Quantitative estimation of natural recharge due to monsoon rainfall using the principle of information theory in the area of Ghorawal block of Sonebhadra district, U.P., India

Lazarus G. Ndatuwong, G. S. Yadav
1- Banaras Hindu University, Department of Geophysics, 221005 – Varanasi, India
2- Adamawa State University, Department of Physics, PMB 25-Mubi, Adamawa State–Nigeria
ndatuwong@gmail.com
doi: 10.6088/ijes.2012030133005

ABSTRACT

The ground water recharge rate due to monsoon rainfall is estimated for the area of Ghorawal block in Sonebhadra district of Uttar Pradesh (U.P.), India, using the entropy method from the principle of information theory. Average water level measured in ten (10) different locations in Ghorawal block and the monsoon rainfall recorded in the district from year 2007 - 2010 was used for the analysis. The contribution due to the monsoon rainfall during this period was obtained by calculating the marginal and mutual entropies of the monsoon rainfall and the depth to water level. A ratio of mutual entropy to marginal entropy of rainfall was used as a measure of the fractional amount of rainfall that recharges the ground water resources in the study area. An estimated rate of 12.1% of the rainfall measured during the study period was calculated.

Keywords: Information theory, Entropy, water table, monsoon rainfall, recharge, hard rock.

1. Introduction

Recharge commonly means different things to different people. Someone concerned with crop growth may consider recharge as the amount of water that moves beneath the root zone, thus adding to the soil moisture. Someone who is interested in determining the characteristics of an aquifer may view recharge as the amount of water that reaches and contributes to the storage of the aquifer (Younger, 2007). It is the principal causes of the rises in water table. Water that reaches the water table and contributes to the net input of ground water storage within the regional ground water flow system requires a measurement of recharge on a regional scale. Such measurements are helpful in determining the amount of water that is necessary to maintain a safe yield within an aquifer system (de Silva, 2004).

Ground water recharge is a key component in any model of ground water flow or contaminant transport. To properly manage and protect the valuable resources, accurate quantification of recharge rate is needed. Estimating the rate of aquifer replenishment is probably the most difficult of all measurements in the evaluation of ground water resources (Israil et al., 2006). Estimates are normally and almost inevitably subject to large errors. No single comprehensive estimation technique can yet be identified from the spectrum of those available, which gives reliable results (Scanlon et al., 2002). Estimating recharge rate can be based on a wide variety of models which are designed to represent the actual physical...
processes. The methods, commonly in use for estimation of natural ground water recharge, include ground water balance method, soil water balance method, zero flux plane method, one-dimensional soil water flow model, inverse modeling technique, isotope and solute profile techniques (Healy 2010).

Ground water occurrence in the crystalline basement aquifer has been noted to be characterized by the presence of a shallow water table and recharge is mainly from rainfall. The water levels have been observed to follow a seasonal fluctuation pattern influenced by the rainfall pattern. Athavale et al., 1992 carried out a study in twenty (20) river basins in India in order to assess the effect of low infiltration rates of rainfall in to the soil. They suggest that about 15 – 20% of seasonal rainfall contributes to ground water recharge in the Indo-Gangetic plains, figures that falls to only 5 – 10% in peninsular hard rock regions. The most commonly used methods for estimation of natural ground water recharge in India include empirical methods and ground water level fluctuation method. Based on the studies undertaken by different scientists and organizations regarding correlation of ground water level fluctuation and rainfall, some empirical relationships have been derived for computation of natural recharge to ground water from rainfall (GEC, 1997). This study however explores the use of the principle of information theory known as the entropy method to estimate the recharge rate in the study area.

2. Locations and geology of the study area

![Map of study area](image_url)

**Figure 1:** Map of study area.

*Lazarus G. Ndatuwong, G. S. Yadav*
International Journal of Environmental Sciences Volume 3 No.3, 2012
The study area is part of the Kaimur series of the Upper Vindhyan Super group located in the southern part of Uttar Pradesh State, (figure 1). The area falls within the survey of India toposheets 63L/9, 63L/10, 63L/13, 63L/14.

The Kaimur series have the maximum horizontal and vertical extension in the eastern part of the Vindhyans towards Mirzapur and Sasaram. The Kaimur sandstones and quartzites are prominent horizons with wide distribution. It is more quartzites than sandstone. All the major sandstone horizons form the scrap while the shale horizons form gentle slopes. The lower kaimur consists of fine to medium grained quartzites and sandstones of pinkish colour characterized by cross bedding and ripple marks. In Mirzapur district, the upper part of the lower Kaimur series is made up of Bijaigarh shales. These are fine grained carbonaceous, micaceous, frequently pyriteous and contained occasional iron stones bands (Bose et al., 2001). Ground water may occur in the weathered and fractured sand stone provided the zone is connected with recharging sources. The Kaimur Group of the Vindhyan Supergroup is of special significance because it consists dominantly of siliciclastic rocks lying unconformably over the carbonate- rich Semri Group–Lower Vindhyans. Therefore, the rocks from the Kaimur Group hold strong evidence regarding changing environment of deposition, climatic conditions, tectonics and weathering conditions, during Mesoproterozoic (Mishra and Sen, 2010).

2.1 Hydrogeology

The area is covered by thin surface soil with varying thickness ranging from 1-4m and few exposures of sandstones rocks (Krishnan, 1982). The main streams that ran across the study area are the Belan Nadi and Bakhar Nadi with some minor tributaries. The Belan rises in the Bijaygarh east of Robertsganj and flows in the western direction. The bedrock (Vindhyan Super Group of rocks) is expected below the surface soil cover since the study area is lying in close vicinity of the Vindhyan exposures. The area is mainly flat and gently undulating terrain except in few parts. The occurrence and movement of groundwater is mainly restricted within the weathered and fractured sandstone/shale (Amaresh and Prakash, 2003). The amount of water that can be extracted from the fractured zones depends on the size and location of fractures.

2.2 Climate and rainfall

The climate of the district is characterized by three distinct seasons; the winter season which begin at the end October till the end of February with temperature range of about 3°C to 30°C. The winter season is followed by the hot summer which extends up to middle of June. The summer is usually hot with temperature going up to 47°C. The raining season commenced from middle of June and last up to the middle of October. The area receives rainfall from both the monsoon currents i.e., the Bay of Bengal and Arabian Sea. The rainfall is uncertain and erratic. The average annual rainfall is about 1530mm. The recharging source of the area is mainly through the rainfall during the monsoon period.

2.3 Principle of entropy

Information theory provides intuitive tools to quantify the uncertainty of random quantities, or how much information is shared by a few of them (Cover and Thomas, 1991). Shanon in 1948 proves that Boltzman’s entropy is the only function which satisfies’s the requirements for a function to measure the uncertainty in a message. He described entropy as the amount of
uncertainty in any probability distribution. Since the pioneer work of Shannon and Weaver in 1948 and the principle of maximum entropy (POME) by Jaynes in 1957, 1982, the concept of entropy has been undergoing a rapid development with promising results.

The entropy theory has been applied in many different fields, such as ecology, biology, data mining, economics and financial time-series analysis (e.g. Darbellay and Wuertz, 2000; Carranza et al., 2007; Sato, 2008; Karamanos, 2009; Sy 2001; Zhou et al., 2010). It has also been applied extensively in hydrology and water resources for measuring information contents of random variables and models, evaluating information transfer between hydrological processes, evaluating data acquisition systems, and designing water quality monitoring (WQM) networks (Mogheir et al., 2004; Karamouz et al., 2009; Mondal and Singh, 2010).

2.4 Entropy theory

To understand the informational aspect of entropy we consider a set consisting of n events. We view uncertainty as a situation where we do not know which event among n events will occur. Thus, uncertainty is about which one of those events actually occurs. Based on one’s knowledge about the events, the uncertainty can be more or less. For example, the total number of events is a piece of information and the number of those events with non-zero probability is another piece of information. The probability distribution of the events, if known, provides a certain amount of information. Shannon (1948) defined a quantitative measure of uncertainty associated with a probability distribution or the information content of the distribution in terms of entropy, called Shannon entropy or informational entropy. The uncertainty can be quantified with entropy taking into account all different kinds of available information. Thus, entropy is a measure of the amount of uncertainty represented by the probability distribution and is a measure of the amount of chaos or of the lack of information about a system. If complete information is available, entropy = 0. Otherwise, it is greater than zero (Singh, 2000).

Let \( E_i \) stand for an event and \( p_i \) for the probability of event \( E_i \) to occur. Let there be \( n \) events \( E_1, \ldots, E_n \) with probabilities \( p_1, \ldots, p_n \). Shannon (1948) defined a measure of information \( I \) as a logarithmic function expressed as

\[
I(p_i) = \log_2 \left( \frac{1}{p_i} \right)
\]  

(1)

From the \( n \) number of information values \( I \), the expected information content of a probability distribution, called entropy, can be derived by weighing the information values \( I \) by their respective probabilities \( p_i \) as

\[
H = \sum_{i=1}^{n} p_i \log_2 \left( \frac{1}{p_i} \right)
\]  

(2)

where \( H \) stands for entropy. The unit of information measures depends on the logarithmic base used. If a base 2 is used, then the unit is bit; for a logarithmic base 10 the unit is decibel, and it is nat (natural units) if the logarithmic base is e.

Thus the value of entropy, \( H \), varies from infinity to zero for probability distribution ranging from 0 to 1. The function reflects the idea that the lower the probability of an event to occur,
the higher the amount of information of a message stating that the event occurred. It is zero when the random variable is certain to be predicted.

Marginal, joint, conditional and mutual entropies are referred to as information coefficients or measures. They are the main types of entropy that are measured in information.

Given two discrete variables X and Y, with values X_i, i=1,2,……,n; Y_j, j = 1,2,……,m, defined in the same probability space, each of which has a discrete probability of occurrence p(x_i) and/or p(y_j). The marginal entropy $H(X)$ or $H(Y)$ can be defined as the measure of the total amount of uncertainty or the indirect measure of the total amount of information content of a single process, X or Y. It is given as (Mogheir and Singh, 2002)

$$H(X) = \sum_{i=1}^{n} p(x_i) \log_2 \left( \frac{1}{p(x_i)} \right)$$

$$H(Y) = \sum_{j=1}^{m} p(y_j) \log_2 \left( \frac{1}{p(y_j)} \right)$$

The conditional entropy (or equivocation) $H(x|y)$ or $H(y|x)$ quantifies the amount of information needed to describe the outcome of a random variable X or Y given that the value of another random variable Y or X is known. It is given as

$$H(X | Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log_2 \left( \frac{1}{p(x_i|y_j)} \right)$$

$$H(Y | X) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log_2 \left( \frac{1}{p(y_j|x_i)} \right)$$

The joint entropy $H(X,Y)$ is the total information content in both X and Y given as

$$H(X, Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log_2 \left( \frac{1}{p(x,y)} \right)$$

The mutual entropy, $MI(X,Y)$, is the information between X and Y. It is a quantity that measures a relationship between two random variables that are sampled simultaneously. In particular, it measures how much information is communicated, on average, in one random variable about another. Intuitively, we can ask how much one random variable tells us about another. It can be defined as the information content of X which is contained in Y, it can be interpreted as the reduction in uncertainty in X due to the knowledge of the random variable Y, and it is given as

$$MI(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log_2 \left[ \frac{p(x_i,y_j)}{p(x_i)p(y_j)} \right]$$

It can also be expressed as

$$MI(X, Y) = H(X) - H(X | Y)$$

$$MI(X, Y) = H(X) + H(Y) - H(X,Y)$$

$$MI(Y, X) = H(Y) - H(Y | X)$$
Quantitative estimation of natural recharge due to monsoon rainfall using the principle of information theory in the area of Ghorawal block of Sonebhadra district, U.P., India

\[ MI(Y, X) = H(Y) + H(X) - H(Y, X) \]  

(12)

Therefore, considering rainfall as an independent random variable (\( Y \)) and the depth to water level as the dependent variable (\( X \)), the percentage of rainfall, \( R_e (\%) \), contributing to the natural recharge of an unconfined aquifer can be obtained as (Mondal and Singh, 2010)

\[ R_e (\%) = \frac{MI(X, Y)}{H(Y)} \times 100 \]  

(13)

3. Materials and method

The data used for this work were obtained from the Jal Nigam office of Sonebhadra District. It comprises of the depth to water level measured at ten different locations in Ghorawal block of Sonebhadra District during the post-monsoon months (October, November and December) from 2007 – 2010. This is because the contribution to ground water reserve during this period is mostly as a result of the monsoon rainfall. The total monsoon rainfall data recorded from the District for the same period were used to determine the contribution of rainfall to ground water storage of the study area.

3.1 Methodology

To calculate information measures for more than one variable, the joint or conditional probability is required, and this can be obtained using a cross tabulation or contingency table. This is a type of table in a matrix format that displays the (multivariate) frequency distribution of the variables.

It is often used to record and analyze the relation between two or more categorical variables. A cross tabulation of the measured depth to water level and the monsoon rainfall data were constructed (table 1.1). The depth to water level ranging from the minimum to the maximum values with a class interval of 1.00m bgl was used as the dependent variable \( X \), while the rainfall value with class interval of 100 mm measured during the monsoon period were used as the independent variable \( Y \).

The marginal entropy of rainfall and water table are obtained from the probability distribution of the variables. The result is given in table 1.2 and 1.3 respectively. In order to obtain the mutual information, MI between the water level and rainfall the conditional or joint entropy of the two variables was first determined. The conditional entropy of water level, \( X \) given the rainfall, \( Y \) was evaluated by first obtaining conditional probability distribution, \( p(x/y) \) using

\[ p(x/y) = \frac{p(x, y)}{p(y)} \]  

(14)

The result is shown in table 1.4, while the joint entropy, \( H(X,Y) \) of the water level and rainfall was obtained using equation (7) as shown in table 1.5.

The mutual information between rainfall and the depth to water level in the study area was calculated using eq. (9) or (10). The ratio between the mutual information and the marginal entropy of rainfall was calculated using eq. (13) for estimating the recharge due to the monsoon rainfall.
Quantitative estimation of natural recharge due to monsoon rainfall using the principle of information theory in the area of Ghorawal block of Sonebhadra district, U.P., India

Table 1.1: Cross tabulation of data set

<table>
<thead>
<tr>
<th>x(j), DTW (m, bgl)</th>
<th>y(i), Monsoon rainfall (mm)</th>
<th>i = 1</th>
<th>i = 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>j = 1, 8 - 9</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>j = 2, 9 - 10</td>
<td>13</td>
<td>3</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>j = 3, 10 - 11</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>j = 4, 11 - 12</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>24</td>
<td>6</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

DTW: Depth to water level

Table 1.2: Marginal entropy of rainfall

<table>
<thead>
<tr>
<th>p_i</th>
<th>0.8</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>log_2(1/p_i)</td>
<td>0.3</td>
<td>2.3</td>
</tr>
<tr>
<td>p_i log_2 (1/p_i)</td>
<td>0.2</td>
<td>58</td>
</tr>
</tbody>
</table>

H(Y) = 0.722 bits, calculated from eq. (3)

p_i: probability distribution of rainfall

Table 1.3: Marginal entropy of water level

<table>
<thead>
<tr>
<th>p_j</th>
<th>log_2(1/p_j)</th>
<th>p_j log_2 (1/p_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.200</td>
<td>2.322</td>
<td>0.464</td>
</tr>
<tr>
<td>0.533</td>
<td>0.907</td>
<td>0.484</td>
</tr>
<tr>
<td>0.133</td>
<td>2.907</td>
<td>0.388</td>
</tr>
<tr>
<td>0.133</td>
<td>2.907</td>
<td>0.388</td>
</tr>
</tbody>
</table>

H(X) = 1.723 bits, calculated from eq. (4)

p_j: probability distribution of dept to water level

Table 1.5: Conditional entropy of rainfall and depth to water level

<table>
<thead>
<tr>
<th>P_{0j}</th>
<th>0.208</th>
<th>0.167</th>
<th>0.542</th>
<th>0.500</th>
<th>0.083</th>
<th>0.333</th>
<th>0.167</th>
<th>0.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_{0j} log_2 (1/P_{0j})</td>
<td>0.471</td>
<td>0.431</td>
<td>0.479</td>
<td>0.500</td>
<td>0.299</td>
<td>0.528</td>
<td>0.431</td>
<td>0.000</td>
</tr>
</tbody>
</table>

H(X/ Y) = 1.636 bits, calculated from eq. (5)
Quantitative estimation of natural recharge due to monsoon rainfall using the principle of information theory in the area of Ghorawal block of Sonebhadra district, U.P., India

Lazarus G. Ndatuwong, G. S. Yadav

Table 1.5: Joint entropy of rainfall and depth to water level

| \( p_{ij} \) | 0.167 | 0.033 |
| \( p_{ij} \) | 0.433 | 0.100 |
| \( p_{ij} \) | 0.067 | 0.067 |
| \( p_{ij} \) | 0.133 | 0.000 |

\[
\begin{array}{cc}
 p_{ij} \log_2 \left( \frac{1}{p_{ij}} \right) & 0.431 & 0.164 \\
 0.523 & 0.332 \\
 0.260 & 0.260 \\
 0.388 & 0.000 \\
 \end{array}
\]

\( H(X, Y) = 2.358 \) bits, calculated from eq. (7)

\( \text{MI}(X, Y) = 1.723 - 1.636 = 0.087 \) bits, calculated from equation (9).

Or \( 1.723 + 0.722 - 2.358 = 0.087 \) bits, calculated from equation (10).

The estimated recharge rate due to the monsoon rainfall in the study area is \( R_e(\%) = 12.1 \), calculated from equation (13).

4. Results and discussion

The results obtained showed marginal entropies \( H(X) \) and \( H(Y) \) of the depth to water level and rainfall as 1.723 bits and 0.722 bits respectively. These values are the respective information gained from the recorded dataset. In order word, they are the reduction in uncertainty associated with the dataset. The conditional entropy of the depth to water level, when the rainfall is known is 1.636 bits; this is the remaining uncertainty in the depth to water level when the rainfall data is known to us. The total potential information contained in both the rainfall and depth of water level also known as joint entropy is 2.358 bits. Therefore the information transferred from rainfall to water table known as mutual information was obtained as 0.087 bits. The closer the result to unity, the more uncertainty about the rainfall exists. An estimated average natural recharge ratio of about 10.11% of seasonal rainfall for fifteen granitic and gneiss area in varying climatic and hydrogeological providence in India was estimated using tritium injection by Rangarajan and Athavale, 2000. The Ground water Estimation Committee, 1997 estimated a rainfall recharge rate of about 10-12% in weathered granite, gneiss, schist with low clay content and a value of about 5-9% in the same formation having a significant clay content using an empirical method.

The result of this work shows that there exit a dependency between rainfall and depth to water level. The fractional amount of the monsoon rainfall that contributes to the ground water storage was obtained by taking the ratio of the mutual information between depth to water level and the monsoon rainfall to the marginal entropy of the monsoon rainfall. The result obtained is 12.1%. The estimated value is gross and simplified. It does not account for other secondary recharge such as from irrigation and seepage from water bodies.

Acknowledgement

The authors wish to thank the staff of Jal Nigam office Robertsganj for providing the data used in this work.
5. References


