Entropy-Based Approach to Remove Redundant Monitoring Wells from Regional-Scale Groundwater Network *

Chen Zhihua  
Engineering Faculty, China University of Geosciences, Wuhan 430074  
Jing Juanli  
Graduate School, China University of Geosciences, Wuhan 430074  
Chen Gang  
Engineering Faculty, China University of Geosciences, Wuhan 430074

ABSTRACT: An entropy-based approach is applied to identify redundant wells in the network. In the process of this research, groundwater-monitoring network is considered as a communication system with a capability to transfer information, and monitoring wells are taken as information receivers. The concepts of entropy and mutual information are then applied to measure the information content of individual monitoring well and information relationship between monitoring well pairs. The efficiency of information transfer among monitoring wells is the basis to judge the redundancy in the network. And the capacity of the monitoring wells to provide information on groundwater is the point of evaluation to identify redundant monitoring wells. This approach is demonstrated using the data from a regional-scale groundwater network in Hebei plain, China. The result shows that the entropy-based method is recommendable in optimizing groundwater networks, especially for those within media of higher heterogeneities and anisotropies.

KEY WORDS: entropy, groundwater monitoring, network, optimize, redundant.

INTRODUCTION

Redundant Monitoring Wells (RMW) within a groundwater-monitoring network are those observation wells that have contributed little or no information to our understanding of the groundwater system being monitored and the removal of RMW would not significantly affect the ability of the network to provide information. As RMW in the network increase the cost of the network operation, removing these RMW from monitoring network is one of the main purposes of optimizing the monitoring network.

RMW are abundant in the monitoring networks that were founded prior to in 1960's. The reasons for the occurrence of RMW in the networks possibly resulted from two cases: Firstly, as the design of monitoring networks was mainly based on the hydrogeological approach, the number and location of monitoring wells were based on local hydrogeological conditions (Everett, 1980). Obviously, an individual understanding of groundwater laws and hydrogeological conditions impacts, whether the spatial configuration of the network is rational or not. The existence of RMW in the network designed at that time was almost unavoidable due to the limitations of knowledge and information about groundwater system. Secondly, for economic reasons, boreholes that were drilled for geological and hydrogeological investigations, pump testing and other purposes were adopted for many monitoring networks. Such diversity frequently results in unreasonable spatial distributions of monitoring wells. For example, the locations and numbers of monitoring wells were closely related with local permeability of aquifers and accessibility rather than the objectives of monitoring networks and local permeability of aquifers and accessibility rather than the objectives of monitoring networks and local

* This paper is supported by State Planning Committee, China.

Manuscript received March 28, 2002.  
Manuscript accepted May 20, 2002.
hydrogeological circumstances. Hence, dense distribution of wells in an area with high permeability and good accessibility, and sparse distribution of wells in an area with low permeability and/or difficulty to access should not be surprising.

Previous researches related to the rationalization of monitoring networks were widely distributed, but they are more concentrated on the field of hydrology. These works include the application of principal component analysis to analyze an existing precipitation network (Morin et al., 1979). Galeati et al. (1986) also used the principal components analysis but coupled with clustering techniques to optimize a snow data collection network. Burn and Goulter (1991) demonstrated a hierarchical clustering technique combined with users’ judgments to rationalize a streamflow data collection network in Pembina basin, southern Manitoba, Canada. Timothy and Lucila (1990) presented two statistical approaches: the stochastic technique and geostatistical technique for optimizing the groundwater quality network in the Llobregat Delta, Spain. Although the methods were easy to understand and use in practice, these methods relied on a statistical relationship between sampled numbers and standard errors. The selection of wells’ number is based on the expected standard error, which is just the determination of well’s number but not the spatial configuration of these wells. So it is impossible to determine exactly where the RMW are located. The key of removing RMW from the network is the determination of RMW. It is troublesome to overcome this problem using only statistical methods relying on some statistical indexes. Many difficulties arise when conventional statistical methods are applied to the design of spatial data collection network, and lead to the adoption of narrowly defined statistical objectives (William and Tahir, 1980).

At the end of the 1970s, some hydrologists introduced the application of Shannon’s information theory in a hydrologic data collection network. For example, William and Tahir (1980) confirmed that information transmission among different stations is dependent upon the joint entropy of the monitoring station’s output but independent of the spatial estimators. Yang and Donald (1994) illustrated the entropy approach to design gauging stations. Harmancioglu et al. (1992) and Ozkul et al. (2000) demonstrated the entropy-based assessment of water quality monitoring networks in the case study of Mississippi River. All these various applications are focused on the rationalization of gauging stations in catchments, which are fewer in numbers and easier to identify than RMW in groundwater monitoring networks. The methods used for optimization in hydrology and hydrogeology are similar in principle, but in hydrogeology, eliminating RMW from a monitoring network usually needs consideration of the heterogeneities and anisotropies of the aquifer media.

The aim of this research is to extend the entropy-based methods to hydrogeology, and identify RMW in monitoring networks in terms of the information relationships among different monitoring wells. The dataset used in this case study are taken from the water level observations of a regional-scale monitoring network in Hebei plain, China. It includes 140 monitoring wells and their 10-year observation (1981—1990).

**ENTROPY-BASED APPROACH**

Since a monitoring network has an ability to transmit hydrological information, groundwater monitoring networks can be considered as a signal communication system. This concept permits the monitoring network to be redesigned on the basis of its ability to convey hydrologic information (William and Tahir, 1980). When the similarities between the communication system and the network are compared, the groundwater hydrodynamic field plays the role of conducting media for transferring water level signals. And all monitoring wells in the network are the receivers of water level signals. Similarly, three inherent characteristics of the communication system in signal transmission exist in any monitoring network (1) signal transferring; any variations of water level in the hydrodynamic field will be reflected in all monitoring wells, even though the responses may differ in time and/or in extent; (2) signal decaying; information transfer among monitoring wells decay with an increase in the distances between monitoring wells pairs; (3) signal identifying: the variations of groundwater levels are the signals presenting various states of groundwater system, and there should be some differences in the signal characteristics because the monitoring wells are located in different hierarchical groundwater systems and/or different geological formation. These differences in signal characteristics can be described by their probability distributions and used to identify their relationships. In short, since there are many similarities between monitoring networks and communicating systems, we
can adopt the concepts of entropy to quantitatively measure the info-relationships among monitoring well pairs and then to recognize RMW from groundwater monitoring networks.

**ENTROPY AND MUTUAL INFORMATION**

Entropy is a fundamental concept in information and communication theory. It can quantitatively describe the uncertainty of a random event, or in other words, the information contained in the random event through the observations of it (Yang and Donald, 1994). According to Shannon, a random event's information content depends upon the probability distribution of its signal occurrence. A random event with high probabilities has little information and vice versa. Here, the probability of occurrence of a random event is the measurement of its uncertainty (entropy). In this sense, information and uncertainty are dual terms and are often used interchangeably (Ozkul et al., 2000).

For a monitoring well, its information content relies on the statistical structure of its observations. If the fluctuations of water level are very small, its entropy is approximately zero and no information is contained in its observations. If its observations occur with different probabilities, there is an uncertainty in the random event. The more complicated the fluctuations of water level are, the higher the entropy the monitoring well has. Therefore, the entropy can be applied to measure the capability of a monitoring well to provide information of groundwater.

Entropy is formulated in terms of probabilities for a discrete set of observations, i.e., $X = [x_1, x_2, \ldots, x_N]$ representing observations of a monitoring well with a probability function $p(x_i)$. $x_i \in X$. The entropy of the monitoring well expresses an expected information content using the following function

$$H(X) = -\sum_{i=1}^{N} p(x_i) \ln p(x_i)$$

$$\sum_{i=1}^{N} p(x_i) = 1$$

Here, $H(X)$ is entropy, described as the information contained in monitoring well $X$, the unit of $H(X)$ is expressed by Napier for natural logarithms, and $N$ is the number of observations for a random observations series from monitoring well $X$.

Another important concept gained from communication theory is the mutual information $T(X_1, X_2)$, which represents the information transfer between monitoring well pair $X_1$ and $X_2$. The mutual information can be described as the difference between the entropy $H(X_j)$ and conditional entropy $H(X_j | X_i)$

$$T(X_1, X_2) = H(X_1) - H(X_1 | X_2)$$

Equation (3) can be written as

$$H(X_1) = T(X_1, X_2) + H(X_1 | X_2)$$

Equation (4) describes the actual relationship between information transfer $T(X_1, X_2)$ and information loss $H(X_1 | X_2)$. Since entropy $H(X_1)$ is a fixed value derived from $X_1$'s observations, the information transfer and information loss will change inversely. Hence, when the spatial distance $D_{ij}$ between $X_i$ and $X_j$ increases, the information loss rises and the information transfer decay at the same time. Mutual information can also be presented as (Harmancioglu et al., 1992)

$$T(X_1, X_2) = H(X_1 | X_2) + H(X_1) - H(X_1, X_2) = 0$$

This indicates that there is no information transfer between the monitoring well pair $X_1, X_2$, and $X_3$ if there is a statistical relationship between monitoring well pairs $X_1$ and $X_2$, which shows that there is an information transfer between $X_1$ and $X_2$.

The calculation of $T(X_1, X_2)$ is a complicated procedure for multivariable. However, if the statistical distributions of the monitoring well's observations can be described in terms of normal or lognormal distribution, entropy $H(X_j)$ for single variable and joint entropy $H(X_j | X_j)$ for multiple variables can be calculated by the following simplified

$$H(X) = (M/2) \ln 2\pi + (1/2) \ln |C| + M/2 - M \ln(\Delta x)$$

where $M$ is the number of variables, $|C|$ is the covariance matrix, $\Delta x$ is the interval size assumed to be the same for all $M$ variables (Ozkul et al., 2000; Husain and Khan, 1983).

**Identifying RMW**

Before removing RMW from a monitoring network, we need a criterion to identify where the potential RMW are located and which monitoring wells are the RMW. Usually, it is very complicated to lay down criteria since it interrelates many factors in different conditions.

In this investigation, the information transmission $T(X_1, X_j)$ between a pair of monitoring wells has been adopted as a measurement, which is information standard instead of standard error or assembly average. With this method, we can find out accurate-
ly those RMW in a network. For example, there are 5 monitoring wells in a certain area in Fig. 1, the questions are: Is it necessary to include all the monitoring wells? Which are superfluous among these monitoring wells? And which one can replace another while their information transferring is high enough? The procedure of using mutual information $T(X_i, X_j)$ to identify the RMW from a network is described as follows.

Generally, monitoring well X has information transfers $T(X_i, X_j)$ ($i=1,2,3,4$) with its surrounding monitoring wells $X_1, X_2, X_3, X_4$. Since the information transfer among monitoring wells is essentially a function of distances between their locations, i.e., the amount of info-transmission decrease while the distance $D_i$ increases. So, if the aquifer media is homogeneous and $D_1 < D_2 < D_3 < D_4$, it is true for the relation: $T(X, X_1) > T(X, X_2) > T(X, X_3) > T(X, X_4)$.

Suppose $T(X_i, X_j) \geq \eta$, when the information transfer $T(X_i, X_j)$ between well pair is high enough that either of the well can be replaced. Therefore, one of the wells can be recognized as the redundant monitoring well. $\eta$ is the information transfer threshold. For instance in Fig. 1, if $T(X_i, X_j) \geq \eta$, the monitoring wells $X_1$ and $X_2$ are considered as the redundant monitoring wells as they have the same information as the monitoring well X and therefore should be eliminated.

For the monitoring well pairs $X, X_i$, since their information transfer $T(X, X_i) < T(X, X_j) < \eta$, that means the formation from monitoring well X is different from that of monitoring well $X_i$, and new information can be obtained from $X_i$’s observations. So, it is necessary to keep $X_i$ as another information source about the groundwater.

From the procedures mentioned above, we can conclude that the existence of RMW in the network depends mostly on the monitoring well’s capabilities to transfer information within nearby monitoring wells. A high information-transferring capability of a monitoring well implies that it can keep strong information-relationships with its adjacent wells. That is the information from the well about the groundwater has significant representation in a wide coverage. Therefore, evaluating information-transferring capacities of monitoring wells is the foundation for identifying RMW in the network.

Even though the information transfer $T(X_i, X_j)$ is essentially a function of the distance between a monitoring well pair, its decaying characteristic would be diverged in different media. UKayli et al. (UKayli M. Humaid T, Khan H U., 1983, Meteorological Network Design and Optimization for Saudi Arabia, Final Report Prepared for Meteorology and Environmental Protection Administration, Jeddah, 325) described this characteristic in the meteorological network as an exponential curve. Ozkul et al. (2000) investigated the information transfer between hydrological stations of Mississippi River, and their result shows that it is approximately a semi-exponential regression. The relationship between information transfer and distances in groundwater monitoring network would differ from that of air or water media because of the difference in behaviors of the medium. Apart from the distance between monitoring wells, aquifer’s characteristics, including transmissibility, heterogeneity and anisotropy, are also important factors that influence information transferring capability of monitoring wells. Hence, this information relationship in monitoring network should be validated based on the observations from groundwater monitoring wells.

**APPLICATION**

**Background**

The approach is demonstrated by using the data from the monitoring network of Hebei plain, located at the northern part of China (Fig. 2). Hebei plain covers an area over $6.0 \times 10^5$ km$^2$, with a general topography from the western hill-plain declining to the eastern sea level. Geologically, various unconsolidated deposits including diluvial, alluvial and marine compose the main aquifers. Regional aquifers receive recharges mainly from the sub-runoff of western mountain areas and rain infiltrate, and finally discharge to eastern sea. The complicated deposition history of the unconsolidated sediments results in the

![Figure 1. Info-transferring with adjacent monitoring wells.](image-url)
aquifers with high heterogeneities. Previous investigations delineate the hydrogeological characteristics of the regional aquifers; western pluvial aquifers consist of large coarse-grained deposits. The transmissibility is in the range of 500 to 1 000 m²/d. In the middle of Hebei plain, aquifers contain inter-fingered fluvial fine or medium sand and bedded clay with 50—100 m²/d transmissibility. The eastern parts of regional aquifers are composed of fine-sand, clay and silt, predominantly marine deposits with transmissibility less than 50 m²/d (Xiao, 2000; Chen, 1998) (Report from Geo-Environmental Institute in Hebei, 1998. The Countermeasures to the Geo Environ-
mental Problems Resulted from Over-Extracted Groundwater in Hebei Plain (in Chinese)).

![Figure 2. Geological background and distribution of monitoring wells in Hebei plain.](image)

In order to verify the probability distribution forms of the observations from 140 monitoring wells, \( X_2 \) statistical test and plots for the Quartiles of a variable's distribution against the Quantiles of any of a number of test distributions (Q-Q Plot) were carried out (SPSS for Windows Manual). Test results have shown that their probability distributions fit approximately normal distributions under 95% confidence.

### Information Transfer Capability in Aquiferous Media

After calculating using \( T(X_i, X_j) \), using equation (6) and equation (5), statistical techniques (e.g. linear, exponential and logarithmic approximation, power function) have been used to investigate the relations between \( D_{n,j} \) and \( T(X_i, X_j) \). Their statistical relations can be summarized as: (1) all the nonlinear curves could be matched best with logarithmic equations, whatever the decaying of \( T(X_i, X_j) \)
occurs in any geological sediment. This implies that there is log-type relationship between $T(X_i, X_j)$ and $D_{n,j}$; (2) the relationships between $T(X_i, X_j)$ and $D_{n,j}$ become more definite after the data pairs are grouped according to the flow directions and geological properties, this reflects that the heterogeneities of aquiferous media have blurred the differences of the information transferring capabilities; (3) comparing the data pairs from alluvium and diluvium, the latter possesses higher capability to convey information but with a relatively high dispersion. The monitoring wells located at diluvium have higher ability than those in alluvium to transfer information since diluvial sediments have higher transmissibility. Even if there are relative dispersions around these decaying curves, both of them already have statistical significances of 95 % confidence intervals (Table 1).

![Figure 3. Comparisons of the water level contours with different info-transfer thresholds.](image)

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Comparison of information transfer in different geological deposits</th>
</tr>
</thead>
<tbody>
<tr>
<td>geological deposits types</td>
<td>diluvium</td>
</tr>
<tr>
<td>transmissibility</td>
<td>high</td>
</tr>
<tr>
<td>heterogeneity and anisotropy</td>
<td>poorly sorted (high)</td>
</tr>
<tr>
<td>statistical relations decaying coefficients</td>
<td>logarithmic decay 0.4246</td>
</tr>
<tr>
<td>averages of $T(X_i, X_j)$/Napier</td>
<td>1.084 6 (sample size=44)</td>
</tr>
<tr>
<td>distance/km</td>
<td>max 1.717</td>
</tr>
<tr>
<td>statistical averages</td>
<td>20 km 1.4</td>
</tr>
<tr>
<td></td>
<td>40 km 1.2</td>
</tr>
<tr>
<td></td>
<td>60 km 1.0</td>
</tr>
<tr>
<td>dispersion of data pairs</td>
<td>high</td>
</tr>
</tbody>
</table>

**Removing RMW from Network**

Identification of RMW is a searching procedure of RMW judged by the specific threshold $\eta$. The searching radius is the effective distance of a monitoring well to transfer hydro-information among adjacent wells. The searching procedure includes two steps: (1) calculating the $T(X_i, X_j)$ between each pair of monitoring wells within the radius; (2) if the
information transfer $T(X_i, X_j) \geq \eta$, then compare their entropies $H(X_i)$ and $H(X_j)$. Usually, the monitoring well with higher entropy will be remained, but a few of these monitoring wells are selected by means of local hydrogeological conditions.

Generally, adopting a high threshold of information transfer means that the network can have a higher density and accuracy, which is suitable for local and small-scaled purposes, and low criterion is chosen for the regional monitoring network. In this study, we have designed several information transfer thresholds ($\eta = 1.6$, 1.4, 1.2, 1.0, 0.8) to identify the redundant monitoring wells in the Hebei plain network.

After performing this process on the regional monitoring network, we have identified the RMW corresponding to various thresholds and calculated their relative errors when the RMW are removed (Table 2).

We chose the contour map interpolated by 10 m intervals as the typical case to discuss because this accuracy is currently used in the monitoring network of Hebei plain.

When the $T(X_i, X_j) = 1.6$ is info-transfer threshold, the result shows that only 6 out of 140 monitoring wells are identified as the RMW. All of these RMW are distributed in the diluvial sediments. If we lower the threshold to 1.2 with 40 km searching radius in diluvial deposits and 10 km in alluvial deposits, the number of RMW increases to 29. Among these 29 RMW, 18 monitoring wells are distributed in the diluvial deposits and the remaining are in alluvium deposits. If the threshold is lowered further to 0.8, the number of RMW rises to 56, but more than 50 % of the RMW come from the alluvium area. So, after evaluation using the same information threshold, the network densities change with the difference in transmissibility of the media. That is, the densities of monitoring wells match the different geological and hydrogeological backgrounds automatically.

In order to validate whether the information contained in observations of these RMW is significant or not, we have calculated and compared their relative errors under different water level contours interpolated by 5 m, 10 m, and 20 m intervals. The relative errors for each info-transfer threshold are listed in Table 2; it represents an area difference between water level contours, interpolated based on all the monitoring wells and the monitoring wells in which the RMW were removed. Figure 3 shows two contour maps when the threshold equals to 1.4 and 1.2, separately. Comparison of the two maps indicates that both maps are so similar to the original contours that there is no tangible difference in the spatial configurations. It implies that there is no significant change in the information in the regional groundwater-level even when 29 RMW are eliminated. Normally, the total relative error is only 1.7 % or 3.63 %. This comparison confirms that the 29 wells selected from existed monitoring wells by entropy method are really the redundant monitoring wells. The remaining 111 monitoring wells are enough to characterize the spatial states of the regional groundwater in Hebei plain. Most likely, the benefit of saving 20, 7 % monitory costs but losing 3.63 % information is acceptable for a regional scale groundwater-monitoring network.

**CONCLUSIONS AND SUGGESTIONS**

This study shows that the entropy-based approach is a recommendable method to optimize and redesign the existing monitoring networks because of its theoretical and practical advantages. Theoretically, both the stochastic probability techniques and information and communication theory are integrated to evaluate the capability of providing information for individual monitoring well as well as the hydro-information transmissibility within different monitoring wells. The foundation of the method is the quantitative measurement of information from monitoring well rather than a traditionally used statistical averaging or a mean accuracy approach. Practically, this method is very suitable for optimizing groundwater networks, especially in groundwater systems with higher heterogeneities, for this technique can extract hydro-information directly from the observation of monitoring wells rather than using the parameters that describe the nature of groundwater media, which are too difficult to obtain. Therefore, the methodology used in this case study can be applied to any exist-

**Table 2  List of info-transfer thresholds and relative errors (RE)**

<table>
<thead>
<tr>
<th>threshold/</th>
<th>num of</th>
<th>RMW/</th>
<th>RE/</th>
<th>RE/</th>
<th>RE/</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nabier/</td>
<td>RMW</td>
<td>$\Delta=5$ m</td>
<td>$\Delta=10$ m</td>
<td>$\Delta=20$ m</td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td>6</td>
<td>6/140=0.5</td>
<td>1.28</td>
<td>0.57</td>
<td>0.32</td>
</tr>
<tr>
<td>1.4</td>
<td>18</td>
<td>12.6</td>
<td>3.66</td>
<td>1.77</td>
<td>0.53</td>
</tr>
<tr>
<td>1.2</td>
<td>29</td>
<td>20.7</td>
<td>7.34</td>
<td>3.63</td>
<td>1.77</td>
</tr>
<tr>
<td>1.0</td>
<td>39</td>
<td>27.9</td>
<td>9.45</td>
<td>6.64</td>
<td>3.32</td>
</tr>
<tr>
<td>0.8</td>
<td>56</td>
<td>32.8</td>
<td>13.55</td>
<td>9.75</td>
<td>3.51</td>
</tr>
</tbody>
</table>
ing groundwater monitoring network to remove RMW if their observations are available. Another ob-
vious benefit of the entropy-based approach is that it can exactly identify those redundant monitoring wells if they exist in the network. Moreover, it also allows user or expert's judgments in the selection process of monitoring wells. Except for information trans-
fering criterion, various information can be included.

The result of this study shows that 29 out of 140 monitoring wells are identified as RMW and can be removed from the present network. It is necessary to state that removing redundant monitoring wells does not mean improving the accuracy of the network, for except for the redundant information existing in the network, some other deficiencies, such as inelastic spatial distributions or configurations of network can result in a lack of information. So, it is necessary to evaluate the potential lack of information of a network before a complete groundwater-monitoring net-
work is formed.

Finally, the information transfer criterion can be adjusted easily to match those different motoring net-
works with various purposes and/or scales. How-
ever, as to the issue of which standard is suitable for a regional scale or what is appropriate for a local scale monitoring and the corresponding relationships be-
tween criteria and network's purposes and scales need further study. Therefore, the determination of the information-transferring criterion is still an issue that needs further research in hydrology and hydrogeo-
logy.

ACKNOWLEDGMENTS

The authors thank Dr Ji Chen in the University of Illinois at Champaign & Urbana for many useful discussions. The manuscript benefits from thought-
ful reviews by Laura Woodhouse.

REFERENCES CITED

drology, 122: 71–91


nectady. General Electric Company

drology Science Journal, 31(1): 93–108

Harmancioglu N B, Necdet A. 1992. Water Quality Moni-
toring Network Design; A Problem of Multi-Objective De-


Pan G C. 1995. Related Information Measures for the Associa-
tions of Earth-Science Variables. Mathematical Geology, 27(5): 5–12


ning and Management Division. 106: 503–520

Xiao G Q. 2000. Research on the Evolution of Quaternary Groundwater System in Hebei Plain under Human Activi-
ties. [Dissertation]. Wuhan: China University of Geosci-
ces. (in Chinese)

Yang Y, Donald H B. 1994. A Entropy Approach to Data Col-