Research Article

An Efficient Prediction Model for Water Discharge in Schoharie Creek, NY

Katerina G. Tsakiri,[1,2] Antonios E. Marsellos,[3,4] and Igor G. Zurbenko[5]

[1] *School of Computing, Engineering and Mathematics, University of Brighton, Brighton BN2 4GJ, UK*

[2] *Division of Math, Science and Technology, Nova Southeastern University, 3301 College Avenue, Fort Lauderdale-Davie, FL 33314, USA*

[3] *School of Environment and Technology, University of Brighton, Brighton BN2 4GJ, UK*

[4] *Department of Geological Sciences, University of Florida, 241 Williamson Hall, P.O. Box 112120, Gainesville, FL 32611, USA*

[5] *Department of Biometry and Statistics, State University of New York at Albany, One University Place, Rensselaer, NY 12144, USA*

Correspondence should be addressed to Katerina G. Tsakiri; ktsakiri@gmail.com

Flooding normally occurs during periods of excessive precipitation or thawing in the winter period (ice jam). Flooding is typically accompanied by an increase in river discharge. This paper presents a statistical model for the prediction and explanation of the water discharge time series using an example from the Schoharie Creek, New York (one of the principal tributaries of the Mohawk River). It is developed with a view to wider application in similar water basins. In this study a statistical methodology for the decomposition of the time series is used. The Kolmogorov-Zurbenko filter is used for the decomposition of the hydrological and climatic time series into the seasonal and the long and the short term component. We analyze the time series of the water discharge by using a summer and a winter model. The explanation of the water discharge has been improved up to 81%. The results show that as water discharge increases in the long term then the water table replenishes, and in the seasonal term it depletes. In the short term, the groundwater drops during the winter period, and it rises during the summer period. This methodology can be applied for the prediction of the water discharge at multiple sites.

1. Introduction

It has been shown that there is a connection between increased river flooding and climate change [1–3]. In addition to this increased flooding activity, it has also been proven that global atmospheric, surface, and troposphere temperatures are rising [4, 5], coincident with increasing amounts of water vapor in the atmosphere [6]. Consequently, storms with increasing supplies of moisture produce more intense precipitation events therefore increasing the risk of flooding [7]. Temperature that increases during the winter period is known to break up the river's icy surface and hence initiate "ice jams." An ice jam is a localized accumulation of ice that can create blockages of river flow and increase the probability of subsequent upstream flooding [8].

Over the last decade, Schoharie Creek has experienced a significant increase in water discharge [9–11]. This increase

can be explained by the increase of the winter melt events associated with warming temperatures [9].

Water discharge depends on the watershed area and the tributaries that drain into the stream. Schoharie Creek (hydrologic unit: 02020005 with a drainage basin of approximately 886 mi² (2300 km²) in New York, USA, flows north 93 miles (150 km) from the foot of Indian Head Mountain in the Catskill Mountains through the Schoharie Valley to the Mohawk River. It is one of the two principal tributaries of the Mohawk River, with the other being West Canada Creek. Eight further tributaries drain into the Schoharie Creek. It is critical to model flooding in order to help local governments and communities to take precautionary steps to minimize or to prevent damage on property from flooding.

Several studies on flooding in Schoharie or in different locations (e.g., in Germany, Indus River in India, or in

Elbe River in Czech [1–3, 12]) show a moderate explanation of the water discharge by the climatic variables. The main reason is the presence of uncertainties in all the variables of the time series. In particular, it has been shown that uncertainties in basic variables or from the effects of scales related with different frequencies of the variables can change the inferences of the analysis between the variables [13–15]. Moreover, it has been shown that separation of time series into different components is necessary to avoid interference from different covariance structures existing between the components of the time series. An absence of separation of scales may lead to erroneous estimated parameters of the linear regression model or other multivariate models (principal components, canonical pairs). For this reason, the time series decomposition is essential before performing analysis since absence of separation of scales can lead to misinterpretation [16].

In previous studies regarding water discharge, the time series of the water discharge variable has rarely been decomposed into different components [1–3, 12]. Specifically, Kotlarski et al. investigate the Elbe River flooding in August 2002, using the precipitation variable [3]. In their model, the authors determined a correlation pattern of 0.75 yielding an R^2 value of 0.56. A moderate interpretation of the water discharge by the climatic variables might be due to a mixed interference of scales of the time series. The explanation of the water discharge can be improved by using the decomposition of the time series.

The main purpose of this paper is to present a novel methodology to explain and improve the predictive capability of river water discharge time series by using climatic and hydrological variables. We introduce a new model which incorporates separate treatments of frequency scales. Those scales have been determined based on spectral and cospectral analysis, and they provide a physical explanation of the hydrology of this area. In particular, we separate the data into seasonal and long and short term components by using the Kolmogorov-Zurbenko (KZ) filter [17]. The KZ filter provides effective separation of frequencies and its properties are described in the following section. For this paper, we use two models for the prediction of the water discharge time series to avoid mixed interferences. This is especially necessary in high latitudes where winter and summer seasons are prolonged. Prolonged seasons, such as winter (extended by a month or more) in this latitude, may provide a mixed interference due to their different correlation structure, and they must be separated as evidenced in standard statistical approaches.

One model is designed for the winter period and another model for the summer period. This is because flooding in Schoharie Creek may be caused by heavy rainfall during the summer or rapid snowmelt or "ice jam" during the prolonged winter period [18]. For both models, we first separate the different scales of the time series of water discharge, the depth of water level below land surface (referred as groundwater level), and the climatic variables by different frequency bands. We then design different multivariate models for each scale of the water discharge time series. Different frequencies

are always uncorrelated and as a result multivariate models can be designed for each scale of the water discharge time series, separately. The long term component and the seasonal component are the same for the winter and summer models. However, the short term component has been designed using two different models for the prediction of the water discharge time series. Physical phenomena (such as an ice jam or flooding) are a result of contributions from all components (seasonal and long and short term) and not simply short term variations. We prove that by using the KZ filter, we have improved the explanation of the water discharge time series up to 81%.

2. Methods

2.1. Data. Daily time series of water discharge, groundwater level, tides, and climatic variables have been obtained through the online public database of National Oceanic and Atmospheric Administration (NOAA; http://www.ncdc .noaa.gov/data-access/land-based-station-data/land-based-datasets/quality-controlled-local-climatological-data-qclcd/; Quality Controlled Local Climatological Data, Albany, Hudson River, NY) and the U.S. Geological Survey (USGS; http://waterdata.usgs.gov/nwis/, Schoharie Creek at Burtonsville, NY, for water discharge and for depth of water level data below surface). All monitoring stations are located in an area nearby the Schoharie Watershed. In particular, daily time series of air temperature (°F), wind speed (m/sec), total rainfall and snowfall precipitation (mm/hr), and tide (feet) data from a nearby station of Schoharie Creek have been obtained by the National Oceanic and Atmospheric Administration for January 2005–February 2013 period. Moreover, daily time series of water discharge and groundwater level have been analyzed for the same area and time period.

2.2. Decomposition of Time Series. In several studies, it has been shown that the separation of a time series into different components is essential in order to avoid contributions from different covariance structures between the components of the time series [13–16, 19]. The time series of a variable can be expressed by

$$X(t) = L(t) + \text{Se}(t) + \text{Sh}(t), \tag{1}$$

where $X(t)$ represents the original time series of a variable, $L(t)$ is the long term trend component, $\text{Se}(t)$ is the seasonal component, and $\text{Sh}(t)$ is the short term component. The long term component describes the fluctuations of a time series defined as being longer than a given threshold, the seasonal component describes the year-to-year fluctuations, and the short term component describes the short term variations. The long term trend component incorporates information regarding the trend component together with the cyclical component [20]. The cyclical component of a time series describes the fluctuations around the trend. The cyclical component can be viewed as those fluctuations in a time series which are longer than a given threshold, for example, one and a half years but shorter than those attributed to the trend.

The KZ filter, which separates long term variations from short term variations in a time series [17], provides a simple design and the smallest level of interferences between scales (seasonal and long and short term components) of a time series. It allows a physical interpretation of the scales [21, 22]. Furthermore, the KZ filter provides effective separation of frequencies for application directly to datasets containing missing observations [23–25]. KZ filtration is also known to provide the best and closest results to the optimal mean square of error [17, 22, 26]. As examples of its use, the KZ filter has been applied effectively for the explanation and prediction of the ozone problem and for the explanation of the water use time series in Gainesville, Florida [19, 24, 27–30].

Specifically, the KZ filter is a low pass filter, defined by p iterations of a simple moving average of m points. The moving average of the KZ can be expressed by

$$Y_t = \left(\frac{1}{m}\right) \sum_{j=-k}^{k} X_{t+j}, \tag{2}$$

where $m = 2k + 1$. The output of the first iteration becomes the input for the second iteration, and so on. The time series produced by p iterations of the filter described in expression (2) is denoted by

$$Y_t = \text{KZ}_{m,p}(X_t). \tag{3}$$

The parameter m in expression (3) has been determined to provide the maximum explanation of the water discharge time series by the climatic variables. In this study, first, we examine the periodograms of all the time series. By using the periodograms we verify that the decomposition of the time series is essential for the prediction of the water discharge time series using the climatic variables. We design different multivariate models corresponding to each scale of the time series. For the decomposition of the time series, we use the KZ filter. The multivariate models describe the projection of the water discharge time series in the space defined by the groundwater variable and the climatic variables. This projection can be expressed through linear regression between the water discharge time series and the hydrological and climatic variables. Therefore, we decompose all the time series of the variables using the KZ filter. After the application of the KZ filter, we need to predict each component of the water discharge time series, separately. For this reason, we use multivariate models for the prediction of each component of the water discharge time series. For the application of the multivariate model, we select different climatic variables to explain the components. Finally, we estimate the water discharge time series using the climatic variables for the winter and summer period.

In the following sections, we describe the method for prediction of the seasonal and long and short term components of the water discharge time series using Schoharie Creek in New York as an example of application area.

2.2.1. Raw Data Analysis of the Water Discharge Time Series. Here, we apply the previous methodology for the decomposition of the time series to explain and predict the water
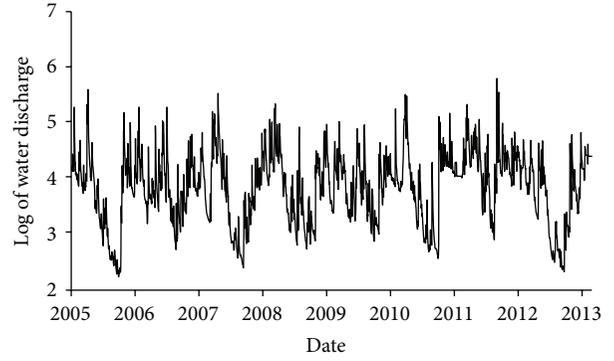

FIGURE 1: Time series of logarithm of water discharge data (ft^3/sec) during the period January 2005–February 2013.

discharge time series of Schoharie Creek. For the explanation and prediction of the water discharge, we use the climatic variables and the groundwater level. The logarithm of the water discharge from the study area was measured from January 2005 to February 2013, and it has been presented in Figure 1. We use the logarithm transform for the water discharge time series in order to stabilize the variance.

Table 1 presents the correlation matrix between the raw data of the variables. It can be noted that based on the significance level, the correlations between the water discharge and the climatic variables are statistically significant. To explain the water discharge time series using the raw data of the climatic variables, a linear regression is performed with regression coefficient, R^2, equal to 0.590. This model yields a moderate R^2 for the water discharge time series (expression (4)). Specifically, the time series of the water discharge raw data can be expressed through a linear regression as follows:

$$\text{WD}(t) = -0.378\text{T}(t) + 0.113\text{TD}(t) + 0.049\text{WS}(t)$$
$$- 0.551\text{GWL}(t) + 0.118\text{PR}(t) + \varepsilon(t), \tag{4}$$
$$R^2 = 0.590,$$

where $\text{WD}(t)$, $\text{T}(t)$, $\text{TD}(t)$, $\text{WS}(t)$, $\text{GWL}(t)$, and $\text{PR}(t)$ denote the raw data of the logarithm of water discharge, temperature, tide, wind speed, groundwater level, and precipitation, respectively. Furthermore, in expression (4), $\varepsilon(t)$ represents the residuals of this relationship and R^2 is the square of the correlation coefficient.

The relationship of the water discharge with the climatic variables can be strengthened by separating the seasonal and long and short term variations in all the time series. The separation of scales in the time series can also be verified through examination of the coherence between the variables. Table 2 shows the main periods of all the variables derived by the periodograms of the variables. It can be observed that almost all the variables consist of a 365-day period describing the seasonal component of the variables. Some variables consist of short periods (7 days, 4 days, etc.) which are related to the short term components of the variables. In particular, the precipitation variable consists mostly of short periods and for this reason this variable contributes

TABLE 1: Correlation matrix of the raw data between the water discharge, the climatic variables, and the groundwater level.

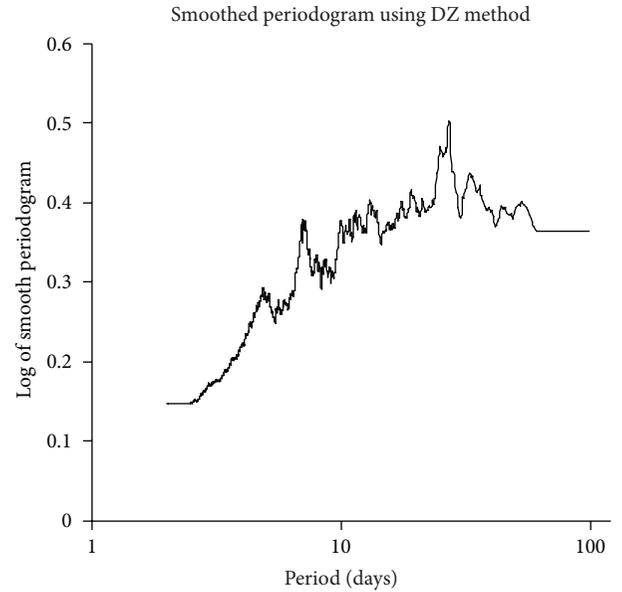| Variables	Water discharge	Groundwater level	Temperature	Tide	Wind speed	Precipitation
Water discharge	1	−0.65	−0.481	0.254	0.203	0.109
Groundwater level	−0.65	1	0.178	−0.277	−0.094	0.033
Temperature	−0.481	0.178	1	0.041	−0.193	−0.007
Tide	0.254	−0.277	0.041	1	0.105	−0.008
Wind speed	0.203	−0.094	−0.193	0.105	1	0.139
Precipitation	0.109	0.033	−0.007	−0.008	0.139	1

TABLE 2: Periods derived by the graph of periodograms of the water discharge variable and the climatic variables.

Variable	Period (days)		
	First peak of period	Second peak of period	Third peak of period
Water discharge	365	594	
Groundwater level	365	991	198
Tide	365	185	14
Precipitation	27	7	4
Wind speed	365		
Temperature	365		
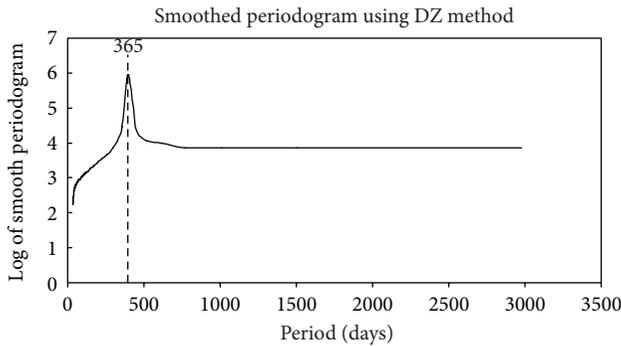
FIGURE 2: Periodogram of the temperature variable. The y-axis describes the spectral density using the KZFT (Kolmogorov Zurbenko Fourier Transform) method.

FIGURE 3: Periodogram of the precipitation variable. The y-axis describes the spectral density using the KZFT (Kolmogorov Zurbenko Fourier Transform) method.

mostly to the short term component of the water discharge. Figures 2 and 3 show two examples of the periodograms for the temperature and the precipitation variable using the DZ (DiRienzo and Zurbenko [31]) algorithm in R software. From Figures 2 and 3 and Table 2, we can conclude that it is essential to decompose the time series into different components to avoid the contribution of different frequencies between the scales of the time series.

For the decomposition of the time series, we use the $KZ_{33,3}$ filter (length is 33 with 3 iterations) to provide a physical based explanation of the water discharge time series. Following Rao et al. [24], the parameters of the KZ filter are chosen to provide the optimal solution for our study. In particular, the parameters have been estimated in order to sustain the properties of each time series component and reduce the short term variations displayed in the long and seasonal component of the time series.

The KZ filter is applied to the logarithm of daily water discharge and produces a time series devoid of short term variations and consisting only of the long term variations of the time series ($L(t)$). The same filter is applied to the variables of daily temperature, tide, wind speed, precipitation, and groundwater level.

2.2.2. Prediction of the Long Term Component of Water Discharge.
To explain the long term component of water discharge and its relationship with groundwater level and climatic variables, we examine the filtered daily temperature, tide, wind speed, and precipitation of the sum of four continuous days. For the analysis, we denote the long term components of the water discharge, temperature, tide, wind speed, precipitation, and groundwater level time series with $WD_{KZ}(t)$, $T_{KZ}(t)$, $TD_{KZ}(t)$, $WS_{KZ}(t)$, $PR_{KZ}(t)$, and $GWL_{KZ}(t)$, respectively.

The correlation between the long term component of the water discharge and the precipitation is weak (Table 3). This can be verified by the peaks in the periodogram

TABLE 3: Correlation matrix between the long term components of the water discharge, the climatic variables, and the groundwater level.

Variable	Water discharge	Groundwater level	Temperature	Tide	Wind speed	Precipitation (4 days)
Water discharge	1	−0.744	−0.641	0.39	0.596	0.175
Groundwater level	−0.744	1	0.26	−0.412	−0.43	−0.224
Temperature	−0.641	0.26	1	0.033	−0.691	0.103
Tide	0.39	−0.412	0.033	1	0.462	−0.114
Wind speed	0.596	−0.43	−0.691	0.462	1	−0.131
Precipitation (4 days)	0.175	−0.224	0.103	−0.114	−0.131	1

of the precipitation variable which occur over short periods (Table 2). Consequently, the precipitation variable contributes mostly to the short term component of the water discharge while groundwater level contributes mostly to the long term component of the water discharge (Table 3). The groundwater level is the only variable that consists of time periods greater than a year (991 days; Table 2). This period may be explained by larger time scale atmospheric episodes or from river flow changes due to anthropogenic impacts (e.g., dams, power projects, lock stations, bridges, etc.).

For the prediction of the long term component of water discharge, a linear regression is performed using the filtered logarithm of the water discharge, the filtered climatic variables, and the groundwater level time series with an R^2 value of 0.833. The long term component of the logarithm of the water discharge time series can then be expressed by

$$\mathrm{WD}_{\mathrm{KZ}}(t) = -0.704\mathrm{T}_{\mathrm{KZ}}(t) + 0.337\mathrm{TD}_{\mathrm{KZ}}(t) - 0.241\mathrm{WS}_{\mathrm{KZ}}(t)$$

$$- 0.494\mathrm{GWL}_{\mathrm{KZ}}(t) + 0.143\mathrm{PR}_{\mathrm{KZ}}(t) + \varepsilon(t),$$

$$R^2 = 0.833.$$

$$(5)$$

It can be noted that the P values for the coefficients of the variables of the linear regression model are equal to zero. Thus, the coefficients of the linear regression are statistically significant. The scale of $\varepsilon(t)$ corresponds to percent change in the long term component of water discharge data due to effects other than the climatic variables.

From expression (5), we can conclude that 83.3% of the long term fluctuations of water discharge can be explained by long term groundwater level fluctuations and major climatic variables. If we were to consider additional climatic variables such as sea level, wind direction, or relative humidity, the coefficients related to those variables would not be statistically significant (the P value associated with the t-test is greater than the significance level). For this reason, we did not consider additional climatic variables in the model.

Because expression (5) is given for the natural logarithm of water discharge, the additive term $\varepsilon(t)$ has a multiplicative effect, $\exp(\varepsilon(t))$, in the original data. Since the term $\varepsilon(t)$ is sufficiently small, we can obtain that

$$\exp(\varepsilon(t)) \approx 1 + \varepsilon(t); \text{ that is,}$$

$$\ln(1 + \varepsilon(t)) \approx \varepsilon(t).$$

$$(6)$$

Therefore, $\varepsilon(t) * 100$ corresponds to percent changes in the long term of water discharge unexplained by the groundwater level and the climatic variables.

2.2.3. Prediction of the Seasonal Component of Water Discharge. The seasonal component of a time series represents the year-to-year fluctuations of the corresponding variable. In order to predict the seasonal component of the water discharge time series, we use seasonal components of the climatic variables and the groundwater level. In particular, the seasonal component of the water discharge, $\mathrm{WD}_{\mathrm{Se}}(d)$, can be defined by

$$\mathrm{WD}_{\mathrm{Se}}(d) = \frac{1}{Y}\sum_{y=1}^{Y}\left(\mathrm{WD}(t) - \mathrm{WD}_{\mathrm{KZ}}(t)\right), \qquad (7)$$

$d = 1, \ldots, 365$, $d = 1$ represents January, and so forth

$$y = 1, \ldots, Y, \qquad (8)$$

where d represents the days of a year, y is the number of years of the observed values, $\mathrm{WD}(t)$ is the raw data of the time series of the water discharge, and $\mathrm{WD}_{\mathrm{KZ}}(t)$ is the long term component of the water discharge time series derived by the application of the KZ filter. Similarly, we define the seasonal components of the climatic variables and the groundwater level. In particular, the seasonal components of the temperature, tide, wind speed, the sum of four continuous days of precipitation, and groundwater level variables are denoted by $\mathrm{T}_{\mathrm{Se}}(d)$, $\mathrm{TD}_{\mathrm{Se}}(d)$, $\mathrm{WS}_{\mathrm{Se}}(d)$, $\mathrm{PR}_{\mathrm{Se}}(d)$, and $\mathrm{GWL}_{\mathrm{Se}}(d)$, respectively. The maximum correlation between the water discharge and the precipitation variable occurs when we consider the sum of the precipitation variable for four days.

To investigate the relationship between the seasonal component of the water discharge and the climatic variables, we estimate their correlation matrix (Table 4). The correlation between the seasonal component of the water discharge and two of the climatic variables (temperature and tide) is strong as is the case with the groundwater level. Moreover, all examined variables have a period of 365 days (Table 2). The most correlated variable with the seasonal component of the water discharge is the seasonal component of tide (Table 4).

To predict the seasonal component of the water discharge, we perform linear regression using the seasonal components of the climatic variables and the groundwater level. The coefficient of determination, R^2, is equal to 0.912.

TABLE 4: Correlation matrix between the seasonal components of the water discharge, the climatic variables, and the groundwater level.

Variable	Water discharge	Groundwater level	Temperature	Tide	Wind speed	Precipitation (4 days)
Water discharge	1	0.63	0.579	0.893	0.418	0.27
Groundwater level	0.63	1	0.492	0.831	0.406	0.012
Temperature	0.579	0.492	1	0.627	0.3	−0.046
Tide	0.893	0.831	0.627	1	0.428	0.089
Wind speed	0.418	0.406	0.3	0.428	1	0.044
Precipitation (4 days)	0.27	0.012	−0.046	0.089	0.044	1

In particular, the seasonal component of the logarithm of the water discharge can be estimated by

$$\mathrm{WD_{Se}}(d) = 0.042\mathrm{T_{Se}}(d) + 1.103\mathrm{TD_{Se}}(d) + 0.061\mathrm{WS_{Se}}(d)$$
$$- 0.335\mathrm{GWL_{Se}}(d) + 0.174\mathrm{PR_{Se}}(d) + \varepsilon(t),$$
$$R^2 = 0.912.$$
(9)

Thus, 91.2% of the variability of the seasonal component of the water discharge can be explained by the climatic variables and the groundwater level as described in expression (9). The P values for the coefficients of the variables of the linear regression are equal to zero. Thus, the coefficients of the linear regression are statistically significant. The addition of extra climatic variables such as sea level, wind direction, and relative humidity in expression (9) does not change the value of R^2.

2.2.4. Prediction of the Short Term Component of Water Discharge. For the prediction of the short term component of the water discharge, we consider the short term components of the climatic variables and groundwater level. The short term component of the water discharge time series, $\mathrm{WD_{Sh}}(t)$, can be defined as follows:

$$\mathrm{WD_{Sh}}(t) = \mathrm{WD}(t) - \mathrm{WD_{KZ}}(t) - \mathrm{WD_{Se}}(d), \quad (10)$$

where $\mathrm{WD}(t)$ represents the raw data of the water discharge time series, $\mathrm{WD_{KZ}}(t)$ represents the long term component of water discharge, and $\mathrm{WD_{Se}}(d)$ is the seasonal component of the water discharge time series. Similarly, we can define the short term components of the remaining variables.

For the short term component of the water discharge, we consider two different models. One describes the prediction of the short term component during the summer period (May through September) and the other during winter (December through March). We consider two different models because flooding in the rivers in New York State is caused by extensive precipitation (e.g., extensive rainfall or tropical storms during the summer period) or by rapid snowmelt (e.g., ice jams during the prolonged winter period). The overall explanation of the winter model is maximum during the prolonged winter period (December to March), while the summer model shows maximum explanation during the prolonged summer period (May to September).

2.2.5. Prediction of the Short Term Component of Summer Water Discharge. To predict the short term component of the water discharge time series, we perform a linear regression by using the short term components of daily temperature, precipitation, the sum of four continuous days of daily precipitation, and daily groundwater level. The short term components of the above variables are denoted by $\mathrm{T_{Sh}}(t)$, $\mathrm{PR_{Sh}}(t)$, $\mathrm{PR4_{Sh}}(t)$, and $\mathrm{GWL_{Sh}}(t)$, respectively. Table 5 shows the correlation matrix between the short term components of the variables for the summer period (May through September).

For the prediction of the short term component of the water discharge for the summer period, a linear regression is performed by using the short term components of the above climatic variables and the groundwater level with coefficient of determination, R^2, equal to 0.447. Specifically, the short term component of the water discharge can be expressed through a linear regression as follows:

$$\mathrm{WD_{Sh}}(t) = -0.097\mathrm{T_{Sh}}(t) - 0.377\mathrm{GWL_{Sh}}(t) + 0.199\mathrm{PR_{Sh}}(t)$$
$$+ 0.453\mathrm{PR4_{Sh}}(t) + \varepsilon(t),$$
$$R^2 = 0.447.$$
(11)

The P values for the coefficients of the variables of the linear regression model are equal to zero.

2.2.6. Prediction of the Short Term Component of Winter Water Discharge. Flooding in the rivers during the winter period occurs through rapid snowmelt due to the increase of air temperature. For this reason, the average temperature is used for the prediction model in the linear regression. The maximum correlation between water discharge and average temperature occurs for the average temperature over four days. Moreover, for the prediction model, we consider the variables: tide, wind speed, groundwater level, and the sum of four days of precipitation. The short term components of the above variables are denoted by $\mathrm{T4_{Sh}}(t)$, $\mathrm{WS_{Sh}}(t)$, $\mathrm{PR4_{Sh}}(t)$, and $\mathrm{GWL_{Sh}}(t)$, respectively. Table 6 shows the correlation matrix for the short term components of the variables.

To predict the short term component of the water discharge time series, we perform a linear regression by using the short term components of the above climatic variables with resultant coefficient of determination, R^2, equal to 0.719.

TABLE 5: Correlation matrix between the short term components of the water discharge, the climatic variables, and the groundwater level for the summer period.

	Water discharge	Groundwater level	Temperature average (4 days)	Precipitation	Precipitation (4 days)
Water discharge	1	−0.379	−0.247	0.124	0.502
Groundwater level	−0.379	1	0.099	0.194	−0.069
Temperature average (4 days)	−0.247	0.099	1	0.004	−0.249
Precipitation	0.124	0.194	0.004	1	−0.004
Precipitation (4 days)	0.502	−0.069	−0.249	−0.004	1

TABLE 6: Correlation matrix between the short term components of the water discharge, the climatic variables, and the groundwater level for the winter period.

	Water discharge	Groundwater level	Temperature average (4 days)	Tide	Wind speed	Precipitation (4 days)
Water discharge	1	0.398	0.57	0.732	0.25	0.229
Groundwater level	0.398	1	0.408	0.597	0.12	−0.005
Temperature average (4 days)	0.57	0.408	1	0.444	0.136	0.038
Tide	0.732	0.597	0.444	1	0.181	0.117
Wind speed	0.25	0.12	0.136	0.181	1	0.155
Precipitation (4 days)	0.229	−0.005	0.038	0.117	0.155	1

The short term component of water discharge can be expressed as follows:

$$\text{WD}_{\text{Sh}}(t) = 0.323\text{T4}_{\text{Sh}}(t) + 0.629\text{TD}_{\text{Sh}}(t) + 0.087\text{WS}_{\text{Sh}}(t)$$
$$- 0.117\text{GWL}_{\text{Sh}}(t) + 0.130\text{PR4}_{\text{Sh}}(t) + \varepsilon(t),$$
$$R^2 = 0.719.$$

(12)

The variables considered for the prediction of the short term component of water discharge during the winter period are different to those explaining the summer period. Therefore, the separation of scales of the time series is essential since different components are related to different physical phenomena (rapid snowmelt and "ice jam" for the winter period while extensive rainfall or tropical storms for the summer period). The P values for the coefficients of the variables of the linear regression model are equal to zero.

3. Results

A river can act as a gaining stream, receiving water from the groundwater system, or as a losing stream, losing water to the groundwater system. The water table's height reflects a balance between the rate of replenishment, through precipitation, and removal through discharge and withdrawal. Any imbalance either raises or lowers the water table, acting with the opposite effect to the groundwater level value (because groundwater level value is measured as the distance of the water level depth below surface). In the summer period, the water river discharge of the Schoharie Creek increases as it flows downstream as tributaries and groundwater contribute additional water (recalling that eight tributaries contribute to Schoharie Creek before it reaches Mohawk River).

During summer, groundwater resources appear to increase to ample levels (recharge). Natural replenishment will decrease the measured value of the groundwater level and enhance the river water discharge. Hence, it produces a negative association.

In regions where there is a prolonged winter period, the rainfall to replenish the water table is often scarce, and the rate of water table recharge will be less than the river's water discharge. The groundwater will drop and may result in very low water level (depletion). Groundwater level value will increase (water level drops), and water discharge in the stream increases (primarily through snow melt), which produces a positive association.

Due to the separation of scales in the time series, the long term component of water discharge shows a negative correlation with the long term component of the groundwater level and temperature (Table 3), while the correlation between the seasonal components of those variables is positive (Table 4). A positive correlation has also been observed between the short term components of those variables during the winter (Table 6), while a negative correlation exists during the summer period (Table 5).

The negative correlation between the long-term components of the water discharge and the groundwater level is due to the increased rainfall during the last decade in Schoharie Creek area. This phenomenon has also been observed for the short term components during the summer period when precipitation is intense. A positive correlation between the water discharge and the groundwater level takes place during the prolonged winter period.

The correlation between the short term component of the water discharge and precipitation during the winter is lower than the short term components of those variables during the summer. This is because in areas that experience prolonged

TABLE 7: Summary of variances and the coefficient of determinations for the seasonal and long and short term component for the summer and winter period.

	Long term	Seasonal term	Short term
	Summer period		
Variance	0.588	0.178	0.232
R^2	0.833	0.912	0.447
	Winter period		
Variance	0.531	0.162	0.307
R^2	0.833	0.912	0.719

winter seasons, rainfall does not contribute to water table in the same rate as it would replenish the water table during the summer period.

To estimate the total explanation of the model for the water discharge of the summer period, we combine expressions (5), (9), and (11) to represent the seasonal and long and short term components (Table 7). The contribution of the long term component of the climatic variables and the groundwater level in the time series of the water discharge is 49% (0.588 × 0.833) from expression (5). Furthermore, 16.2% (0.178 × 0.912) is the contribution from the seasonal component using expression (9), while 10.4% (0.232 × 0.447) is the contribution of the short term component from expression (11). By combining expressions (5), (9), and (11), in a similar way to expression (1), we can then explain 75.6% of the total variance of the water discharge time series using the climatic variables and groundwater level during the summer period. Figure 4 shows the raw data of the time series of the water discharge data (blue line) along with the prediction model (purple line) derived by expressions (5), (9), and (11) for the year 2006 (summer flood in June 29; widespread flooding in the Mohawk and Hudson basins, and Catskills was observed). For other years, similar graphs can be presented for the remaining years during the summer period.

By combining expressions (5), (9), and (12), we can estimate the total explanation of the water discharge time series from climatic variables for the winter period (Table 7). In particular, 44.2% (0.531 × 0.833) is the contribution of the long term components of the climatic variables and groundwater level to the time series of the water discharge as described in expression (5). 14.8% (0.162 × 0.912) is the contribution of the seasonal components of the climatic variables as described in expression (9), while 22.1% (0.307 × 0.719) is the contribution of the short term component by expression (12). Consequently, 81.1% is the total explanation of the water discharge time series using the climatic variables and the groundwater level during the winter period. As an example, Figure 5 shows the raw water discharge time series data (blue line) for the year 2010 along with the prediction model (purple line) described by expressions (5), (9), and (12). Our model accounts for the "ice jam" event of 2010 (January 25-26; widespread flooding occurred across east central New York and adjacent western New England from a combination of rain, snowmelt, and frozen ground).
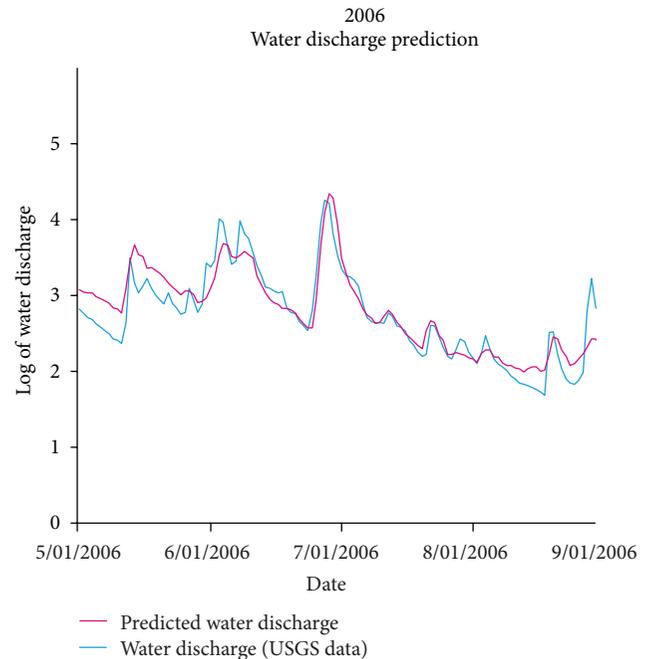

FIGURE 4: Raw water discharge data (blue line) along with the prediction model (purple line) for the summer period (May through September 2006).

As a consequence of summer and winter model results, we prove that the decomposition of the time series improves our ability to describe and predict the time series variations of water discharge by approximately two times. In particular, the unexplained variance derived by the raw data described in expression (4) is 41%, while the unexplained variance derived from the decomposition method of the winter water discharge time series is 19%. Similar results can be derived for the summer period providing an unexplained variance of 24.4%. This method can be applied in other locations as well. In such cases, the coefficients associated with the above expressions will be different.

4. Discussion

This study focuses on predicting the daily water discharge time series using available climatic variables and the groundwater level. We prove that the decomposition of the time series is essential due to the presence of short term variations in the time series. We use the KZ filter to decompose the time series into the seasonal and long and short term components to provide a physical based explanation for the time series of the water discharge. The long term component is associated with long term changes, the seasonal component with year-to-year fluctuations, and the short term component with short term variations.

In this study, we prove that the seasonal and long and short term components of water discharge can be explained through consideration of climatic variables and groundwater level. The results show that as water discharge increases in the long term the water table replenishes, while in the seasonal
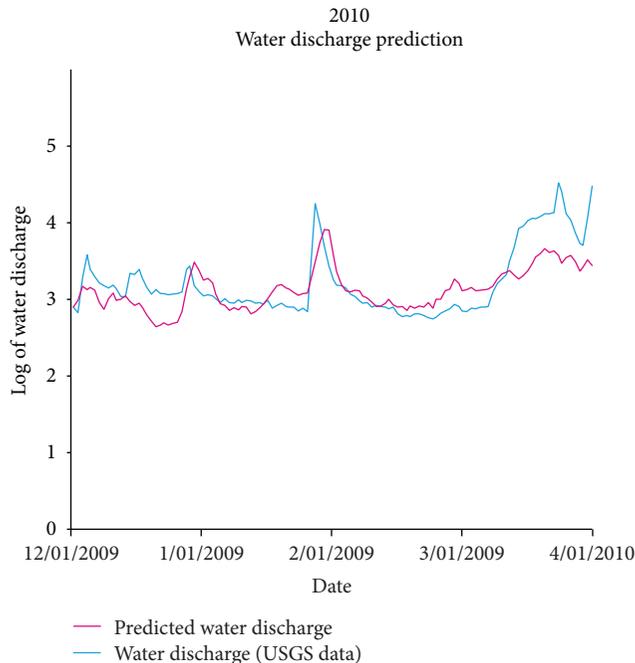
2010
Water discharge prediction

FIGURE 5: Raw water discharge data (blue line) along with the prediction model (purple line) for the winter period (December of 2009 through March of 2010).

term it depletes. In the short term, the groundwater drops during the winter period and rises during the summer period. The short term component of the water discharge is related to synoptic weather fluctuations and short term effects of rain, storms, and cyclones during the summer period and the rapid increase in temperature as well as storms during the winter period. As it is described in expressions (5), (9), (11), and (12), the selection of variables and coefficients of the multivariate models is different for the explanation of the seasonal and long and short term components of the water discharge time series. This requires the separation of the different scales of the water discharge time series in order to avoid erroneous results [14–16].

After the application of the KZ filter and the decomposition of all time series, we apply different multivariate models to each component of the water discharge time series. Different scales of time series are associated with different correlation structures (Tables 3–6) and provide different coefficients for the multivariate models [14–16]. Furthermore, for the explanation of the short term component of water discharge, we design two different models. We use different variables to explain the short term component during the winter and summer period and as a result, we need to use two different multivariate models for the explanation of the short term component of water discharge (Tables 5 and 6). This is because, in the specific location, the local climate is controlled mainly by two prolonged seasons instead of four as a typical temperate climate represents.

In our paper, we show that the accuracy of the water discharge time series prediction can be increased up to 81.1% for the winter period and 75.6% for the summer period by

incorporating the KZ filter in a separation of the scales of the time series [17]. Moreover, the coefficient of determination of the water discharge time series is very strong for all scales of water discharge time series, and it exceeds those of the raw data. This model can be used for the prediction of critical levels of water discharge ahead of time as long as other variables will be received prior to the event and are applicable to other locations. Due to the variability of local hydrological and climatic characteristics, it would be likely that other locations would produce variable numeric values but maintain the relative improvements in prediction accuracy.

5. Conclusions

The prediction of the daily water discharge time series using the climatic variables and groundwater level can be substantially improved through the decomposition of the time series. The decomposition of the different components (scales), which are the seasonal and long and short term components, avoids erroneous results and approximately doubles the prediction accuracy of the water discharge time series relative to raw data. The resulting isolation of the short term variations by the decomposition of the time series shows a summer period with a water table replenishment and a winter period with a water table depletion. The design of multivariate models (winter and summer) can improve the prediction of flooding caused by storms, rapid snowmelt, and ice jams.
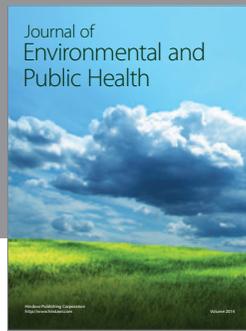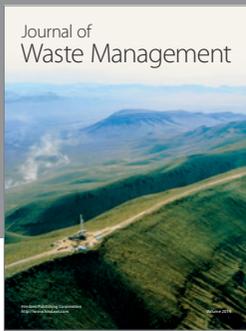
Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] A. Bronstert, "River flooding in Germany: influenced by climate change?" *Physics and Chemistry of the Earth*, vol. 20, no. 5-6, pp. 445–450, 1995.

[2] H. Hartmann and L. Andresky, "Flooding in the Indus river basin-A spatiotemporal analysis of precipitation records," *Global and Planetary Changes*, vol. 107, pp. 25–35, 2013.

[3] S. Kotlarski, S. Hagamann, P. Krahe, R. Podzun, and D. Jacob, "The Elbe river flooding 2002 as seen by an extended regional climate model," *Journal of Hydrology*, vol. 472-473, pp. 169–183, 2012.

[4] P. Brohan, J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, "Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850," *Journal of Geophysical Research*, vol. 11, no. 12, 2006.

[5] T. M. Smith and R. W. Reynolds, "A global merged land-air-sea surface temperature reconstruction based on historical observations (1880–1997)," *Journal of Climate*, vol. 18, no. 12, pp. 2021–2036, 2005.

[6] K. E. Trenberth, J. Fasullo, and L. Smith, "Trends and variability in column-integrated atmospheric water vapor," *Climate Dynamics*, vol. 24, no. 7-8, pp. 741–758, 2005.

[7] K. E. Trenberth, "Changes in precipitation with climate change," *Climate Research*, vol. 47, no. 1-2, pp. 123–138, 2011.

[8] S. Beltaos, "Onset of river ice breakup," *Cold Regions Science and Technology*, vol. 25, no. 3, pp. 183–196, 1997.

[9] A. L. Kern, *Study of 20th century trends in stream flow for West Canada and schoharie creeks of the Mohawk-Hudson Rivers watershed [Ph.D. thesis]*, Department of Geology, Union College, Schenectady, NY, USA, 2008.

[10] A. E. Marsellos, J. I. Garver, J. M. H. Cockburn, and K. Tsakiri, "Determination of historical channel changes and meander cut-off points using LiDAR and GIS in Schoharie creek, NY," in *Proceedings of the Mohawk Symposium*, pp. 28–33, Schenectady, NY, USA, 2010.

[11] A. E. Marsellos and J. I. Garver, "Channel incision and landslides identified by LiDAR in the lower reaches of Schoharie creek, New York," *Geological Society of America Abstracts with Programs*, vol. 42, no. 1, p. 73, 2010.

[12] S. A. Johnston and J. I. Garver, "Record of flooding on the Mohawk River from 1634 to 2000 based on historical archives," *Geological Society of America Abstracts with Programs*, vol. 33, no. 1, 2001.

[13] I. G. Zurbenko and M. Sowizral, "Resolution of the destructive effect of noise on linear regression of two time series," *Far East Journal of Theoretical Statistics*, vol. 3, pp. 139–157, 1999.

[14] K. G. Tsakiri and I. G. Zurbenko, "Destructive effect of the noise in principal component analysis with application to ozone pollution," in *Proceedings of the Joint Statistical Meetings*, pp. 3054–3068, Statistical Computing Section, American Statistical Association, Alexandria, Va, USA, 2008.

[15] K. G. Tsakiri and I. G. Zurbenko, "Model prediction of ambient ozone concentrations," in *Proceedings of the Joint Statistical Meetings*, pp. 3054–3068, Statistical Computing Section, American Statistical Association, Alexandria, Va, USA, 2009.

[16] K. Tsakiri and I. G. Zurbenko, "Effect of noise in principal component analysis," *Journal of Statistics and Mathematics*, vol. 2, no. 2, pp. 40–48, 2011.

[17] I. G. Zurbenko, *The Spectral Analysis of Time Series*, Statistics and Probability, North Holland, Amsterdam, The Netherlands, 1986.

[18] J. I. Garver and J. M. H. Cockburn, "A historical perspective of ice jams on the lower Mohawk River," in *Proceedings of the Mohawk Symposium*, pp. 25–29, 2009.

[19] K. Tsakiri and I. G. Zurbenko, "Explanation of fluctuation in water use time series," *Journal of Environmental and Ecological Statistics*, vol. 20, no. 3, pp. 399–412, 2013.

[20] OECD, "Section 4: guidelines for the reporting of different forms of data," in *Data and Metadata Reporting and Presentation Handbook*, OECD, Paris, France, 2005.

[21] Wikipedia, "Kolmogorov-Zurbenko Filters," http://en.wikipedia.org/wiki/Kolmogorov%E2%80%93Zurbenko_filter.

[22] W. Yang and I. Zurbenko, "Kolmogorov-Zurbenko filters," *Wiley Interdisciplinary Reviews*, vol. 2, no. 3, pp. 340–351, 2010.

[23] R. E. Eskridge, J. Y. Ku, S. Trivikrama Rao, P. Steven Porter, and I. G. Zurbenko, "Separating different scales of motion in time series of meteorological variables," *Bulletin of the American Meteorological Society*, vol. 78, no. 7, pp. 1473–1483, 1997.

[24] S. T. Rao, I. G. Zurbenko, R. Neagu, P. S. Porter, J. Y. Ku, and R. F. Henry, "Space and time scales in ambient ozone data," *Bulletin of the American Meteorological Society*, vol. 78, no. 10, pp. 2153–2166, 1997.

[25] W. Yang and I. Zurbenko, "Nonstationarity," *Wiley Interdisciplinary Reviews*, vol. 2, no. 1, pp. 107–115, 2010.

[26] B. Close and I. Zurbenko, "Kolmogorov-Zurbenko adaptive filters," version 3, R-software, 2013, http://cran.r-project.org/web/packages/kza/index.html.

[27] S. T. Rao and I. G. Zurbenko, "Detecting and tracking changes in ozone air quality," *Journal of the Air and Waste Management Association*, vol. 44, no. 9, pp. 1089–1092, 1994.

[28] C. Hogrefe, S. T. Rao, I. G. Zurbenko, and P. S. Porter, "Interpreting the information in ozone observations and model predictions relevant to regulatory policies in the Eastern United States," *Bulletin of the American Meteorological Society*, vol. 81, no. 9, pp. 2083–2106, 2000.

[29] K. G. Tsakiri and I. G. Zurbenko, "Prediction of ozone concentrations using atmospheric variables," *Journal of Air Quality, Atmosphere, and Health*, vol. 4, no. 2, pp. 111–120, 2011.

[30] K. G. Tsakiri and I. G. Zurbenko, "Determining the main atmospheric factor on ozone concentrations," *Journal of Meteorology and Atmospheric Physics*, vol. 109, no. 3-4, pp. 129–137, 2010.

[31] A. G. DiRienzo and I. G. Zurbenko, "Semi-adaptive nonparametric spectral estimation," *Journal of Computational and Graphical Statistics*, vol. 8, no. 1, pp. 41–59, 1999.

Journal of
Waste Management

The Scientific
World Journal

Journal of
Environmental and
Public Health

Scientifica

International Journal of
Ecology

Journal of
Ecosystems

International Journal of
Oceanography

International Journal of
Atmospheric Sciences

Journal of
Marine Biology

Hindawi

Submit your manuscripts at
http://www.hindawi.com

International Journal of
Biodiversity

Journal of
Geological Research

International Journal of
Forestry Research

Applied &
Environmental
Soil Science

Journal of
Climatology

Journal of
Earthquakes

Journal of
Computational
Environmental Sciences

International Journal of
Geophysics

Advances in
Meteorology

Advances in
Environmental
Chemistry