Combined risk assessment of nonstationary monthly water quality based on Markov chain and time-varying copula

Wei Shi and Jun Xia

ABSTRACT

Water quality risk management is a global hot research linkage with the sustainable water resource development. Ammonium nitrogen (NH$_3$-N) and permanganate index (COD$_{Mn}$) as the focus indicators in Huai River Basin, are selected to reveal their joint transition laws based on Markov theory. The time-varying moments model with either time or land cover index as explanatory variables is applied to build the time-varying marginal distributions of water quality time series. Time-varying copula model, which takes the non-stationarity in the marginal distribution and/or the time variation in dependence structure between water quality series into consideration, is constructed to describe a bivariate frequency analysis for NH$_3$-N and COD$_{Mn}$ series at the same monitoring gauge. The larger first-order Markov joint transition probability indicates water quality state Class Vw, Class IV and Class III will occur easily in the water body of Bengbu Sluice. Both marginal distribution and copula models are nonstationary, and the explanatory variable time yields better performance than land cover index in describing the non-stationarities in the marginal distributions. In modelling the dependence structure changes, time-varying copula has a better fitting performance than the copula with the constant or the time-trend dependence parameter. The largest synchronous encounter risk probability of NH$_3$-N and COD$_{Mn}$ simultaneously reaching Class V is 50.61%, while the asynchronous encounter risk probability is largest when NH$_3$-N and COD$_{Mn}$ is inferior to class V and class IV water quality standards, respectively.

Key words | Huai River Basin, land cover, Markov joint transition probability, non-stationarity, time-varying copula, water quality

INTRODUCTION

Water quality degradation have become a hot focus for sustainable water environment management around the world. Water quality variation are not only affected by precipitation, runoff, temperature and other natural factors, but also anthropogenic activities such as economic restructuring and planning, industrial and agricultural wastewater emissions, dams and floodgates constructions. Therefore, the evolution is a complex, nonstationary dynamic process characterized by trend, season, period, etc., and even mutation presented in the hydrological series. If non-stationarity in water quality series is not fully considered, the results of the traditional water quality risk assessment based on the stationarity assumption would be invalid in practice (Jiang et al. 2015). Quantitative risk assessment of different water quality indicators considering non-stationarity existing in each water quality series, will provide a significant reference for water quality prediction and pollution control in-depth study of the evolution of non-stationary water quality time series, which is an integrated response of the water environment systems to the natural factors and human activities.

Many previous researches of water quality risk analysis focused on the univariate time series such as dissolved oxygen (DO), ammonium nitrogen (NH$_3$-N) and permanganate index (COD$_{Mn}$), but seldom conducted the combined risk assessment and joint transition risk probability by constructing joint distribution function of nonstationary water quality series based on time-varying copula and Markov...
chain, respectively. A large number of statistical methods (i.e. Markov Switching Auto Regressive Model: MS-AR, neural network-based genetic algorithm and generalised regression: GA-ANN and GRNN, Monte Carlo: MC) had been used to analyse the univariate time series of water quality at different spatial and temporal scales under normality assumption of distribution functions in the previous study (Whitehead & Young 1979; Jayawardena & Lai 1989; Worrell & Burt 1999; May et al. 2009; Jiang et al. 2013; Niu et al. 2013; Burchard-Levine et al. 2014; Arya & Zhang 2015). These researches were almost based on stationary water quality time series, or the univariate time series may be transformed to follow one distribution empirically. However, the stationarity assumption of water quality series will be invalid under changing environment, especially for the highly regulated Huai River, China. Besides, the marginal distribution employed for describing the water quality series should also take the non-stationarity into consideration.

Zou & Yu (1996) conducted the dynamic factor model to analyse the non-stationary process of six water quality indicators and consider time-lagged correlations and the linear trends. Young & Young (1992) considered the non-stationarity of environment series over the observation interval, and found out that the important environmental time-series CO₂ exhibited an obvious non-stationarity. In recent years, many researches on nonstationary time series such as flood frequency analysis were widely developed around the world as the effect of climate, land use, human interventions (dams, floodgates and reservoirs) intensifying (Milly et al. 2005; Villarini et al. 2009; Gilroy & McCuen 2012; Xiong et al. 2015; Sarhadi et al. 2016). For describing the non-stationarity as well as the joint distribution of multiple random variables, time-varying moment model has been widely applied in many studies (Strupczewski et al. 2002a, 2002b; Khalid et al. 2006; Jiang et al. 2015). In the time-varying moment model, the distribution parameters of the series were usually expressed as functions of time to reflect the non-stationarities of the series (Khalid et al. 2006). In addition to the direct usage of time as the explanatory variable of the distribution parameters, some other time-varying variables related to the time series had also been employed. For example, the distribution parameters of flood, rainfall, temperature or water quality series were usually modelled as functions of some climate or environmental indices (El Adlouni et al. 2007; Kwon et al. 2008; Villarini et al. 2010; López & Francés 2013; Du et al. 2015). In order to reflect reasonably the water quality risk faced by Huai River Basin (HRB), we need to study the joint probability distribution of water quality parameters NH₃-N and CODMn, which are the core indicators for water quality management in the study area (Wu et al. 2015). Copula function is a good choice and extensively used for describing the dependence structure and the joint probability distribution of multiple random variables (Genest & Favre 2007; Salvadori et al. 2007; Schweizer 2014), and further for quantitatively evaluating the combined risk probability of water quality exceeding standard (GB 3838-2002; Anderson-Cook 2006; Ganguli & Reddy 2012; Lian et al. 2013; Niu et al. 2013; Borgomeo et al. 2014). Due to the lack or limited water quality data monitored, multivariate water quality risk analysis based on the copula method has rarely considered the possible time variation of the dependence structure or the joint distribution between the water quality random variables.

Thus making full use of the long time series of monitoring data, this study aims to reveal the joint transition laws of water quality risk based on Markov method, and investigate the non-stationary of univariate water quality time series using the time-varying marginal distributions taking the time and land cover index as the explanatory variables based on the time-varying moment model. Through the optimal time-varying marginal distributions of two core water quality indicators in HRB, the time-varying copula model is constructed to describe the dependence structure and quantitatively assess the combined risk of NH₃-N and CODMn, which will provide a scientific foundation for water quality improving management and decision making for sustainable water resource management. This study consists of the following components. Firstly, the study area and water quality monitoring data are introduced. Secondly, the methodologies of Markov joint transition probability and time-varying copula model are described in this paper. Thirdly, the results including the joint transition probabilities revealed, based on Markov theory, the nonstationary characteristic identified by time-varying marginal distributions and combined risk-assessed quantitatively with time-varying copula, are presented. Finally, some conclusions of the above analysis are summarised.

**STUDY AREA AND DATA**

HRB (30°55’ – 36°36’N, 111°55’ – 120°45’E), as one of the seven largest river basins in China, flows through five provinces, i.e. Hubei, Henan, Anhui, Shandong and Jiangsu (Figure 1). The HRB with average annual rainfall and runoff of 911 mm and 45.2 billion m³ from 1956 to 2000
controls a catchment area of 270,000 km², accounting for approximately 3% of the national area. As the important agricultural base with farmland area of 12,192 km², the proportion of agricultural population accounts for 86% in HRB. The main land use types are farmland, forest and urban blocks, followed by grassland and waters. In order to solve water shortage and flooding of HRB, which is the first river implemented with most comprehensive management measures, approximately 11,000 reservoirs and sluices were constructed by 2000, which obviously altered the hydrological regime (Jiang et al. 2011). A large number of water conservancy projects highly regulate the runoff, and resulted in the uneven seasonal distribution of water storage capacity, which was extremely unfavourable for self-purification and dilution of water pollutants. Due to the highly distributed reservoirs, dams and floodgates, pollutants accumulation and centralized pollution emissions in flooding season had led to pollution events in the local time, which destructed the water ecological system, and it was difficult to restore. Thus it is very necessary to recognize current water quality condition and assess the combined risk of nonstationary water quality time series in highly regulated HRB.

The monthly monitoring data of water quality parameters NH₃-N and COD₅ᵢᵦ at Bengbu Sluice over the period 1986–2014 were obtained from Huai River Water Resources Protection Bureau (Figure 2). Bengbu Sluice (BBS) (32°57’N, 117°17’E) was mainly built for irrigation in agriculture but it is also used for navigation and electricity generation. This sluice, built in 1962 with a basin area of 120,000 km², is located at the middle reach of the main stream in HRB (Figure 1). It consists of 30 check gates, most of which have a width of 10 m and a height of 7.5 m; one hydropower station with six electric generators (Zhao et al. 2010). Domestic and
industrial water for the Bengbu City, an important city of Anhui province in China, is mainly drawn from the reach upstream of BBS. Besides, a digital elevation model was obtained from the Shuttle Radar Topography Mission (SRTM) 3 s Digital Elevation Database of USGS/NASA. The land cover data set for years 1990, 1995, 2000, 2005 and 2010 was obtained by Chinese Academy of Sciences. Land cover was aggregated for six major land types – farmland, forest, grassland, waters, urban blocks and unused land.

METHODS

The classification of water quality

NH$_3$N and COD$_{Mn}$, as the commonly used indicators in HRB which are monitored following the national standard of NH$_3$N and COD$_{Mn}$ testing from the monitoring center of Huai River Water Resources Protection Bureau, are selected to conduct the water quality variation and risk analysis (GB 3838-1988; HJ 355-2009; Zhang et al. 2013a, 2013b; Zhai et al. 2014; Dou et al. 2015). As shown in Table 1, five standard thresholds of NH$_3$N and COD$_{Mn}$ are based on ‘Environmental Quality Standard for Surface Water’ (GB 3838-2002) issued by State Environmental Protection Administration of China. Furthermore, this study divides the water quality state into six classes: Class I, Class II, Class III, Class IV, Class V, Class V$_w$, and the situation probability can be calculated combing with NH$_3$N and COD$_{Mn}$ as Table 2. When the water quality state is Class V$_w$, it represents that the concentration of NH$_3$N or COD$_{Mn}$ is greater than 2 and 15 mg/L, respectively.

Markov process

Let a stochastic process denote $X(t)$, and its state space is $J = \{1, 2, \ldots, \}$. $S(t)$ is the state variable at time $t$. Based on the Markov Process theory, the next state has no correlation with the other past state, and decided by the current state. Markov chain is created by conditioning probability of $X(t)$ with the following characteristics:

$$P(X(n + 1) = s | X(n) = r, X(n - 1) = r_{n-1}, \ldots, X(1) = r_1, X(0) = r_0) = P[X(n + 1) = s | X(n) = r] = P(1)$$

Combining with the time varying, the moving path of the next states are decided by a series of transition intensities $p_{rs}(t, z(t))$ depending on the time $t$, where $r$ and $s$ are the states of the pair. $z(t)$ is the time-varying explanatory variable. The instantaneous risk of the state $s$ transferring to $r$ is calculated by the transition intensities.

$$p_{rs}(t, z(t)) = \lim_{\delta t \to 0} \frac{P(S(t + \delta t) = s | S(t) = r)}{\delta t}$$

The transition intensity matrix $P$ denotes the intensity. Its rows sum to 0 and diagonal entries can be calculated by the formula $p_{rr} = - \sum_{s \neq r} p_{rs}$. The first-order transition probability is $P(1) = \begin{pmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{pmatrix}$. 

Table 1 | Standard thresholds of water quality classification (GB 3838-2002)

<table>
<thead>
<tr>
<th>Indicators</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH$_3$N</td>
<td>$\leq$ 0.15</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td>2</td>
</tr>
<tr>
<td>COD$_{Mn}$</td>
<td>$\leq$ 2</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td>15</td>
</tr>
</tbody>
</table>

*aThe unit of standard value is mg/L.*
Table 2 | Water quality state classification combining with NH3-N and CODmn

<table>
<thead>
<tr>
<th>Serial number</th>
<th>State (NH3-N-CODmn)</th>
<th>Situation probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I</td>
<td>( p_1 = P(X \leq x_1, Y \leq y_1) )</td>
</tr>
<tr>
<td>2</td>
<td>II</td>
<td>( p_2 = P(x_1 &lt; X \leq x_2, Y \leq y_2) + P(X \leq x_2, y_1 &lt; Y \leq y_2) - P(x_1 &lt; X \leq x_2, y_1 &lt; Y \leq y_2) )</td>
</tr>
<tr>
<td>3</td>
<td>III</td>
<td>( p_3 = P(x_2 &lt; X \leq x_3, Y \leq y_3) + P(X \leq x_3, y_2 &lt; Y \leq y_3) - P(x_2 &lt; X \leq x_3, y_2 &lt; Y \leq y_3) )</td>
</tr>
<tr>
<td>4</td>
<td>IV</td>
<td>( p_4 = P(x_3 &lt; X \leq x_4, Y \leq y_4) + P(X \leq x_4, y_3 &lt; Y \leq y_4) - P(x_3 &lt; X \leq x_4, y_3 &lt; Y \leq y_4) )</td>
</tr>
<tr>
<td>5</td>
<td>V</td>
<td>( p_5 = P(x_4 &lt; X \leq x_5, Y \leq y_5) + P(X \leq x_5, y_4 &lt; Y \leq y_5) - P(x_4 &lt; X \leq x_5, y_4 &lt; Y \leq y_5) )</td>
</tr>
<tr>
<td>6</td>
<td>( V_w )</td>
<td>( p_6 = P(x_5 &lt; X) + P(y_5 &lt; Y) - P(x_5 &lt; X, y_5 &lt; Y) )</td>
</tr>
</tbody>
</table>

Time-varying copula

Time-varying copula model

As an effective mathematical tool for describing the various inter-variable structure including linear and nonlinear, symmetry and asymmetry, and tail dependence (Jammazi et al. 2015), copulas are considered flexible tools for constructing multivariate distributions and modelling the dependence structure between correlated variables (Sklar 1959). The popularity of copulas is due to their flexibility in forming dependence between variables using any type of marginal distribution, which will not lose any information contained in the marginal distribution. More detailed introductions about copulas can be seen from the literature (Salvadori et al. 2007; Schweizer 2014; Sarhadi et al. 2016).

Due to the lack and limited water quality data, the copula-based combined risk analysis is difficult to be conducted, and then the relevant research is rare. Moreover, the past multivariate copula models were widely used in the hydrological frequency analysis, drought and rainstorm (Salvadori & De Michele 2007; Corbella & Stretch 2012; Chen et al. 2013; Zhang et al. 2013a, 2013b; Aghakouchak 2013), and most of their distribution parameters of marginal statistics and copula model were considered as a constant. However, whatever the hydrological series or water quality monitoring series, the non-stationarity should be taken into account under changing climate, land use and human interventions in the highly regulated rivers. Suppose that \( Y_t = (Y_{t1}, Y_{t2}) \) is a pair of water quality variables whose dependence structure is defined by a copula function. A general form of a time-varying joint distribution can be built at any time \( t \) using time-varying copula as follows:

\[
Y - F(y_1 | \theta_1)
\]

\[
F(y_1 | \theta_1) = C(F_1(y_{1t} | \theta_{1t}), F_2(y_{2t} | \theta_{2t}) | \theta_{ct})
\]

\[
F(y_1 | \theta_1) = \text{C}(u_{1t}, u_{2t} | \theta_{ct})
\]

where \( F(\cdot) \) and \( C(\cdot) \) are the cumulative distribution function and copula function, respectively; \( \theta_{1t} \) and \( \theta_{2t} \) are time-varying parameters of marginal distribution; \( u_{1t} \) and \( u_{2t} \) are the marginal probabilities of the built time-varying copula in the unit hypercube with uniform \([0,1]\) marginal distributions. Comparing with nonstationary behaviour considered both in the marginal distribution and copula model, we experiment their parameters treated as a constant in this study.

Archimedean copulas as the most popular family of copula are extensively applied in different fields, i.e. hydrology, meteorology and water environment, because of the advantages of their simple structure, diversity, adaptability and so on (Nelsen 2006). Archimedean copulas, i.e. Gumbel-Hougaard (GH), Frank and Clayton copula, which have been used to build the joint distribution of water quality variables, are selected as the candidates in modelling the time-varying dependence between the two water quality series (Zhang et al. 2011; Kuchment & Demidov 2013). Suppose that \( \theta_{ct} \) is the copula parameter, \( t = 1, 2, \ldots, n \) the sample size of observed water quality series, \( l \) the percent land cover of farmland and urban blocks representing the main land development (79.6–87.1%) and \( l \in (0, 1) \), \( \alpha_i = (i = 0, 1, \ldots, k) \) are the link function parameters, \( x_{il} = (i = 1, 2, \ldots, k) \) are the explanatory variables. The link function \( L(\cdot) \) of the explanatory variables time \( (t) \) and percent land cover of farmland and urban blocks \( (l) \), respectively, can be expressed as follows:

\[
L_t(\theta_{ct}) = a_0 + \sum_{i=1}^{k} \alpha_i x_{it}
\]

\[
L_l(\theta_{cl}) = a_0 + \sum_{i=1}^{k} \alpha_i x_{il}
\]
Time-varying marginal distribution

Different types of independent distributions including three two-parameter distributions (Logistic, Gumbel and Normal) and two three-parameter distributions (Pearson type III distribution and T Family) (Rigby & Stasinopoulos 2009), are selected to construct the marginal distribution whose parameters are time-dependent in a nonstationary water quality process. Three parameters (location, scale and shape) of the marginal distribution are generally denoted by a vector \( \theta = (\mu, \sigma, \nu)^T \) corresponding to the symbol used in Equations (4) and (5). The time-varying marginal distribution of each water quality series is established using the time-varying moment model. Rigby & Stasinopoulos (2005) developed a general class of univariate regression models called Generalized Additive Models in Location, Scale and Shape (GAMLSS), which has been widely applied in frequency analysis of nonstationary series (Xiong et al. 2014; Jiang et al. 2015).

For the GAMLSS-based time-varying moment model, \( \mu, \sigma \) and \( \nu \) are the location parameter, scale parameter and shape parameter, and indicate the mean, variance and the skewness coefficient, respectively. If the observation of the response variable \( y_i \) at time \( t \) \( (t = 1, 2, \ldots, n) \) follows a distribution with probability density function \( f_Y(y_i; \mu_t, \sigma_t, \nu_t) \), and then each distribution parameter can be expressed as a linear function of the explanatory variable \( x_{it} = (i = 1, 2, \ldots, k) \) which has the monotonic characteristic via a link function as follows:

\[
L_1(\mu_t) = \beta_{10} + \sum_{i=1}^{k} \beta_{1i}x_{it} \\
L_2(\sigma_t) = \beta_{20} + \sum_{i=1}^{k} \beta_{2i}x_{it} \\
L_3(\nu_t) = \beta_{30} + \sum_{i=1}^{k} \beta_{3i}x_{it}
\]

where \( \beta_{ij} (j = 1, 2, 3, i = 0, 1, \ldots, k) \) are the GAMLSS parameters. Besides, \( j \) is the number label of distribution parameters and \( k \) is the number of explanatory variables. Inversion of Kendall’s tau, maximum likelihood, maximum pseudo-likelihood (Genest et al. 1995), inversion of Spearman’s rho and inference functions for margins (Joe 2005) are the commonly used parameter estimation methods. For two-dimensional Archimedean copula function, the parameters value of time-varying joint distribution and time-varying marginal distributions can be estimated by the IMF method proposed by Joe (2005).

Model selection

The goodness-of-fit of time-varying copula models are examined by analyzing the fitness of the copula model and time-varying marginal distributions, which are tested by the Kolmogorov-Smirnov (KS) test (Wang et al. 2011) at the 95% confidence significance level. Worm plot of the residuals from the fitted model to the water quality data is also employed to test the goodness-of-fit (Van Buuren & Fredriks 2001). The final marginal distribution and copula models are selected by comparing the value of the Corrected Akaike information criterion (AIC), which is the widely used evaluation method of goodness-of-fit (Akaike 1974). The AIC value for a model can be calculated as:

\[
AIC = N \ln \left( \frac{1}{N - M} \sum_{i=1}^{N} (p_{ei} - p_i)^2 \right) + 2M
\]

where \( p_{ei} \) and \( p_i \) represent the empirical and theoretical probabilities, respectively, \( N \) is the sample size of water quality series, and \( M \) is the number of the estimated parameters of time-varying marginal distribution and time-varying copula model.

RESULTS AND DISCUSSION

The joint transition probability of NH\(_3\)-N and COD\(_{\text{Mn}}\)

According to Equations (1) and (2) and Table 2, the joint transition probability of water quality state combing with NH\(_3\)-N and COD\(_{\text{Mn}}\) are calculated based on the Markov Chin theory (seen in Table 3). Thus, a typical water quality

### Table 3 | Transition probability of NH\(_3\)-N and COD\(_{\text{Mn}}\)

<table>
<thead>
<tr>
<th>Initial state</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>V(_w)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.25</td>
<td>0.27</td>
<td>0.14</td>
<td>0.19</td>
<td>0.08</td>
<td>0.07</td>
</tr>
<tr>
<td>II</td>
<td>0.19</td>
<td>0.02</td>
<td>0.20</td>
<td>0.28</td>
<td>0.19</td>
<td>0.13</td>
</tr>
<tr>
<td>III</td>
<td>0.08</td>
<td>0.20</td>
<td>0.28</td>
<td>0.25</td>
<td>0.01</td>
<td>0.17</td>
</tr>
<tr>
<td>IV</td>
<td>0.04</td>
<td>0.15</td>
<td>0.31</td>
<td>0.11</td>
<td>0.03</td>
<td>0.36</td>
</tr>
<tr>
<td>V</td>
<td>0.05</td>
<td>0.12</td>
<td>0.10</td>
<td>0.15</td>
<td>0.31</td>
<td>0.27</td>
</tr>
<tr>
<td>V(_w)</td>
<td>0.15</td>
<td>0.06</td>
<td>0.09</td>
<td>0.19</td>
<td>0.16</td>
<td>0.36</td>
</tr>
</tbody>
</table>
environment situation in state 1, Class I, has a probability of 0.27 of being Class II from now, a probability of 0.25 being still Class I, a probability of 0.19 being Class IV, a probability 0.14 of being Class III, a probability 0.08 and 0.07 of being Class V and Class Vw, respectively. Class II state has the largest joint transition probability. Similar results of transition probability from different initial states are given in Table 3.

**Time-varying marginal distributions**

**Stationary marginal distributions**

Under stationary assumption, the marginal distribution parameters of NH$_3$-N and COD$_{Mn}$ series are estimated at the 95% confidence significance level as shown in Table 4. The AIC values of five fitting models for NH$_3$-N and COD$_{Mn}$ series under stationary assumption indicate that P-III distribution model performs better goodness-of-fit for two series (Figure 3). The results of fitting water quality data also show that Gumbel distribution and logistic distribution have the largest AIC values than the other three distributions (P-III distribution, t family distribution and normal distribution).

<table>
<thead>
<tr>
<th>Series</th>
<th>P-III</th>
<th>Logistic</th>
<th>t Family</th>
<th>Gumbel</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH$_3$-N</td>
<td>1357.58</td>
<td>1400.70</td>
<td>1379.74</td>
<td>1405.85</td>
<td>1377.74</td>
</tr>
<tr>
<td>COD$_{Mn}$</td>
<td>689.14</td>
<td>694.50</td>
<td>690.92</td>
<td>771.30</td>
<td>689.19</td>
</tr>
</tbody>
</table>

**Time-varying marginal distributions**

*The explanatory variable of time.* The results of time-varying marginal distributions with time as the explanatory variable of the distribution parameters at the 95% confidence significance level are summarized in Table 5. According to AIC values, Gumbel distribution and P-III distribution are the best fitting model for NH$_3$-N and COD$_{Mn}$ series, respectively. For describing the nonstationary NH$_3$-N series, the location and scale parameter of Gumbel distribution both suggest a downward trend. As for the nonstationary COD$_{Mn}$ series, the location parameter of P-III distribution is constant, while its scale parameter presents an upward trend. Besides, compared to the constant stationary model, the results show the time-varying nonstationary model has a better fitness to NH$_3$-N and COD$_{Mn}$ data. As we know, location parameter of distribution models presents a mean situation. The results of NH$_3$-N and COD$_{Mn}$ presenting a downward trend, are consistent with the real situation which large-scale pollution control measures have been implemented in HRB since the 1990s.

*The explanatory variable of land cover index.* Similarly as Table 5, Table 6 shows the statistics results of time-varying marginal distribution models, taking the land cover index that percent land cover of farmland and urban blocks as the explanatory variable at the 95% confidence significance level. According to the AIC values, Gumbel distribution and P-III distribution are the best fitting model for NH$_3$-N and COD$_{Mn}$ series, respectively. For describing the nonstationary NH$_3$-N series, the location and scale parameter of Gumbel distribution both suggest a downward trend with the explanatory variable of land cover index. On the contrary, the location and

![Figure 3](https://iwaponline.com/wst/article-pdf/75/3/693/455784/wst075030693.pdf)
scale parameter of P-III distribution for nonstationary CODMn series both show an upward trend. Comparing with the constant stationary model in Table 4, the results of the AIC values in Table 6 also show the nonstationary model has a better fitness to NH3-N and CODMn data. Also, it is indicated that, in terms of the AIC values, the time-varying marginal distribution with time as explanatory variable is better than that with land cover index as the explanatory variable in modelling the both two nonstationary water quality series. As shown in Figure 4, the goodness-of-fit tests of two nonstationary marginal models illustrate a good fitting performance.

Time-varying copula-based combined risk assessment

Time-varying copula

Based on the above analysis, it indicates that the parameters of marginal distributions are time-varying, and P-III distribution is selected to construct the time-varying joint distribution function. Using the explanatory variable time to describe dependence structure compared with the constant parameter, the statistical results of three popular copula models, i.e. GH, Clayton and Frank at the 95% confidence significance level are summarized in Table 7. As Table 7 shows, Frank copula has a better modelling to the dependence structure between NH3-N and CODMn, and followed by Clayton and GH copula for constant parameter copula. In time-varying copula, the dependence parameter varies with the time and presents a downward trend. Similar with the constant parameter copulas, time-varying copulas all have better modelling of the dependence structure between NH3-N and CODMn. The downward trend of dependence parameter suggests that the dependence

| Table 5 | Statistical results of five time-varying marginal distributions with the explanatory variable time at the 95% confidence significance level (t = 1, 2, ... 348) |
|---------|---------------------|---------------------|---------------------|---------------------|
| Series  | Distributions       | Parameters           | AIC                |
| NH3-N   | P-III               | \exp(0.69 - 0.0022)  | 1.06               | 1188.29             |
| Logistic| 1.08 - 0.0028       | \exp(-0.13 - 0.0011) | -                 | 1200.18             |
| t Family| 0.98 - 0.00261      | \exp(0.39 - 0.00095) | 16.71              | 1193.47             |
| Gumbel  | 1.66 - 0.0031t      | \exp(0.17 - 0.00086) | -                 | 1157.52             |
| Normal  | 0.97 - 0.00261      | \exp(0.39 - 0.00095) | -                 | 1191.47             |
| CODMn   | P-III               | 6.23                | 0.86               | 528.13              |
| Logistic| 7.10                | \exp(-0.75 + 0.00111) | -                | 531.21              |
| t Family| 6.89                | \exp(-0.78 + 0.00018) | 5.36              | 530.14              |
| Gumbel  | 8.94                | \exp(-0.75 - 0.00022) | -                | 569.86              |
| Normal  | 6.89                | \exp(-0.77 + 0.00016) | -                | 530.30              |

| Table 6 | Statistical results of five time-varying marginal distributions with the explanatory variable of percent land cover of farmland and urban blocks at the 95% confidence significance level (0 < \theta < 1) |
|---------|---------------------|---------------------|---------------------|---------------------|
| Series  | Distributions       | Parameters           | AIC                |
| NH3-N   | P-III               | \exp(9.31 - 10.84l)  | 1.348              | 1187.98             |
| Logistic| 10.88l              | \exp(3.22 - 4.22l)   | -                 | 1214.69             |
| t Family| 10.94l              | \exp(3.22 - 3.82l)   | 13.35              | 1201.41             |
| Gumbel  | 12.89l              | \exp(2.32 - 2.74l)   | -                 | 1179.09             |
| Normal  | 10.95l              | \exp(3.42 - 3.82l)   | -                 | 1199.41             |
| CODMn   | P-III               | \exp(0.69 - 0.00022) | 1.158              | 575.36              |
| Logistic| 1.08 - 0.00028      | \exp(-0.13 - 0.00039) | -                | 585.69              |
| t Family| 1.08 - 0.00038      | \exp(-0.75 + 0.000111) | 5.36              | 531.21              |
| Gumbel  | 1.18 - 12.89l       | \exp(-0.78 + 0.00023) | -                | 530.14              |
| Normal  | 1.18 - 12.95l       | \exp(-0.75 - 0.00022) | -                | 569.86              |

Downloaded from https://iwaponline.com/wst/article-pdf/75/3/693/455784/wst075030693.pdf by guest
between NH$_3$-N and COD$_{Mn}$ will be weakened as the time varies.

**Combined risk assessment**

According to the classification of water quality parameters and time-varying copula, the encounter risk probability combining with NH$_3$-N and COD$_{Mn}$ are calculated. As shown in Table 8, it indicates the following. (1) Synchronous encounter risk probability (SERP) of NH$_3$-N and COD$_{Mn}$, simultaneously reaching Class II, Class III, Class IV, Class V are 15.92%, 30.42%, 43.23% and 50.61%, respectively. The largest SERP indicates that BBS faced serious pollution in the past 30 years. (2) Asynchronous encounter risk probability (AERP) of NH$_3$-N and COD$_{Mn}$ (reaching Class II, Class III, Class IV and Class V) shows that it is largest when NH$_3$-N is inferior to class V standard and COD$_{Mn}$ is inferior to class IV standard, and followed by NH$_3$-N (Class IV) and COD$_{Mn}$ (Class V). (3) The encounter risk probability of NH$_3$-N (Class III) and COD$_{Mn}$ (Class IV) and NH$_3$-N (Class III) and COD$_{Mn}$ (Class V) are 38.20% and 40.53%, respectively. Thus if the COD$_{Mn}$ concentration discharge is limited in the upstream of BBS when NH$_3$-N concentration is superior to Class III, the compliance rate of water quality will be improved to 7.78% and 10.11%, respectively.

As Figure 5 shows, the contour plot of encounter risk probability between NH$_3$-N and COD$_{Mn}$ simultaneously reaching Class II, Class III, Class IV, Class V are 15.92%, 30.42%, 43.23% and 50.61%, respectively. The largest SERP indicates that BBS faced serious pollution in the past 30 years. (2) Asynchronous encounter risk probability (AERP) of NH$_3$-N and COD$_{Mn}$ (reaching Class II, Class III, Class IV and Class V) shows that it is largest when NH$_3$-N is inferior to class V standard and COD$_{Mn}$ is inferior to class IV standard, and followed by NH$_3$-N (Class IV) and COD$_{Mn}$ (Class V). (3) The encounter risk probability of NH$_3$-N (Class III) and COD$_{Mn}$ (Class IV) and NH$_3$-N (Class III) and COD$_{Mn}$ (Class V) are 38.20% and 40.53%, respectively. Thus if the COD$_{Mn}$ concentration discharge is limited in the upstream of BBS when NH$_3$-N concentration is superior to Class III, the compliance rate of water quality will be improved to 7.78% and 10.11%, respectively.

As Figure 5 shows, the contour plot of encounter risk probability between NH$_3$-N and COD$_{Mn}$ is given. Taking the NH$_3$-N concentration 1.0 mg/L and COD$_{Mn}$ concentration 6.0 mg/L for example, the connection point of two red dotted lines represents the probability of the condition that NH$_3$-N concentration less than 1.0 mg/L and COD$_{Mn}$ concentration less than 6.0 mg/L. Also, it can be seen that the contour lines is denser between the probability value 0 and 0.40 than the others.

### Table 8 | Combined risk assessment of NH$_3$-N and COD$_{Mn}$ exceeding the standard thresholds

<table>
<thead>
<tr>
<th>Combined risk probability (%)</th>
<th>NH$_3$-N</th>
<th>COD$_{Mn}$</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>II</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Class</td>
<td>II</td>
<td>III</td>
<td>IV</td>
<td>V</td>
</tr>
<tr>
<td>I</td>
<td>15.92</td>
<td>18.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>24.77</td>
<td>30.42</td>
<td>33.51</td>
<td>33.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>29.93</td>
<td>38.20</td>
<td>43.23</td>
<td>46.32</td>
<td>46.84</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>31.35</td>
<td>40.53</td>
<td>46.32</td>
<td>50.61</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### CONCLUSIONS

1. The first-order Markov joint transition probabilities of water quality state combining with NH$_3$-N and COD$_{Mn}$...
were different depend on initial water quality state. The greater first-order Markov joint transition probability of two water quality indicators tends to Class Vw, Class IV and Class III state which indicated that would occur easily in the water body of BBS, so the local government should make some effective pollution control measures for improving water quality conditions.

(2) The non-stationarity had been revealed both in marginal distributions and copula models between the NH3-N and CODMn series. Time and land cover index were employed as the explanatory variable for estimating the nonstationary distribution parameters. In terms of the AIC values, the time-varying marginal distribution with the explanatory variable time is better than that with land cover index as the explanatory variable in modelling both nonstationary water quality series. Water quality variation was not only affected by land use, but also the other influence factors such as the intensive dams and floodgates, water temperature, topography and extreme events and so on.

(3) Whatever time-varying copula or constant parameter copula, Frank copula had a better performance for describing the dependence structure. In modelling the dependence structure changes, time-varying copula had an obvious advantage compared to constant parameter copula. Besides, the downward trend presented in dependence parameter illustrated that the dependence between NH3-N and CODMn may weaken with time.

(4) The encounter risk were quantitatively assessed based on time-varying copula. The largest SERP was 50.61\% while they simultaneously reached Class V. The AERP was largest when NH3-N and CODMn was inferior to Class V and Class IV standard, respectively. It suggested that the BBS of HRB still faced severe water quality risk, although a lot of pollution control plans had been implemented.

ACKNOWLEDGEMENTS

This research was supported by the National Grand Science and Technology Special Project of Water Pollution Control and Improvement (No. 2014ZX07204-006).

REFERENCES


El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R. & Bobée, B. 2007 Generalized maximum likelihood estimators for the

Environmental Quality Standards for Surface Water of People’s Republic of China (GB 3838-2002); 2002 [in Chinese].


Genest, C. & Favre, A. C. 2007 Everything you always wanted to know about copula modeling but were afraid to ask. *American Society of Civil Engineers* **12** (4), 347–368.


Nelsen, R. B. 2006 *An Introduction to Copulas*. Springer, New York, NY, USA.


The State Environmental Protection Standards of the People’s Republic of China: water quality-determination of permanganate index (GB 11892-1989); 1989 [in Chinese].


First received 12 June 2016; accepted in revised form 17 November 2016. Available online 30 November 2016.