International Conference "Water, Environment, Energy and Society" (WEES–2009)
New Delhi, 12–16 January 2009

A Factor Analytic Model for Delineating Groundwater Source Areas

O.P. Dubey[1] and Krishna Kumar
UP Water and Land Management Institute
Walmi Bhawan, Utretia, Lucknow - 226 025, INDIA
E-mail: [1]ompdubey@yahoo.com

ABSTRACT: About 80% of the total water demand is being fulfilled by developing groundwater resources. For optimal sitting of groundwater developmental activities and sustained agricultural growth it is essential to assess the available resource. Resource evaluation by conventional methods is costly and time consuming. Data required for groundwater resource evaluation are not sufficient. When they are available their integration and decision making is difficult. In this study an attempt has been made to represent the groundwater potential system by a Factor Analytic Model (FAM). Input to the model has been given through synergistic use of remote sensing and ancillary data. The model can be used to delineate the groundwater source areas. In these areas detailed investigations can be carried out for optimal development of the available resources. The model compresses the data by 70% that makes the analysis simple.

BACKGROUND

Groundwater is a dependable source of water in rural areas. Occurrence of the groundwater in an area depends upon a wide range of above surface, surface and subsurface environmental parameters. The conventional method of delineation of groundwater source areas comprises of two categories of investigation. The surface investigations intend to investigate the surface parameters influencing the groundwater occurrence, whereas, the subsurface investigations aim to investigate the subsurface parameters influencing the groundwater occurrence. Delineation of ground water source areas involves decision-making keeping in view multiple interwoven criteria (Hamil *et al.,* 1996).

Generally, the data required for groundwater source delineation is either not available or inadequate (Howe, 1960, Rango 1974, Champati Rai *et al.,*. 1993, Dubey *et al.,* 1984, 1987). It is worth mentioning that the remotely sensed data has proven capability in providing many above surface, surface and sub surface characteristics of a land unit. Synergistic use of remote sensing and ancillary data can be made for the development of the database required for the delineation of the groundwater source areas. Further the capabilities of the GIS can be used to store, process and retrieve the developed database (Howe 1960, Dubey *et al.,* 1984, 1985, 1986, 1987, Champati Rai *et al.,* 1993, Hamil *et al.,* 1996n Chi K *et al.,* 1994).

Field engineers involved with the groundwater study find the collection of all prominent data its storage, processing and subsequent analysis difficult, costly, and time consuming job. Even if adequate data is available one of the problems arises as to how to integrate the available information to delineate the groundwater source areas in integrated manner so that detailed investigations can be carried out in the promising areas. In the present study an attempt has been made to propose a simple approach of groundwater source areas delineation by integrating influential parameters. For this an Analytic Model has been developed. The model has been applied in a part of Ganga Yamuna Piedmont.

THE ANALYTICAL MODEL

Occurrence of groundwater is influenced by conglomeration of a variety of parameters (Howe, 1960, Champati, 1993, Chi and lee, 1994, Walton, 1994, Hamil *et al.,* 1996, Dubey, 1986). Field engineers at the time of groundwater source delineation have to concurrently analyze a variety of data covering different aspects of influencing domain. It is a widespread observation that decision making considering all the influential parameters in one go is really a Herculean task. To simplify the efforts the analysis is generally based on many assumptions. In turn these simplifying assumptions only help in computational effort but drag down the accuracy of the work.

In the present study, surface and sub-surface characteristics of the land has been taken in to account while delineating the groundwater source areas. It has been hypothesized that the influential land surface characteristics are represented by Land use, land slope, distance from paleo channels, distance from flood plain, surface soil characteristics, and distance from urban areas. The subsurface characteristics of the land have been assumed to be represented by the aquifer media and permeability in vertical direction. Integration or blending of these land characteristics has been done by a weighing scheme (Rango, 1974, Dubey *et al.,* 1984, Dubey, 1985). The weights have been decided following an analytical approach (Satty 1996). The step by step procedure is being described below:

1. As mentioned the decision making process have been decomposed in to a set of cascades. The top level (Table 1) of the cascade is the goal of the analysis that is the groundwater source areas. The elements of the lower level include the attribute such as objectives perhaps even more redefined attributes follows at the next lower level-until the last level.
2. In the second phase, pair wise comparisons of the attributes or elements at a particular cascade level relative to their contribution or significance to the elements of the next higher cascade level is made. This phase constitutes much of the evaluation (qualitative) or assessment (quantitative) of the decision making process. The general principle of comparative judgments is applied in order to construct pair wise comparisons of the relative importance of elements. The scale assumed for the pair-wise comparisons is given in Table 2.

Table 1: Relative Variable Loading (RIW)

Goal	*Level*					
	Level I		*Level II*		*Level III*	
	Feature	*RIW*	*Feature*	*RIW*	*Feature*	*RIW*
G W P	Surface	(0.65)	Land use	(0.44)	Agriculture	0.40
					Water Body	0.25
					Barren Land	0.18
					Thin forest	0.10
					Thick forest	0.05
					Settlement	0.02
			Land Slope	(0.26)	Low Slope	0.72
					Mild Slope	0.21
					Milder Slope	0.07
			Distance from Paleo Channel	(0.18)	Less than 50 m	0.90
					More than 50 m	0.10
			Distance from Flood Plain	0.04	Up to 50 m	0.90
					More than 50 m	0.10
			Soil	0.06	Sand	0.56
					Sandy loam	0.27
					Loamy sand	0.13
					Clay	0.04
			Distance from Urban areas	0.02	Less than 0.5 km	0.65
					0.5km – 1.0 m	0.28
					More than 1.0 km	0.07
	Sub Surface	0.24	Aquifer Media	(0.5)	Sand and Boulder	0.63
					Sand Boulder and Clay	0.28
					Sand and Clay	0.09
			Permeability in Vertical Direction	(0.5)	High	0.90
					Low	0.04
	Ground Water	(0.11)	Groundwater Depth	(0.60)	< 5 m	0.73
					5–15 m	0.19
					> 15 m	0.08
			Rainfall Recharge	(0.32)	High	0.65
					Medium	0.28
					Low	0.07
			Water Quality	(0.08)	SAR Value Low	0.75
					SAR Value High	0.25

Table 2: Pair-Wise Comparisons of the Relative Importance of Elements

Relative Importance of a Parameter in Comparison to Another	*Number Assigned to Linguistic Expression*
Equal Preference of Indifference	1
Weak Preference	3
Strong Preference	5
Demonstrated Preference	7
Absolute Preference	9
Intermediate Values	2, 4, 6, 8

The scale has been validated for effectiveness. Based on this scale, comparative judgments are expressed as ratios resulting in a square reciprocal matrix as follows as shown below:

$$A = \{a_{jj}\} = \begin{array}{ccccc} : & & & & \\ 0 & 1 & 2 & & n \\ 1 & b_1/\mathrm{b}_1 & b_1/b_2 & \dots & b_1/b_n \\ 2 & b_2b_1 & b_2b_2 & \dots & \mathrm{b}_2b_n \\ n & b_n/b_1 & b_n/b_2 & \dots & b_n/b_n \end{array}$$

Where the *b/s* are the pair wise comparisons In pair wise comparison the dominant element is assigned an integer value, and the dominated element is assigned a fractional value. That is,

$a_{ji} = 1/a_{ij}$ for all i and j

Hence, diagonal elements in matrix A are always equal to 1. Also, the lower triangular elements of the matrix A are the reciprocal of the upper triangular elements. Therefore, the pair wise comparison data are required for only half of the matrix elements.

3. In the third phase, the pair wise matrix is decomposed spectrally. For this following assumptions have been made (1) the parameters are linearly related. (2) Composite parameter and unique factors have mean zero and standard deviation unity. (3) Common factor 'c' and unique factor 'e' is independent. The decomposition proceeds by imposing conditions that allows to uniquely estimate the eigen values and eigen vector or the parameter loading. Spectral decomposition provides an estimate of the Relative Influence Weight (RIW) of the elements at a particular cascade.

   For achieving the above mentioned objectives a linear mixing modeling has been done. Let us consider a multivariate system consisting of 'p' responses described by the observable random variables $X_1, X_2, X_3 \dots X_p$. The observable random vectors have mean x and co variance 'S'. The Linear Mixing Model (LMM) postulates that 'X' is linearly dependent upon few unobservable variables $F_1, F_2, \dots F_m$ called Linear Composite (LC) and additional source of variations $e_1, e_2, \dots\dots e_p$ called specific factors. Hence the model in matrix form may be written as,

   $$X - x = L\,F + e$$

   Where, $(X–x)$ is a vector having p elements containing deviations of observed variable X and its mean value x, L is matrix of LC loading having prows and m column, C is vector of Composite having m rows and e is error vector having p elements. From the above equation it is evident that "p" deviations $(X_1 - x_1) \dots (x_p - x_p)$ are expressed in terms of $(p + m)$ random variables, $c_1, c_2, \dots c_m, e_1 \dots e_p$. With so many unobservable quantities a direct solution of LMM from the observations on $x_1, x_2 \dots x_p$ is difficult. However, with the help of above mentioned assumptions the model reduces to simple and easy form. The model proceeds by imposing conditions that allow one to uniquely estimate the loading and the specific variance matrix. The loading matrix is then rotated, where the rotation is determined by some, 'ease of interpretation', method. Once the loading and the specific variance matrix are obtained composites are identified and estimated values for the composites themselves (called composites scores) are frequently constructed. The composites are determined so as to account for maximum variance of all the observed variables. The residual terms (i.e. specific factors e_1) are assumed to be small in this method,

   $$\mathrm{Min}\ [\{(X–x) - LF\}^T\ \{(X–x)–LF\}]^{-1}$$

   Subject to $\Sigma fi = 1, f_i > 0$

   A comparison was made between various themes, land elements on a common scale and a confusion matrix representing their relative importance was developed. The confusion matrix was decomposed spectrally in to components. The first component accounts about 90% variation in the data. The vector corresponding to this component represents the weights to different land characteristic that were considered influencing the decision. For data mixing RIW at different cascade level is shown in the Table 1.

   At this stage if the variable loading is not easily interpretable it may requires some rotation. The amount of rotation is determined by some, 'ease of interpretation', method.

   Once the loading and the specific variance matrix are obtained composites are identified and estimated values for the composites themselves (called composites scores) are frequently constructed. The composites are determined so as to account for maximum variance of all the observed variables. The residual terms (i.e. specific factors e_1) are assumed to be small in this method.

4. The Relative Importance Weight (RIW) corresponding to each level is determined by normalizing the eigen vector. The proposed system appears to be mathematically sound, but the pair wise comparison between various themes may differ from one investigator to another depending on their experience and familiarity of the area.

The developed model was calibrated using historical database consisting of 200 memory records. The database was subjected to Factor Analysis. Weights or membership function each cascade level for different land characteristics has been developed (Table 1). The importance of a particular land characteristic was decided on the basis of a linguistic measure of importance. The entire processing has been carried out through a model consisting above mentioned phases of analysis

Through modeling a decision support system for weighting a particular land characteristic keeping in view its influence on the occurrence of the ground water has been developed. The model finally ranks a particular land unit in to a predefined groundwater potential Class, based on its attributes. So that finally groundwater source areas can be delineated for further development. The developed model was calibrated in a part of Ganga Yamuna Piedmont. For this historical database consisting of 200 observed records has been used. A comparison was made between various themes, land elements on a common scale and a confusion matrix representing their relative importance was developed. The confusion matrix was decomposed spectrally in to components. The first component accounts about 90% variation in the data. The vector corresponding to this component represents the weights to different land characteristic that were considered influencing the decision. For data mixing RIW at different cascade level is shown in the Table 1.

CONCLUSIONS

For over all developments of a region reliable estimate of groundwater quality and quantity is of paramount importance. Generally sufficient data required for groundwater pollution potential mapping are not available for Indian watersheds. Satellite data can be analyzed to generate database required for GWPP studies. Generated database can be put to analytical model for extracting the most influential composite and subsequently the variable loading. The model efficiency was tested by carrying out field surveys and found to above 80 percent. The model can be used for evaluating the GWP in any area after calibration. The added advantage of the proposed approach is that it compresses the data up to 70% that helps in efficient analysis and prediction.

ACKNOWLEDGEMENTS

This research work is a small part of a research program for Bazada Land funded by MoWR through INCOH. The authors are highly thankful to the MoWR, INCOH in particular to Shri Masood Hussain and Dr Ramakar Jha, for their help and support in conducting the research program. Authors are also thankful to the organizations and individuals for their help in data sharing and data analysis.

REFERENCES

Champati Rai, Singh, L.M. and Rai, A.K. (1993). "Application of remote sensing and Geographic system in Hydrologic investigations of Imphal Valley (Manipur)." *Proc National symposium on remote sensing application for resource management with spécial emphasis on North eastern region Gauhati*, pp. 143–148.

Chi, K.H., Lee, B.J. (1994). "Extracting Groundwater Potential areas Using Remotely Sensed dat and GIS Techniques." *Proc. Integrated Applications of Remote sensing For Land and Water Resources Management* pp. 64–69.

Dubey, O.P., Sriniwas and Awasthi, A.K. (1984). "Analysis of Remote Sensed Data for Groundwater studies In Piedmont Zone." *Proc. Asian Confe. On Remote sensing,* Kathmandu.

Dubey, O.P. (1985). "Aquifer Response Modelling Using Satellite Data." *J. Hydrology*.

Dubey, O.P., Sriniwas, Parvez Ahmed (1986). Deciphering Groundwater Using Remote sensing Techniques.

Dubey, O.P., Sriniwas and Awasthi, A.K. (1987). "Groundwater Recharge Evaluation Using Satellite Data." *Proc. Integrated Water Resources management for drinking agriculture & industry*, pp. 37–40.

Hamil, L. and Bell, F.G. (1996). *Groundwater Resources Development.* Butterworth, London.

Howe, R.H. (1960). "Application of Aerial Photographic Interpretation of Hydrologic problems." *Photogram Engg.*

Rango, A. (1974). "New Dimensions in Satellite Hydrology." *Water Resour.* Vol. 106.

Saaty, Thomas L. (1996). *Decision Making with Dependence and Feedback*, RWS Publications, Pittsburgh, PA, 1996.

Walton, W.C. Groundwater Resources Evaluation. Mc graw Hill Book Co.