The quantitative analysis of the influence of environmental factors on the water yield capacity: a study in Haihe river basin, China
Qiufen Zhang, Jiakai Liu, Lihua Chen and Xinxiao Yu

ABSTRACT
Many studies have qualitatively analyzed the response of hydrological characteristics to climate change in Haihe river basin, but quantitative research was rare, which is essential for water resource management. To evaluate and quantitatively analyze the relationship between catchment runoff capacity and environmental factors, principal component analysis, step regression analysis, and sensitivity analysis were conducted. The results show that the runoff capacity of Haihe river basin was mainly controlled by vegetation types and soil texture; catchments with lower runoff capacity in were mainly distributed in the upstream/northwest regions. In the catchments with middle runoff capacity, a 10% increase in precipitation (PRE), potential evapotranspiration (PE), and plant-available water coefficient (PAWC) would result in a 23.6% increase, 12.9% decrease, and 5.1% decrease in annual runoff, respectively, whereas in low runoff capacity catchments, a 10% increase in slope and leaf area index (LAI) would result in a 17.8% increase and 10.5% decrease in annual runoff, and in high runoff capacity catchments, a 10% increase in normalized differential vegetation index (NDVI) would result in a 12.6% increase in annual runoff. Soil conditions and vegetation configuration improvement in the upstream of Haihe river basin may contribute to the improvement of available water resources.

Key words | annual runoff, climate variability, Haihe river basin, land use/cover conditions, runoff capacity

INTRODUCTION
The Haihe river basin is located in northern China and covers the entire Beijing–Tianjin–Hebei region, the political and cultural center of China. The per capita water availability in this area is less than 150 m³/y (Yang et al. 2003; Xia et al. 2012); a lack of water resources has become one of the most important issues for regional social and economic development (Wang & Yu 2014; Zhao et al. 2015). Furthermore, many observations and studies have shown that the amount of water resources in the Haihe river basin significantly decreased over the last half of the 20th century (Xu et al. 2014), and the reduction magnitude of runoff in the basin has been the highest among China’s major rivers over the past 40 years. The annual runoff from the mountain region has sharply decreased since the 1970s (Bao et al. 2012a; Wang et al. 2015). Hence, this area has received much more attention recently, and many studies have focused on analyzing the reasons of runoff decline in this area.

Climate change and human activities are the most common reasons for stream flow changes (Sriwongsitanon & Taesombat 2011; Renner & Bernhofer 2012). Many studies have analyzed the sensitivity and/or effect of climate changes and human activities on the runoff in the Haihe river basin. Wang & Yu (2015) studied 42 watersheds in
the mountainous area and noted that a 10% increase in precipitation or potential evapotranspiration (PE) would lead to a 32.1% increase or 22.1% decrease in runoff. As their study was based on the climate change impact hypothesis (Renner et al. 2012), it might overestimate the sensitivity in the area. Similarly, Xu et al. (2014) studied 33 catchments in the Haihe river basin and found that a 10% increase in precipitation, PE, and land use and land cover conditions (LULC) would result in a 23.3% increase, 13.5% decrease, and 15.8% decrease in stream flow, respectively, and the contribution rates of climate change (including precipitation and PE) and LULC to runoff change were 22.4% and 77.6%, respectively. Previous studies on the annual stream flow of catchments in the Haihe river basin concluded that human activities contribute to 57–73.9% of the runoff reduction, whereas climate change only accounts for 23.4–43% (Wang et al. 2009; 2013; Wang & Hejazi 2011; Bao et al. 2012b). Therefore, as the main manifestation of human activity, land use and cover change (LUCC) has a great influence on the runoff capacity in this region.

However, the parameters used to represent LULC characteristics were often simple or inexplicit. Zhao et al. (2014) used the area variation of each land use type to represent the LUCC and found that terrace and forest areas have a positive correlation with the annual runoff. While this simplified analysis provides a first step to a better understanding of the effect of LULC on runoff, other LULC factors such as slope gradient and soil texture require further study. Xu et al. (2014) used a simplified parameter \( n \) mainly related to soil, topography, and vegetation properties (Yang et al. 2008) to represent the LUCC conditions and predicted that every 10% increase in \( n \) would lead to a 27% decrease in annual runoff. However, the used parameters were too obscure to provide an understanding of the relationship between water yield and each individual environmental variable. Moreover, previous studies mainly analyzed the contribution or sensitivity of climate changes or human activities to runoff (Bao et al. 2012b; Wang et al. 2013); this is insufficient to propose feasible suggestions for Haihe water resource management. A quantitative analysis of the relationship between runoff and environmental variables according to the long-time measured data in catchments with different runoff capacities should be more practically instructive but seldom conducted.

Various methods have been used to quantitatively study the effects of climatic change on streamflow. Among these methods, statistical analysis, climate elasticity, and hydrological modeling have been widely used in different regions of the world (Legesse et al. 2003; Ma et al. 2010; Bao et al. 2012a; Wang et al. 2013; Xu et al. 2014). Hydrological modeling method is physically sound but has uncertainties in the structure and parameters of the model, and model calibration is laborious and subjective (Legesse et al. 2003). In contrast, many studies based on statistical analysis and elastic coefficient method of streamflow have been evaluated and determined to be more objective, effective, and simpler than hydrological modeling (Sankarasubramanian et al. 2001; Fu et al. 2007).

In this study, statistical analysis combined with elastic coefficient method was used to quantitatively determine the relationship between runoff coefficient (RC) and environmental variables. The aim of this study was to investigate the relationship between RC and environmental variables, establish the quantitative relationship between runoff and environment variables in catchments with different runoff capacities, and estimate the sensitivity of runoff to environmental variables.

MATERIALS AND METHODS

Study area

The Haihe river basin has an area of approximately 318,200 km², accounting for 3.3% of the national land area. Whereas the population of Haihe river basin is almost 150 million, it accounts for more than 10% of the total population of China, and the GDP of this area is almost 12% of the GDP of China. The studied rocky mountainous area (35°4′– 44°11′ N and 111°41′–120° 17′ E) of the Haihe river basin has an elevation of 0–2,940 m above sea level (Figure 1). This region is semiarid because of the semihumid continental monsoon climate. The average annual precipitation in all the studied catchments ranges from 320 mm to 774 mm. Approximately 75% of
precipitation occurs in the rainy season (between June and September), and the mean aridity index (AI) is 2.03. The annual mean temperature is 8.1°C. To exclude the anthropogenic interference on runoff, 57 eligible catchments without water conservation projects (e.g., dams) were selected; the LULC area was almost unchanged during the study period. The total area of Haihe river rocky mountainous area is approximately 186,400 km², accounting for 58.6% of the total Haihe river basin. More detailed information about all the studied catchments can be found in the Appendix, Table A1.

Environmental variables and data

Data source

Ten environment factors, precipitation (PRE), runoff (R), PE, AI, available soil water capacity (ASWC), LAI, normalized differential vegetation index (NDVI), average catchment slope (slope), plant-available water coefficient (PAWC), and annual RC, were considered in this study. Numerous studies have indicated that the runoff in Haihe river basin changed abruptly after 1979 under the
influence of human activities and showed a significant decrease trend (Sun & Ren 2015; Wang et al. 2015). Therefore, the annual runoff data (1979–2000) after the abrupt change point (1979) of Haihe river basin were extracted from the Hydrological Year Book of the People’s Republic of China, which is not an open source but can be applied to for access; the meteorological data (precipitation and evapotranspiration) for the same period were obtained from the China Meteorological Data Sharing Service System (http://cdc.cma.gov.cn/). The digital elevation model was downloaded from the National Geospatial Data Cloud Platform (http://www.gscloud.cn/) to delineate catchment boundaries and calculate average catchment slopes. The LULC data (including soil and vegetation data) were derived from the National Earth System Science Data Sharing Infrastructure (http://www.geodata.cn). The resolutions of digital elevation model and LULC data were 90 m and 300 m, respectively. Furthermore, the original soil component data, including particle size, bulk density, organic matter (OM), etc., were obtained from a soil profile survey (Zhou et al. 2005) across the Haihe river basin in 2000 and used to calculate ASWC.

**Data calculation**

According to the available data of hydrological stations, 57 catchments without reservoirs in Haihe river basin were selected in this study. The water reduction in each outlet was calculated by the water conservancy department to eliminate the variation in water quantity caused by human activities, reducing the analysis error. The PE and PRE in each catchment were calculated using the Thiessen polygon method (Han & Bray 2006; Yang et al. 2018); the values of ASWC, LAI, NDVI, and PAWC in each catchment were calculated using the weighted average method with the areas used as weights. In addition, the average annual NDVI and LAI data from 1970 to 1980 were used. It should be emphasized that PAWC and ASWC are different parameters with totally different physical meanings. ASWC was used to quantify the soil texture and composition in a catchment (Zhou et al. 2005), whereas PAWC represents the water use characteristics of various types of vegetation. The ASWC was calculated using the following empirical formula (Zhou et al. 2005):

$$\text{ASWC} = 54.509 - 0.132\text{sand} - 0.032(\text{sand})^2 - 0.055\text{clay}$$

$$- 0.006(\text{clay})^2 - 0.738\text{silt} - 0.007(\text{silt})^2 - 2.688\text{OM}$$

$$+ 0.501(\text{OM})^2 (r^2 = 0.925, P < 0.001)$$

(1)

where OM is the organic matter.

Moreover, according to the long-term water balance equation,

$$\text{PRE} = \text{ET} + R$$

(2)

and the Zhang equation (Zhang et al. 2001):

$$\frac{\text{ET}}{\text{PRE}} = \frac{1 + \text{PAWC} \times \frac{\text{PE}}{\text{PRE}}}{1 + \text{PAWC} \times \frac{\text{PE}}{\text{PRE}} + \left(\frac{\text{PE}}{\text{PRE}}\right)^{-1}}$$

(3)

where PRE, ET, PE, and R are the precipitation, evapotranspiration, potential evapotranspiration, and annual runoff, respectively. Then, PAWC can be calculated as follows:

$$\text{PAWC} = \frac{\text{PRE}^3 - \text{PRE}^2 \times \text{PE} - \text{PRE} \times \text{PE} \times R}{\text{PE} \times R}$$

(4)

**K-means clustering and principal component analysis (PCA)**

K-means clustering is a commonly used data clustering method for performing unsupervised learning tasks (Ding & He 2004); it has been widely used to study water engineering issues (Chang et al. 2008; Nosrati & Eeckhaut 2012; Javadi et al. 2017). It is a useful technique to identify groups that both minimize the within-group variation of data in a cluster and maximize the between-group variation of data to identify potential differences between clusters (Chang et al. 2008). Especially in large river basins, proper classification of subwatersheds is helpful for the management of water resources. In this study, the clustering number was decided by cascade clustering and the simple structure index (SSI) method (Carroll 1953), and the catchments were clustered by annual RC.

Principal component analysis (PCA) has been widely used in many fields such as ecology (Borcard et al. 2011) and hydrology (Mathboust et al. 2017; Niu et al. 2017; Znachor...
to analyze the relationships between observations and several intercorrelated quantitative independent variables by reducing dimensionality (Jolliffe 2011) and to display them in a plan. PCA is structured to derive causal associations between dependent (response) and independent (explanatory) variables. Similar causal analysis methods include: Grow-Shrink (GS) (Margaritis & Thrun 1999); a variant of Incremental Association Markov Boundary (IAMB), inter IAMBnPC (Tsamardinos et al. 2003); Local Causal Discovery, LCD2 (Cooper 1991); HITON Markov Blanket, HITON-MB (Aliferis et al. 2003); and First Order Utility, FOU (Brown 2003). Ssegane et al. (2017) compared the accuracy, consistency, and predictive potential of variables selected by PCA and the above methods based on 26 Mid-Atlantic Piedmont watersheds with 111 watershed descriptors and 19 streamflow percentiles, and confirmed that PCA still had absolute superiority in variable selection.

The details of PCA and K-mean clustering have been described in previous studies (Carroll 1954; Legendre & Legendre 1998; Borcard et al. 2011; Jolliffe 2011) and can be represented as:

$$x_{ij} = z_{ij} - ar{z}_{j}$$  \hspace{1cm} (5)

where $z_{ij}$ represents the primary data in catchments, and $X = [1]$ is the new normalized matrix ($P \times M$). Then, the inner product matrix $S$ can be expressed as

$$S = XX^T$$  \hspace{1cm} (6)

Then, the eigenvalues $\lambda$ of $S$ can be calculated as:

$$|S - \lambda I| = \begin{bmatrix} (S_{11} - \lambda_i) & S_{12} & L & S_{1P} \\ S_{21} & (S_{22} - \lambda_i) & L & S_{2P} \\ M & M & O & M \\ S_{P2} & S_{P2} & L & (S_{PP} - \lambda_i) \end{bmatrix} = 0$$  \hspace{1cm} (7)

The number of eigenvalue is $P$; then, the eigenvectors can be calculated as follows:

$$|S - \lambda_i I|U_i = \begin{bmatrix} (S_{11} - \lambda_i) & S_{12} & L & S_{1P} \\ S_{21} & (S_{22} - \lambda_i) & L & S_{2P} \\ M & M & O & M \\ S_{P2} & S_{P2} & L & (S_{PP} - \lambda_i) \end{bmatrix} U_i = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$  \hspace{1cm} (8)

The number of $U_i$ is $P$; the eigenvectors can now be used to construct a new matrix:

$$U = \begin{bmatrix} U_{11} & L & U_{1P} \\ M & O & M \\ U_{P1} & L & U_{PP} \end{bmatrix}$$  \hspace{1cm} (9)

Next, the sorting coordinate $Y$ can be calculated as:

$$Y = UX$$  \hspace{1cm} (10)

Finally, the contribution of variables can be calculated as:

$$C_{ij} = \sqrt{\lambda_j} \times U_{ij}(i = 1, 2 \ldots, P)$$  \hspace{1cm} (11)

where $C_{ij}$ is the contribution of variable $i$ to the principal component $j$ (PC). In this study, PC1 and PC2 and a double sequence diagram were used to represent the relationship between RC and the variables; the RStudio1.1.383 and R packages were used to perform the above analysis.

**Stepwise regression analysis**

Stepwise regression analysis has been widely used in multivariate statistics. This includes a regression model where the choice of predictive variables was carried out using an automatic procedure by conducting F-test and fitting the final selected model followed by reporting estimates and confidence intervals (Hocking 1976; Øhlenschlæger et al. 2016), and the R was used to conduct the analysis in this study.

**Sensitivity analysis**

Sensitivity coefficient can be calculated as:

$$\varepsilon_x = \frac{\partial y}{\partial x} \times \frac{x}{y}$$  \hspace{1cm} (12)

where $y$ is the dependent variable; in this study, it would be runoff or RC. Here, $\varepsilon_x$ is the sensitivity coefficient of runoff.
or $RC$ to the independent variables, and $x$ is the independent variable. This type of sensitivity analysis has been widely used in hydroclimate studies (Xu et al. 2014; Zhao et al. 2014; Wang & Yu 2015). A sensitivity coefficient of 1 indicates that a 10% increase in the value of independent variable would result in a 10% increase in the value of dependent variable.

RESULTS

Descriptive statistics

Prior to multivariate analysis, univariate statistical descriptions were used to determine the distribution of variables such as topography, climate, soil, and vegetation in the selected catchments for 57 samplings. Minimum, maximum, range, median, mean, standard deviation, and coefficient of variation were used to describe them (Table 1). The drainage area was in the range of 40.96–15,736.29 km², exhibiting a significant scale difference in the spatial scale. The average slope of study catchments was 1.25–6.47°. The overall terrain of the study area was relatively flat on the average level; in fact, the slope of some areas might exceed the given range. Watershed type is usually defined as water-limited ($AI > 1.35$), equitant ($0.76 < AI < 1.35$), and energy-limited ($AI < 0.76$) environments based on dryness index (McVicar et al. 2003). In this study, $AI$ was in the range of 1.3–3.4, indicating that the study area was mainly distributed in the water-limited area. The ranges of $PRE$ and $PE$ were 320.06–744.32 mm/a and 813.4–1,257.4 mm/a, respectively, and showed a lower variability. However, the ranges of $R$ and $RC$ were 5.34–214.39 mm/a and 0.01–0.38, respectively, exhibiting a higher variability. Therefore, during this period, the climatic factors might not be the main factors influencing runoff changes. In contrast, $PAWC$, $LAI$, $ASWC$, as well as topographic factors might play important roles in higher variability.

Relationship between $RC$ and environmental variables

The results of cascade k-means clustering are shown in Figure 2; nine initial k values (2, 3, 4, 5, 6, 7, 8, 9, and 10) were selected. According to the minimum SSI, the studied catchments can be divided into five groups by $RC$. The $RC$ values of each group were significantly different from those of the other groups (Table 1). The distribution of catchments divided by $RC$ is shown in Figure 3. The catchments with a low runoff capacity, group 1 ($RC = 0.027 \pm 0.010$) and group 2 ($RC = 0.069 \pm 0.012$), were mainly distributed upstream or northwest of the Haihe river basin, while the areas with high runoff capacity were mainly distributed downstream.

$RC$ represents the runoff capacity of catchments (Grillone et al. 2014); thus, a high $RC$ is an asset for water resource development, especially in arid and semiarid
Different RCs originate from climate change and by varying the LULC conditions (Viglione et al. 2009; Sriwongsitanon & Taesombat 2014), such as the AI and soil-available water content. Based on this analysis, PCA was conducted, and the environmental variables of the groups were compared.

Figure 4(a) shows the eigenvalues of each principal component. The cumulative proportion of the first two eigenvalues was 70.02%, and they are all higher than the average eigenvalues, indicating that PC1 and PC2 can explain the relationship between RC and environmental factors. In Figure 4(b), the numbers are the RC values, the arrow lengths represent the effect of different environment factors, and the angle between two variable arrows indicates the relationship between the two variables. According to Figure 4(b), the PAWC, AI, precipitation, and NDVI substantially contributed to the distribution of RCs. In addition, NDVI, LAI, ASWC, annual runoff, and precipitation were positively correlated; they were negatively correlated with AI and PAWC. In Figure 4(b), lower RC values are mainly distributed at higher AI and PAWC areas, whereas in the area with a higher precipitation, ASWC, and NDVI, the RCs were relatively higher.

Figure 5 shows a comparison of the environmental factors associated with different groups; as shown in Figure 5, the PE, LAI, and slope were not significantly different among groups ($p > 0.05$). The precipitation in groups 4 and 5 was significantly higher than that in groups 1 and 2 ($p < 0.001$). The NDVI of group 5 was significantly higher than that of other groups ($p < 0.001$), and those of groups 2 and 3 were significantly higher than that of group 1. Moreover, the runoff and ASWC significantly increased with increased RC; however, the AI decreased as RC increased. In summary, the precipitation, AI, ASWC, NDVI, and PAWC differed significantly in catchments with different water yield capacities.

Quantitative analysis of runoff capacity and environmental factors

PCA only analyzes the relationship between environmental variables and research objects qualitatively (Legendre &
Legendre 1998). Regression and sensitivity analysis define the relationship between RC and specific environmental variables. Using stepwise regression analysis, RC can be expressed as follows:

$$RC = 0.261 \log_{10} \text{PRE} + 0.488 \text{ASWC} + 0.017 \text{LAI} - 0.01 \text{PAWC} (r^2 = 0.729, p < 0.001)$$

Figure 3 | Distribution of runoff capacity in Haihe mountainous area.
The regression equations of $R$ for different groups and their sensitivity coefficients are shown in Tables 2 and 3, respectively. For this analysis, groups 4 and 5 were merged because the catchment amount of each group was too small to serve as a basis for regression analysis.

According to the regression equation and sensitivity coefficients of $RC$, $RC$ was controlled by precipitation, ASWC, LAI, and PAWC; a 10% increase in precipitation, ASWC, and LAI would result in 26.7%, 6.3%, and 2.2% increases in $RC$, respectively, whereas a 10% increase in PAWC would decrease $RC$ by 3.3. This result is similar to the PCA result. Both NDVI and LAI represent the amount of vegetation, and they have similar meanings in the water cycle process. The use of LAI in the equation was just out of

---

The regression equations of $R$ for different groups and their sensitivity coefficients are shown in Tables 2 and 3, respectively. For this analysis, groups 4 and 5 were merged because the catchment amount of each group was too small to serve as a basis for regression analysis.

According to the regression equation and sensitivity coefficients of $RC$, $RC$ was controlled by precipitation, ASWC, LAI, and PAWC; a 10% increase in precipitation, ASWC, and LAI would result in 26.7%, 6.3%, and 2.2% increases in $RC$, respectively, whereas a 10% increase in PAWC would decrease $RC$ by 3.3. This result is similar to the PCA result. Both NDVI and LAI represent the amount of vegetation, and they have similar meanings in the water cycle process. The use of LAI in the equation was just out of
Table 2 | Regression formulations of each group (mean ± standard deviation)

<table>
<thead>
<tr>
<th>Group</th>
<th>RC</th>
<th>Regression equation</th>
<th>$r^2$</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.027 ± 0.010</td>
<td>$R = 6.506 \text{slope} - 10.353 \text{LAI} + 2.306$ (14)</td>
<td>0.912</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.069 ± 0.012</td>
<td>$R = 0.170 \text{PRE} - 0.042 \text{PE} - 6.14 \text{PAWC} + 12.545$ (15)</td>
<td>0.925</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>3</td>
<td>0.131 ± 0.021</td>
<td>$R = 0.284 \text{PRE} - 0.086 \text{PE} - 20.54 \text{PAWC} + 36.518$ (16)</td>
<td>0.923</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>4</td>
<td>0.226 ± 0.021</td>
<td>$R = -42.88 + 318.325 \text{NDVI}$ (17)</td>
<td>0.587</td>
<td>0.028</td>
</tr>
<tr>
<td>5</td>
<td>0.326 ± 0.043</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$R$, mean annual runoff; PRE, mean annual precipitation.

Table 3 | Sensitive coefficients $R$ of RC in different groups (mean ± standard deviation)

<table>
<thead>
<tr>
<th>Equation number</th>
<th>Sensitive coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>(13)</td>
<td>$\epsilon_{(\text{LAI})} = 0.63 \pm 0.36$</td>
</tr>
<tr>
<td>(14)</td>
<td>$\epsilon_{(\text{slope})} = 1.78 \pm 0.24$</td>
</tr>
<tr>
<td>(15)</td>
<td>$\epsilon_{(\text{PAWC})} = -1.05 \pm 0.31$</td>
</tr>
<tr>
<td>(16)</td>
<td>$\epsilon_{(\text{PE})} = 0.22 \pm 0.36$</td>
</tr>
<tr>
<td>(17)</td>
<td>$\epsilon_{(\text{NDVI})} = 1.26 \pm 0.58$</td>
</tr>
</tbody>
</table>

Table 4 | Results of runoff sensitivity to climate variable comparison among relevant studies and this study

<table>
<thead>
<tr>
<th>Studies</th>
<th>$\epsilon_{(\text{PRE})}$</th>
<th>$\epsilon_{(\text{PE})}$</th>
<th>Based hypothesis</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang et al. (2013)</td>
<td>2.8</td>
<td>-1.4</td>
<td>Zhang et al. (2001)</td>
<td>Hydrological sensitive method</td>
</tr>
<tr>
<td>Xu et al. (2014)</td>
<td>2.33</td>
<td>-1.35</td>
<td>Budyko hypothesis</td>
<td>Water-energy balance equation</td>
</tr>
<tr>
<td>Zhao et al. (2014)</td>
<td>3.5</td>
<td>-1.1</td>
<td>Zhang et al. (2001)</td>
<td>Hydrological sensitive method</td>
</tr>
<tr>
<td>Wang et al. (2015)</td>
<td>3.21</td>
<td>-2.21</td>
<td>CCUW</td>
<td>Climate elasticity method</td>
</tr>
<tr>
<td>Current study</td>
<td>2.36</td>
<td>-1.29</td>
<td>Zhang et al. (2001)</td>
<td>Multivariate statistics</td>
</tr>
</tbody>
</table>

Note: CCUW is the climate change impact hypothesis proposed by Renner et al. (2012).
variables is an effective way to regulate runoff. As the result in Figure 5 shows, precipitation, AI, ASWC, NDVI, and PAWC have the greatest influence on runoff in catchments with different water yield capacities. However, climate conditions such as precipitation and AI are often beyond our control, whereas parameters such as NDVI, ASWC, and PAWC can be adjusted by changing the type of LULC (Wang & Yu 2014). According to the above-mentioned analysis, the increase on NDVI and ASWC and decrease in PAWC can increase the RC values in semiarid areas. A higher NDVI reflects a higher vegetation cover rate (Farrar et al. 1994; Thana-pura et al. 2007). ASWC quantifies the soil structure (Zhou et al. 2005) and can be improved by plant measures; the PAWC is controlled by vegetation types: woodlands have higher PAWC than shrublands and grasslands (Zhang et al. 2001), whereas different plant species have different PAWC (Sun et al. 2008). Thus, a reasonable vegetation allocation is the key to regulate runoff. In summary, to increase RC, namely, to increase the water yield of Haihe river mountain area, especially the upper stream of the region, it is important to increase the vegetation cover, and vegetation restoration projects should select species with a lower PAWC preferentially, for example, restoration of shrubs and herbs rather than large-scale afforestation.

Some studies argued that a higher RC of catchments might not indicate a higher production of water yield, because some approaches aimed at increasing the RC such as afforestation or deforestation might change the number of times of moisture recycles within a catchment and hence decrease the precipitation, particularly in semiarid and arid areas (Savenije 1996; Berbet & Costa 2005). In a recent study, rainfall was shown to be negatively correlated with PAWC, indicating that it should be negatively correlated with forest areas based on Zhang’s hypothesis (Zhang et al. 2001). However, Zhang’s hypothesis was not applicable to every catchment (Sun et al. 2008), and further studies are required for more precise conclusions. Thus, it is suggested to improve soil structures and adjust plant species to increase the RC and obtain a higher water yield.

Sensitivity of runoff to climate change

The Haihe river basin is the water source of Beijing–Tianjin–Hebei region; thus, much research on the relationship between runoff and climate change has been carried out on this area (Bao et al. 2012a; Wang et al. 2013; Xu et al. 2014; Zhao et al. 2014; Wang & Yu 2015). In Table 4, some previous study results for streamflow sensitivity to precipitation and PE in the Haihe river basin in recent years are compared with the results obtained in this study. The results of Wang et al. (2015) are based on the climate change hypothesis (Renner et al. 2012; Wang & Yu 2015), and this hypothesis was applied to the areas where the AI is close to 1. In the Haihe river basin, the AI is 2.03 ± 0.49; under such conditions, the results are overestimated (Renner & Bernhofer 2012). Specifically, the average $\epsilon_{\text{PRE}}$ should be less than 3.21, and the $\epsilon_{\text{PE}}$ should be higher than −2.21. Moreover, the sum of $\epsilon_{\text{PRE}}$ and $\epsilon_{\text{PE}}$ should be close to 1 (Peel et al. 2010). Thus, the results of Wang et al. (2015) and Zhao et al. (2014) might have some deviation due to the uncertainty of PAWC parameter. Based on the results of Zhang et al. (2001), the PAWC should be 2.0 for trees and 0.5 for shrubs and grasslands; however, the actual value of this parameter is affected by LULC, vegetation species, and forest structures. In this study, the PAWC of each catchment was calculated using the measured parameters, and thus the uncertainty was reduced. Xu et al. (2014) estimated the sensitivity of 33 catchments in Haihe river basin based on the Budyko hypothesis (Budyko 1974) and water-energy balance equation (Yang et al. 2008). Although their results were very close to those obtained in this study, the parameter n used by them to represent LULC might also cause uncertain error because of its ambiguity. In general, the results of this study are likely to be more accurate than those obtained in previous studies, and the sensitivity of runoff to PRE and PE was quantified.

Sensitivity of runoff to LULC conditions

In addition to climate factors, runoff is also affected by LULC conditions. Few studies have reported the sensitivity of runoff to LULC conditions. This is because, on one hand, LULC conditions are a comprehensive feature that is...
difficult to quantify with an exact index; on the other hand, few precise models or equations have been developed to quantify it. In this study, regression equations provide the \( \text{slope}, \text{LAI}, \text{PAWC}, \) and \( \text{NDVI} \) to represent LULC conditions in different runoff capacity catchments (Table 2). Xu et al. (2014) used the water balance equation to estimate a landscape parameter \( n \) \((-2.15 \pm 0.42)\) to represent the LULC condition; the values obtained using that method were higher than those obtained in this study. Zhao et al. (2014) used the area changes of different LULC to estimate the sensitivity; those results indicated that the runoff sensitivity coefficient was \(-2.0\) in the forest area and \(1.55\) in the sloped land area. This result is similar to those obtained in our analysis, where the sensitivity to \( \text{PAWC} \) and \( \text{slope} \) are \(-0.52 \pm 0.51\) and \(1.78 \pm 0.24\), respectively, and \( \text{PAWC} \) increases with the increase in forest.

Uncertainty and limitations

In this study, the \( r^2 \) of \( RC \) equation is 0.792; hence, it is reasonable that the regression and sensitivity analyses accurately quantify the relationship between \( RC \) and environmental variables. In the \( R \) equations of different runoff capacity catchments, the \( r^2 \) values of Equations (14), (15), and (16) are greater than 0.9, and the model prediction should be precise, whereas the \( r^2 \) of Equation (17) is only 0.587. This indicates that in the high runoff capacity catchments, the runoff prediction results have a relatively large error, and the \( \epsilon_{(\text{NDVI})} \) might not be precise. All the parameters used in this study were measured except the \( \text{ASWC} \), which was obtained from an empirical equation (Equation (1)) (Zhou et al. 2005) using the proportions of soil components. The soil component data were obtained from quadrat study; because of the heterogeneity and nonlinear distribution of soil, the calculated \( \text{ASWC} \) might be different from the actual situation. The \( \text{PAWC} \) values used in this study were obtained using Zhang’s empirical model (Zhang et al. 2001) and might also have an uncertain effect on the result (Pike 1964; Budyko 1974).

Another notable issue is that the environmental variables selected in this study were not completely independent of each other. For example, \( \text{PE} \) can be affected by \( \text{PAWC} \), and the \( \text{AI} \) is decided by \( \text{precipitation} \) and \( \text{PE} \). This might affect the regression results and increase uncertainty. However, the variables selected by stepwise regression in the equations are not significantly correlated with each other according to the Pearson coefficient; thus, the errors have been controlled in an acceptable range. In addition, the recommendations proposed in this study are only for the regulation of annual runoff but may not be applicable to seasonal events such as floods or low flow, which is very important in disaster prediction and water resources assessment. The response mechanism to the environment in seasonal runoff is different from annual runoff and worthy of further study.

**CONCLUSIONS**

In this study, PCA and elasticity methods based on regression equations are proposed to quantify the sensitivity of annual runoff and runoff capacity to environmental variables.

First, the runoff capacity is mainly controlled by precipitation, vegetation type, soil texture, and land cover conditions. Low runoff capacity catchments have relatively higher \( \text{AI} \) and \( \text{PAWC} \) and lower \( \text{ASWC} \), whereas high runoff capacity catchments usually have relatively higher \( \text{ASWC} \) and lower \( \text{AI} \) and lower \( \text{PAWC} \). Improvement of the soil structure and vegetation composition by proper vegetation restoration and allocation, especially in low runoff capacity catchments, is suggested to increase the water yield in the Haihe river basin.

Second, the runoff in different runoff capacity catchments is mainly controlled by different environmental variables. In low runoff capacity catchments \( (RC < 0.05) \), the runoff is positively correlated with \( \text{slope} \) and negatively correlated with \( \text{LAI} \). A 10% change in mean \( \text{slope} \) and \( \text{LAI} \) would lead to 17.8% and \(-10.5%\) changes in runoff, respectively. In middle runoff capacity catchments \( (0.05 < RC < 0.2) \), the runoff is positively correlated with \( \text{precipitation} \) and negatively correlated with \( \text{EP} \) and \( \text{PAWC} \). Every 10% change in \( \text{precipitation}, \text{PE}, \) and \( \text{PAWC} \) would result in 23.6%, \(-12.9\%), and \(-5.1\%) \) changes in \( \text{annual runoff} \), respectively. Finally, in catchments with a high runoff capacity \( (RC > 0.2) \), the \( \text{annual runoff} \) is related to \( \text{NDVI} \), and every 10% increase in \( \text{NDVI} \) would lead to a 12.6% increase in \( \text{annual runoff} \).
ACKNOWLEDGEMENTS

This research was supported by the National Key R&D Program of China (2016YFC0500802) and Beijing Municipal Education Commission of China (Project No. CEFF-PXM2018-014207-000043). Qiufen Zhang composed the paper. Qiufen Zhang and Jiakai Liu analyzed the data. Lihua Chen and Xinxiu Yu supervised this work and they contributed the same to this work. The authors declare no conflict of interests. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

REFERENCES


Margaritis, D. & Thrun, S. 1999 Bayesian Network Induction via Local Networks. [Research paper]/Carnegie Mellon University, School of Computer Science, School of Computer Science Carnegie Mellon University, Pittsburgh, PA.


First received 5 December 2018; accepted in revised form 15 March 2019. Available online 30 April 2019