A fuzzy inference method based on association rule analysis with application to river flood forecasting

Chi Zhang, Yilun Wang, Lili Zhang and Huicheng Zhou

ABSTRACT

In this paper, a computationally efficient version of the widely used Takagi-Sugeno (T-S) fuzzy reasoning method is proposed, and applied to river flood forecasting. It is well known that the number of fuzzy rules of traditional fuzzy reasoning methods exponentially increases as the number of input parameters increases, often causing prohibitive computational burden. The proposed method greatly reduces the number of fuzzy rules by making use of the association rule analysis on historical data, and therefore achieves computational efficiency for the cases of a large number of input parameters. In the end, we apply this new method to a case study of river flood forecasting, which demonstrates that the proposed fuzzy reasoning engine can achieve better prediction accuracy than the widely used Muskingum–Cunge scheme.

Key words | association rule analysis, flood forecasting, fuzzy reasoning

INTRODUCTION

In the field of civil and hydraulic engineering, the information of a river flood discharge is related to flood control, and may be used to evaluate the performance of water resources planning and management. Therefore, it becomes a crucial issue how to accurately analyze the river flood discharge based on measured records at a specified flood domain.

Current river flood forecasting (RFF) methods have been mainly classified into three categories. The first one is the family of the physics based forecasting models, which are based on the clear understanding of the underlying physical behavior of the system (hydrological cycle) (Aronica et al. 1998; Liu et al. 2010; Nayak 2010). The second one is the family of conceptual forecasting models. ‘The conceptual models consider the physics of the underlying process in a simplistic manner. Various water resource management activities require conceptual models that are simple in structure and nature, yet capable of modeling the complex and dynamic nature of the rainfall runoff process, and have a limited number of parameters that can be estimated easily’ (Ponce et al. 1996; Srinivasulu & Jain 2009). The third one is the family of black-box (data-driven) forecasting models that put more efforts on the analysis and simulation on the historical flood data, but less on underlying flood causing factors, for example, neural networks (Zhu et al. 2005). The black-box forecasting model can simulate flood spreading process through sophisticated nonlinear relationships established based on historical data, and is able to recognize some river routing rules by self-learning (Jacquin & Shamseldin 2006).

Unlike other black-box methods, the data-driven fuzzy rule-based modeling approach behaves in a more logical and transparent way (Zadeh 1965). Therefore, it is being used in many fields including the analysis of hydrologic systems. ‘The main advantages of the fuzzy applications are that the fuzzy theory is more logical and scientific in describing the properties of an object. A fuzzy rule-based model of a system is a qualitative logical description of its behavior using variables that are expressed linguistically by means of labels such as ‘low’, ‘medium’, and ‘high’ (Nayak 2010). In this paper, we can focus on a specific kind of fuzzy models, named Takagi-Sugeno (T-S) fuzzy model (Takagi & Sugeno 1985), and the reason why it is chosen is explained in the next section. T-S fuzzy model has been used in hydrology, in many aspects (Xiong et al. 2001; Jacquin & Shamseldin 2006; Altunkaynak & Sen 2007; Nayak 2010).

Despite the great advances in T-S fuzzy reasoning models, it is still hard to simulate some river systems.
where branches and trunk streams have many inter-
junctions. In such cases, the number of input variables of dif-
fuzzy reasoning could be big, and the number of the fuzzy
rules grows exponentially as the number of input variables
increases. Moreover, for each fuzzy rule, we need to cali-
brate its own parameters by solving a minimization
problem, and therefore if the number of the rules is too
large, the total computational cost could be prohibitive.

In order to overcome this computational difficulty, we
propose to combine the T-S fuzzy reasoning forecasting
method with the association rule analysis on historical
flood data. The association rule analysis helps to identify
and combine the most relevant fuzzy rules for our specific
purpose and therefore greatly reduces the number of the
total IF–THEN rules used in our reasoning machine.

Then we set up a fuzzy reasoning flood forecasting model to forecast the future discharge data based on these fuzzy relations.

TAKAGI-SUGENO (T-S) FUZZY REASONING

In the T-S fuzzy model, each fuzzy implication rule is
expressed by a polynomial, i.e., fuzzy rule consequents are
assumed to be linear combination of the input variables,
and the output is a convex combination of consequents,
with coefficients that are the grades of membership function
of the inputs in the antecedents (Yu 2009). Comparing with
Mamdani fuzzy model, where the consequent of rules is lin-
guistic term, such as high, medium, etc, the consequent of
des of a linear mathematical relationship, and T-S fuzzy reasoning is not that
transparent (Monjezi & Rezaei 2011). However, the architec-
ture of T-S fuzzy model provides the feasibility of stability
analysis, and reduces the computational efforts of fuzzy
logic. This is the main reason why T-S fuzzy model is
selected to run river flood forecasting in this paper.

As mentioned before, a main advantage of the fuzzy
control is to use the more scientific and logical methods to illustrate some fuzzy concepts (Mamdani & Assilian 1975;
Takagi & Sugeno 1985). In this paper Takagi-Sugeno (T-S)
fuzzy reasoning method (TSFR, for short) is selected to
run river flood forecasting, because it gives an explicit
numerical result, which is the combination of multiple
inputs. Moreover, the T-S model structure resembles the
river flood situation well, due to the fact that the river
flood is caused by the upstream branch flood spreading
and lateral valley rainfall runoff.

The control rule of fuzzy reasoning machine of the
temporary method is as follows:

\[ R_r: \ IF \left( x_1 \text{ is } A_r^{(1)}, \ x_2 \text{ is } A_r^{(2)}, \ldots, \ x_p \text{ is } A_r^{(p)} \right) \ THEN \]

\[ y_r = f_r(x_1, x_2, \ldots, x_p) \]  

(1)

In the equation, \( R_r \) stands for the \( r \)-th rule, \( x_1, \ldots, x_p \) stand
for \( p \) input variables, \( A_r^{(1)}, \ldots, A_r^{(p)} \) are fuzzy sub-sets of each
input variable. \( y_r \) is output of the \( r \)-th reasoning rule and it
can be expressed as linear combination of each input:

\[ y_r = f_r(x_1, x_2, \ldots, x_p) = b_{r1}(0) + b_{r1}(1)x_1 + \ldots + b_{r1}(p)x_p \]  

(2)

The final output of total fuzzy reasoning system can be
obtained by gravity center method as follows:

\[ y = \frac{\sum_{r=1}^{k} \alpha_r y_r}{\sum_{r=1}^{k} \alpha_r} = \frac{\sum_{r=1}^{k} \alpha_r f_r(x_1, x_2, \ldots, x_p)}{\sum_{r=1}^{k} \alpha_r} \]  

(3)

Here \( \alpha_r \) refers to closing degree between the \( r \)-th fuzzy
rule and the model inputs, and its calculation will be shown in the later sections.

SIMPLIFY FUZZY RULES VIA ASSOCIATION RULE
ANALYSIS

A big issue about the fuzzy reasoning is that along with increas-
ing of the number of input variables, the number of fuzzy rules will grow exponentially, resulting in a too com-
plicated fuzzy model with very high computing burden
due to the so called ‘curse of dimensionality’. Notice that a lot of rules have no remarkable effects on improving fore-
casting precision. In this paper, we presented an efficient
way to reduce the number of the fuzzy rules and still guar-an-
tee the fuzzy reasoning precision by making use of the
association rule analysis. The associate rule analysis,
which executes forecasting knowledge mining from histori-
cal data of river flood, tries to get reasonable discharge
pertinence information of each upstream branch and down-
stream trunk stream, and sets up a much simpler fuzzy rule
set.

The association rule analysis was first presented by
Agrawal et al. (1993) when they analyzed market shopping
basket problem. The basic concepts of the associated rules
are: let \( I = \{i_1, i_2, \ldots, i_m\} \) be an item set, and \( D \) be a set of data-
base objects, each of which is a subset of \( I \). Each object has
an associated statistical measure called support. For a given object \( T \subset I \), support(\( T \)) is defined as the fraction of objects in \( D \) containing \( T \). An associated rule is an implication formula like \( A \Rightarrow B \), where \( A \subset I \), \( B \subset I \) and \( A \cap B = \Phi \). \( A \) is the antecedent of the rule, and \( B \) is the consequent of the rule. Rule \( A \Rightarrow B \) can make sense in the set \( D \) and has support degree \( (A \Rightarrow B) \), which is the percentage of the objects of \( D \) containing \( A \cup B \), i.e., probability \( P(A \cup B) \). Confidence of rule \( A \Rightarrow B \) is defined as the percentage of the objects of \( D \) containing \( B \) among all objects of \( D \) containing \( A \), i.e., conditional probability \( P(B/A) \). So we have:

\[
\text{support}(A \Rightarrow B) = P(A \cup B) \quad (4)
\]

\[
\text{confidence}(A \Rightarrow B) = P(B/A) \quad (5)
\]

Rules that satisfy minimal support degree (MSD) threshold and minimal confidence degree (MCD) threshold are called strong rules. It can often be used to predict that if \( A \) occurs in a transaction, then \( B \) will likely also occur in the same transaction. While support corresponds to the statistical significance of the given rule, confidence is a measure of the rule’s strength. Usually, these two thresholds are determined by experts, and in practice, their scopes are initially determined by experiences and then further adjusted by trials.

RFF MODEL BASED ON FUZZY REASONING AND ASSOCIATION RULE ANALYSIS

The RFF model based on fuzzy reasoning mainly consists of three parts: the first part is the data pre-processing stage, i.e., processing river historical discharge or level data in order to make them ready for the subsequent forecasting; the second part is to build the fuzzy rule set using the association rule analysis and calculate corresponding rule parameters, based on the historical data and the relevant optimization methods for parameter calibration; the third part is to apply the resulted fuzzy reasoning machine upstream river discharge (level) to predict the downstream discharge (level) and obtain the forecasting results, according to the input upstream discharge data. The flow chart is plotted in Figure 1.

Here we take a simple example of river flood spreading to show how the model works. Suppose that the discharge of river \( a \) is mainly influenced by \( n \) rivers’ discharges label by number \( 1,2,\ldots,n \) at upstream. We have the historical data of these \( n+1 \) rivers, and the discharges of \( n \) upstream rivers control stations before a certain period have been known. Our purpose is to forecast the discharge of river \( a \) after this period (for simplicity of computing, we assume that flood spreading time of each branch to river \( a \) are the same as one period). The procedure of our forecasting method is as follows:

1. Determining the membership functions of all fuzzy subsets of the \( n \) upstream rivers control stations. We divide discharges into three fuzzy subsets corresponding to low discharge, medium discharge and high discharge, respectively. After the clustering analysis of historical data, we denote the fuzzy membership degree functions as follows:

\[
f_i(Q) = \left\{ \begin{array}{ll}
m_{l_i}(Q) & \text{if } Q < Q^l_i \\
m_{m_i}(Q) & \text{if } Q^l_i \leq Q \leq Q^h_i \\
m_{h_i}(Q) & \text{if } Q > Q^h_i \\ \end{array} \right.
\]

where the \( Q^1, Q^2,\ldots,Q^n \) are the discharges corresponding to the control stations at 1st branch, 2nd branch, \ldots, \( n \)-th branch and trunk \( a \), respectively, and \( m_{l_i}(Q) \), \( m_{m_i}(Q) \) and \( m_{h_i}(Q) \) are the membership degree functions of low, medium and high discharge at \( i \)-th branch control station, respectively.

2. Determining the set of the fuzzy rules. Now we have \( n \) rivers and three fuzzy states, i.e., \( 3^n \) fuzzy rules. As simple demonstrations, some of them are listed as follows:

\[
\begin{align*}
R_1: & \text{ IF } (Q^1 \text{ is low discharge, } Q^2 \text{ is low discharge,} \ldots, \text{ and } Q^n \text{ is low discharge}) \\
& \text{ THEN } Q^n \text{ is... ;}
\end{align*}
\]

Figure 1 | River flood forecasting model based on fuzzy reasoning.
Based on the maximal membership degree principle, we apply membership degree Equation (6) to the historical data of these \( n + 1 \) rivers and put the post-processed data into Table 1, where each record corresponds to a historical time point \( t \), total discharges of \( n \) branches at a time \( t \), and discharge of river \( a \) coming right after \( t + \Delta t \).

We applied the a priori algorithm of the association rule analysis (Takagi & Sugeno 1985) to mine the fuzzy rules from the data in Table 1. The main steps are as follows:

A. Let \( I = \{i_{11}, i_{12}, i_{13}, i_{21}, i_{22}, i_{23}, \ldots, i_{n1}, i_{n2}, i_{n3}, i_{01}, i_{02}, i_{03}\} \) be a set of items, and each item represent the attribution of high, medium and low discharge of the 1st branch, 2nd branch, \( \ldots \), \( n \)-th branch and river \( a \). Then we put the previously organized records of Table 1 into a dataset denoted as \( D \), each record \( T_j \) of which is an item set, i.e., \( T_j \subseteq I \) and \( j \in \{1, \ldots, N\} \), where \( N \) is the number of the total records. For example, the first record \( T_1 = i_{13}, i_{23}, \ldots, i_{n3}, i_{03} \) in Table 1 represents a high discharge in 1st branch at time \( T_1 \), a medium discharge in 2nd branch, a high discharge in \( n \)-th branch and river \( a \). Then we put the previously organized records of Table 1 into a dataset denoted as \( D \), each record \( T_j \) of which is an item set, i.e., \( T_j \subseteq I \) and \( j \in \{1, \ldots, N\} \), where \( N \) is the number of the total records. For example, the first record \( T_1 = i_{13}, i_{23}, \ldots, i_{n3}, i_{03} \) in Table 1 represents a high discharge in 1st branch at time \( T_1 \), a medium discharge in 2nd branch, a high discharge in \( n \)-th branch and a high discharge in river \( a \) after time \( \Delta t \). As mentioned before, there are total \( N \) records \( \{T_1, \ldots, T_N\} \) in dataset \( D \).

B. Designate the constrained conditions for the fuzzy rules to be mined, i.e., set the form of the rules. In this paper, we designate the latter part of a rule as the discharge of river \( a \). We also need to determine MSD, \( \text{min}_\text{sup} \) and MCD, \( \text{min}_\text{conf} \).

C. Iterate in layers and obtain the sets of items with different frequencies. Based on MSD, we first scan all records and find out the set of frequency 1 items, which is denoted as \( L = \{i_h\} \), i.e., satisfying the condition \( \sum T_{kl}/N > \text{min}_\text{sup} \), where \( \sum T_{kl} \) is the number of records that include \( i_{hl} \) and \( k \in \{1, \ldots, n\} \), and \( l \in \{1, 2, 3\} \). Then based on the set of frequency 1 items, we scan all records and generate the sets of items with higher frequencies and denote them as \( L_1, \ldots, L_m \), where \( m \) is the number of concentrating items in frequency items.

D. For every frequency items set \( l \) we have all nonempty subset of \( l \). For each nonempty subset \( s \) of \( l \), if \( \text{support}(l)/\text{support}(s) \geq \text{min}_\text{conf} \), then the output rule is \( s \Rightarrow (l_s) \), where \( \text{support}(l) \) and \( \text{support}(s) \) represent the number of records that contain these two items, respectively.

E. Evaluate all obtained rules and analyze the relevance of the hydrological information.

The main objective of the association rule analysis is to find out the reasoning rules under which high discharge of river \( a \) is more likely to appear and remove or combine those reasoning rules which only have minor effects on flood spreading. Subsequently the total \( 3^n \) rules can be reduced to only \( M \) rules with \( M < 3^n \).

\( (x) \) Determine the mathematical formulas of fuzzy rules which produce the discharge \( Q^a \) of river \( a \). From Equation (2) \( y_i = f_i(x_{1i}, x_{2i}, \ldots, x_{ni}) = b_i(0) + b_i(1)x_{1i} + \ldots + b_i(p)x_{ni} \) we can see that, in the model of TSFR, the discharge of river \( a \) mainly consists of the conflux of the \( n \) upstream branches and lateral rainfall conflux from gauging station of branches to river \( a \), where \( p = n \). To simplify problems, we suppose that lateral rainfall conflux can be ignored, i.e., constant item \( b_i(0) = 0 \). Then Equation (2) can be simplified and rewritten as:

\[
\begin{align*}
R_1: & \quad Q^a_i = x_{11}Q^i_1 + x_{12}Q^i_2 + \ldots + x_{1n}Q^i_n \\
R_2: & \quad Q^a_i = x_{21}Q^i_1 + x_{22}Q^i_2 + \ldots + x_{2n}Q^i_n \\
& \quad \ldots \\
R_m: & \quad Q^a_i = x_{M1}Q^i_1 + x_{M2}Q^i_2 + \ldots + x_{Mn}Q^i_n
\end{align*}
\]

The above are the simplified expressions of the \( M \) fuzzy reasoning rules, where \( l, m \) and \( h \) refer to low, medium and high discharge, respectively, and \( Q^i_h \) to high discharge of \( i \)-th branch, \( x_{ij} \) to weight parameter of \( j \)-th branch in \( i \)-th rule.

The left side of each formula is the discharge of river \( a \) under the corresponding fuzzy rule.

\( (4) \) Organize historical dataset. We organized the historical dataset into such a set as \( \{Q^1_1, Q^1_2, \ldots, Q^i_h, Q^2_1, \ldots, Q^n_h, Q^3_1, \ldots, Q^n_h\} \), where \( Q^1_1, Q^2_1, \ldots, Q^h_1 \) stand for the discharges of the 1st branch, 2nd branch, and \( n \)-th branch at time \( t \) in \( i \)-th historical record, \( Q^i_h \) is

<table>
<thead>
<tr>
<th>No.</th>
<th>Branch 1st</th>
<th>Branch 2nd</th>
<th>Branch nth</th>
<th>River a</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>2</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>3</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Table 1 | Rivers historical discharge record
the real discharge of river \( a \) at time \( t + \Delta t \) in \( i \)-th historical record, \( N \) is the number of total historical records, and \( a_{ki} \) is the membership degree of the \( k \)-th rule in the \( i \)-th record, \( k \in \{1, ..., M\} \). \( a_{ki} \) can be obtained by Equation (8):

\[
a_{ki} = \min \{a_{1ki}, a_{2ki}, \ldots, a_{aki}\} \quad (8)
\]

where \( a_{1ki}, a_{2ki}, \ldots, a_{aki} \) stand for the membership degree of the discharge of each branch that belongs to the \( k \)-th rule and can be obtained by Equation (6).

Establish the optimization model for each rule in order to calculate the involved discharge weight parameters of each branch. For example, for the \( k \)-th rule (\( k = 1, \ldots, M \)), build a constrained optimization model under the minimal distance square as follows:

\[
\begin{align*}
\{ & \min \sum_{i=1}^{N} a_{ki}(Q^1_i - x_{ki}Q^1_i - x_{ki}Q^2_i - \ldots - x_{ki}Q^n_i)^2 \quad (9) \\
\end{align*}
\]

By solving Equation (9), we get the branch discharge weight parameters \( x_{ki}, i = 1, 2, \ldots, n \) of the \( k \)-th rule, i.e., the parameters at the \( k \)-th row of Equation (7). Similarly, we can also obtain all other weight parameters in Equation (7).

Apply the model for forecast. The input information for the model is discharges \( Q^1, Q^2, \ldots, Q^n \) of each branch at time \( t \). Then using Equations (6) and (8) to get \( a_1, a_2, \ldots, a_n \); substitute values of \( \{a_1Q^1, a_2Q^2, \ldots, a_nQ^n\}, \ldots, \{a_1Q^n, a_2Q^2, \ldots, a_nQ^1\} \) and parameters in step 3 into formulas of Equation (7), and the values of \( Q_{fl}^1, Q_{fl}^2, \ldots, Q_{fl}^n \) can be obtained respectively. Substitute \( a_1, a_2, \ldots, a_n \) and \( Q_{fl}^1, Q_{fl}^2, \ldots, Q_{fl}^n \) into Equation (3), and we can obtain the result \( Q^a \), which is the final forecasting result.

**APPLICATIONS**

We forecast the discharge of Tongmeng Station at trunk downstream on Nenjiang River, based on the measured discharges of Nenjiang River valley’s three river control stations: Guchengzi Station at Nuomin River, Dedu Station at Namoer River and Ayanqian Station at trunk upstream of Nenjiang River (see Figure 2).

The incoming flow of Tongmeng Station mainly comes from the flood spreading of upstream trunks and branches, and therefore we select the above three representative stations at trunk and branches. Compared with valleys of Nuomin River, Namoer River and upstream of Nenjiang River, the valley area between Guchengzi, Ayanqian, Dedu and Tongmeng Station is much smaller, and the effect of lateral rainfall on Tongmeng Station can be ignored. We select these four stations’ runoff data of the flood season in the following high flow years: 1973, 1978, 1979, 1980, 1984 and 1988 as the training data to construct forecasting model, and select four stations’ runoff data of the flood season in small flood year 1982, medium flood year 1983, great flood year 1989 and extreme high flood year 1969 year to validate the model. Part of the training data is listed in Table 2. The lead time is the average spreading time of river flood spreading and the length of forecasting period is one day.

The paper applies the cluster analysis to the runoff of Ayanqian, Guchengzi and Dedu in training data and applies C-means method (Jiawei 2000) to divide these three stations’ runoff into three classes of low, medium and high discharge, respectively. The cluster centers are listed in Table 3.

**Table 2** | Four stations’ discharge data (unit: m³/s)

<table>
<thead>
<tr>
<th>Time</th>
<th>Dedu (three days before)</th>
<th>Ayanqian (1.5 days before)</th>
<th>Guchengzi (1 day before)</th>
<th>Tongmeng station</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980–7–19 8:00</td>
<td>30.6</td>
<td>883</td>
<td>619</td>
<td>1,580</td>
</tr>
<tr>
<td>1980–7–20 8:00</td>
<td>28.7</td>
<td>971</td>
<td>759</td>
<td>1,800</td>
</tr>
<tr>
<td>1980–7–21 8:00</td>
<td>29.7</td>
<td>1,120</td>
<td>821</td>
<td>2,080</td>
</tr>
</tbody>
</table>

Figure 2 | Spatial distribution of Guchengzi, Ayanqian, Dedu and Tongmeng in Nenjiang river basin.
Taking the obtained cluster centers in Table 3 as the references, we classify the three stations’ runoff into three fuzzy subsets corresponding to low, medium and high discharge, respectively. The three stations’ runoff membership degree functions are expressed as linear functions plotted in Figure 3.

As we know, different runoff changing trends of upstream rivers will result in different flood features of downstream combinations. Therefore it is reasonable to take runoff changing factors into consideration and compare each river runoff with last time runoff. The increasing of runoff above 5%, that between 5% and –5% and that less than –5% will be set as rise, equal and drop of runoff, respectively. The runoff level of former three stations is determined by maximal membership degree principle, and Tongmeng station’s runoff level will be obtained by the cluster method (Jiawei 2003). The processed data are shown in Table 4.

So we applied the above data to run the associated rule mining and find out when the Tongmeng station will have high runoff. MSD and MCD might be different in different applications. Based on features of flood spreading, we tried different combinations of MSD and MCD to execute the mining and find out that when MSD are 5% and MCD are 60%, the set of mined rules is better. Part of results is showed as follows:

| Medium runoff of Ayanqian and high runoff of Guchengzi ⇒ high runoff of Tongmeng |
| High runoff of Ayanqian and medium runoff of Guchengzi ⇒ high runoff of Tongmeng |
| High runoff of Ayanqian and high runoff of Guchengzi ⇒ high runoff of Tongmeng |
| Medium runoff of Ayanqian and medium runoff of Guchengzi ⇒ high runoff of Tongmeng |
| High runoff of Dedu and high runoff of Ayanqian and low runoff of Guchengzi ⇒ high runoff of Tongmeng |
| Ayanqian swelling and Guchengzi swelling ⇒ high runoff of Tongmeng |

Here some characteristics are observed: (a) due to small runoff level, runoff of Dedu station has little effect on the peak of Tongmeng; (b) peaks combination of Ayanqian and Guchengzi often lead to peaks of Tongmeng; (c) in tread combinations, trends of Dedu can be ignored.

### Table 3 | Cluster center of three stations’ runoff (unit: m³/s)

<table>
<thead>
<tr>
<th>Control station</th>
<th>Runoff</th>
<th>Ayanqian</th>
<th>Cluster center</th>
<th>Approximation</th>
<th>Guchengzi</th>
<th>Cluster center</th>
<th>Approximation</th>
<th>Dedu</th>
<th>Cluster center</th>
<th>Approximation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td>562.35</td>
<td>600</td>
<td>321.63</td>
<td>300</td>
<td>38.76</td>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td></td>
<td></td>
<td>1,003.52</td>
<td>1,200</td>
<td>689.65</td>
<td>700</td>
<td>110.44</td>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td>1,601.87</td>
<td>1,800</td>
<td>1,211.91</td>
<td>1,100</td>
<td>199.76</td>
<td>200</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Figure 3 | Membership degree functions of three stations.

### Table 4 | Treated runoff data of four stations (unit: m³/s)

<table>
<thead>
<tr>
<th>Time</th>
<th>Dedu (3 days before)</th>
<th>Ayanqian (1.5 days before)</th>
<th>Guchengzi (1 day before)</th>
<th>Tongmeng Runoff</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Runoff</td>
<td>Trend</td>
<td>Runoff</td>
<td>Trend</td>
</tr>
<tr>
<td>1980-7-25 8:00</td>
<td>Low</td>
<td>Equal</td>
<td>High</td>
<td>Rise</td>
</tr>
<tr>
<td>1980-7-26 8:00</td>
<td>Low</td>
<td>Rise</td>
<td>High</td>
<td>Rise</td>
</tr>
<tr>
<td>1980-7-27 8:00</td>
<td>Low</td>
<td>Rise</td>
<td>High</td>
<td>Equal</td>
</tr>
</tbody>
</table>
and most combination of Ayanqian and Guchengzi are either simultaneous rise or simultaneous drop, and then combinations of rise-equal and drop-equal, finally very few rise-drop and drop-rise.

Based on these characteristics, the rule set can be simplified as follows: (a) resetting the runoff levels of Dedu into two categories: one is medium-low runoff and the other one is high runoff, meanwhile, the runoff trends of Dedu are not taken into consideration due to its very little effect on downstream; (b) reducing runoff combinations of Ayanqian and Guchengzi into six kinds of combinations such as: high–medium, medium–high, high–high, medium–medium, medium–low and other; (c) reducing runoff trend combinations of Ayanqian and Guchengzi into five kinds of combinations as: rise–rise, drop–drop, equal–drop or rise–equal, equal–rise or drop–equal and other. According to above three principles that only \(2 \times 6 \times 5 = 60\) rules will be adopted for the following analysis, far less than the original 729 rules that are determined by three branches, three runoff levels and three trends. Part of the 79 rules is listed in Table 5.

Here we employ the restricted optimization model (9) to determine model parameters. Part of the rules is listed in Table 6.

### Table 5 | Fuzzy reasoning rules set (rules 6–10 only shown)

<table>
<thead>
<tr>
<th>Rules</th>
<th>Dedu</th>
<th>Runoff</th>
<th>Trend</th>
<th>Ayanqian</th>
<th>Runoff</th>
<th>Trend</th>
<th>Guchengzi</th>
<th>Runoff</th>
<th>Trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>High</td>
<td>Medium</td>
<td>Drop</td>
<td>High</td>
<td>Medium</td>
<td>Drop</td>
<td>High</td>
<td>Drop</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>High</td>
<td>Medium</td>
<td>Drop, equal</td>
<td>High</td>
<td>Equal, rise</td>
<td>High</td>
<td>Equal, rise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>High</td>
<td>Medium</td>
<td>Equal, rise</td>
<td>High</td>
<td>Drop, equal</td>
<td>High</td>
<td>Drop, equal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Medium or low</td>
<td>High</td>
<td>Rise</td>
<td>Medium or low</td>
<td>High</td>
<td>Rise</td>
<td>Medium</td>
<td>Rise</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Medium or low</td>
<td>High</td>
<td>Drop</td>
<td>Medium or low</td>
<td>High</td>
<td>Drop</td>
<td>Medium</td>
<td>Drop</td>
<td></td>
</tr>
</tbody>
</table>

### Table 6 | Flood forecasting rules under fuzzy reasoning (rules 17–20 only shown)

<table>
<thead>
<tr>
<th>Rules</th>
<th>IF Dedu</th>
<th>Ayanqian</th>
<th>Guchengzi</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>(Q_1)</td>
<td>Medium–low</td>
<td>High (rise)</td>
</tr>
<tr>
<td>18</td>
<td>(Q_1)</td>
<td>Medium–low</td>
<td>High (drop)</td>
</tr>
<tr>
<td>19</td>
<td>(Q_1)</td>
<td>Medium–low</td>
<td>High (drop, equal)</td>
</tr>
<tr>
<td>20</td>
<td>(Q_1)</td>
<td>Medium–low</td>
<td>High (equal, drop)</td>
</tr>
</tbody>
</table>

We select data in the following 4 flood seasons of 1982 with peak values of 1,500–2,000 m³/s, 1983 with 3,000–4,000 m³/s, 1989 with 6,000–8,000 m³/s and 1969 with 8,000 m³/s as model checking data. The fuzzy rules set was applied to run fuzzy reasoning on Tongmeng runoff in these four flood seasons and obtain results are shown in Figure 4. Forecasting results of subsection Muskingum–Cunge scheme and actual measurement runoff values are also given in Figure 4 for comparison. The forecasting results by these two methods are listed in Table 7. In Table 7, each item in the column named ‘Flood no.’ is composed of four parts in the form of ‘YYYYMMDD’, for example, ‘19820707’ means the flood date for 7 July 1982.

From Table 7 and Figure 4 we can see that, for forecasting high runoff, the precision of the fuzzy reasoning method is better than Muskingum–Cunge scheme. However, for low runoff, the precision of the two methods almost the same and an important reason is that the simplifying of the fuzzy rule set decreases the forecasting precision on low frequency and low runoff. Muskingum–Cunge scheme is based on water-balance equation and storage equation. Due to instability of the parameter determined by historical data and the instinct problems of the adopted parameter calculating methods, some system errors in calculating runoff...
process occur and behave as some lag features. Note that Muskingum–Cunge scheme is simple and easy to implement, therefore this scheme is still used in many practical applications, especially when we have little errors in determination of parameters and few valley's river inter-junctions.

Table 7 | Forecasting results of fuzzy reasoning and Muskingum–Cunge scheme

<table>
<thead>
<tr>
<th>Peak scope (m³/s)</th>
<th>Flood no.</th>
<th>Fuzzy reasoning method</th>
<th>Muskingum–Cunge scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Peak relative error (%)</td>
<td>Relative error of peak time (period number)</td>
</tr>
<tr>
<td>0–3,000</td>
<td>19820707</td>
<td>15.55</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>19820804</td>
<td>5.73</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>19830610</td>
<td>5.02</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>19890605</td>
<td>7.96</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>19890813</td>
<td>11.28</td>
<td>0</td>
</tr>
<tr>
<td>3,000–5,000</td>
<td>19830706</td>
<td>–2.52</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>19830802</td>
<td>–4.74</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>19890629</td>
<td>2.82</td>
<td>0</td>
</tr>
<tr>
<td>&gt;5,000</td>
<td>19690811</td>
<td>–4.75</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>19890717</td>
<td>–3.67</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 4 | Forecasting results contrast of Tongmeng runoff for the flood seasons of 1982, 1983, 1989 and 1969.
CONCLUSION

In this paper, a RFF model based on fuzzy reasoning and the association rule analysis is presented. It first constructs fuzzy reasoning machine according to fuzzy control theory, then applies association rule analysis method for associated rule mining to simplify the set of the adopted fuzzy reasoning rules and solve the issue of too many rules in the fuzzy reasoning machine. The physical meaning of the model is clear, especially for forecasting runoff and level of river or conflux networks, which are difficult to be described by other common alternative hydraulic models.

ACKNOWLEDGMENTS

We are pleased to acknowledge the financial support for this project from the National Natural Science Foundation of China (grant no. 50809010 and 51079014).

REFERENCES


Liu, W., Chen, W., Hsu, M. & Fu, J. 2010 Dynamic routing modeling for flash flood forecast in river system. *Natural Hazards* 52, 519–537.


First received 19 December 2011; accepted in revised form 13 June 2012