

Statistical and Probability Quantification of Hydrologic Dynamics in the Lake Tuscaloosa Watershed, Alabama, USA

Shawn Dawley¹, Yong Zhang^{1*}, Xiaoting Liu², Peng Jiang³, Lin Yuan², Hongguang Sun²

¹Department of Geological Sciences, University of Alabama, Tuscaloosa, Alabama, USA

²College of Mechanics and Materials, Hohai University, Nanjing, China

³State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, China

Email: *yzhang264@ua.edu

How to cite this paper: Dawley, S., Zhang, Y., Liu, X.T., Jiang, P., Yuan, L. and Sun, H.G. (2018) Statistical and Probability Quantification of Hydrologic Dynamics in the Lake Tuscaloosa Watershed, Alabama, USA. *Journal of Geoscience and Environment Protection*, 6, 91-100.
<https://doi.org/10.4236/gep.2018.65008>

Received: April 27, 2018

Accepted: May 20, 2018

Published: May 23, 2018

Abstract

Interconnected components of water cycle, including surface water, groundwater, and precipitation, can exhibit complex hydrologic dynamics. This study investigates dynamics embedded in surface water, groundwater, and precipitation time series data in the Lake Tuscaloosa watershed located in northern Alabama, using standard statistics and non-stationarity analysis. Standard statistics analysis shows that less water is available in this watershed over time. A significant correlation between different data sets is found, and groundwater is found to be slower evolving than its nearby surface systems. Non-stationarity analysis based on time scale-local Hurst exponents calculated by the multi-fractal detrended fluctuation approach shows that, on one hand, the stream system exhibits non-stationarity properties similar to precipitation, as expected. On the other hand, groundwater and lake stage non-stationarity is found to be influenced by the seasonal variation in rainfall and the long-term anthropogenic factors. Therefore, sustainability of surface water and aquifer may be affected by natural input and/or anthropogenic activity, both of which can evolve non-stationary in different time scales.

Keywords

Surface Water and Groundwater, Statistics, Probability, Non-Stationary Evolution

1. Introduction

Water is one of the most important natural resources on the planet, with the United Nations (UN) estimating 40% of people already being affected by scarcity and projecting that number to rise. Areas where water was once abundant now

have shortages due to many factors, including changes in climate, increasing demand and changing land use [1] [2]. Climate records and predictions for the southeastern United States (US) show either a steady or decreasing trend of annual precipitation [3]. If precipitation decreases, less water will be available in the future. Groundwater is an especially important reservoir during drought conditions when it is the primary source for stream flow (*i.e.*, under base flow conditions) and surface reservoirs. One of the major concerns for hydrogeologists is how to effectively evaluate the long-term evolution, which is the sustainability assessment, of groundwater quantity and its response to a changing climate.

There are many different methods and software, such as GMS and GSFLOW, to model the interaction between surface water and groundwater at a watershed scale through physical-based, or deterministic, processes. Effectively modeling a watershed's surface/subsurface hydrologic process using physical laws, however, requires a significant amount of geological information that may be prohibitively expensive to acquire at all relevant scales, and even with ideal input data, uncertainty may still be present in the models. To avoid excess spending on data acquisition and high uncertainty in physical process based models, probability and/or statistics based models can be used. Since abundant data for groundwater level and related surface hydrological processes are often freely available through various government agencies for many locations in the US, stochastic/statistical investigations are possible without building a large-scale integrated physical model when evaluating hydrologic dynamics in, and interactions between components of water cycle.

This study aims to develop probability/statistics approaches to interpret the complex hydrologic dynamics, especially the long-term evolution, embedded in water cycle. Three steps are proposed to reach this goal. First, historical (time series) datasets need to be collected (and filled carefully using adjacent observation stations if there are any data gaps) for surface water, groundwater, and precipitation. Second, we analyze the hydrologic data for basic statistics, including the first several moments and their correlations. Third, we calculate multifractal statistics such as the Hurst exponent to investigate their scaling behavior and response to changes in climate and urbanization.

2. Study Site and Methodology

2.1. Three Sets of Time Series Data

This study is based around Lake Tuscaloosa located in northern Alabama, US, which is an artificial lake created by the damming of North River. The watershed contains two primary surficial aquifers, the Pottsville and Coker, which are primarily sandstones with some interbedded shale, siltstone and gravel. Data are collected from various stations shown in **Figure 1**.

Data are collected from two sources, including the United State Geological Survey's National Water Information Center (USGS NWIS) and the National

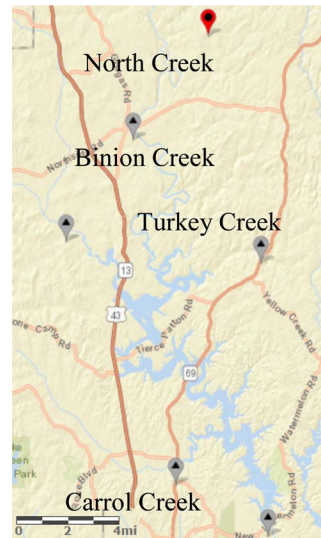


Figure 1. A map of the study area - Lake Tuscaloosa as well as the gauge stations used in this work. The red marker represents the groundwater well, the southernmost gray point represents lake stage and the other four gray points representing stream gauges. All of the above measurement points are measured at a daily resolution. Precipitation gauges are slightly outside the study area, ~13 km to the south west.

Oceanic and Atmospheric Administration National Center for Environmental Information (NOAA NCEI). The USGS measurements are all measured at a daily resolution and reported as depth to water, discharge or lake stage above mean sea level for groundwater, streams and lake respectively. The data sets are all of varying length with some measurements from as early as 1938. The data sets generally become more complete (*i.e.* less missing days) closer to the present and from late 1997 to present all data sets have nearly complete coverage. The NOAA measurements are taken at both daily and hourly resolutions and reported as depth of precipitation. Precipitation is measured as early as the 1950's.

2.2. Basic Probability/Statistics

Initially, data are analyzed for their basic statistics including probability density functions (PDFs), six typically used statistical values (including mean, median, minimum or min, maximum or max, standard deviation, and variance), and correlation. PDFs are plotted first for the duration of the available data. This "global" PDF is then compared with "local" PDFs at various scales, such as annual and decadal, to determine the change in distribution over time. The PDFs are plotted at different scales depending on the range of the data sets. Lake stage and groundwater are plotted on a linear axis, while stream flow and precipitation are plotted on log-log plots to capture the values across multiple orders of magnitude.

The standard statistics of mean, median, and others are calculated as annual values across the entire range of the data set. These statistics are then fitted with a linear trend line and compared across data sets. Each of the data sets have the six statistics mentioned above (min and median are excluded for precipitation)

and the trends compared across years and data sets to qualitatively draw conclusions on correlation. Correlation coefficients are also calculated for the different measurement stations.

2.3. Multifractal Analysis Using Time-Dependent Hurst-Coefficient

The next step is to calculate multifractal embedded in the time series. The primary index used will be the Hurst Exponent (denoted by H), which may change with time due to nonstationary evolution of the driving mechanisms. The Hurst Exponent H is a measure of long-term memory in a series that was first developed for use in hydrology by Harold Hurst in 1951 [4]. It has since been modified and applied in many signal processing applications, ranging from economics to the sequencing of DNA as in Peng *et al.* [5]. Peng *et al.* were the first to use the detrended fluctuation analysis (DFA) method to calculate Hurst exponent. In this study, we will adopt the DFA method as presented in [6] using the following four equations:

$$F(s) \approx s^H \quad (1)$$

$$Y(j) = \sum_{k=1}^j [X_k - \langle x \rangle] \quad (2)$$

$$F_k^2(S) = \frac{1}{S} \sum_{j=1}^L (Y_{j,k} - P_{j,k}^n)^2 \quad (3)$$

$$F(S) = \left[\frac{1}{m} \sum_{k=1}^m F_k^2(S) \right]^{1/2} \quad (4)$$

Equation (1) gives the scaling function, denoted as $F(s)$, which is approximately equal to the scale (s) raised to the Hurst exponent H . **Equation (2)** develops a cumulative sum, denoted as Y , where X_k is a specific value and $\langle x \rangle$ is the series mean. **Equation (3)** determines the variance of each section by subtracting a best fit polynomial of order n . Finally, **Equation (4)** finds the average variance for all segments which defines the scaling function $F(s)$. For this study, Hurst exponent calculations are conducted in MATLAB, utilizing the code for multifractal DFA written and made available in [7]. In addition to characterizing the memory of the time series, the Hurst exponent also has different meanings as show in **Table 1**.

Table 1. Range and physical meaning of the Hurst coefficient H .

Hurst Exponent Value	Meaning
$0 < H < 0.5$	Anti-Persistent Noise
$H = 0.5$	White Noise
$0.5 < H < 1$	Persistent Stationary Noise
$H = 1$	Pink Noise
$H > 1$	Non-Stationary Noise
$H = 1.5$	Brownian Motion

3. Results

Basic statistics for each data set are calculated on a yearly basis and plotted over time as shown in **Figure 2**, using the groundwater level as an example. Across all of the data sets a clear trend emerged for mean, median, min, and max all generally shows a decreasing tendency (*i.e.*, less water stored/discharged) the closer they are to the present. One notable exception is groundwater, which deviates from this trend with the annual minimum value increasing over time, or in other words, the annual maximum depth to water decreases over time. The trend across PDFs is less consistent than the statistics, but generally it showed a decrease in the frequency of high water events (*i.e.*, the aquifer receives recharge from the infiltrated rainfall) and an increase in low water events (*i.e.*, periods of groundwater loss). This can be seen most clearly in the increasing density of low water values with increased time in the precipitation, stream discharges, and lake stage. Correlation coefficients and a linear regression is shown for various data sets in **Figure 2**. Since rainfall does not instantaneously affect terrestrial water systems, the correlation is plotted across different systems and lag values as shown in **Figure 3**. The autocorrelation function (ACF) of depth to groundwater and precipitation calculated at different lag values is shown in **Figure 4**.

The Time-Scale Local Hurst Exponent (TS-LHE) is calculated for each measurement station from 1/1/1998 to 12/31/2016. The three streams show similar distributions of TS-LHE with peaks at $H = 0.2$, 0.4 and 0.55 with a positive heavy tail shown in **Figure 5** shown the minimum, maximum and mode values of H are shown in **Table 2**. The lake stage has a peak near 0.5 with most values falling between 0.5 and 1.0 . The groundwater plot shows a peak slightly above 0.5 and a negative heavy tail with another peak near 0.1 as shown in **Figure 6**. Precipitation exhibits a symmetric distribution with a peak at 0.86 and a weak heavy positive tail. Groundwater and lake stage also show a cyclical trend.

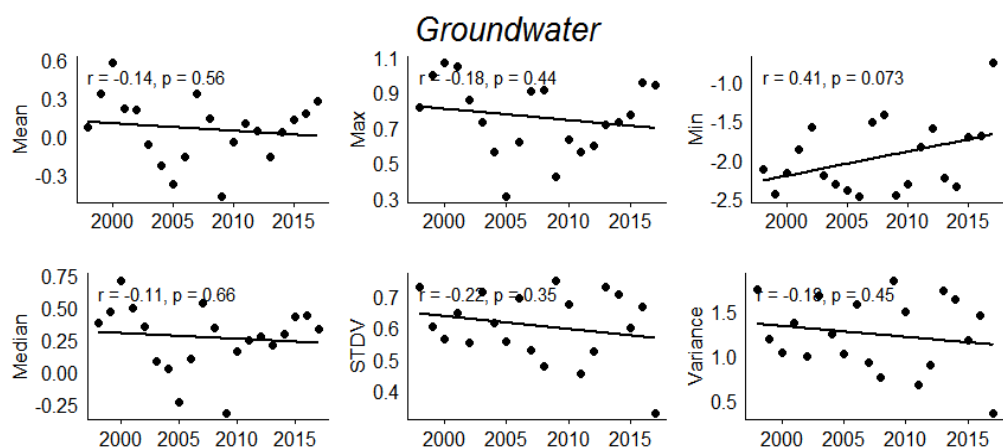


Figure 2. Statistics calculated for the groundwater fluctuation from average in meters at an annual resolution. These statistics generally show a weak negative correlation with significant noise. There are a few exceptions to this trend with groundwater minimum and much of the Turkey creek's statistics showing weak positive correlation.

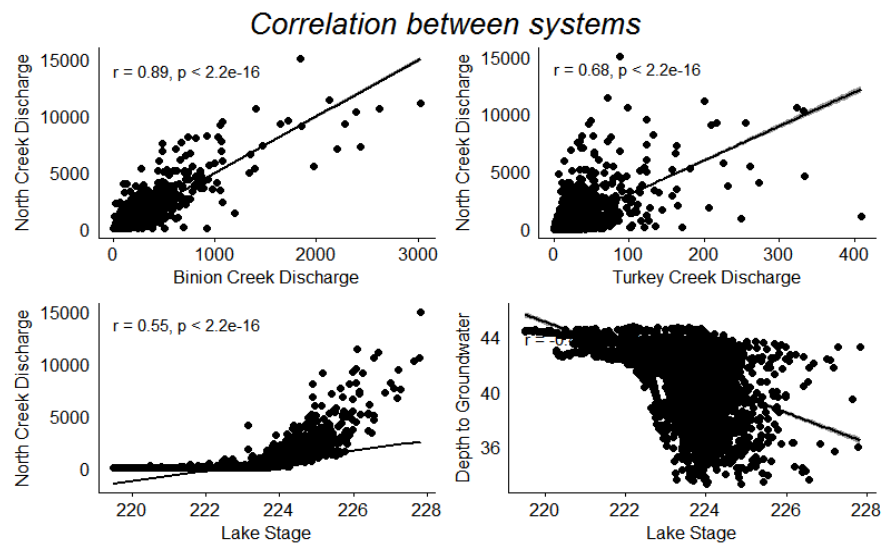


Figure 3. Scatter plots comparing correlation between discharge to discharge ((A) and (B)), discharge and lake stage (C), and depth to groundwater and lake stage (D). The r values are $r = 0.89$ (A), $r = 0.68$ (B), $r = 0.55$ (C), and $r = 0.47$ (D), respectively.

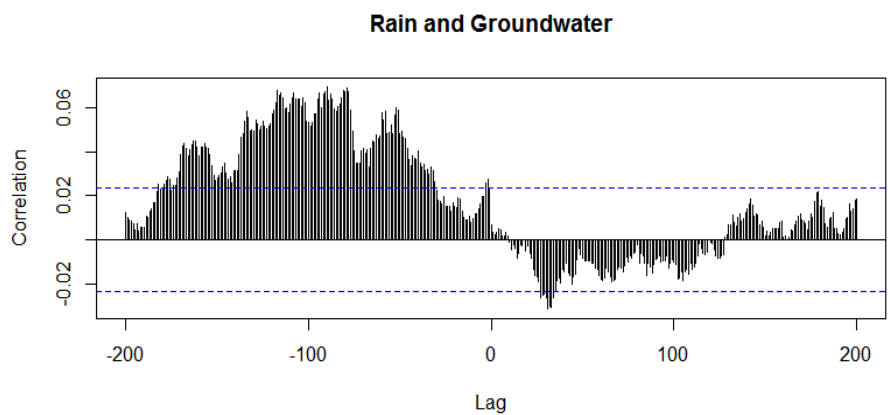


Figure 4. The correlation (autocorrelation function or ACF) of depth to groundwater and precipitation calculated at different lag values, with “1” representing one day. It is worth noting that a negative correlation for groundwater here represents higher water tables with increase precipitation/lake stage.

4. Discussion

4.1. Basic Statistics

The trend in the basic statistics shows that the amount of water available in the watershed changes with time. This is also shown by the PDF for each measurement, with an increase in the low water events and decrease in frequency of high water events. This trend is likely a result of the decreasing precipitation over the study period. The coefficient of correlation (CC) between two streams is relatively large, since they are similar systems with the same precipitation input and probably similar surface runoff. The lake stage and stream discharge show a moderately strong correlation which is expected since the streams contribute the lake. Groundwater and lake stage also have a moderate correlation, which however

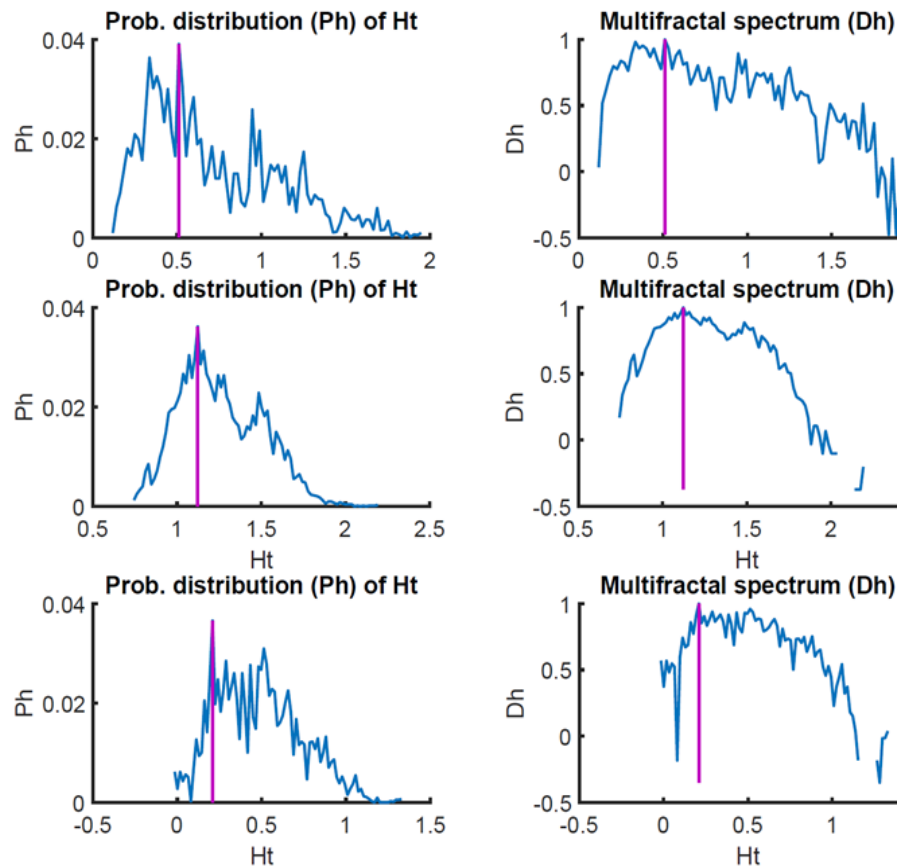


Figure 5. The PDF of TS-LHE (left), and the relationship between TS-LHE and multifractal spectrum (Dh) (right) for North Creek (the top row), Binion Creek (the middle row) and Turkey Creek (the bottom row).

Table 2. Values of mode, maximum and minimum TS-LHE for each of the three analyzed streams and groundwater.

Station	Statistics	Value
NorthSreek	Mode	0.510
	Maximum	1.959
	Minimum	0.107
Binion Creek	Mode	0.209
	Maximum	1.337
	Minimum	-0.023
Turkey Creek	Mode	0.383
	Maximum	1.349
	Minimum	-0.054
Groundwater	Mode	0.566
	Maximum	0.750
	Minimum	-0.0112

unexpected since they may not have direct hydraulic connection (note that the groundwater well is located ~16 km upstream from the lake).

Since precipitation does not instantaneously affect the terrestrial water system, the CC was calculated at different lag values to determine the offset with the highest correlation. The two surface water systems (lake and streams) have a peak correlation at a lag of 28 days, much faster than the groundwater response which peaks around 100 days and has a wider spread of the correlation values near its peak. The surface water should have a much faster impact from precipitation since overland flow is a much faster process than infiltration, and since infiltration will occur over several days or even weeks. However, this peak is obscured since there may be several rainfall events before the maximum correlation, so this is not necessarily to represent the actual time for a single rainfall event.

4.2. Nonstationary Statistics for Hydrologic Systems

The three streams all primarily show an anti-persistent characteristic (**Figure 5**), which is likely because when precipitation occurs, a relatively rapid increase in stream discharge will also occur. This rapid response to precipitation causes the time series to not be as dependent on its own past values. After the peak discharge however, there will be a slower decrease when the stream returns to base flow conditions, during which the series will be negatively correlated and thus self-dependent. The various peaks observed in the distribution of TS-LHE may represent the alternating processes of rainfall/overland flow and return to base flow conditions; however, a further study is needed to determine the above hypothesis.

Lake stage has a strong peak above $H = 0.5$ and most values between 0.5 and 1.0, in the persistent stationary noise range. This is expected since the lake is a very large system with a huge water storage compared to the input of streams

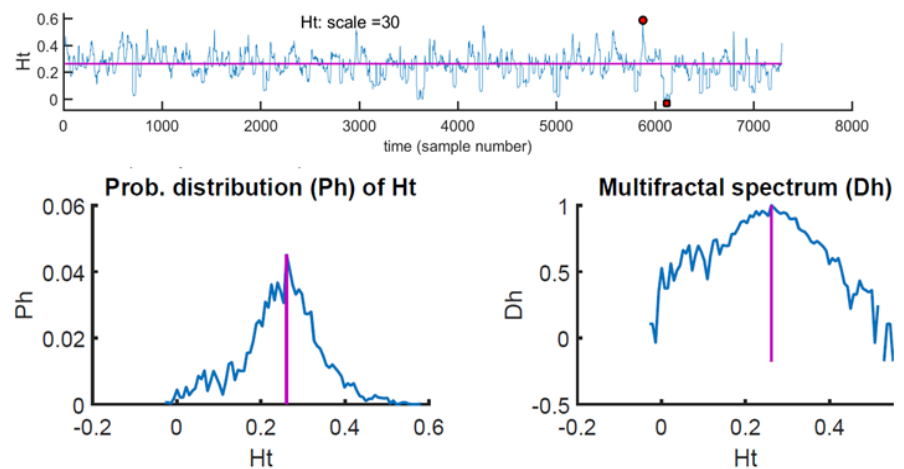


Figure 6. The plots of the groundwater TS-LHE as a time series (top), the PDF of TS-LHE (bottom left), and the relationship between TS-LHE and multifractal spectrum (D_h) (bottom right).

and precipitation. Hence the lake stage does not change rapidly due to external sources/sinks.

Groundwater has a strong peak slightly above 0.5, representing stationary noise, and several smaller peaks between 0 and 0.5, representing anti-persistent noise. These different values likely represent the different seasonal effects from the climate, as shown by the cyclical nature of the Ht plot (**Figure 6**). The strongest peak being in the persistent stationary noise range implies that groundwater is a slowly evolving system with certain memory, but it can also fluctuate depending on the seasonal input signals causing the anti-persistent peaks. Precipitation had a very strong peak with the stationary persistent noise range at 0.86. This is likely caused by the repeating days of zero precipitation.

Two factors may cause the non-stationarity of these systems, which are the seasonal variation of precipitation and the change in land use and cover. Seasonal variations in precipitation can be seen by the monthly variation of precipitation with high values in the spring, transitioning to lower values in the fall. This change in the precipitation patterns causes the systems to be in a constant stage of fluctuation. The change in land use can be seen through satellite imagery with a significant increase in anthropogenic activity in the watershed between 1984 and 2016. This increased land use has affected the measured systems, generally causing less water to be available, and influenced the non-stationarity of the systems.

5. Conclusions

This study quantifies the basic trends and long-term dynamics of various possibly interconnected hydrologic systems using both the standard statistical techniques and the no-stationarity/multifractality analysis through DFA and TS-LHE. The study site is the Lake Tuscaloosa Watershed, northern Alabama, representing watersheds in the southeastern US. From these data and calculations, the following four conclusions are made.

First, there is less water available over time, likely due to the changing climate, as shown by the basic statistics and PDF's. There is no indication that this decreasing trend will stop.

Second, there is a clear correlation between the surface water systems and the other surface/groundwater systems. The maximum influence of precipitation on surface water occurs around 30 days after rainfall and 100 days for groundwater. This discrepancy is likely affected by different precipitation events, but it also illustrates that groundwater is a much slower response system than the surface water to precipitation.

Third, the streams exhibit similar TS-LHE distributions because they are supplied by the same storms through similar land properties.

Fourth, the TS-LHE for groundwater and lake stage is highly cyclical, which is likely caused by the seasonal variation in precipitation. The seasonal variation causes the non-stationarity evolution of the terrestrial hydrologic system in shorter scales. The change in land use and land cover over time is the main

anthropogenic driver in the long-term variability of the terrestrial systems.

Acknowledgements

This work was funded partially by the National Natural Science Foundation of China under grants 41628202, 41330632 and 11572112. This paper does not necessarily reflect the views of the funding agency.

References

- [1] Collins, T.W. and Bolin, B. (2007) Characterizing Vulnerability to Water Scarcity: The Case of a Groundwater-Dependent, Rapidly Urbanizing Region. *Environmental Hazards*, **7**, 399-418. <https://doi.org/10.1016/j.envhaz.2007.09.009>
- [2] Assouline, S., Russo, D., Silber, A. and Or, D. (2015) Balancing Water Scarcity and Quality for Sustainable Irrigated Agriculture. *Water Resources Research*, **51**, 3419-3436. <https://doi.org/10.1002/2015WR017071>
- [3] Yigzaw, W. and Hossain, F. (2015) Inferring Anthropogenic Trends from Satellite Data for Water-Sustainability of US Cities Near Artificial Reservoirs. *Global and Planetary Change*, **133**, 330-345. <https://doi.org/10.1016/j.gloplacha.2015.09.013>
- [4] Hurst, H.E. (1951) Long-Term Storage Capacity of Reservoirs. *Transactions of the American Society of Civil Engineers*, **116**, 770-799. <https://ci.nii.ac.jp/naid/10011004012>
- [5] Peng, C.K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E. and Goldberger, A.L. (1994) Mosaic Organization of DNA Nucleotides. *Physical Review E*, **49**, 1685-1689. <https://doi.org/10.1103/PhysRevE.49.1685>
- [6] Habib, A., Sorensen, J., Bloomfield, J., Muchan, K., Newell, A. and Butler, A. (2017) Temporal Scaling Phenomena in Groundwater-Floodplain Systems Using Robust Detrended Fluctuation Analysis. *Journal of Hydrology*, **549**, 715-730. <https://doi.org/10.1016/j.jhydrol.2017.04.034>
- [7] Ihlen, E.A. (2012) Introduction to Multifractal Detrended Fluctuation in Matlab. *Frontiers in Physiology*, **3**.