CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Texas A&M Repository

entropy

MDPI

Article

Hot Spots and Persistence of Nitrate in Aquifers Across Scales

Dipankar Dwivedi [1,*] and Binayak P. Mohanty [2]

Received: 29 May 2015; Accepted: 5 January 2016; Published: 13 January 2016
Academic Editor: Kevin H. Knuth

1 Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
2 Biological and Agricultural Engineering Department, Texas A&M University, College Station, TX 77843, USA; bmohanty@tamu.edu
* Correspondence: DDwivedi@lbl.gov; Tel.: +1-510-486-4005

Abstract: Nitrate-N (NO_3^--N) is one of the most pervasive contaminants in groundwater. Nitrate in groundwater exhibits long-term behavior due to complex interactions at multiple scales among various geophysical factors, such as sources of nitrate-N, characteristics of the vadose zone and aquifer attributes. To minimize contamination of nitrate-N in groundwater, it is important to estimate hot spots (>10 mg/L of NO_3^--N), trends and persistence of nitrate-N in groundwater. To analyze the trends and persistence of nitrate-N in groundwater at multiple spatio-temporal scales, we developed and used an entropy-based method along with the Hurst exponent in two different hydrogeologic settings: the Trinity and Ogallala Aquifers in Texas at fine (2 km × 2 km), intermediate (10 km × 10 km) and coarse (100 km × 100 km) scales. Results show that nitrate-N exhibits long-term persistence at the intermediate and coarse scales. In the Trinity Aquifer, overall mean nitrate-N has declined with a slight increase in normalized marginal entropy (NME) over each decade from 1940 to 2008; however, the number of hot spots has increased over time. In the Ogallala Aquifer, overall mean nitrate-N has increased with slight moderation in NME since 1940; however, the number of hot spots has significantly decreased for the same period at all scales.

Keywords: entropy analysis; nitrate-N in groundwater; temporal variability; spatial variability; multi-scale analysis; Hurst exponent

1. Introduction

Globally, more than 1.5 billion people rely on groundwater as their primary source of drinking water [1]. Nitrate-N is one of the most ubiquitous contaminants in groundwater. Nitrate-N (NO_3^--N) is susceptible to reaching groundwater by leaching through soils with infiltrating water as a consequence of its high solubility and mobility [2]. Furthermore, nitrate-N can persist in groundwater for a long time (years to decades); it can further increase to high levels as more nitrogen is used at the land surface due to anthropogenic activities. Several studies have noted that high nitrate-N (>10 mg/L of NO_3^--N) concentration in drinking water is a threat to human health, particularly for infants and pregnant women [3–6]. To mitigate health risks, cleaning up nitrate-contaminated groundwater is an expensive and infeasible solution, especially for large aquifers [7]. Therefore, to minimize contamination by developing optimal management strategies, it is desirable to analyze hot spots and hot moments or the long-term behavior of nitrate in groundwater.

The long-term behavior of nitrate-N in groundwater shows variability in space and time. The spatio-temporal variability of nitrate-N in groundwater is influenced by the interaction among multiple geophysical factors, such as source availability, precipitation pattern, thickness and composition of the vadose zone [8], types of aquifers (confined or unconfined) and aquifer heterogeneity [9]. Varying sources of nitrate-N in space and time that render spatio-temporal

variability include fertilizers applied to crops, irrigation source water, animal manure, air emissions, atmospheric deposition, municipal effluent, food processing and wastewater treatment plants, recharge from manure storage lagoons, leachate from septic system drain fields and percolation from urban parks, lawns and golf courses [9–15]. Several studies have demonstrated that precipitation patterns impact the extent of agricultural nitrate-N leaching [16–20]. The thickness, composition of the vadose zone and aquifer heterogeneity [9] that vary across space in aquifers are significant factors that can affect nitrate-N delivery to groundwater because of active biogeochemical processes in this zone [21,22]. Furthermore, types of aquifers (confined or unconfined) affect the level of contamination, as unconfined shallow aquifers are more prone to nitrate-N contamination [23]. Additionally, groundwater pumping impacts the long-term behavior of nitrate-N in groundwater by local flow in the area of influence near wells. These factors also exhibit variability across seasons to years to decades and across catchments to watersheds to the aquifer scale. Evaluating the consequences of the interactions among different geophysical factors and processes on nitrate contamination in groundwater is challenging. However, it is possible to develop best management practices based on the spatio-temporal characterization of nitrate in groundwater. For example, strategic monitoring and sampling locations can be identified based on hot spots and hot moments of nitrate in groundwater across scales. Additionally, adaptive land use decisions can be made to minimize nitrate leaching from the land surface and surface water bodies, such as, streams, rivers and lakes. Therefore, it is essential to explore the spatio-temporal variability and persistence of nitrate in groundwater to develop optimal groundwater monitoring, management and remediation strategies.

There are several approaches to quantify the trends and persistence of a system variable (e.g., nitrate-N) across spatio-temporal scales. For instance, Assaf and Saadeh [24] analyzed groundwater nitrate-N levels using a geostatistical approach and demonstrated maps of the persistence of nitrate-N contamination in groundwater in Upper Litani Basin in Lebanon [24]. Similarly, a principal component analysis (PCA) and K-means clustering technique was used to analyze the temporal evolution of groundwater composition [25]. Likewise, several others characterized the spatial variability of groundwater quality using the entropy theory by measuring transinformation and the information transfer index among spatial datasets [26–28]. However, geostatistical methods are limited by the assumption of stationarity, which may not hold well across spatio-temporal scales. Moreover, the interpolated value may be different from the measurement itself [29]. Similarly, PCA requires certain assumptions, such as linearity and large variances being the only important structure in the dataset [30]. On the other hand, entropy is a non-parametric approach and a robust measure of variability. Moreover, entropy does not change drastically by small changes in data [31–33]. Additionally, the Hurst exponent is a robust statistical measure of the long-term behavior or persistence of a phenomenon [34]. Therefore, we have applied an entropy-based approach to quantify the spatio-temporal variability of nitrate-N in groundwater. To ascertain the persistence of nitrate-N in groundwater, we have used the Hurst exponent.

As a result, we have developed an entropy-based approach jointly with the Hurst exponent to explore the trends and persistence of nitrate-N in groundwater to inform and develop optimal groundwater sampling and remediation strategies, in two different hydrogeologic settings: the Ogallala and Trinity Aquifers in Texas. The specific objectives of this study are to (1) analyze the persistence of nitrate-N contamination and examine decadal variability in nitrate-N at three spatial scales (fine, intermediate and coarse) in the Ogallala and Trinity Aquifers and (2) develop and present a new metric for identifying hot spots of nitrate (NO_3^--N >10 mg/L).

2. Methodology

To analyze the trends and persistence of nitrate-N in groundwater, we utilized the entropy theory and the Hurst exponent. We further developed an index that quantifies the distribution of hot spots (>10 mg/L of NO_3^--N) of nitrate-N in groundwater at different scales. Entropy is a probabilistic approach for measuring the variability, randomness, uncertainty or the information contained in a

random variable. The term entropy or Shannon entropy usually quantifies bits of the information contained in a dataset [35]. Similarly, the Hurst exponent measures the irregularity, dependence or persistence of occurrence of a variable in time. We use the entropy concept and the Hurst exponent jointly for analyzing the spatio-temporal variability of nitrate-N in the Ogallala and Trinity Aquifers in Texas. A brief summary of the entropy concept and the Hurst exponent is provided below.

2.1. Entropy

For a variable with a probability density function (PDF), $f_X(x)$ the information entropy is [35]:

$$E = -\int_{-\infty}^{+\infty} f_X(x) log(f_X(x)) dx \tag{1}$$

where X is a random variable. A discrete form of entropy $E(X)$ is given as:

$$E = -\sum_{n=1}^{n=N} p(x_n) log(p(x_n)) \tag{2}$$

where n is a discrete data interval, x_n is an outcome corresponding to interval n, $p(x_n)$ is the probability of x_n and N is the number of data points. To measure entropy, the first step is to create a histogram. The entire range of data is divided into a series of small intervals, and how many values fall into each interval are counted. These intervals are also known as "bins". There are several ways the number of bins can be chosen [36–38]; we used Scott's choice, as it can take into account the integrated squared error of the density estimate. Scott's choice of the number of bins is given as [36]:

$$\begin{aligned} b &= \lceil \frac{max(X) - min(X)}{h} \rceil \\ h &= \frac{3.5\sigma}{N^{\frac{1}{3}}} \end{aligned} \tag{3}$$

where b is the number of bins, and braces indicate the ceiling function; σ is the sample standard deviation. It is also worth mentioning that the number of bins for the time series from 1940 to 1950, 1940 to 1960, and so on, had nearly the same number of bins across all datasets.

The function $log(.)$ in Equation (2) can be used with the user's choice of base, such as 2, e or 10. We used 2 as the base of the log function in this paper.

E measures the relative variability or randomness of a variable (e.g., nitrate-N) with respect to complete randomness, *i.e.*, the randomness associated with a uniform distribution. The uniform distribution is a non-informative distribution. The value of the entropy reflects how much information is contained in random variables of interest with respect to the information contained in the non-informative probability distribution of the system. Therefore, E is maximum if all states are equiprobable (uniform distribution). To a certain event, there is only one outcome. Hence, the probability is 1, and the entropy value is 0. There is no upper bound for the entropy value; because $log(.)$ can tend to infinity if the probability of any or some values in the dataset tends to zero. Therefore, for removing the effect of different numbers of data points and comparing entropy values across datasets, we defined a normalized measure of entropy:

$$\begin{aligned} E_N &= \frac{max(E) - E}{maxE}; 0 \geq E_N \leq 100 \\ E_N &= \frac{log(k)) - E}{log(b)}; \end{aligned} \tag{4}$$

where E_N is the normalized marginal entropy (NME) of X; b is the number of bins. Maximum entropy is the log of the number of bins (b). The higher the NME, the lower is the entropy value and, so, the lesser is the variability.

2.2. Normalized Risk Index

We define a new index, the "normalized risk index" (NRI), for quantifying the distribution of hot spots (nitrate-N >10 mg/L). NRI can be extended to any variable depending on the criterion of hot spots for that variable. For nitrate-N, if there are k spatial locations (grids) in an aquifer, there will be k time series datasets. Let N_0 be the total number of cases having nitrate-N >10 mg/L at a given spatial scale and n_{0i} be the total number of cases for a particular dataset or time series (wells).

$$N_0 = \sum_{i=1}^{i=k} n_{0i} \tag{5}$$

where k is the total number of grids or time series datasets. Risk entropy (RE) is thus given as:

$$RE = -\sum_{i=1}^{i=k} \frac{n_{0i}}{N_0} log(\frac{n_{0i}}{N_0}) \tag{6}$$

where RE measures the density of hot spots in an aquifer. If there is only one time series dataset that has all cases (samples having nitrate-N $>$ 10 mg/L), RE will be zero, and if all of the k time series datasets have an equiprobable number of cases, then RE will be maximum, *i.e.*, $log(k)$. For eliminating the effect of the number of data points across datasets and comparing these entropy values, we defined a normalized measure of risk entropy: the normalized risk index (NRI):

$$\begin{aligned} NRI &= \frac{RE}{max(RE)} \times 100; or \\ NRI &= \frac{RE}{log(k)} \times 100 \end{aligned} \tag{7}$$

The range of NRI is between 0 and 100. The higher the NRI, the higher is the number of hot spots in the aquifer.

2.3. Hurst Exponent

Harold E. Hurst was a British hydrologist, who first used the range analysis of time series data in hydrology. He analyzed the time series for trends and persistence by dividing the time series into shorter sub-time series and rescaling ranges of each sub-time series [34]. The Hurst exponent has been an indicator of the irregularity, dependence or persistence, such as to analyze fractal properties of river networks [39], financial markets for price fluctuations [40] and digital signal processing [41]. The range of the Hurst exponent (H) is from 0 to 1; H is equal to 0.5 for random processes, e.g., Brownian motion. A Hurst exponent can be used to analyze the trends of nitrate-N in groundwater. A Hurst exponent value that ranges from 0 to 0.5 ($0 < H < 0.5$) indicates the "anti-persistent behavior". Anti-persistent behavior implies that an increase in nitrate-N concentration (in a well for a particular year) will follow a decrease in nitrate-N concentration, in the same well, in future times. This phenomenon is also known as "mean-reversion". The intensity of the mean-reversion increases as H tends to 0. A Hurst exponent H between 0.5 and 1 indicates the "persistent behavior", which means that the time series has a strong trend (either increasing or decreasing). The larger is H, the more persistent is the trend. It is easier to predict time series ($H \geq 0.5$) that show persistent trends as compared to time series ($H \leq 0.5$) that fall in the other category (mean-reversion) [34].

There are various methods (such as rescaled range, regression on periodogram, Whittle estimator, aggregated variances) to estimate the Hurst coefficient in the literature [42,43]. We chose the rescaled range method to calculate the Hurst exponent, because rescaled range statistics is relatively robust with data having a long-tailed probability density function [44]. In our case, nitrate-N data follow the Weibull distribution that is positively skewed. A detailed description of how the Hurst

exponent is calculated is provided elsewhere [45]. A time series of length N_h is split into shorter time series of smaller lengths, such as $n_h = N_h, N_h/2, N_h/4 \cdots$. For each value of n_h, the average range is rescaled. To calculate the average rescaled range for each partial time series of length n_h, $X_h = X_1, X_2, \cdots, X_n$, the following steps are outlined:

1. Compute the mean $m_h = \frac{1}{n_h} \sum_{i=1}^{i=N_h} X_i$.
2. Detrend the series by subtracting mean $Y_t = X_t - m$; for $t = 1, 2, \cdots, n_h$.
3. Calculate the summation of all detrended series $Z_t = \sum_{i=1}^{t} Y_i$; for $t = 1, 2, \cdots, n_h$.
4. Rescale the range, by dividing the range by the standard deviation.
5. Calculate the mean of the rescaled range for all sub-series of length n_h; $(\frac{R}{S})_{n_h} = \frac{1}{n_h} \sum_{i=1}^{n_h} (\frac{R_i}{S_i})$.
6. Finally, the value of the Hurst exponent is obtained using an ordinary least square regression with $log(n_h)$ as the independent variable and $n_h log(\frac{R}{S})$ as the dependent variable. The gradient of the fit is the estimate of the Hurst exponent.

3. Study Site

We conducted this study in two different hydro-geologic settings: (1) the Ogallala Aquifer, which is an unconfined aquifer and a principal source of water for agricultural, municipal and industrial development; and (2) the Trinity Aquifer, which is partially confined and a principal source of water in four densely inhabited urban centers: San Antonio, Austin, Fort Worth and Dallas metropolitan areas in Texas. Previous studies (such as [2,46]), have suggested that the Ogallala and Trinity Aquifers have a history of nitrate-N contamination. A short description of study sites is provided below.

3.1. Ogallala Aquifer

The Ogallala Aquifer is the largest groundwater system in North America. The Ogallala Aquifer spreads across eight states: South Dakota, Nebraska, Wyoming, Colorado, Kansas, Oklahoma, New Mexico and Texas [47] In this study, we critically examine the Ogallala Aquifer in the Texas regions that extends over 91,815 km^2 (20% of the aquifer) and provides water to all or part of 46 counties. The Ogallala is an unconfined aquifer that is formed mostly by sand, gravel, clay and silt deposited during the Tertiary Period. The Canadian River, Prairie Dog Town Fork Red River and Colorado River are major rivers that flow across the Ogallala Aquifer in Texas regions (Figure 1). The Ogallala Aquifer is an important water supply in the High Plains of Texas, particularly for agriculture, as 90% of the water pumped is used to irrigate crops. Previous studies suggest that the Ogallala Aquifer has a median nitrate-N concentration exceeding 10.0 mg/L [2].

3.2. The Trinity Aquifer

The Trinity Group Aquifer is a prime water-bearing entity in north-central, central and southwest-central Texas, extends over more than 106,190 km^2 and provides water to all or part of 52 counties. The Trinity Aquifer is a partially-confined aquifer. The significance of the Trinity Aquifer is immense, as the outcrop and sub-crop areas of the Trinity Aquifer supply drinking water for almost seven million people in this region apart from irrigation and industrial purposes. The Red, Trinity, Brazos, Colorado, San Antonio, Guadalupe and Medina are major rivers that flow across the Trinity Aquifer (Figure 1). The Trinity Aquifer has a history of nitrate-N contamination [46]

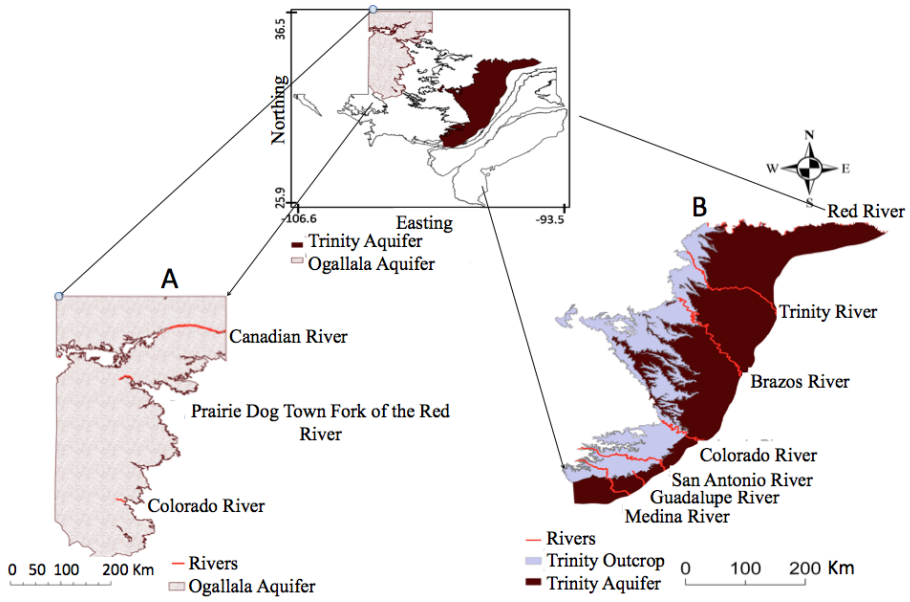

Figure 1. Map showing the Ogallala Aquifer (**A**) and Trinity Aquifer and its outcrop area (**B**). The Ogallala Aquifer is an unconfined aquifer, primarily composed of sand, gravel, clay and silt, whereas the Trinity Aquifer is a sandstone-carbonate rock aquifer and partly confined.

4. Data Analysis

The nitrate-N data were obtained from the Texas Water Development Board [48]. Data used in this study were from 1940 to 2008. There is vast information of nitrate-N in groundwater, but this information is limited, as it lacks the continuity of sampling in individual wells. In other words, nitrate-N concentrations in a particular well, for example, were available only at five instances from 1940 to 2008, which limits the use of these data in their totality. However, for other times, nitrate-N concentrations were available in neighboring wells of the well under consideration. Therefore, we reconstructed nitrate-N data by binning them in three scales (fine, intermediate and coarse). Each scale had grids of different resolutions as fine (2 km $\times$ 2 km), intermediate (10 km $\times$ 10 km) and coarse (100 km $\times$ 100 km) grids. The choice of scales was based on flow systems—local, intermediate and coarse—that exist in aquifers depending on the spatial locations of recharge and discharge zones. A local flow system of groundwater, where recharge and discharge areas are adjacent to each other, leads to the higher mixing of nitrate-N. On the contrary, an intermediate flow system of groundwater, where recharge and discharge areas are disconnected by topographic highs and lows, leads to intermediate mixing of nitrate-N in groundwater. Finally, a regional flow system of groundwater, where recharge and discharge areas are disconnected by groundwater divides, leads to the lower mixing of nitrate-N in groundwater [49,50]. Therefore, various flow systems that exist in aquifers also impact the long-term behavior of nitrate-N in groundwater. In the Trinity and Ogallala Aquifers, it is assumed that the local, intermediate and coarse flow system spans an area of 2 km $\times$ 2 km, 10 km $\times$ 10 km and 100 km $\times$ 100 km, which is in accordance with the previous studies [49,50]. Subsequently, we combined all wells falling in a particular grid as one time series dataset and analyzed the intra-decadal variability, trend and persistence of nitrate-N annually.

5. Results and Discussion

The Trinity and Ogallala Aquifers are two contrasting hydrogeologic settings, as shown in (Figure 1). Nitrate-N occurs in varying concentrations in the Trinity and Ogallala Aquifers. There are 25% and 16% of total wells exceeding 10 mg/L of nitrate-N in the Ogallala and Trinity Aquifers, respectively (Figure 2). Higher concentrations or hotspots (nitrate-N $>$ 10 mg/L) were found on

the outcrop and the southern part (downdip) of the Trinity Aquifer. In the Ogallala Aquifer, hotspots (nitrate-N $>$ 10 mg/L) were found almost everywhere, except the region between the Canadian River and the Prairie Dog Town Fork Red River (Figure 2). A detailed analysis of the trends and persistence of nitrate-N for the two aquifers is described in the following sections.

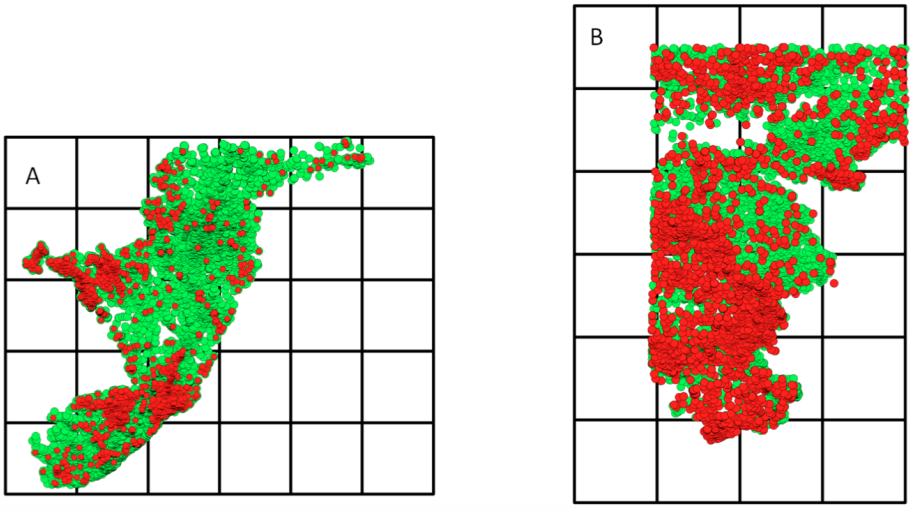

Figure 2. The nitrate-N concentrations (1940 to 2008) are shown in the (**A**) Trinity and (**B**) Ogallala Aquifers. In both aquifers, three different spatial scales, namely fine (2 km $\times$ 2 km), intermediate (10 km $\times$ 10 km) and coarse (100 km $\times$ 100 km), are used to analyze the trends and persistence of nitrate-N. Red and green dots show wells where nitrate-N was measured in both aquifers. Additionally, red dots show wells that have nitrate-N $>$ 10 mg/L (hot spots) in these aquifers.

5.1. Trend and Persistence of Nitrate-N across Scales

Figure 3 shows the map of Hurst exponents (H) in the Ogallala and Trinity Aquifers at three spatial scales: fine, intermediate and coarse. Although there are small-scale variations of H, trends are more persistent at the intermediate and coarse scales in both aquifers. There are regions where more variation (anti-persistence) is observed in both the aquifers at the coarse scale. It is evident from Figure 3 that H of nitrate-N is scattered over the small spatial scale in the Ogallala and Trinity Aquifers. At fine and intermediate scales, both aquifers show variability across different river basins. Especially in the Trinity Aquifer, the H transitions around the Red, Brazos, Colorado and Guadalupe Rivers (see Figure 1), which indicates the significance of rivers at fine and intermediate scales for nitrate-N contamination in groundwater.

Figure 4 shows the probability distribution (PDF) of H at the fine, intermediate and coarse scales in the Trinity and Ogallala Aquifers. PDFs illustrate how H is spatially distributed in two aquifers across scales. In the Trinity Aquifer, PDFs show that H ranges from 0.4 to 0.8 and is distributed around (peaks of the PDF) 0.5 and 0.7 at the fine scale. These values indicate that nitrate-N in groundwater shows random to persistent behavior at the fine scale. Moreover, the bimodal behavior of nitrate-N, in the Trinity Aquifer at the fine scale, also indicates different physical controls (more than one) at that scale. Likewise, H ranges from 0.45 to 0.85 and is distributed around 0.65 at the intermediate scale, which means nitrate-N values in groundwater show more persistence at the intermediate scale as compared to the fine-scale distribution of nitrate-N. At the coarse scale in the Trinity Aquifer, H ranges from 0.5 to 0.8 and is distributed around multiple modes (e.g., 0.5, 0.6). Multimodal distribution of H signifies multiple processes (e.g., river-groundwater interactions, regional flow system, *etc.*) controlling nitrate-N in groundwater at different times from 1940 to 2008 at the coarse scale.

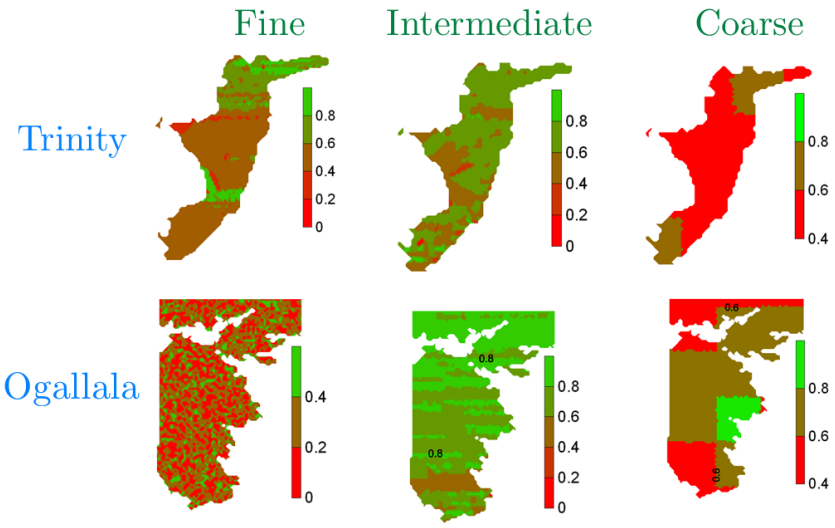

Figure 3. The Hurst exponent (H) of nitrate-N in both aquifers at small, intermediate and coarse scales. The Hurst exponent varies from 0 to 1. H close to 0, 0.5 and 1 shows anti-persistence, random behavior and persistence, respectively.

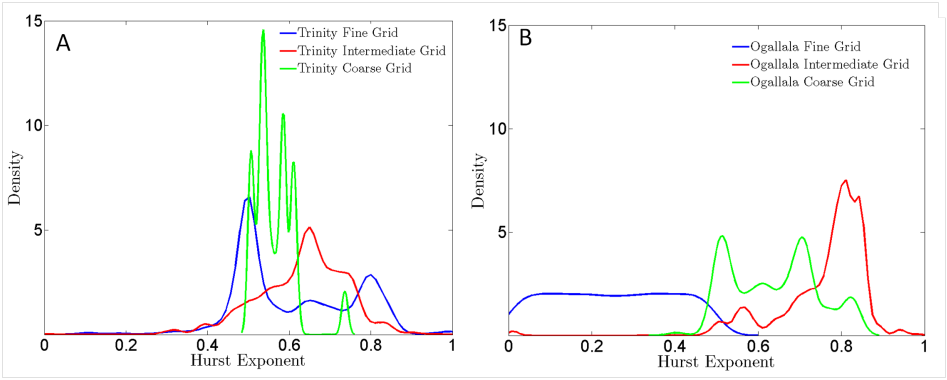

Figure 4. Probability distribution functions (PDFs) of the Hurst exponent of nitrate-N in the (**A**) Trinity and (**B**) Ogallala Aquifers across different spatial scales (fine, intermediate and coarse).

In the Ogallala Aquifer, PDFs show that H ranges from 0 to 0.5 and is distributed around (peak of the PDF) 0.3 at the fine scale, which indicates that nitrate-N values in groundwater show a mean-reversion phenomenon to random behavior at the fine scale. On the other hand, H ranges from 0.6 to 1 and is distributed around multiple modes (e.g., 0.5, 0.6, 0.7, 0.8) at the intermediate scale, which means nitrate-N values in groundwater show slight to strong persistence at the intermediate scale in contrast to the fine scale distribution of nitrate-N for the Ogallala Aquifer. At the coarse scale in the Ogallala Aquifer, H ranges from 0.6 to 0.9 and is distributed around 0.8, which indicates that nitrate-N values in groundwater have strong persistence at the coarse scale.

Further analysis of H values shows that more than 50% of grids at the fine scale, 70% at the intermediate scale and 10% at the coarse scale show persistence in the Trinity Aquifer. On the other hand, more than 80% of grids at the small scale, 90% at the intermediate scale and 70% at the coarse scale show persistence in the Ogallala Aquifer. The constantly changing behavior of nitrate-N at the small scale in the Ogallala Aquifer, as shown in Figure 3 (in contrast to the Trinity Aquifer), may be explained by the presence of prevalent local flow systems. A local flow system is typically present in sandy and unconfined aquifers, and such aquifers respond quickly to increased groundwater

recharge. At the coarse scale in both aquifers, PDFs show multimodality. In the Trinity Aquifer, these modes are located around 0.5 and 0.6, which means rapidly changing behavior of the variability of nitrate-N. However, in the Ogallala Aquifer, these modes are located around 0.5, 0.6 and 0.7, which means there are regions signifying the persistence of nitrate-N. The Ogallala Aquifer is intensive agricultural land, so the use of fertilizer may be a reason for the persistence.

5.2. Temporal Variability of Nitrate-N

For understanding the temporal variability of nitrate-N over different decades, an inter-decadal variation of nitrate-N was analyzed (Figures 5 and 6). Decadal mean (μ), standard deviation (SD), percent of samples having nitrate-N $>$ 10 mg/L and normalized marginal entropy (NME) were plotted for different scales (fine, intermediate and coarse). In the Trinity Aquifer, overall mean nitrate-N (at the fine scale from 12 mg/L to 6 mg/L; at the intermediate scale from 12 mg/L to 5.8 mg/L; at the coarse scale from 7 mg/L to 5 mg/L) has declined with a slight increase in the normalized marginal entropy (NME) (at the fine scale from 93 to 98; at the intermediate scale from 88 to 90; at the coarse scale from 55 to 89) over the decades between 1940 and 2008. However, percent samples having nitrate-N >10 mg/L have increased across times (at the fine scale from 79% to 89%; at the intermediate scale from 80% to 87%; at the coarse scale from 77% to 90%). In other words, overall decadal mean nitrate-N has declined at all scales. The standard deviation has also decreased with time, implying lesser variability (Figure 5).We also note that standard deviations are relatively high in the decades of 1940 to 1950 and 1960 to 1970. The high standard deviations may be because the 1940 to 1950 decade experienced enhanced use of fertilizer that resulted in a larger spread (or standard deviation) of nitrate values. The larger spread can be attributed to nitrate-N increasing in groundwater starting from a small or negligible concentration. In contrast, the 1960 to 1970 decade experienced the lesser use of fertilizer and lesser return flow from irrigation, which resulted in a decrease in nitrate values from a high concentration of nitrate-N in groundwater. Therefore, nitrate-N values show a large spread. Furthermore, NME also suggests a slight decline in the temporal variability of nitrate-N at the coarse scale (Figure 6). However, percent samples having nitrate-N >10 mg/L are almost constant over the last seven decades (1940 to 2008). Based on these findings, it seems that nitrate-N levels in groundwater have stabilized in the Trinity Aquifer. It can be inferred from these results that overall consumption of fertilizers (or nitrate-N emanating from other sources, as well) has gone down, but there are hot spots in the Trinity Aquifer.

In the Ogallala Aquifer, overall decadal mean nitrate-N has increased at all scales from 1940 to 1970, which may be because of a higher use of fertilizers to improve agricultural productivity since the 1940s. After 1970, overall mean nitrate-N has decreased at both fine and coarse scales (Figure 5). The standard deviation has increased in the Ogallala Aquifer at all scales (Figure 5). However, percent samples having nitrate-N $>$ 10 mg/L decreased significantly over the last seven decades. This decrease is suggested to be because of more efficient water management methods employed in the irrigation practices of the region, particularly after the early 1980s. Hence, the irrigation return flow, which contributes significant amounts of recharge to the Ogallala Aquifer, has declined through time. In the Ogallala Aquifer, at all scales, overall mean nitrate-N has increased because of enhanced use of fertilizer since 1940 (at the fine scale from 9 mg/L to 22 mg/L; at the intermediate scale from 10 mg/L to 22 mg/L; at the coarse scale from 10 mg/L to 22 mg/L). However, the percent of samples having nitrate-N $>$ 10 mg/L have significantly decreased (at the fine scale from 71% to 32%; at the intermediate scale from 71% to 38%; at the coarse scale from 66% to 27%) over the last seven decades due to the use of more efficient methods of irrigation. This is also reflected in the normalized marginal entropy that has gone down in the Ogallala Aquifer at all scales (at the fine scale 98 to 89; at the intermediate scale 98 to 89; at the coarse scale 98 to 39).

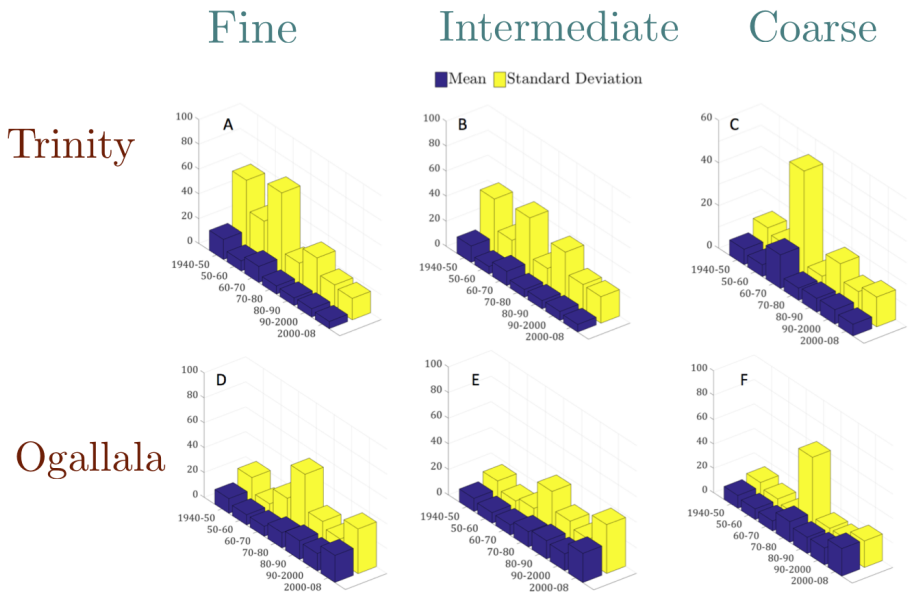

Figure 5. Decadal analysis of the mean and standard deviation (SD) for nitrate-N in the Trinity Aquifer across (**A**) fine, (**B**) intermediate and (**C**) coarse scales and in the Ogallala Aquifer across (**D**) fine, (**E**) intermediate and (**F**) coarse scales.

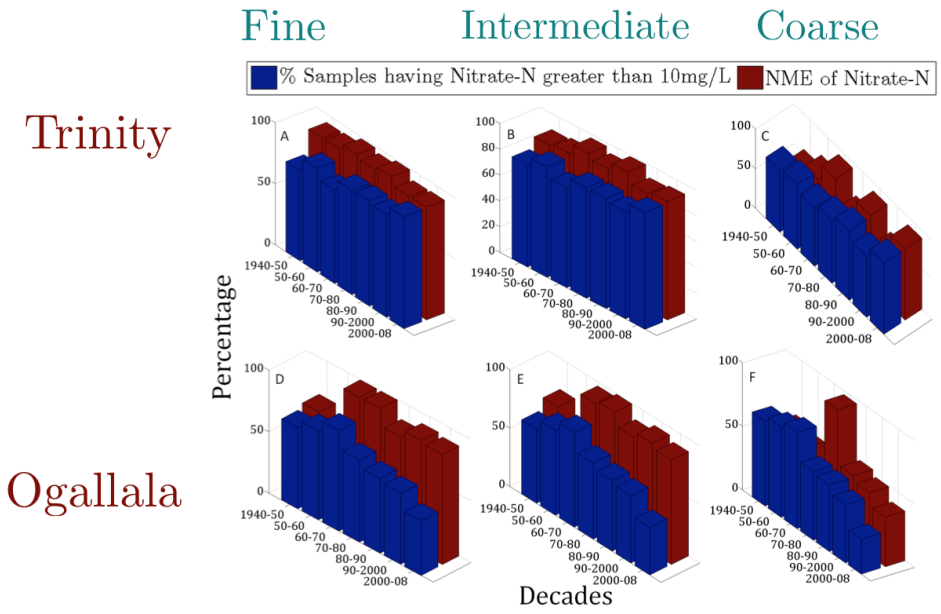

Figure 6. Decadal analysis of normalized marginal entropy (*NME*) and hot spots for nitrate-N in the Trinity Aquifer across (**A**) fine, (**B**) intermediate and (**C**) coarse scales and in the Ogallala Aquifer across (**D**) fine, (**E**) intermediate and (F) coarse scales.

5.3. Hot Spots of Nitrate-N

The normalized risk index (*NRI*) measures the distribution of hot spots (nitrate-N > 10 mg/L) in aquifers. The higher the *NRI*, the higher are the number of hot spots in the aquifer at a particular spatial scale. In Table 1, the *NRI* values are given for both the aquifers at different

scales. The results indicate that the number of hot spots of nitrate-N contamination in groundwater decreases with an increase in the spatial scale (from fine to coarse grids). This finding suggests that there are concentrated hot spots in both the aquifers where nitrate-N contamination is localized. Moreover, comparing NRI values with standard deviations across scales (Figure 5) also corroborates these results. NRI values decreased across scales. Similarly, standard deviations also decreased across scales.

Table 1. Normalized risk index (NRI) in the Ogallala and Trinity Aquifers across scales.

Aquifers	Grid	Normalized Risk Index (%)
Ogallala	Fine	67.8
	Intermediate	64.2
	Coarse	39.1
Trinity	Fine	75.3
	Intermediate	61.3
	Coarse	40.9

Results show that a larger number of hot spots of nitrate-N contamination exist at the fine scale as compared to intermediate or coarse scales. Thus, at the fine scale (particularly around the transitional zones), automated sampling that helps to provide data at a smaller interval should be used to monitor nitrate-N contamination in groundwater. Furthermore, trends of nitrate-N in groundwater are more persistent at intermediate and coarse scales. This result suggests that a few sampling sites at a relatively larger interval will be needed at these scales to monitor groundwater quality. In other words, these results can, therefore, be used to design metrics for optimal groundwater monitoring, management and remediation strategies for nitrate-N. An example of such an application is provided below.

6. An Application: Designing Monitoring Strategies

In this section, an example of a practical field application of the proposed entropy approach is presented. Although this example is focused on designing sampling strategies for nitrate contamination in groundwater, the following approach is generic enough in nature for designing the optimal sampling strategy to track groundwater quality. The first step in any monitoring well network design requires collecting preliminary data on potential sites or hot spots of nitrate in aquifers. Various approaches have been suggested in the literature for identifying strategic locations for monitoring groundwater. For example, Al-Zabet [51] suggested a monitoring strategy based on aquifer vulnerability to contamination potential [51], whereas Scot [52] developed a random site-selection approach [52], and others have used an overlay model using multiple geographic information system (GIS) data layers [53] for monitoring. These approaches are data intensive (e.g., land use land cover data, water well density, aquifer vulnerability) and usually general for all contaminants. However, different contaminants show varying levels of mobility and reactivity in groundwater [47]; a general approach cannot describe hot spots of each contaminant. In contrast, the entropy approach presented in this paper is specific for each contaminant and does not require multiple GIS data layers. In this study, the entropy approach identified hot spots of nitrate contamination in groundwater across various scales (Figure 3, Table 1). These hot spots can be used as strategic sampling locations.

The second step in designing the optimal sampling strategy is to determine sampling frequency for various contaminants. Because data collection, processing and analysis can be expensive, it is desirable that the sampling frequency be cost effective. There are various criteria, such as land use change and groundwater fluctuations, that are used to determine sampling frequency for monitoring. However, it is essential to find redundancies in observations and to outline the sampling frequency

without losing any significant information. To test the redundancy in the sampling frequency, we computed NME in Ogallala and Trinity Aquifers for the 2000-onwards decade at the fine scale by removing nitrate data points. We chose the 2000-onwards decade at the fine scale because both time series (nitrate time series in Trinity and Ogallala Aquifers for 2000-onwards) had more than 500 nitrate values. We removed part of the data (10%, 20%, 50%, 75%) from the nitrate time series and computed NME. To make sure that there is no systematic bias in the analysis, part of the data was removed randomly from the time series using "randperm" function in MATLAB. Subsequently, for comparison, we also removed one alternate and two alternate data points from both time series and computed NME. As shown in Table 2, for the Ogallala Aquifer, when sample points were removed by 50% (removing one alternate sample), NME did not change. However, when sample points were removed by 66% (removing two alternate samples), NME changed significantly. In addition, when sample points were removed by 10 to 50%, NME values did not change much from the base case. However, NME values changed significantly, when samples were removed randomly by 75%. Therefore, NME values suggest that sampling frequency can be reduced by 50% in future sampling without losing much information in the Ogallala Aquifer. On the contrary, NME values showed little change when samples were removed randomly by 10 to 75% or one or two alternate samples in the Trinity Aquifer. These results are also substantiated by Figure 3. It is evident from Figure 3 that Trinity Aquifer shows higher persistence as compared to the Ogallala Aquifer at the fine scale. Therefore, removing sample points (>50%) from the time series of the Ogallala Aquifer changed NME significantly, but removing sample points (up to 75%) from the time series of the Trinity Aquifer did not change NME values significantly. These results suggest that persistence and NME together can be an effective approach to design the optimal sampling strategy. We also want to emphasize that each monitoring program has different objectives; therefore, the reader can creatively apply this entropy-based method in designing their own sampling strategies.

Table 2. Change in normalized marginal entropy (NME) in Ogallala and Trinity Aquifers for the decade 2000-onwards.

Aquifers	Sampling Strategy	NME
Ogallala	Base case	89.3
	Removing one alternate sample	89.3
	Removing two alternate samples	63.8
	Removing 10% (randomly)	88.7
	Removing 20% (randomly)	88.7
	Removing 50% (randomly)	85.2
	Removing 75% (randomly)	60.0
Trinity	Base case	92.4
	Removing one alternate sample	92.2
	Removing two alternate samples	91.9
	Removing 10% (randomly)	91.7
	Removing 20% (randomly)	91.7
	Removing 50% (randomly)	91.9
	Removing 75% (randomly)	90.4

7. Summary and Conclusions

Nitrate contamination in groundwater shows multi-scalar variability in space and time. However, a systematic approach to characterize the spatio-temporal variability of nitrate in groundwater has been lacking. This study uses entropy theory and the Hurst exponent to identify the trends and persistence of nitrate-N at different spatial scales (fine, intermediate and coarse) in the Trinity and Ogallala Aquifers of Texas.

Results suggest that nitrate-N in groundwater shows a scale phenomenon in both space and time. The trends of nitrate-N variability show long-term persistence at the intermediate and coarse scales. At the fine scale, there is a fluctuation in the trends of nitrate-N, especially in the transitional areas, where the interaction between rivers and aquifer is prominent, or in the zones that are characterized by the presence of local flow systems. Furthermore, agricultural lands are more prone to nitrate-N contamination than urban areas due to the application of fertilizers. Furthermore, outcrop or unconfined aquifers become more susceptible to contamination of nitrate-N if inorganic sources of nitrate-N (e.g., fertilizers) are present.

This study also highlights how entropy techniques and the Hurst exponent can be used to design decision-making tools for water quality monitoring and management. An improved monitoring of nitrate-N contamination of groundwater can be achieved by having densely-located sampling sites and collecting samples at smaller time intervals in transitional areas, such as the river-aquifer interface. In contrast, at intermediate and coarse scales, sparsely-located sampling sites that collect nitrate-N samples at larger time intervals are adequate. Furthermore, we presented an example application for designing monitoring strategies for nitrate in groundwater. The example demonstrated how the entropy-based approach along with the Hurst exponent can be used to identify strategic sampling locations and outline cost-effective sampling frequency by capturing the characteristic distribution of nitrate in groundwater, without losing any significant information. Although we applied the entropy technique and the Hurst exponent to understand the multi-scalar behavior of nitrate-N in groundwater, this approach should readily be transferable to other contaminated aquifers and catchments.

Acknowledgments: This research was supported by the EPA 319(h) grant for Total Maximum Daily Load (TMDL) in Texas streams and partly supported by the National Institute of Environmental Health Sciences (Grant 5R01ES015634), the Texas Water Resources Institute and Texas A&M support account number 02130003. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

Author Contributions: Dipankar Dwivedi did this work as part of his PhD dissertation, and Binayak P. Mohanty provided guidance. Both authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Alley, W.M.; Healy, R.W.; LaBaugh, J.W.; Reilly, T.E. Flow and storage in groundwater systems. *Science* **2002**, *296*, 1985–1990.
2. Hudak, P. Regional trends in nitrate content of Texas groundwater. *J. Hydrol.* **2000**, *228*, 37–47.
3. Brender, J.; Olive, J.; Felkner, M.; Suarez, L.; Hendricks, K.; Marckwardt, W. Intake of nitrates and nitrites and birth defects in offspring. *Epidemiology* **2004**, *15*, doi:10.1289/ehp.1206249.
4. Brender, J.D.; Weyer, P.J.; Romitti, P.A.; Mohanty, B.P.; Shinde, M.U.; Vuong, A.M.; Sharkey, J.R.; Dwivedi, D.; Horel, S.A.; Kantamneni, J.; *et al.* Prenatal nitrate intake from drinking water and selected birth defects in offspring of participants in the National Birth Defects Prevention Study. *Environ. Health Perspect.* **2013**, *121*, 1083–1089.
5. Lundberg, J.O.; Govoni, M. Inorganic nitrate is a possible source for systemic generation of nitric oxide. *Free Radical Biol. Med.* **2004**, *37*, 395–400.
6. Showers, W.J.; Genna, B.; McDade, T.; Bolich, R.; Fountain, J.C. Nitrate contamination in groundwater on an urbanized dairy farm. *Environ. Sci. Technol.* **2008**, *42*, 4683–4688.
7. McMahon, P.B.; Böhlke, J.; Kauffman, L.J.; Kipp, K.L.; Landon, M.; Crandall, C.A.; Burow, K.R.; Brown, C.J. Source and transport controls on the movement of nitrate to public supply wells in selected principal aquifers of the United States. *Water Resour. Res.* **2008**, *44*, doi:10.1029/2007WR006252.
8. Williams, A.; Johnson, J.; Lund, L.; Kabala, Z. Spatial and temporal variations in nitrate contamination of a rural aquifer, California. *J. Environ. Qual.* **1998**, *27*, 1147–1157.
9. Spalding, R.F.; Exner, M.E. Occurrence of nitrate in groundwater—A review. *J. Environ. Qual.* **1993**, *22*, 392–402.

10. Harter, T.; Davis, H.; Mathews, M.C.; Meyer, R.D. Shallow groundwater quality on dairy farms with irrigated forage crops. *J. Contam. Hydrol.* **2002**, *55*, 287–315.
11. Gormly, J.R.; Spalding, R.F. Sources and Concentrations of Nitrate-Nitrogen in Ground Water of the Central Platte Region, Nebraskaa. *Groundwater* **1979**, *17*, 291–301.
12. Keeney, D.; Olson, R.A. Sources of nitrate to ground water. *Crit. Rev. Environ. Sci. Technol.* **1986**, *16*, 257–304.
13. Domagalski, J.L.; Ator, S.; Coupe, R.; McCarthy, K.; Lampe, D.; Sandstrom, M.; Baker, N. Comparative study of transport processes of nitrogen, phosphorus, and herbicides to streams in five agricultural basins, USA. *J. Environ. Qual.* **2008**, *37*, 1158–1169.
14. Green, C.T.; Böhlke, J.K.; Bekins, B.A.; Phillips, S.P. Mixing effects on apparent reaction rates and isotope fractionation during denitrification in a heterogeneous aquifer. *Water Resour. Res.* **2010**, *46*, doi:10.1029/2009WR008903.
15. Nolan, B.T.; Hitt, K.J.; Ruddy, B.C. Probability of nitrate contamination of recently recharged groundwaters in the conterminous United States. *Environ. Sci. Technol.* **2002**, *36*, 2138–2145.
16. Wick, K.; Heumesser, C.; Schmid, E. Groundwater nitrate contamination: Factors and indicators. *J. Environ. Manag.* **2012**, *111*, 178–186.
17. Boumans, L.; Fraters, B.; van Drecht, G. Nitrate in the upper groundwater of "De Marke" and other farms. *NJAS-Wageningen J. Life Sci.* **2001**, *49*, 163–177.
18. Elmi, A.A.; Madramootoo, C.; Egeh, M.; Liu, A.; Hamel, C. Environmental and agronomic implications of water table and nitrogen fertilization management. *J. Environ. Qual.* **2002**, *31*, 1858–1867.
19. Fraters, D.; Boumans, L.J.; van Drecht, G.; de Haan, T.; Wim, D. Nitrogen monitoring in groundwater in the sandy regions of the Netherlands. *Environ. Pollut.* **1998**, *102*, 479–485.
20. Salo, T.; Turtola, E. Nitrogen balance as an indicator of nitrogen leaching in Finland. *Agric. Ecosyst. Environ.* **2006**, *113*, 98–107.
21. Pabich, W.J.; Valiela, I.; Hemond, H.F. Relationship between DOC concentration and vadose zone thickness and depth below water table in groundwater of Cape Cod, USA. *Biogeochemistry* **2001**, *55*, 247–268.
22. Arora, B.; Mohanty, B.P.; McGuire, J.T.; Cozzarelli, I.M. Temporal dynamics of biogeochemical processes at the Norman Landfill site. *Water Resour. Res.* **2013**, *49*, 6909–6926.
23. Hatfield, J.L.; Follett, R.F. *Nitrogen in the Environment*; Elsevier: San Diego, CA, USA, 2008.
24. Assaf, H.; Saadeh, M. Geostatistical assessment of groundwater nitrate contamination with reflection on DRASTIC vulnerability assessment: The case of the Upper Litani Basin, Lebanon. *Water Resour. Manag.* **2009**, *23*, 775–796.
25. Helena, B.; Pardo, R.; Vega, M.; Barrado, E.; Fernandez, J.M.; Fernandez, L. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res.* **2000**, *34*, 807–816.
26. Mogheir, Y.; de Lima, J.; Singh, V. Characterizing the spatial variability of groundwater quality using the entropy theory: I. Synthetic data. *Hydrol. Processes* **2004**, *18*, 2165–2179.
27. Husain, T. Hydrologic Uncertainty Measure and Network Design. *J. Am. Water Resour. Assoc.* **1989**, *25*, 527–534.
28. Bueso, M.; Angulo, J.; Cruz-Sanjulian, J.; García-Aróstegui, J. Optimal spatial sampling design in a multivariate framework. *Math. Geol.* **1999**, *31*, 507–525.
29. Mariethoz, G.; Renard, P.; Froidevaux, R. Integrating collocated auxiliary parameters in geostatistical simulations using joint probability distributions and probability aggregation. *Water Resour. Res.* **2009**, *45*, doi:1029/2008WR007408.
30. Dwivedi, D.; Mohanty, B.P.; Lesikar, B.J. Estimating *Escherichia coli* loads in streams based on various physical, chemical, and biological factors. *Water Resour. Res.* **2013**, *49*, 2896–2906.
31. Chapman, T.G. Entropy as a measure of hydrologic data uncertainty and model performance. *J. Hydrol.* **1986**, *85*, 111–126.
32. Mishra, A.K.; Özger, M.; Singh, V.P. An entropy-based investigation into the variability of precipitation. *J. Hydrol.* **2009**, *370*, 139–154.
33. Singh, V.P.; Marini, G.; Fontana, N. Derivation of 2D power-law velocity distribution using entropy theory. *Entropy* **2013**, *15*, 1221–1231.
34. Hurst, H.E. Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.* **1951**, *116*, 770–808.

35. Shannon, C.E. A mathematical theory of communication. *ACM Sigmobile Mob. Comput. Commun. Rev.* **2001**, *5*, 3–55.
36. Scott, D.W. On optimal and data-based histograms. *Biometrika* **1979**, *66*, 605–610.
37. Sturges, H.A. The choice of a class interval. *J. Am. Stat. Assoc.* **1926**, *21*, 65–66.
38. Freedman, D.; Diaconis, P. On the histogram as a density estimator: L 2 theory. *Probab. Theory Relat. Fields* **1981**, *57*, 453–476.
39. Maritan, A.; Rinaldo, A.; Rigon, R.; Giacometti, A.; Rodríguez-Iturbe, I. Scaling laws for river networks. *Phys. Rev. E* **1996**, *53*, doi:10.1103/PhysRevE.53.1510.
40. Bak, P.; Paczuski, M.; Shubik, M. Price variations in a stock market with many agents. *Phys. A Stat. Mech. Appl.* **1997**, *246*, 430–453.
41. Goldberger, A.L.; Amaral, L.A.; Hausdorff, J.M.; Ivanov, P.C.; Peng, C.K.; Stanley, H.E. Fractal dynamics in physiology: Alterations with disease and aging. *Proc. Natl. Acad. Sci.* **2002**, *99*, 2466–2472.
42. Szolgayová, E.; Laaha, G.; Blöschl, G.; Bucher, C. Factors influencing long range dependence in streamflow of European rivers. *Hydrol. Processes* **2014**, *28*, 1573–1586.
43. Taqqu, M.S.; Teverovsky, V.; Willinger, W. Estimators for long-range dependence: An empirical study. *Fractals* **1995**, *3*, 785–798.
44. Montanari, A.; Rosso, R.; Taqqu, M.S. Some long-run properties of rainfall records in Italy. *J. Geophys. Res.* **1996**, *101*, 29431–29438.
45. Sakalauskienė, G. The Hurst phenomenon in hydrology. *Environ. Res. Eng. Manag.* **2003**, *3*, 16–20.
46. Ground Water Atlas of the United States, Oklahoma, Texas, HA 730-E. Available online: http://pubs.usgs.gov/ha/ha730/ch_e/E-text8.html (accessed on 1 May 2015).
47. McCarthy, J.F.; Zachara, J.M. Subsurface transport of contaminants. *Environ. Sci. Technol.* **1989**, *23*, 496–502.
48. Texas Water Development Board. Available online: http://www.twdb.state.tx.us (accessed on 1 May 2015).
49. Toth, J. A theoretical analysis of groundwater flow in small drainage basins. *J. Geophys. Res.* **1963**, *68*, 4795–4812.
50. Sophocleous, M. Interactions between groundwater and surface water: The state of the science. *Hydrogeol. J.* **2002**, *10*, 52–67.
51. Al-Zabet, T. Evaluation of aquifer vulnerability to contamination potential using the DRASTIC method. *Environ. Geol.* **2002**, *43*, 203–208.
52. Scott, J. *Computerized Stratified Random Site-Selection Approaches for Design of a Ground-Water-Quality Sampling Network*; Technical Report; Department of the Interior, US Geological Survey: Oklahoma City, OK, USA, 1990.
53. Monitoring-Well Network and Sampling Design for Ground-Water Quality, Wind River Indian Reservation, Wyoming. Available online: http://pubs.usgs.gov/sir/2005/5027/ (accessed on 1 May 2015).

© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).