Quantifying the contribution of flood intensity indicators with the projection pursuit model
Lina Wang, Yanqing Lian and Xiaohong Chen

ABSTRACT

Identifying the various factors that affect the intensity of a flood event, such as its duration and volume, is essential for strategic planning and flood management. Further, quantifying the impacts of these major factors on flood intensity using the contribution rate is essential, but technically challenging. In this study, the authors have adopted the projection pursuit model to quantify the contribution rates of peak flood stage and peak flood discharge, flood duration, and total flood volume (the maximum 12-, 24-, and 72-hour flood volumes) in the Wujiang River in Southern China. This study showed that peak flood discharge and total flood volume were the two dominant factors impacting flood intensity. Although flood duration can be a major factor for some flood events, it contributed the least to flood intensity for most of the historic flood events studied. Likewise, the maximum 24-hour and 72-hour flood volumes contributed little to flood intensity. Findings from this study not only demonstrated the successful adoption of the projection pursuit model for contribution rates, but also provided critical information for planning and managing the regional hydraulic resources in the Wujiang River.

Key words | contribution rate, flood intensity, key driving factors

INTRODUCTION

Floods can be considered catastrophes, causing lost lives and properties, and natural environments that are damaged beyond repair (Kiss et al. 2014). The severity of damage from a flood event can be assessed by determining the flood intensity. Factors that contribute to the intensity include the peak flood discharge, volume and duration, and flow velocity. Ahmed & Mirza (2000) suggested using the flood intensity index to account for multiple factors such as the flood duration, destructive capacity, and the area affected by floods. Many studies have examined flood intensity from different perspectives. For example, Adamowski (2000) pointed out that the peak flood discharge is the most important feature of a flood event because the maximum discharge rather than the average discharge is the main concern. French & Miller (2011) stated that the flood intensity is determined by flow depth and velocity. Swades & Surajit (2012) used three variables, including flood stage, flood frequency, and the flood stagnation period, to evaluate the flood intensity. Wang et al. (2015) proposed using different indicators to rank flood intensity. Javelle et al. (2005) pointed out that flood severity is not only defined by its peak value, but also by its volume and duration. An extensive amount of research has examined the impact of climate factors on flood intensity (Droegemeier et al. 2000; Changnon et al. 2001; O’Connor & Costa 2003, 2004). Flood volume may have the largest impact on flood intensity for large river basins, although the impacts of multiple factors on flood intensity would vary with respect to hydrologic, hydraulic, and climate conditions.
While peak flood discharge, peak stage, and flood volume are regarded as the three major contributors to flood intensity, for practical purposes the peak flood discharge and its corresponding stage are often used to quantify the flood intensity (Jumadar et al. 2008; Singh & Singh 2011; Swades & Surajit 2012; Ward & Paulus 2013). The impacts of peak flood discharge and discharge on a flood event can be quantified by using contribution rates, or the weight of the variables that contributed to flood intensity. The contribution rate is an important, comprehensive value used for flood control. Quantifying contribution rates of factors that could significantly impact flood intensity not only provides information to identify the dominating factors for flood events, but also is crucial for flood risk assessment and water resource planning (Bai et al. 2015). The evaluation of a contribution rate should be as accurate as possible given the implications for designing hydraulic structures, water management, and estimation of scour at a hydraulic structure (Browne & Hoyt 2000; Carolan 2007; Shabri et al. 2014).

Although the principal component analysis (PCA) model has been used to calculate the contribution rate (Grimalt et al. 1990; Domingo et al. 2000; Kharmouz et al. 2013; Tomašić et al. 2014; Miloudi et al. 2017), the PCA method itself can only quantify the contribution rate of principal components. The principal components often consist of several variables, thus it is difficult to determine the contribution rate of a single or a specific variable (Hu et al. 2007; Noori et al. 2010, 2011). The projection pursuit algorithm is a statistical technique for finding the most ‘interesting’ possible projections in multidimensional data. The idea of projection pursuit is to locate the projection or projections from high-dimensional space to low-dimensional space that reveal the most details about the structure of the data set. It has been widely used in many fields such as multivariable prediction, cluster, water resources assessment, environmental protection, etc. (Kennedy & Basu 2000; Wang et al. 2004; Rajeevan et al. 2007; Chi & Dong 2009; Sun et al. 2012; Balke et al. 2013; Zhao et al. 2014; Jie et al. 2015; Durocher et al. 2016). Wang & Zhang (2010) developed the projection pursuit dynamic cluster model, which is based on the projection pursuit principle, to analyze the carrying capacity of water resources in China. Huang & Zhang (2011) proposed a flood disaster classification assessment method, which adopted the multi-swarm system particle swarm optimization method to optimize the parameters of the projection index functions. Liu et al. (2016) evaluated water resource resilience based on a projection pursuit classification model, with the help of artificial fish-swarm algorithm (AFS), which optimizes the projection index function.

The main objective of this study was to adopt the projection pursuit model to quantify contributions from multiple indicators, such as peak level, peak flood discharge, flood volume, flood duration, and the maximum t-hour volume. This study used historic flood event data to demonstrate the application of this method in the Wujiang River.

**STUDY AREA AND DATA**

The Wujiang River is one of the largest tributaries of the Beijiang River in the Pearl River basin in southern China. The Wujiang River basin is located south of the Wuling mountains between the latitudes of 24°46’ to 25°41’ N and longitudes of 112°23’ to 115°36’ E (Figure 1). Its total drainage area is about 7,097 km² with a mainstream spanning 260 km. The climate of the Wujiang River is dominated by the southwest and southeast Asian monsoons, which results in high humidity in the summer and an uneven distribution of precipitation throughout the year. The mean annual rainfall within the basin is approximately 1,450 mm. The Lishi stream-gaging station is located near the mouth of the river, monitoring conditions for 98.2% of the drainage area, or 6,976 km² of the Wujiang River basin (Figure 1). The stage and flow data from 1955 to 2007 at the Lishi station were provided by the Shaoguan Branch of the Guangdong Provincial Bureau of Hydrology. According to this agency, the discharge data were computed from the direct flow velocity measurements at the gaging station and the water level was measured by using a stage gage at the same location. The flood event with the largest peak flood discharge in each calendar year was selected for the study analysis. The method for extracting flood events was introduced in detail by Wang et al. (2015).
METHODOLOGIES

Indicators of flood events

The peak flood discharge, peak flood stage, total flood volume, and flood duration are often used as indicators for flood intensity. In addition, the maximum 12-h, 24-h, and 72-h flood volumes were also used as major indicators in this study. Wang et al. (2018) provided calculation details for the flood duration, the maximum 24-h and 72-h flood volume, the total flood volume, and other factors. Wang et al. (2018) showed that the average flood duration in the Wujiang River basin was 3 days.

Calculation of contributing rate

The projection pursuit model is designed to determine the projection from high-dimensional spaces to low-dimensional spaces that reveal the most details about the structure of the data set. Details of the projection pursuit model for the contribution rate of indicators can be found in Friedman & Tukey (1974). The steps of the projection pursuit model for contribution rate of indicators are as follows (Friedman & Tukey 1974; Jin et al. 2005; Huang & Zhang 2011; Zheng & Lin 2012).

The projection pursuit model is made to convert the p-dimensional data \( \{x_{ij}\}_{j=1}^{p} \) for \( i=1,2,\ldots,m \) into one-dimensional data \( \{Z_i\} \), referred to as the projection value with the projection direction \( a'=(a_1,a_2,\ldots,a_p) \) (Qin & Lin 2012). According to the projection component value \( Z_i \) of the indicator \( j \)-th of \( i \)-th flood process, the indicator contribution rate can be expressed by \( Z_{ij}/Z_i \), where \( Z_i \) synthesizes the projection value, and \( Z_{ij} \) is the projection component value receptivity (Wang et al. 2004).

Step 1. Establishment of indicator

Given the matrix \( X_{mp}^0 = \begin{bmatrix} x_{11}^0 & x_{12}^0 & \cdots & x_{1p}^0 \\ x_{21}^0 & x_{22}^0 & \cdots & x_{2p}^0 \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1}^0 & x_{m2}^0 & \cdots & x_{mp}^0 \end{bmatrix} \) present the indicator values of flood process samples, where \( p \) is
the number of flood process indicator variables, \( m \) is the number of indicator \( j \)-th of \( i \)-th flood process. In this study, \( m = 53, p = 7 \) and

\[
x_{i1}^0 = \{P_{11}, P_{21}, \ldots, P_{i1}, \ldots, P_{m1}\},
\]

\[
x_{i2}^0 = \{H_{12}, H_{22}, \ldots, H_{i1}, \ldots, H_{m2}\},
\]

\[
x_{i3}^0 = \{(V_{12})_{13}, (V_{12})_{23}, \ldots, (V_{12})_{i3}, \ldots, (V_{12})_{m3}\},
\]

\[
x_{i4}^0 = \{(V_{24})_{14}, (V_{24})_{24}, \ldots, (V_{24})_{i4}, \ldots, (V_{24})_{m4}\},
\]

\[
x_{i5}^0 = \{(V_{72})_{15}, (V_{72})_{25}, \ldots, (V_{72})_{i5}, \ldots, (V_{72})_{m5}\},
\]

\[
x_{i6}^0 = \{V_{16}, V_{26}, \ldots, V_{i6}, \ldots, V_{m6}\},
\]

\[
x_{i7}^0 = \{T_{17}, T_{27}, \ldots, T_{i7}, \ldots, T_{m7}\}.
\]

### Step 2. Normalization of data

Units for the seven variables selected in this study are different. Four types of data pretreatment methods including the mean centering, the differentiation, normalization, and auto-scaling can be used to eliminate dimensions of different process variables (Amrhein et al. 1996). The normalization method was adopted in this study and is given below:

\[
x_{ij}^0 = \frac{x_{ij0} - x_{j\min}^0}{x_{j\max}^0 - x_{j\min}^0}
\]

where \( x_{j\min}^0 \) and \( x_{j\max}^0 \) are the initial minimum and maximum values of the \( j \)-th indicator, respectively.

After normalization, the matrix \( X_{mp}^0 \) could be replaced by matrix

\[
X_{mp} = \begin{bmatrix}
x_{11}^0 & x_{12}^0 & \cdots & x_{1p}^0 \\
x_{21}^0 & x_{22}^0 & \cdots & x_{2p}^0 \\
\vdots & \vdots & \ddots & \vdots \\
x_{m1}^0 & x_{m2}^0 & \cdots & x_{mp}^0 
\end{bmatrix}
\]

### Step 3. Analysis of linear projection

The \( p \)-dimensional data \( (X_{mp}) \) can be converted into one-dimensional data \( (Z_i) \) of the projection direction by the following formula:

\[
Z_i = \sum_{j=1}^{p} a_{ij} x_{ij} \quad (i = 1, 2, \ldots, m, j = 1, 2, \ldots, p)
\]

where \( Z_i \) is the synthesizing projection value.

### Step 4. Establishment of objective function

The projection index function can be constructed as follows:

\[
Q_a = S_a \cdot d_a
\]

where \( S_a \), which equals the standard variance of the projection value \( Z_i \), is the distance function between clusters. \( d_a \) is the density function in clusters, which means that the local density of the projection value is \( Z_i \).

\[
S_a = \sqrt{\frac{1}{m-1} \sum_{i=1}^{m} (Z_i - \bar{Z})^2}
\]

\[
d_a = \frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{m} (R - r_{ik}) f(R - r_{ik}), \quad (i, k = 1, 2, \ldots, m)
\]

where \( \bar{Z} = 1/m \sum_{i=1}^{m} Z_i \), \( R = 0.1S_a \). \( R \) is the window radius of density function in clusters, which is related to data characteristics, while the selection of an average number of window projection points should not be too small, and the average deviation of glide should not be too large. \( r_{ik} \) is the absolute distance between the random projection eigenvalues, and \( r_{ik} = |Z_i - Z_k| \). \( f(R - r_{ik}) \) is the unit step function, when \( R > r_{ik}, \) \( f(R - r_{ik}) = 1 \), or vice versa, \( f(R - r_{ik}) = 0 \).

### Step 5. Optimization of objective function

The projection index function proposed by Friedman & Tukey (1974) can be written as:

\[
\max (Q_a) = S_a \cdot d_a \quad \sum_{j=1}^{p} a_{ij}^2 = 1
\]
where \( a \) is the projection axis vector, also called the projection vector.

In this study, the artificial fish-swarm algorithm (AFS) was adopted to find the optimizing value. AFS was used to solve the non-linear optimization problem of the projection pursuit model. An important difference between AFS and other swarm intelligence algorithms is that the AFS can search for the global optimum effectively and has an adaptive ability for the search space. According to Yazdani et al. (2010) and Mehdi et al. (2012), the AFS algorithm is one of the best optimization methods among the swarm intelligence algorithms. The AFS is a simulation behavior and population-based optimization method, which was initially developed by Tu (1999) and has recently been widely used in practice (Yazdani et al. 2012; Reza & Bojnourd 2014).

**Step 6. Calculation of contribution rate**

The contribution rate can then be computed by using the following expression:

\[
C_{ij} = \frac{Z_{ij}}{Z_i} = \frac{a_i x_{ij}}{Z_i}, \quad (i = 1, 2, \ldots, m; \; j = 1, 2, \ldots, p)
\]

in which, \( a_i \geq 0, \; (j = 1, 2, \ldots, p) \) to make sure that all of the optimization projection directions are positive.

**RESULTS AND DISCUSSION**

**Flood intensity**

Shown in Figure 2 are the relationships among the seven variables or indicators selected in this study. It can be seen, in general, that the flood peak discharges correlate well with flood peak stage, the 12-hour and the 36-hour flood volumes but not very well with the 72-hour flood volume and not well with the total flood volume and flood duration. Peak flood stages also correlate well with the 12-hour and 36-hour flood volumes but not well with the 72-hour and total flood volumes and not with the flood duration either. Examination of data show that 1975 was the seventh largest flood event according to the peak flood discharge but was the tenth largest flood in terms of the flood peak stage and 1993 flood was the eighth based on the peak flood discharge but was the seventh in terms of the flood peak stage. Figure 2 and data examination show that flood intensity is characterized by multiple facts and further justifies the use of several variables in this analysis.

In order to facilitate comparison and understand the difference between different flood events, cluster analysis has been used to compare the intensity of various flood events (Bhaskar & O’Connor 1989; Dong et al. 2007; Cheng et al. 2009; Latt et al. 2015; Wang et al. 2015). Clustering is a method for dividing scattered groups of data into several groups. Thus, the object of flood clustering is to sort flood into groups, so that the flood characteristic is similar between members of the same cluster and dissimilar between members of different clusters. Clustering analysis of flood intensity aims at the identification of groups of floods with common characteristics. Factors that affect flood intensity contain different units. Therefore, it is impossible to compare them by using their absolute values. Instead, dimensionless values were processed using Equation (1), introduced earlier for the cluster analysis.

Initial values need to be set in order to compute the projection value of each flood event. In this study, the initial population size was set at 30, the artificial fish perception scope was visible as 0.3, the largest number of tests of each move try number was 20, the crowd fact was 0.3, the maximum number of iterations was 60, and the moving step length was 0.1. The optimization projection direction \( a \), obtained from the AFS method, would be:

\[
a = (0.369 \; 0.444 \; 0.398 \; 0.2830 \; 0.301 \; 0.565 \; 0.122).\]

From the optimization results, the projected characteristic value (\( Z_i \)) and the cluster results can be calculated. Larger projection values indicate higher flood intensities.

Figure 3 shows that the projection results matched very well with actual floods in the Wujiang River. Analyses from the projection pursuit model also indicated that the 2006 flood was the largest flood ever recorded in the Wujiang River, which was consistent with the previous analysis by Wang et al. (2015).
Contribution rate

The calculation of contribution rate proposed in this study was the ratio of projection value for indicator to the projection value of all indicators. The projection value for a single indicator is the product of projection direction and its normalized value of that indicator. The optimized values for the projection direction showed that the peak flow stage and the total flood volume had relative larger values which ultimately resulted in higher contribution rates for both among the seven indicators used in the analysis. Figure 4 shows the contribution rates of these seven impact factors in the 53-year period. It is apparent that the contribution rates for the same indicator were different among different flood events. The annual contribution rate for the peak flood discharge, peak stage, maximum 12-h, 24-h, and 72-h volume showed a slight increasing trend, whereas the annual contribution rate for the total flood volume and flood duration presented a slight decreasing trend, as shown in Figure 4(f) and 4(g), respectively.

Figure 5 shows that the peak flood discharge stage was consistently one of the two major factors for all years except for 1963. The total volume of floods was the other major contributor for all years other than 1963, 1979, and
1991, during which the duration of flood events weighed more than the total flood volume. The extended length of the flood event was the major cause of flooding in 1963. The 1963 flood occurred in late March and early April, with a lowest peak flood discharge of 630 m$^3$/s. While this study did not show that the peak flood discharge was the major contributing factor or equally as important as the peak stage, the total discharge of a flood event was consistently the major contributor.

**Characteristics of contribution rates**

Figure 6 shows the indicators and their maximum contribution rates in each calendar year. Peak flood stage had the largest contribution rate for 29 of the 53 flood events, or 55% of the flood events. Total flood volume had the largest contribution rates for 21 of 53, or 40% of the flood events. In about 6% of the years, the flood duration had the largest contribution rates. Peak flood discharge, the
maximum 12-h, 24-h, and 72-h volume never had the largest contribution rate in all 53 years.

Although interest and concern are often focused on factors that contribute the most to flood events, it is still useful to determine the factors that have had the least impact on flood events. Figure 7 shows that the flood duration had the lowest contribution rate for 28 out of 53 flood events, or 53%. The maximum 72-h and 24-h flood volume had the lowest contribution rates for 14 and 10 flood events, respectively.

A significant point is that the results, as presented in Figures 6 and 7, show that the peak level and the total flood volume are the dominant factors of high flood intensity, although flood duration can become a major factor.

Figure 5 | The characteristics of the contribution rate. P: peak flood discharge; H: peak level; V12: the maximum 12-h volume; V24: the maximum 24-h volume; V72: the maximum 72-h volume; V: the total flood volume; T: flood duration.

Figure 6 | The indicator corresponding to the maximum contribution rate in the same flood process. P: peak flood discharge; H: peak level; V12: the maximum 12-h volume; V24: the maximum 24-h volume; V72: the maximum 72-h volume; V: the total flood volume; T: flood duration.

Figure 7 | The indicator corresponding to the minimum contribution rate in the same flood process. P: peak flood discharge; H: peak level; V12: the maximum 12-h volume; V24: the maximum 24-h volume; V72: the maximum 72-h volume; V: the total flood volume; T: flood duration.
for some flood events. The flood duration apparently had the lowest contribution for flood intensity for most of the historic flood events, although it was a major factor for 6% of the flood events. Wujiang River is a typical mountainous river with a rapid flow, large hydraulic gradient, and a quick rise and fall of flood events. This study showed that the peak flood stages and the total flood volume are the two major factors in the decision-making process for planning and management of the regional hydraulic resources.

The maximum 24-h and 72-h flood volume contributed the least at 26% and 19% of flood events, respectively, and did not occur as a major contributor to the flood intensity; thus, these volumes may not need to be considered for flood management in the Wujiang River.

**Assessment of the key factors**

Table 1 lists the statistical median, maximum, and minimum of the seven factors used in the analysis. For the 53 flood events, the contribution rates of the peak flood stages ranged from 0% in 1963 to 34.29% in 1991, and the contribution rates of the corresponding peak flood discharges ranged from 0% in 1963 to 21.28% in 1977. The minimum and maximum contribution rate values of flood duration are 0% in 2007 and 72.26% in 1963, respectively. The contribution rate of the total flood volume varied between 0% in 1991 and 40.70% in 1997. The minimum and maximum contribution rates from the maximum 12-h volume are 0% in 1963 and 17.75% in 1993, respectively. Results presented in Table 1 clearly showed that the flood intensity needs to be quantified by multiple factors.

### CONCLUSIONS

Quantifying the factors that would significantly contribute to flood intensity is crucial for flood protection and management. Variables involved in flood events are hardly independent. The analysis results might be slightly different if the number of variables chosen for the analysis were different. Nonetheless, it is expected that the relative significance in terms of contribution rates for the selected variables can be quantified. This study was based on the commonly used peak flood discharge, the peak flood stage, total flood volume, and four other variables including the flood duration, and 12-h, 24-h, and 72-h flood volumes.

This study has successfully used the projection pursuit model to estimate contribution rates of the seven factors that would have affected flood intensity in the Wujiang River by using 53 years of hydrologic data from the Lishi station. The application of the projection pursuit model showed that the 2006 flood was the largest flood ever recorded in the Wujiang River. Study results also showed that indicator contribution rates for the same indicator were different in different flood events. For the 53 flood events in the Wujiang River, the contribution rate of peak flood stage ranged from 0% in 1963 to 34.29% in 1991, and contribution rates of their corresponding peak flood discharges ranged from 0% in 1963 to 21.28% in 1977.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>H</th>
<th>V</th>
<th>V12</th>
<th>P</th>
<th>T</th>
<th>V24</th>
<th>V72</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ave_Cr(%)</td>
<td>22.24</td>
<td>22.18</td>
<td>13.74</td>
<td>13.16</td>
<td>10.11</td>
<td>9.54</td>
<td>9.03</td>
</tr>
<tr>
<td>δ</td>
<td>23.40</td>
<td>20.67</td>
<td>14.26</td>
<td>13.45</td>
<td>8.28</td>
<td>9.81</td>
<td>9.15</td>
</tr>
<tr>
<td>X_min</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>X_max</td>
<td>34.29</td>
<td>40.70</td>
<td>17.75</td>
<td>21.28</td>
<td>72.26</td>
<td>12.80</td>
<td>13.81</td>
</tr>
</tbody>
</table>

Ave_Cr: the mean contribution rate. P: peak discharge; H: peak level; V12: the maximum 12-h volume; V24: the maximum 24-h volume; V72: the maximum 72-h volume; V: the total flood volume; T: the flood duration.

The ranks are given in descending order by the mean contribution rate. δ = median; X_min = minimum value; X_max = maximum value.
This study has shown that peak flood stage and the total flood volume are the dominant factors affecting high flood intensity. Although flood duration can become a major factor for some flood events, it contributed the least to flood intensity for most of the historic flood events. Wujiang River is a typical mountainous river with a rapid flow, large hydraulic gradient, and quick rise and fall of flood events. The Le Chang Valley Hydro-junction project, with a storage capacity of $3.74 \times 10^8$ m$^3$ in 2011, was constructed in the territory of Shaoguan City. In order to better control the flood intensity and risk, decisions are based not only on peak levels, but also on the total flood volume in the Wujiang River. The results are expected to attract more attention from the Le Chang Valley Hydro-junction project regarding flood intensity control and flood risk management.

The maximum 24-h and 72-h flood volumes contributed the least to 26% and 19% of the flood events, respectively, and never occurred as a major contributor to flood intensity; thus, these factors may not even need to be considered for flood management in the Wujiang River. It is worth noting that although the peak flood discharges had good correlation with peak flood stages, the contribution rates for all 53 flood events were much smaller than the peak flood stages had shown. Even though reasons for the low contribution rates of peak flood discharges remain to be explored in future studies, this study showed that peak flood stages and the total flood volume would be the two major factors for the decision-making process in managing regional hydraulic resources.

ACKNOWLEDGEMENTS

This research is financially supported by the National Natural Science Foundation of China (Grant nos. 41501021, 51210013, 51479216), the Public Welfare Project of Ministry of Water Resources (Grant nos. 201201094, 201301002-02), and the project from Guangdong Science and Technology Department (Grant no. 2010B050500010), the project from Guangdong Water Resources Department (Grant no. 2009-39), Key cultivation projects in key reform areas of South China Normal University. Thanks go to Mrs Lisa Sheppard of the Prairie Research Institute at the University of Illinois at Urbana-Champaign for carefully editing the manuscript. The authors would like also to express sincere thanks to the Editor-in-Chief, Professor Chong-Yu Xu, and the anonymous reviewers for their advice for the improvement of this paper.

REFERENCES


First received 3 March 2017; accepted in revised form 18 July 2017. Available online 13 September 2017.