Establishment of Water Quality Prediction Model in Pingdingshan Region Based on Stochastic Theory - Markov Process

KeKe Zou1,*, ZhongYuan Li2, Kai Xu3, XiaoLing Mu3 and TieSheng Li4

1 Pingdingshan Hydrology and Water Resources Survey Bureau of Henan Province, Pingdingshan 467000, P. R. China
2 Hydrology and Water Resources Bureau of Henan Province, Zhengzhou 450003, P. R. China
3 Zhengzhou Hydrology and Water Resources Survey Bureau of Henan Province, Zhengzhou 450000, P.R. China
4 The College of Chemistry and Molecular Engineering, Zhengzhou University, Zhengzhou 450052, P. R. China

*Corresponding author e-mail: zkk13619891880@163.com

Abstract. The stochastic theory - Markov process is used to predict the water quality of Zhan River in Pingdingshan City, and the water quality monitoring data of Zhan River in Pingdingshan City over the past ten years are used to verify the reliability of the model, which has provided some theoretical basis for the management decision team in the Pingdingshan region to make the water quality clear in Huai River Basin in 2010.

Keywords: Stochastic Theory - Markov Process, Pingdingshan Region, Prediction, Transition Probability Matrix

1. Introduction
Zhan River is a polluted river in Pingdingshan City. After leaving Pingdingshan City, it merges into the Shahe River. The Shahe River is the upper tributary of the Yinghe River, and the Yinghe River is the largest tributary of the Huai River. Therefore, the quality of the Zhan River in 2010 will directly affect the country's goal of making the water of the Huai River clear [1-2].

This paper uses stochastic theory-Markov process theory to study the changing trend of Zhan River water quality around 2000 and establishes a transition probability matrix of Zhan River water quality
change state to predict the probability of Zhan River water quality status at a specific time in the future and verify the prediction model.

2. Water Quality Prediction Modeling
The stochastic theory-Markov process is a stochastic process that studies the state of an event and the patterns of transition between states [3]. It studies the change trend of the state at time \((\Delta t + t_0)\) by analyzing the initial probability of the different states of the event at time \(t_0\) and the transition probability relationship between states. The river water quality has a specific range of random variability, and its change pattern is related to a certain state now. Hence, it conforms to the random theory-Markov process change law. The state of river water quality is the type of water quality. For example, the water body is divided into 4 levels: clean, light pollution, pollution, and severe pollution. At this point, the state space \(E = \{1, 2, 3, 4\}\), where level 1 means clean, level 2 indicates light pollution, level 3 indicates pollution, and level 4 indicates heavy pollution [4].

In the Markov chain, the state transition probability is only related to the transition starting state \(i\), the number of transition steps \(k\), and the transition to state \(j\), and has nothing to do with the starting point of the transition. The transition probability is denoted as \(P_{ij}(k)\), as shown in equation (1):

\[
P_{ij}(k) = P_i(n, n + k) = P\{X(n + k) = j | X(n) = i\}, k \geq 1
\]

(1)

Where \(X(n)\) is the observed value at the corresponding time \(t_n\).

When \(k = 1\), \(P_{ij}(1)\) is called the one-step transition probability, referred to as \(P_{ij}\). At this time, \(P_{ij} = P_i(1) = P\{X(n + 1) = j | X(n) = i\}\). For finite state space \(E = \{1, 2, 3, \ldots, N\}\), there is the following one-step state transition matrix.

\[
P = \begin{bmatrix}
P_{11} & P_{12} & \cdots & P_{1N} \\
P_{21} & P_{22} & \cdots & P_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
P_{N1} & P_{N2} & \cdots & P_{NN} \\
\end{bmatrix}
\]

(2)

Where \(P_{ij} = n_{ij}/n_i\), in the equation, from time \(n_i\) to time \(n_i + 1\), the number of samples from state \(i\) to state \(j\), \(n_i\) the total number of samples from state \(i\), \(n_i = \sum_{j=1}^{\infty} n_{ij}\).
Assuming that the probability of being in state i at t=0 is \( P_i(0) \), the probability of transitioning from state i to state j is \( P_{ij} \), and the probability of being in state j at t=1 is \( P_j(1) \), then \( P_j(1) = P_i(0) \times P_{ij} \)

, similarly, \( P_j(n+1) = P_j(n) \times P_{ij} \).

For the environmental quality division in question, where 1, 2, 3, \ldots, \( N \) states at t=n, the probability of transitioning to state j at t=n+1 is as shown in equation (3):

\[
P_j(n+1) = P_j(n) \times P_{ii} + P_i(n) \times P_{ij} + \cdots + P_N(n) \times P_{nj} = \sum_{i=1}^{N} P_i(n) \times P_{ij} \tag{3}
\]

The matrix representation is shown in equation (4):

\[
P(n+1) = P(n) \times \begin{bmatrix}
P_{11} & P_{12} & \cdots & P_{1N} \\
P_{21} & P_{22} & \cdots & P_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
P_{N1} & P_{N2} & \cdots & P_{NN}
\end{bmatrix} = P(n) \times P \tag{4}
\]

Where \( P(n) = [P_1(n), P_2(n), \ldots, P_N(n)] \), \( P(n+1) = [P_1(n+1), P_2(n+1), \ldots, P_N(n+1)] \).

Based on the n-step transition probability equation, there are \( P(n) = P(0) \cdot P^n \), where \( P \) is a one-step transition probability matrix. The above equation is referred to as a Markov chain comprehensive water quality prediction model. It can predict water quality in a specific state.

Based on the monitoring data from the Pingdingshan City Environmental Monitoring Center in the past ten years on the Zhan River out of the Pingdingshan urban section of Hanzhuang Bridge, the comprehensive water quality prediction model of the Markov chain is used to predict the water quality of the Zhan River. The monitoring items include pH, DO, COD, BOD5, nitrous nitrogen, nitrous nitrogen, non-ionic ammonia, volatile phenol, cyanide, arsenic, mercury, chromium, lead, and cadmium. The monitoring data are calculated based on the dry season and the wet season, and the Nemero Water Quality Index is used for evaluation according to the “Surface Water Environmental Quality Standards” (GHZB1-1999)\(^{[5-6]}\).

Nemero water quality index is shown in equation (5):

\[
P_S_i = \frac{\left(S_i\right)_{\text{maximum}}^2 + \left(S_i\right)_{\text{average}}^2}{2}
\]

Where \( S_i : i \) pollutant single item water quality standard index.

The internal Merlot indexes of water quality in dry and wet seasons are calculated. Subsequently, the annual surface water chemical pollution index is calculated. The statistical results are shown in Table 1. Table 2 shows the water environment quality classification. The comprehensive water quality
pollution index divides water quality based on Table 2. At this time, the state space \(E = \{1, 2, 3, 4\}\), that is, \(N = 4\). Level 1 indicates cleanliness, Level 2 indicates light pollution, Level 3 indicates pollution, Level 4 indicates severe pollution.

Based on the data in Tables 1 and 2, the water quality grades for each year are divided, as shown in Table 3.

**Table 1.** Statistics results of pollution index of Zhan River water from 2009 to 2019

| Years     | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|-----------|------|------|------|------|------|------|------|------|------|------|------|
| Dry season| 1.12 | 5.6  | 1.8  | 4.7  | 15.8 | 5.9  | 14.6 | 7.4  | 4.5  | 4.3  | 7.2  |
| Flood season| 5.9  | 3.7  | 1.3  | 8.1  | 1.4  | 4.5  | 7.2  | 8.5  | 4.4  | 2.7  | 2.4  |
| Annual synthesis| 2.7  | 4.8  | 1.8  | 5.8  | 11.1 | 5.4  | 12.1 | 7.9  | 4.3  | 3.6  | 5.5  |

**Table 2.** Water environment quality classification

| Pollution degree classification | Clean | Light pollution | Pollution | Heavy pollution |
|---------------------------------|-------|-----------------|-----------|-----------------|
| Comprehensive Water Quality    | <1    | 1~5             | 5~10      | >10             |
| Pollution Index                 |       |                 |           |                 |

**Table 3.** Status of water quality from 2009 to 2019

| Year | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|------|------|------|------|------|------|------|------|------|------|------|------|
| Level| 2    | 2    | 3    | 4    | 3    | 4    | 5    | 2    | 2    | 2    | 3    |

According to Table 3, count the number of times each state appears, and then count the number of times that other states appear after each state. In this way, the one-step transition rule between states is obtained, as shown in Table 4.

**Table 4.** One-step transition rules for each state

| State \(i\) | 1 | 2 | 3 | 4 | \(n_i\) |
|-------------|---|---|---|---|--------|
| 1           | 0 | 0 | 0 | 0 | 0      |
| 2           | 0 | 3 | 2 | 0 | 5      |
| 3           | 0 | 1 | 0 | 2 | 3      |
| 4           | 0 | 0 | 2 | 0 | 2      |

The state transition probability \(P_{ij} = \frac{n_{ij}}{n_i} (i, j = 1, 2, 3, 4)\) is calculated based on Table 4, and obtain the one-step transition probability matrix of each state of Zhan River water quality as shown in equation (6):

\[
P = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0.600 & 0.400 & 0 \\
0 & 0.333 & 0 & 0.667 \\
0 & 0 & 1 & 0
\end{bmatrix}
\]

From the \(n\)-step transition probability equation \(P(n) = P(0) \times (P)^{n-1}\), the transition probability from state \(i\) to step \(j\) through step \(1, 2, \cdots, n\) can be obtained.

The Markov chain integrated water quality prediction model is used to predict the water quality of Zhan River water from 2013 to 2016. Based on 2012, since the water quality status in 2012 is 3, its
state vector is expressed as \( P(0) = (0, 0, 1, 0) \), using the model to recur in 2011, the state probability vector of the river water quality is as shown in equation (7):

\[
P(1) = P(0)P = (0010) \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.600 & 0.400 & 0 \\ 0 & 0.333 & 0 & 0.667 \\ 0 & 0 & 1 & 0 \end{bmatrix} = (00.333, 0.667) \quad (7)
\]

The state probability vector in 2013 is shown in equation (8):

\[
P(2) = P(1)P = P(0)P^2 = (0010) \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.600 & 0.400 & 0 \\ 0 & 0.333 & 0 & 0.667 \\ 0 & 0 & 1 & 0 \end{bmatrix}^2 = (00.200, 0.800, 0)
\]

Similarly, the state probability vectors for 2010 and 2011 can also be derived, as shown in Table 5.

**Table 5. Probability vector of Zhan River water prediction status from 2010 to 2013**

| Prediction year | P(n) | 1     | 2     | 3     | 4     | Maximum probability state |
|-----------------|------|-------|-------|-------|-------|--------------------------|
| 2010            | P(1) | 0     | 0.345 | 0     | 0.665 | 4                        |
| 2011            | P(2) | 0     | 0.203 | 0.802 | 0     | 3                        |
| 2012            | P(3) | 0     | 0.387 | 0.083 | 0.533 | 4                        |
| 2013            | P(4) | 0     | 0.257 | 0.684 | 0.154 | 3                        |

**3. Model Verification and Prediction**

To verify the reliability of the prediction of the model, 2010 was selected as the benchmark, and the state of water quality change from 2011 to 2014 was reversed and compared with the actual value.

**Table 6. Comparison of predicted and actual values from 2010 to 2014**

| Year | P(n) | State probability vector | Actual state | Comparison of the predicted maximum probability state and the actual state |
|------|------|--------------------------|--------------|--------------------------------------------------------------------------|
| 2010 | P(1) | 0.601 0.402 0            | 2            | Consistent                                                               |
| 2011 | P(2) | 0.494 0.244 0.268        | 2            | Consistent                                                               |
| 2012 | P(3) | 0.375 0.464 0.162        | 3            | Consistent                                                               |
| 2013 | P(4) | 0.382 0.312 0.313        | 4            | Basically, the same                                                     |
| 2014 | P(5) | 0.331 0.441 0.206        | 3            | Consistent                                                               |

Table 6 shows that the predicted state based on the model is compared with the actual state. Except for a slight deviation in 2013, the other years are completely consistent, but the probability of states 2, 3, and 4 in 2012 is basically the same, so the predicted state and the actual state are also basically the
same. It can be seen that the prediction model established by using the Markov chain to the water of the Zhan River has high reliability and is entirely feasible.

An analysis of the prediction results for 2010-2013 in Table 5 shows that the water quality of the Zhan River in 2010-2013 was in a state of pollution and severe pollution, which failed to meet the national water quality requirements of the Huai River Basin. Therefore, to improve the water quality of the Zhan River, the usual management measures are not enough. More powerful measures must be taken. This has attracted great attention from the Pingdingshan Municipal Government and environmental protection departments. The establishment of the sewage treatment plant is intensifying. After the project is completed, the Zhan River water will be significantly improved, and a new prediction model will be established for water quality in Zhan River.

4. Conclusions
As the stochastic theory-Markov process is a random process without considering the influence of other factors, it has its convenience in use. However, under special circumstances, the model will change in the presence of significant factors. Therefore, the model should be used to make predictions based on a thorough understanding of the local situation. Select factors that have little effect on the time of the prediction. The establishment of the model based on data for a period can improve prediction accuracy.

References
[1] Katarina Juselius. Using a Theory-Consistent CVAR Scenario to Test an Exchange Rate Model Based on Imperfect Knowledge[J]. Discussion Papers, 2017, 5(3):30-43.
[2] Emel Savku, Gerhard-Wilhelm Weber. A Stochastic Maximum Principle for a Markov Regime-Switching Jump-Diffusion Model with Delay and an Application to Finance[J]. Journal of Optimization Theory & Applications, 2018, 179(2):696-721.
[3] Rajbir Kaur, Navroop Kaur, Sandeep K. Sood. Security in IoT network based on stochastic game net model[J]. International Journal of Network Management, 2017, 27(4):1-13.
[4] Yongbao Wu, Bingdao Chen, Wenzhe Li. Synchronization of stochastic coupled systems via feedback control based on discrete-time state observations[J]. International Journal of Systems Science, 2017, 26(2):68-85.
[5] Meiping Yun, Wenwen Qin, Xiaoguang Yang. Estimation of urban route travel time distribution using Markov chains and pair-copula construction [J]. Transportmetrica B, 2019, 7(1):1521-1552.
[6] Tomer Shushi. Randomness in modified General Relativity theory: The stochastic f(R) gravity model[J]. Canadian Journal of Physics, 2018, 96(11):1-12.