1997).

Step-by-Step Procedure for Network Evaluation

The following is the procedure for network evaluation.

Step I–Select a search radius for principal component analy-

Table 1
Selected Monitoring Stations with Corresponding PD Wells and Its Closest Wells

Station Number (1)	PD Well Number (2)	Total Number of Wells Within a Radius of 10 km with the Well in Column 2 as Center (PD Well Number[s]) (3)
1	8	8 (26, 84, 41, 37, 32, 40, 28, 21)
21	46	10 (16, 10, 77, 4, 23, 22, 1, 24, 35, 80)
32	13	4 (14, 11, 29, 33)
40	47	5 (60, 48, 50, 65, 49)
55	24	14 (44, 27, 4, 23, 16, 35, 20, 5, 46, 21, 17, 10, 7, 30)
57	29	10 (3, 30, 38, 12, 33, 13, 27, 36, 35, 5)
69	32	7 (31, 26, 8, 28, 49, 84, 63)
85	48	6 (47, 63, 60, 49, 31, 50)
103	81	2 (53, 54)

Table 2
Rotated Factor Loading Matrix for Well PD47

Well Number	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
PD60	.57686	.79763	.17612	–.00312	–.00025
PD48	.31424	.94834	–.04348	.00293	.00033
PD50	.92348	.38332	.01395	.00323	.00659
PD65	.93982	.33614	.06019	–.00944	–.00592
PD49	.79771	.60026	–.04701	.03369	.00053

sis. For each monitoring well i in Figure 2, carry out steps II to IV.

Step II–Identify all the p wells that are within the search radius of well i.

Step III–Carry principal component analysis using p wells and develop the factor loading matrix.

Step IV–Perform a varimax rotation on the factor loading matrix to infer the principal wells q, where $q < p$. For this study, the significant correlation considered was 0.9. The wells that do not have significant correlation with any of the factors were not considered principal wells.

Step V–Calculate the rank of a well by dividing the number of times this well appeared as principal well by the number of times it was used in a principal component analysis. The rank determines the importance of this monitoring well in explaining the fluctuation of ground water levels in the aquifer.

Results and Discussion

Selection of Search Radius

The search radius for principal component analysis is the distance within which a fluctuation in water level at one well location is expected to significantly affect the fluctuation at a neighboring well location. This concept is analogous to the concept of effective radius of influence (McWhorter and Sunada 1977) and discrete kernels used in the generation of response function for aquifer flow simulation (Illangasekare et al. 1984). To estimate the search radius for this study, continuous pumping for 10 years at a constant rate of 50,000 m^3/d (arbitrary and not actual pumping at the site) was introduced at selected locations in the model domain. At the end of 10 years of pumping it was observed that the drawdown, expressed in terms of the percentage of the maximum drawdown at the location of pumping, reduced on average to 30% at a distance of 10 km (two grid cells in the finite-difference model) from the excitation point. It is also obvious that as the distance from the excitation cell increases, the drawdown percentage will also diminish. At distances of 15 km and 20 km from the pumping point the average drawdown percentage was found to be approximately 20% and 10%, respectively. Thus from a physical perspective and practical purpose, in the principal component analysis the search radius for this study was chosen to be 10 km; and there is little need to consider that the fluctuation in ground water level at one well will significantly impact the ground water level at wells beyond this radius.

Since principal component analysis is based on correlation analysis, it is necessary to ensure that only correlations with physical significance are included in this analysis. The search radius provides a means to identify, approximately, the neighboring subset of wells that have correlated observations. This correlation is associated with the complex hydrogeological conditions (Premchitt and Das Gupta 1981), e.g., transmissivity variation, in the particular area. Although one may find some correlation between yearly hydraulic head changes among remotely distant wells, say wells 100 km apart, this correlation can be purely coincidental and cannot be interpreted based on the dynamics of fluid flow. Indeed, in the beginning of this research, it was attempted, with no success, to run principal component analysis on all the wells in the aquifer. It was found that to satisfactorily explain the head variation, all the wells need to be retained. So, principal component analysis did not help in this case. Intuitively and obviously, wells that are close to each other will be more correlated than those farther away. This notion is fundamental in the stochastic theories of subsurface hydrology. In the present analysis, it was therefore essential to establish a distance within which head values are correlated and the search radius is an essential part of the present method.

Table 3
Comparison of Observed and Predicted Water Levels for PD Wells

PD Well Number	Water Level in 1988 (m below MSL) Observed	Predicted	Principal PD Well Number(s) Used in Prediction
13	37.15	37.75	14
24	30.82	30.63	23, 16, 5, 46, 30
29	33.72	33.56	30, 28, 27
32	24.86	24.79	26, 8, 28, 49, 84
46	30.82	30.73	10, 23, 1, 24
48	19.61	20.17	63, 50

Testing Principal Component Analysis with Multiple Regression

Selected monitoring stations with the corresponding PD wells, along with the number of closest wells, are shown in Table 1. Table 2 shows an example of rotated factor loading matrix for well PD47. Five wells are within the search radius of PD47. These

Table 4
Comparison of Observed and Predicted Water Level Using Principal Wells and Wells Within the Search Radius

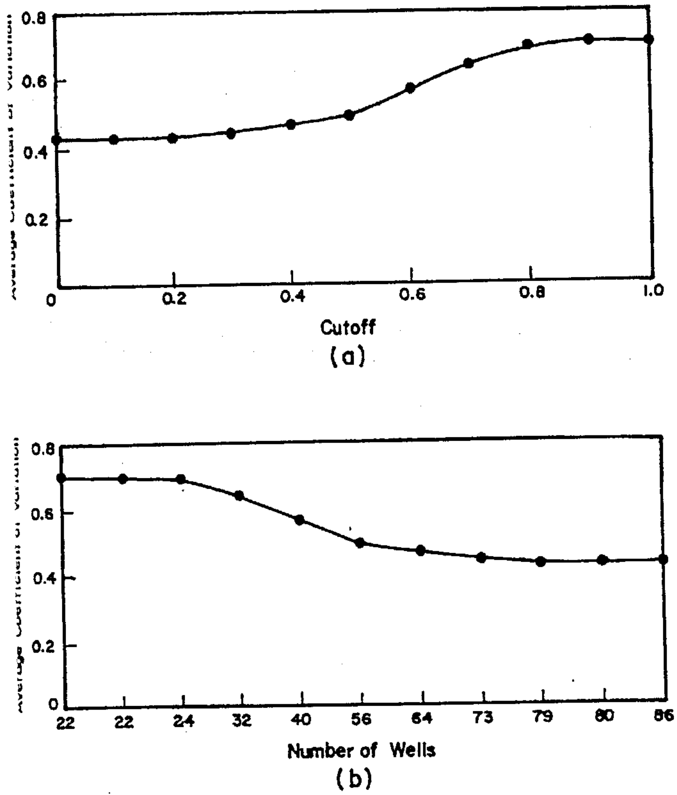
PD Well Number	Water Level in 1998 (m below MSL)		
	Observed	Predicted Using	
		Principal Wells (Number)	Wells Within Search Radius (Number)
32	24.86	24.79 (5)	24.83 (7)
48	19.61	20.17 (2)	19.56 (6)

Table 5
Ranks of PD Wells Derived from Principal Component (PC) Analysis

Serial Number (1)	PD Well Number (2)	Number of Times as Principal Well (3)	Number of Times as a Close Well (4)	Rank (Col 3/Col 4) (5)
1	86	1	1	1.000
2	74	1	1	1.000
3	68	1	1	1.000
4	61	1	1	1.000
5	56	3	3	1.000
6	55	1	1	1.000
7	51	2	2	1.000
8	42	4	4	1.000
9	1	4	4	1.000
10	26	8	9	0.889
11	8	7	8	0.875
12	38	7	9	0.778
13	65	3	4	0.750
14	34	5	7	0.714
15	31	5	7	0.714
16	6	5	7	0.714
17	46	7	10	0.700
18	28	7	10	0.700
19	23	7	10	0.700
20	84	4	6	0.667
21	48	4	6	0.667
22	25	6	9	0.667
23	17	6	9	0.667
24	15	2	3	0.667
25	82	3	5	0.600
26	36	6	10	0.600
27	30	6	10	0.600
28	27	7	12	0.583
29	40	4	7	0.571
30	39	4	7	0.571
31	33	4	7	0.571
32	3	4	7	0.571
33	83	3	6	0.500
34	81	1	2	0.500
35	62	1	2	0.500

Table 5
Ranks of PD Wells Derived from Principal Component (PC) Analysis (continued)

Serial Number (1)	PD Well Number (2)	Number of Times as Principal Well (3)	Number of Times as a Close Well (4)	Rank (Col 3/Col 4) (5)
36	57	2	4	0.500
37	44	5	10	0.500
38	43	2	4	0.500
39	29	5	10	0.500
40	24	7	14	0.500
41	14	1	2	0.500
42	10	4	8	0.500
43	2	2	4	0.500
44	32	3	7	0.429
45	35	5	12	0.417
46	5	5	12	0.417
47	63	2	5	0.400
48	45	2	5	0.400
49	16	4	10	0.400
50	9	2	5	0.400
51	4	4	10	0.400
52	77	3	8	0.375
53	22	3	8	0.375
54	7	4	11	0.364
55	79	2	6	0.333
56	78	1	3	0.333
57	60	1	3	0.333
58	53	1	3	0.333
59	50	2	6	0.333
60	37	3	9	0.333
61	49	2	7	0.286
62	21	3	11	0.273
63	20	3	11	0.273
64	80	2	8	0.250
65	13	1	4	0.250
66	47	1	5	0.200
67	12	1	7	0.143
68	85	0	1	0.000
69	71	0	2	0.000
70	59	0	1	0.000
71	54	0	3	0.000
72	41	0	10	0.000
73	11	0	3	0.000
74	76	0	0	Boundary well
75	75	0	0	Boundary well
76	73	0	1	Boundary well
77	72	0	0	Boundary well
78	70	0	1	Boundary well
79	69	0	0	Boundary well
80	67	0	0	Boundary well
81	66	0	0	Boundary well
82	64	0	0	Boundary well
83	58	0	0	Boundary well
84	52	0	2	Boundary well
85	19	0	1	Boundary well
86	18	0	2	Boundary well



igure 3. Plot of average coefficient of variation with (a) cutoff, and b) number of wells.

ells are PD60, PD48, PD50, PD65, and PD49. Based on the otated factor loading matrix, wells PD48, PD50, and PD65 are elected as principal wells because they have a factor loading reater than 0.9. These wells should explain most the variation of vater level at the location of PD47. To test that this analysis is valid, multiple linear regression model was formulated on the principal omponents and observed water levels were compared with predicted values. Data up to 1987 were used to formulate the multiple egression, and the regression equation was then used to predict the water level for year 1988. Table 3 shows the observed and predicted water levels for 1988 for selected well locations. Again, only the subset of principal wells were used in the linear regression and prediction. The match between the observed and predicted water levels was considered good, and in all cases the principal wells explained more than 90% of the variation.

The question is, can one improve the prediction by adding observations from nonprincipal wells into the regression? In other words, will the additional water level observations from nonprincipal wells improve the prediction? To answer this question, two cases are compared. In the first case, only the principal wells are used in formulating the regression and making the prediction. In the second case, regressions are formulated using all close wells within the search radius. Table 4 compares the two predictions with the observed water levels for PD32 and PD48. To evaluate the prediction using principal wells, it is important to know the order of magnitude of annual water level fluctuation in these wells. The average annual water level fluctuation for well PD32 is 1.10 m and that for well PD48 is 0.75 m. In the case of PD32, if seven wells are monitored instead of five, the improvement in prediction of water level is only by 4 cm. For well PD48, the error in water level prediction changed from about 0.5 m using only the two principal wells to 0.05 m using all six wells within the search radius. Because the primary motivation of this study was long term decline in water levels, prediction error of 0.5 m (less than one-half the annual fluctuation) was considered adequate for the present analysis. Thus the worth of monitoring only the principal wells is clearly reflected. The additional observations from nonprincipal wells explain a small fraction of the water level fluctuation for all practical purposes.

Network Rationalization

After performing principal component analysis on all monitoring wells, the wells were ranked. The rank provides a nondimensional representation as to how a well performs in representing the dynamic variation of potentiometric head, and is calculated as the ratio of the number of times it is selected as a principal well to the number of times it has entered the principal component analysis. The ranks of the well in descending order are given in Table 5. High rank indicates that a specific well was significant in explaining the dynamic variation in the aquifer. On the other hand a ranking of zero indicates that a well may not be worth monitoring because the variation in the aquifer is better explained by monitoring other neighboring wells. For the PD aquifer (Figure 2), 13 wells—PD18, PD19, PD52, PD58, PD64, PD66, PD67, PD69, PD70, PD72, PD73, PD75, and PD76—have been designated as boundary wells and are recommended for monitoring. These wells are needed to assign and verify appropriate new boundary conditions for the simulation model.

Following the rank (number of times a well appears as principal well divided by the number of times this well has entered PCA) of each monitoring well and after designating wells as boundary wells (wells which appear in all groups; 13 wells in this case), cutoff values were defined to group wells with rank equal to 1.0, greater than equal to 0.9, 0.8, so on and so forth. The group corresponding to cutoff zero is the case where all the 86 monitoring wells of the PD aquifer have been considered. The number of wells in the groups corresponding to the cutoff values of 0.0, 0.1, ..., 1.0 are 86, 80, 79, 73, 64, 56, 40, 32, 24, 22, and 22, respectively (total 11 groups). For these cutoff groups, mean and standard deviation of potentiometric head values were calculated for each of the 13 years (1978–1990). The mean and standard deviation of head were then used to calculate the coefficient of variation (standard deviation of head for a particular year divided by the mean head of that year) for these 13 years. The results of these calculations are given in Table 6. The coefficient of variation (CV) is a measure of dispersion and hence provides a measure of uncertainty. Therefore, it can be used to quantify the magnitude of error in estimating the mean potentiometric head using wells of a particular cutoff group. Since the time variation of CV for each of the cutoff groups has a similar pattern–maximum CV value with 22 wells (cutoff = 1.0 and 0.9), gradually decrease till number of wells equal 56 (cutoff = 0.5), and practically have little variation until all 86 wells (cutoff = 0.0) are utilized (Table 6); an average CV value (average value from 13 years) was used to represent this temporal variation. In other words, if we plot the CV values versus the number of wells (or cutoff) for each of the 13 years, the variation pattern is the same for any year (Table 6), and the pattern coincides with the variation of the average CV value plotted against cutoff (Figure 3a) or number of wells (Figure 3b).

The CV values given in Table 6 can be used to calculate the difference, the error (expressed in percentage) in CV values between the base case (where all 86 wells are considered), and the cases with less than 86 wells. This difference can be used as a criteria to judge the goodness-of-fit quantitatively for comparing contour

Table 6
Calculation of Potentiometric Head Statistics for Years 1978–1990

Group	Cutoff (Number of Wells)	Statistics	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	Average
1	0.0	E[h]	17.97	19.6	20.95	21.84	22.17	22.24	21.81	20.87	20.75	21.71	22.45	23.41	23.73	—
	(86)	SD[h]	10.37	10.39	9.94	9.88	9.55	9.28	8.73	8.02	7.83	8.16	8.37	9.15	9.4	—
		CV[h]	0.577	0.530	0.474	0.452	0.431	0.417	0.400	0.384	0.377	0.376	0.373	0.391	0.396	0.429
		Error(%)	0	0	0	0	0	0	0	0	0	0	0	0	0	—
2	0.1	E[h]	18.43	20.04	21.34	22.2	22.5	22.54	22.09	21.11	20.99	21.95	22.67	23.64	23.95	—
	(80)	SD[h]	10.5	10.53	10.04	9.98	9.66	9.39	8.85	8.15	7.95	8.29	8.51	9.32	9.56	—
		CV[h]	0.570	0.525	0.470	0.450	0.429	0.417	0.401	0.386	0.379	0.378	0.375	0.394	0.399	0.429
		Error(%)	–1.27	–0.88	–0.84	–0.63	–0.33	–0.16	0.09	0.47	0.37	0.48	0.69	0.87	0.77	–0.12
3	0.2	E[h]	18.36	19.97	21.22	22.07	22.37	22.4	21.96	20.97	20.86	21.81	22.52	23.5	23.79	—
	(79)	SD[h]	10.55	10.58	10.04	9.98	9.66	9.36	8.83	8.1	7.91	8.26	8.45	9.29	9.52	—
		CV[h]	0.575	0.530	0.473	0.452	0.432	0.418	0.402	0.386	0.379	0.379	0.375	0.395	0.400	0.430
		Error(%)	–0.43	–0.06	–0.28	–0.04	0.25	0.14	0.45	0.52	0.49	0.76	0.64	1.14	1.02	0.30
4	0.3	E[h]	18.21	19.79	20.84	21.69	21.99	22.02	21.61	20.64	20.56	21.5	22.26	23.24	23.54	—
	(73)	SD[h]	10.78	10.82	10.18	10.11	9.78	9.47	8.94	8.19	7.99	8.36	8.58	9.44	9.65	—
		CV[h]	0.592	0.547	0.488	0.466	0.445	0.430	0.414	0.397	0.389	0.389	0.385	0.406	0.410	0.443
		Error(%)	2.58	3.14	2.96	3.04	3.25	3.07	3.35	3.26	2.99	3.45	3.38	3.92	3.49	3.19
5	0.4	E[h]	18.01	19.56	20.59	21.45	21.73	21.78	21.36	20.4	20.34	21.31	22.07	23.13	23.43	—
	(64)	SD[h]	11.15	11.19	10.52	10.46	10.11	9.82	9.26	8.49	8.27	8.7	8.92	9.89	10.1	—
		CV[h]	0.619	0.572	0.511	0.488	0.465	0.451	0.434	0.416	0.407	0.408	0.404	0.428	0.431	0.464
		Error(%)	7.28	7.92	7.69	7.80	8.01	8.05	8.31	8.30	7.75	8.62	8.41	9.40	8.82	8.13
6	0.5	E[h]	16.74	18.22	19.37	20.25	20.6	20.69	20.34	19.52	19.49	20.48	21.26	22.51	22.86	—
	(56)	SD[h]	10.93	10.95	10.37	10.31	10.02	9.76	9.23	8.54	8.29	8.78	9.05	10.26	10.51	—
		CV[h]	0.653	0.601	0.535	0.509	0.486	0.472	0.454	0.438	0.425	0.429	0.426	0.456	0.460	0.488
		Error(%)	13.14	13.37	12.84	12.55	12.92	13.05	13.37	13.85	12.72	14.06	14.18	16.61	16.06	13.68
7	0.6	E[h]	13.9	15.41	16.59	17.45	17.9	18.05	17.8	17.1	17.26	18.06	18.8	19.96	20.23	—
	(40)	SD[h]	10.7	10.83	10.49	10.4	10.15	9.87	9.36	8.65	8.51	8.87	9.15	10.19	10.32	—
		CV[h]	0.770	0.703	0.632	0.596	0.567	0.547	0.526	0.506	0.493	0.491	0.487	0.511	0.510	0.564
		Error(%)	33.39	32.58	33.27	31.74	31.64	31.05	31.37	31.63	30.66	30.67	30.54	30.62	28.78	31.51
8	0.7	E[h]	11.95	13.43	14.54	15.34	15.81	16	15.88	15.32	15.73	16.37	17.12	18.21	18.44	—
	(32)	SD[h]	10.62	10.87	10.52	10.36	10.12	9.82	9.37	8.67	8.74	8.97	9.27	10.29	10.29	—
		CV[h]	0.889	0.809	0.724	0.675	0.640	0.614	0.590	0.566	0.556	0.548	0.541	0.565	0.558	0.637
		Error(%)	54.00	52.68	52.49	49.29	48.60	47.09	47.41	47.27	47.24	45.79	45.23	44.57	40.87	48.31
9	0.8	E[h]	7.81	9.18	10.51	11.41	12.04	12.34	12.44	12.21	12.91	13.4	14.13	15.14	15.46	—
	(24)	SD[h]	7.6	8.03	8.23	8.23	8.25	8.06	7.81	7.38	7.92	8.1	8.51	9.37	9.33	—
		CV[h]	0.973	0.875	0.783	0.721	0.685	0.653	0.628	0.604	0.613	0.604	0.602	0.619	0.603	0.690
		Error(%)	68.63	65.01	65.04	59.44	59.07	56.53	56.85	57.29	62.58	60.82	61.54	58.34	52.35	60.68
10	0.9	E[h]	6.61	7.91	9.19	10.09	10.68	11.11	11.3	11.2	12.06	12.54	13.25	14.26	14.61	—
	(22)	SD[h]	6.72	7.09	7.23	7.24	7.17	7.22	7.1	6.84	7.71	7.89	8.33	9.27	9.26	—
		CV[h]	1.017	0.896	0.787	0.718	0.671	0.650	0.628	0.611	0.639	0.629	0.629	0.650	0.634	0.705
		Error(%)	76.17	69.09	65.81	58.61	55.85	55.74	56.97	58.92	69.42	67.40	68.62	66.32	60.00	64.14
11	1.0	E[h]	6.61	7.91	9.19	10.09	10.68	11.11	11.3	11.2	12.06	12.54	13.25	14.26	14.61	—
	(22)	SD[h]	6.72	7.09	7.23	7.24	7.17	7.22	7.1	6.84	7.71	7.89	8.33	9.27	9.26	—
		CV[h]	1.017	0.896	0.787	0.718	0.671	0.650	0.628	0.611	0.639	0.629	0.629	0.650	0.634	0.705
		Error(%)	76.17	69.09	65.81	58.61	55.85	55.74	56.97	58.92	69.42	67.40	68.62	66.32	60.00	64.14

E[h], SD[h], and CV[h] are the mean, standard deviation, and coefficient of variation of potentiometric head, respectively.
E[h] and SD[h] are in meters; potentiometric head values are meters below MSL. The error is the percentage deviation of the CV[h] values from the CV[h] values of group 1.

plots constructed using the different cutoff values with the base case (potentiometric head contour plots using all 86 monitoring wells).

Figures 4a through 4i illustrate the head variation in the aquifer for the years 1978, 1984, and 1990. The potentiometric head values correspond to the data from the latest month in the last quarter of the year. These figures also compare the contour plots, the spatial distribution of potentiometric head when all 86 PD monitoring wells are used with the cases when 22 wells (13 boundary wells, plus nine wells with rank equal to 1.0), 40 wells (13 boundary wells, plus 27 wells with rank greater than equal to 0.6), and 56 wells (13 boundary wells, plus 43 wells with rank greater than equal to 0.5) are used. Each plot also shows the error or the goodness-of-fit for the different cases (Table 6).

To answer the question, how many wells to choose? The cutoff level (and hence the number of wells) from which the average CV does not significantly change upon adding additional wells would indicate the choice of the number of wells that would be required to monitor. In this case the choice of the cutoff level can be 0.5, that is, 56 wells, and the error (in terms of CV) in estimating the spatial distribution of potentiometric head would be approximately 13% to17% (Table 6) from the base case (using all 86 monitoring wells). The estimate of uncertainty measured in terms of the CV when a certain number of wells are considered for monitoring is valuable information for the decision maker

Depending on budget, the decision maker can select the number of wells, and from the present method can infer the well loca-

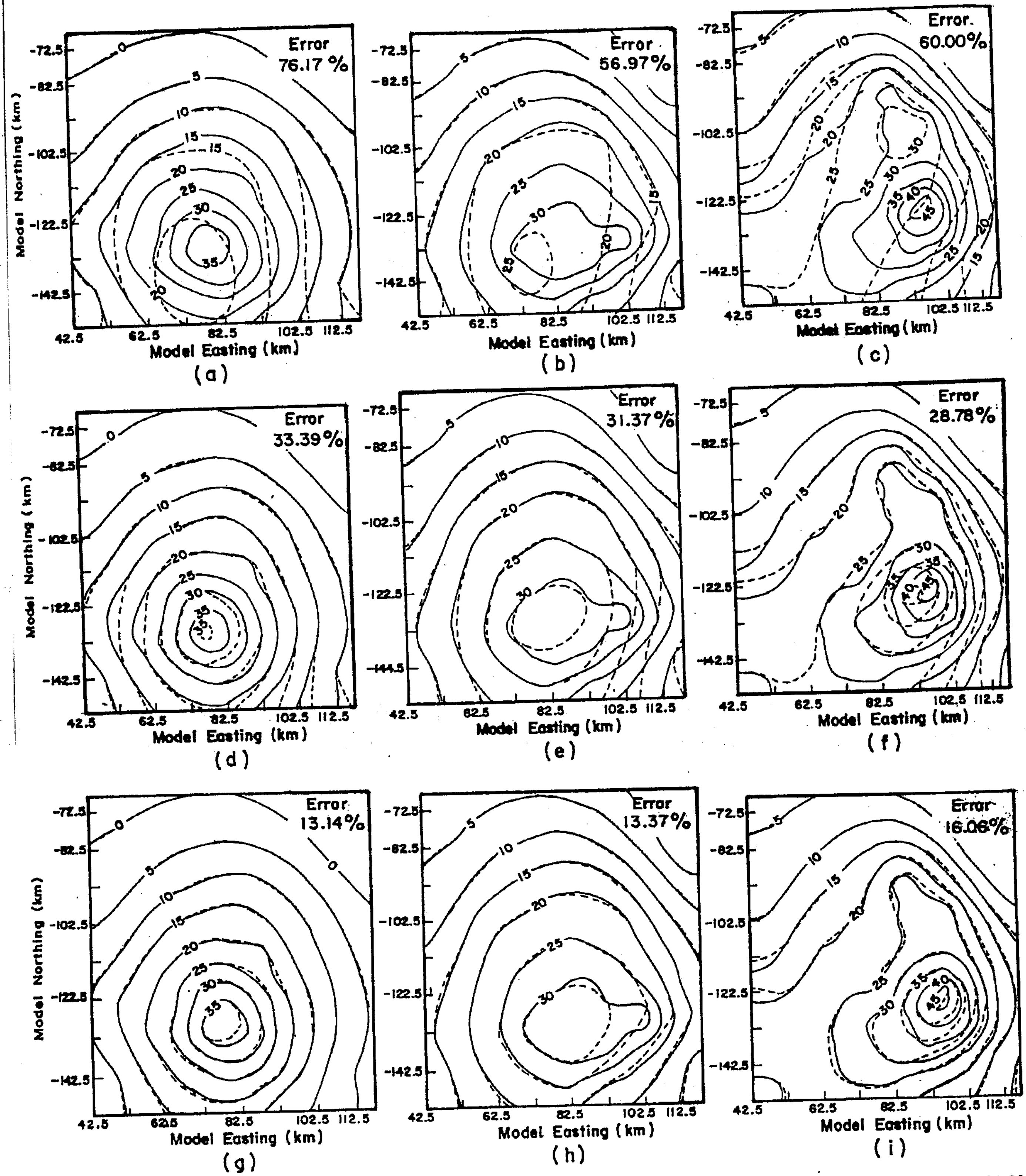


Figure 4. Comparison of contour plots of potentiometric head between the base case (86 wells, solid line) with (a) 22 wells in 1978, (b) 22 wells in 1984, (c) 22 wells in 1990, (d) 40 wells in 1978, (e) 40 wells in 1984, (f) 40 wells in 1990, (g) 56 wells in 1978, (h) 56 wells in 1984, (i) 56 wells in 1990; dashed lines. Head values are in meters below MSL.

tions, and the significance of these wells in representing the potentiometric head variation over time. The difference between existing network design techniques using a criterion such as estimation error (for example, kriging estimation variance) and the proposed method of network evaluation is that, inclusion of wells in the monitoring network is based on how significant the well has been in reflecting the water level variation of the aquifer.

Comparison of Present Method with Random Sampling

The most important advantage of the present method is that the relative importance of an individual well can be identified. This would not be possible if wells are selected based on uniform or random sampling. Within the chosen search radius several combinations of well groups are possible. If say, N wells fall within the selected search radius and groups of n wells are selected then ${}_NC_n$

combinations are possible. Now the question is, which of these combinations can significantly explain the potentiometric head variation over time in the region covered by the N wells? It should also be noted that the principal wells in the present approach are one of these random combinations. Though theoretically it is possible to calculate what fraction of the head variance from N wells is explained by any given combination of n wells, practically, this full enumeration is not plausible. Instead, if only the n principal wells are tested to see what fraction of the total head variance from N wells was being explained, then the choice of these wells can be justified. From the given potentiometric level data set and the subsequent analysis using principal components, 47 cases in which there were at least two principal wells were analyzed. It was found that in only five (about 10%) cases the principal wells could not explain the head variance arising from all the wells within that vicinity being analyzed. Furthermore, the discrepancy in not being able to explain variability in these five cases was between 2% and 8%, which again is not significant. This test confirms the good performance of the principal wells in adequately explaining the potentiometric head variation in the aquifer.

The Method: Its Strength and Weaknesses

In this study, the ground water flow model, based on the MODFLOW code, and calibrated by Gangopadhyay (1997) for the Bangkok ground water system was used to fill head values where historical data were absent. Typically, any ground water monitoring network evolves over time, as is true in this case. So in general, it would not be possible to have head observations for all the well locations during this period of development. As in this case, the 13-year period of analysis, 1978–1990, is the time during which the number of monitoring wells in the study aquifer (Phra Pradaeng Aquifer, Bangkok, Thailand) increased from only a few wells to 86 wells. This increase in the number of observation wells was as a part of the comprehensive program to monitor land subsidence caused by deep well pumping in the area covering Bangkok and its adjoining provinces.

In this case, it is important to note that as with many projects in the developing countries, a large number of these wells in the study area were constructed with donor funds. Within a few years after the donor support ends, the ground water management authority (Department of Mineral Resources [DMR], Ministry of Industry, Royal Thai Government) will be faced with the challenge of which wells to continue to monitor using the limited resources of the Department. In the light of this scenario, this methodology was developed to evaluate the existing ground water monitoring network.

When a flow model has been calibrated, one can use sensitivity coefficients or other statistical measures with the flow model to determine the relative importance of each monitoring well. However, the perception of the present problem has been to establish a methodology for evaluating and rationalizing a ground water monitoring network. The implication of rationalization and evaluation of the existing ground water monitoring network is that the decision maker also has the possibility to exercise expert knowledge in making an informed decision. In spite of having methodologies that combine flow models with sensitivity coefficients, kriging, and Kalman filtering, there is no single accepted methodology to evaluate or design a ground water monitoring network. In this context, the proposed methodology has several important highlights.

In an innovative way—through principal component analysis, the ranking scheme, and using cutoff levels—both the space and time dimensions are combined in the evaluation process. This suggested procedure identifies the performance of the wells to explain the temporal fluctuation of potentiometric head in the aquifer. Cutoff criteria group wells ranked from the principal component analysis, and it can be used to establish the level of uncertainty when a certain group of wells are monitored instead of monitoring all 86 wells. Thus the decision maker can have a feel for the relative importance of a given well, and this also gives the opportunity to exercise expert knowledge.

Also, it is not quite straightforward to incorporate both the space-time components in techniques such as kriging and gradient based methods. Furthermore, the flow simulation model combined with the proposed methodology can also be used to design and expand an existing ground water monitoring network.

The basic idea of using the principal component analysis to select wells for monitoring is a legitimate and efficient one. The idea of monitoring ground water level with this technique is new and unconventional, though the technique itself is not new. The filling in of missing data using a good flow model can also be justified by the fact that the physics of the flow system are better represented than if data filling is carried out using, for example, time series methods or by resorting to geostatistics.

In summary, though the use of the flow model in this case expresses a limited use of a flow model (which, of course, can be put to wider uses), this method holds significant promise. The method is general, technically sound, and practical, and its successful application has been demonstrated through a complex real world case study. For any evaluation or design of ground water monitoring networks, observations are necessary, and the use of a model to obtain missing or additional observations is pertinent.

Summary and Conclusions

An approach for ground water monitoring network evaluation using principal component analysis and ranking scheme is presented. This technique gives the decision maker a tool to rationalize an existing ground water monitoring network, and select the wells that should be continued to be monitored in the near future. The proposed methodology has been applied to the Phra Pradaeng (PD) Aquifer underlying the city of Bangkok and its adjoining provinces. In the analysis, it was assumed that all monitoring wells in this aquifer were present for the whole time period used in analysis (13 years, 1978–1990). All missing data were filled using a calibrated flow simulation model, based on the MODFLOW code.

The method aggregates the monitoring network evaluation procedure through different steps. First, the temporal performance of a well among a cluster of wells within a small spatial unit (area defined by a search radius) is identified. The spatial unit (search radius for this case was 10 km) was established from the physical behavior of the aquifer—the propagation of the response of the aquifer to unit pumping in the flow model. Then, through a nondimensional ranking scheme (the rank of the well was defined as the ratio of the number of times this well emerged as a principal well to the number of times the well occurred as a close well), the performance of this well in representing the overall regional head variation was identified. In this manner, the analysis was extended from a local scale to a regional scale. Furthermore, the method provides opportunity to exercise expert knowledge. In this case, 13 wells were assigned as boundary wells, and will be used to monitor whether a constant head boundary condition specification was a valid one for the regional ground water flow model. As the principal component tool was used to analyze temporal variation, its extension to the spatial domain through the ranking scheme provided a link to select the number and location of the wells to represent the distribution of potentiometric head.

Using multiple regression analysis, some wells were also analyzed to test how successfully the principal wells could predict future water level fluctuation at the dependent well locations. The average head in the PD aquifer dropped by about 6 m in 13 years (average heads in 1978 and 1990 were 18 m and 24 m below MSL, respectively), and the maximum head difference was 8 m. This change in the head values is significant, and the monitoring system has been evaluated based on such a scale of temporal head variation. If the head variation pattern in the aquifer changes significantly, it is recommended that the monitoring network be re-evaluated. With available estimates of future head values (for example, generated using a precalibrated flow model, and by assuming future scenarios of pumping), the proposed methodology can also be used to design a new monitoring network.

Although historical data was not available and a ground water flow model had to be used to fill missing data, and although the selection of search radius was based on unit response in a coarse finite-difference grid (5 km × 5 km), the present approach does provide significant hydrologic advice to the water manager. Most importantly, the method can rank the monitoring wells and identify the wells that would adequately capture the potentiometric head variation in the aquifer. This method also provides a measure of uncertainty (coefficient of variation, CV); the error occurred when a fewer number of wells were monitored. A cutoff level of the rank from which the average CV does not significantly change upon adding additional wells would indicate the choice of the number of wells that would be required to monitor. In this case study, the choice of the cutoff level satisfying this condition was 0.5, and this corresponded to 56 monitoring wells. The deviation of average CV (error) in estimating the spatial distribution of potentiometric head was estimated to be approximately between 13% and 17% from the base case (when all the 86 monitoring wells are considered). This error estimate, when a fewer number of wells are considered for monitoring, is valuable information for the decision maker. The obvious economic consequence would be reduction in the operational costs of the monitoring system. Hence, the proposed methodology can be effectively used by ground water management authorities to evaluate, monitor, plan, and manage an aquifer system for future operations.

Acknowledgments

The research work for this paper, which is a part of the first author's Doctoral of Engineering dissertation was funded by the Federal Republic of Germany through Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ). All necessary data was provided by the Department of Mineral Resources (DMR), Ministry of Industry, Royal Thai Government. We extend our gratitude to R. Loof for supporting this research work, V. Ramnarong for assisting in data collection, and S. Vedula for his suggestions in preparing this paper. We are grateful to the anonymous reviewers for their input which has considerably enhanced the quality of this paper.

References

AIT (Asian Institute of Technology). 1981. Comprehensive report 1978-1981: Investigation of land subsidence caused by deep well pumping in the Bangkok area. Bangkok, Thailand: National Environment Board.

Carrera, J., E. Usunoff, and F. Szidarovszky. 1984. A method for optimal observation network design for groundwater management. *Journal of Hydrology* 73, 147–163.

Cox, J.B. 1968. Research report no. 6. A review of engineering properties of the recent marine clays in Southeast Asia. Bangkok, Thailand: Asian Institute of Technology.

Department of Mineral Resources (DMR). 1994. Mitigation of groundwater crisis and land subsidence in Bangkok (MGL Project) report no. 2: Hydrographs of piezometric levels of groundwater in Bangkok and adjacent provinces. Bangkok, Thailand: DMR, Ministry of Industry, Royal Thai Government.

Gangopadhyay, S. 1997. Deterministic-stochastic modeling in a complex groundwater system. D.Eng. diss., Asian Institute of Technology, Bangkok, Thailand.

Haan, C.T. 1977. *Statistical Methods in Hydrology*. Ames, Iowa: Iowa State University Press.

Hudak, P.F., H.A. Loaiciga, and M.A. Marino. 1995. Regional-scale groundwater quality monitoring via integer programming. *Journal of Hydrology* 164, 153–170.

Heath, R.C. 1976. Design of ground-water level observation-well programs. *Ground Water* 14, no. 2: 71–77.

Illangasekare, T.H., H.J. Morel-Seytoux, and K.L. Verdin. 1984. A technique of reinitialization for efficient simulation of large aquifers using the discrete kernel approach. *Water Resources Research* 20, no. 11: 1733–1742.

Jawad, S.B., and K.A. Hussien. 1988. Groundwater monitoring network rationalization using statistical analyses of piezometric fluctuation. *Hydrological Sciences Journal* 33, no. 2: 181–191.

Journel, A.G., and CH.J. Huijbregts. 1978. *Mining Geostatistics*. London: Academic Press Inc.

Loaiciga, H.A., R.J., Charbeneau, L.G. Everett, G.E. Fogg, B.F. Hobbs, and S. Rouhani. 1992. Review of ground-water quality monitoring network design. *Journal of Hydraulic Engineering* 118, no. 1: 11–37.

Loaiciga, H.A. 1989. An optimization approach for groundwater quality monitoring network design. *Water Resources Research* 25, no. 8: 1771–1780.

Massmann, J., and R.A. Freeze. 1987a. Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy, 1. Methodology. *Water Resources Research* 23, no. 2: 351–367.

Massmann, J., and R.A. Freeze. 1987b. Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy, 2. Results. *Water Resources Research* 23, no. 2: 368–380.

McDonald, M.G., and A.W. Harbaugh. 1988. MODFLOW–A modular three dimensional finite difference ground-water flow model. Reston, Virginia: U.S. Geological Survey, National Center.

McKinney, D.C., and M-D. Lin. 1994. Genetic algorithm solution of groundwater management models. *Water Resources Research* 30, no. 6: 1897–1906.

McWhorter, D.B., and D.K. Sunada. 1977. *Ground-Water Hydrology and Hydraulics*. Fort Collins, Colorado: Water Resources Publications.

Olea, R. 1984. Sampling design optimization for spatial functions. *Mathematical Geology* 16, no. 4: 365–391.

Piancharoen, C., and C. Chuamthaisong. 1976. Groundwater of Bangkok metropolis, Thailand. In *Proceedings of the International Conference on Hydrogeology of Great Sedimentary Basin*, by Hungarian Geological Institute. Budapest: UNESCO.

Premchitt, J., and A. Das Gupta. 1981.Simulation of a complex groundwater system and an application. *Water Resources Research* 17, no. 3: 673–685.

Robert, P., and Y. Escoufier. 1976. A unifying tool for linear multivariate statistical methods: The RV-coefficient. *Applied Statistics* 25, no. 3: 257–265.

Rogers, L.L., and F.U. Dowla. 1994. Optimization of groundwater remediation using artificial neural networks with parallel solute transport modeling. *Water Resources Research* 30, no. 2: 457–481.

Rouhani, S. 1985. Variance reduction analysis. *Water Resources Research* 21, no. 6: 837-846.

Rouhani, S., and T.J. Hall. 1988. Geostatistical schemes for groundwater sampling. *Journal of Hydrology* 81, no. 1: 85-102.

Todd, D.K., R.M. Timlin., K.D. Schmidt, and L.G. Everett. 1976. *Monitoring Groundwater Quality: Monitoring Methodology*. Las Vegas, Nevada: U.S. EPA.

Wagner, B.J., and S.M. Gorelick. 1987. Optimal groundwater management under parameter uncertainty. *Water Resources Research* 23, no. 7: 1162–1174.

Wagner, B.J., and S.M. Gorelick. 1989. Reliable aquifer remediation in the presence of spatially variable hydraulic conductivity: From data to design. *Water Resources Research* 25, no. 10: 2211–2225.

Wagner B.J. 1995, Sampling design methods for groundwater modeling under uncertainty. *Water Resources Research* 31, no. 10: 2581–2591.