The 18th Biennial Conference of International Society for Ecological Modelling

The Over Polluted Water Quality Assessment of Weihe River Based on Kernel Density Estimation

Z.M. Zhang\textsuperscript{a}, X.Y. Wang\textsuperscript{a,}\textsuperscript{*}, Y. Zhang\textsuperscript{a}, Z. Nan\textsuperscript{a}, B.G. Shen\textsuperscript{b}

\textsuperscript{a}College of Resources, Environment & Tourism, Capital Normal University, Beijing 100048, China
\textsuperscript{b} Shaanxi Research and Design Institute for Environmental Science, Xi’an 710054, China

Abstract

Considering the classification assessment of water quality can not fairly reflect the over pollution degree of the water body, this article puts forward a measurement on over pollution based on kernel density estimation, to evaluate water quality of the main stream of Weihe River with the monitoring data of its rainy, dry and normal seasons from 2000 to 2008. Through the comparison between the assessment result by a Standard Category Index Method, and the standard of water function zoning plan, the space distribution of water function zoning is found to be inconsistent with that of status quo of water quality, therefore, from the spatial dimension, the method of PCA-CA is used to classify the water quality situation of 13 existing monitoring sites, and according to ANOVA, the cluster result can significantly distinguish the difference between the actual land use of waterfront; from the time dimension, rainy season is comparatively different from dry season and normal season with these three seasons as time standard which is often used to differentiatate the quantity of flow. This article discusses the characteristic of water quality overload of Weihe River in Shaanxi with the change of seasons, and different clusters of area, also, concludes the over pollution extent of three areas in different seasons based on the classification result from the space and time dimension. The result shows that the biggest overload extent happens in Area1 and Area3 during dry seasons, and Area 2 during normal seasons; overally considering with the 4 indicators, NH\textsubscript{3}-N and BOD\textsubscript{5} are the most seriously polluted indicators in Weihe river.

© 2011 Published by Elsevier B.V. Selection and/or peer-review under responsibility of School of Environment, Beijing Normal University. Open access under CC BY-NC-ND license.

keywords: kernel density estimates; Weihe river; water quality assessment; PCA-CA; inappropriate degree

* Corresponding author. Tel: 86-10-68903968; fax: 86-10-68902332.
E-mail address: cnuwxy@sohu.com.
1. Introduction

Water quality assessment methods with multi-indicator mainly include standard category index
method, fuzzy comprehensive assessment, projection pursuit method, grey system theory, artificial neural
network and probabilistic method, etc. Apart from probabilistic method, these approaches all have water
quality classified based on mean value or mid-value [1]. On one hand, these approaches can not fairly
differentiate water quality of the same sort, especially over-polluted water body, of which the difference
between actual degree of contamination can not be presented; on the other hand, mean value and mid-
value lack of representativeness with regard to water body with a large rangeability of water quality.

At present, there are commonly two indexes to evaluate out-of-limits water quality: exceeding rate and
exceeding times. Exceeding rate is mainly used to reflect duration of water body pollution, while
exceeding times primarily shows the pollution intensity. However, multi-factor overall assessment based
on duration and intensity of over polluted water body is comparatively fewer.

Considering Shaanxi section of Weihe River has seriously exceeded pollution standard with a large
runoff variation, this article reflects pollution degree of the water body by kernel density estimation and
puts forward an approach to evaluate over polluted water body on the basis of kernel density estimation to
accurately quantize the differences between each segment and provide objective assessment in support of
water quality management decision.

2. Overview of study area

Weihe River is the only passageway to receive and let off effluent sewage in the area of the central
Shaanxi plain. Because water pollution treatment facility construction has much lagged behind, a large
amount of untreated or ineffectively treated industrial waste water and urban sewage directly enters into
watercourse, which makes reach below Baojixia, main stream of Weihe River has basically been deprived
of water function. Water function of Shaanxi section of Weihe River includes agricultural water,
landscape water, unloading control, transition zone and industrial water, etc. According to mean value of
four monitoring indicators of COD, DO, BOD$_5$ and NH$_3$-N in nine years from 2000 to 2008 in
13monitoring stations of main streams of Weihe River, only Linjia Village reached the standard of water
quality while the rest of the segments were all worse than Grade V, which is to say, water quality
exceeded the standard badly, failing to reach the water quality standard of water function regionalization
and Weihe River was badly polluted. (Table1)
Table 1. The water quality condition and demand of stream function in different sites.

<table>
<thead>
<tr>
<th>City</th>
<th>No.</th>
<th>Site</th>
<th>Function Zone</th>
<th>Water Function Regionalization Grade</th>
<th>Actually Attained Grade</th>
<th>Indicators that Exceed the Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baoji</td>
<td>1</td>
<td>Linjia Village</td>
<td>agriculture</td>
<td>3</td>
<td>2</td>
<td>none</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Wolongsi Bridge</td>
<td>landscape</td>
<td>4</td>
<td>&gt;5</td>
<td>BOD5, NH3-N</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Guozhen Bridge</td>
<td>unloading control</td>
<td>4</td>
<td>&gt;5</td>
<td>all</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Changxing Bridge</td>
<td>transition, industry, agriculture</td>
<td>3</td>
<td>&gt;5</td>
<td>all</td>
</tr>
<tr>
<td>Xianyang &amp; Xi’an</td>
<td>5</td>
<td>Xingping</td>
<td>agriculture, landscape, industry</td>
<td>3</td>
<td>&gt;5</td>
<td>all</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Nanying</td>
<td>industry</td>
<td>4</td>
<td>&gt;5</td>
<td>all</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Xianyang Railway Bridge</td>
<td>Landscape, unloading control</td>
<td>4</td>
<td>&gt;5</td>
<td>all</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Tianjiangrendu</td>
<td>transition</td>
<td>4</td>
<td>&gt;5</td>
<td>all</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Gengzhen Bridge</td>
<td>transition</td>
<td>4</td>
<td>&gt;5</td>
<td>all</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Xinfeng Bridge</td>
<td>agriculture</td>
<td>4</td>
<td>&gt;5</td>
<td>all</td>
</tr>
<tr>
<td>Weinan</td>
<td>11</td>
<td>Shawangdu</td>
<td>agriculture</td>
<td>4</td>
<td>&gt;5</td>
<td>all</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Shuyuan</td>
<td>agriculture</td>
<td>4</td>
<td>&gt;5</td>
<td>all</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>Tongguan Suspension Bridge</td>
<td>agriculture</td>
<td>4</td>
<td>&gt;5</td>
<td>all</td>
</tr>
</tbody>
</table>

>5 means the water quality is worse than Grade 5

3. Data source

Monitoring data of 13 monitoring sites (Fig. 1) in Shaanxi province during rainy, normal and dry seasons from 2000 to 2008 is selected with COD, DO, BOD5 and NH3-N as monitoring indicators. Thereinto, several vacant data in dry season is complemented by linear interpolation method.
4. Research method

4.1. Kernel density estimation (KDE)

Kernel density estimation, one of the nonparametric testing methods, is used to estimate unknown density function in probability theory. It was put forward by Rosenblatt and Emanuel Parzen, so it is also named as Parzen window. Ruppert and Cline put forward emendatory kernel density estimation based on data set density function clustering algorithm. The KDE can be easily achieved in Matlab7.

4.2. Spatial variation of water quality

4.2.1. Data normalization

The concentration range of water quality indicators of Shaanxi section of Weihe River shows in Fig. 2. The concentration difference of COD is much larger than the other three water quality indicators. To consider comprehensively the water quality characteristics of each segment, the differences between
dimensions should be eliminated and the data needs to be normalized with equation: \( S = S / \text{std}(S) \), where \( \text{std} \) is a function to calculate standard deviation.

![Concentration range of water quality indicators](image1)

**Fig. 2** Concentration range of water quality indicators

### 4.2.2. Monitor sites type distinguishment with PCA-CA

Since there’s certain information redundancies in selected water quality indicators, therefore there’s a need of dimension reduction processing. Principal component analysis is used here to reduce the dimensions with the two principal component variables replacing the original four to do clustering analysis (accumulative contribution rate is 97.19\%) and classify sampling points by the complete linkage method[3]. (See Fig. 3).

![Cluster dendrogram of sampling sites based on water quality indexes of Weihe River](image2)

**Fig. 3** Monitor site type distinguishment with PCA-CA based on water quality indicators of Weihe River

As shown in Fig. 3, segments represented by monitoring sites of main stream can be divided into 3 clusters: site 1-4 belong to the Cluster 1 area, within the municipal district of Baoji, with a low population density and few industrial pollution; 5-10 belong to the Cluster 2 area, within Xi’an and Xianyang city, with a high population density and many industrial point sources, mainly paper mill, juice beverage processing factory, etc. Therefore, it is greatly affected by industrial pollution load and living waste; 11-13 belongs to the Cluster 3 area, with a low population density and few industrial point source. It's mainly agricultural, and there’s comparatively little pollutants discharged from upstream, therefore, it can be seen as an area to degrade pollutants in non-rainy season. According to differences between sample distances from clustering analysis result, Cluster 1 area is distinctly different from Cluster 2 and 3 area, while similarity between Cluster 2 area and Cluster 3 area tells that Cluster 2 area brings a distant change to water quality of the main stream, and the pollutants are not easy to degrade.
Table 2 The result of ANOVA

<table>
<thead>
<tr>
<th>Water quality indicator</th>
<th>F</th>
<th>significant or not</th>
</tr>
</thead>
<tbody>
<tr>
<td>COD</td>
<td>63.367</td>
<td>significant</td>
</tr>
<tr>
<td>DO</td>
<td>153.424</td>
<td>significant</td>
</tr>
<tr>
<td>BOD₅</td>
<td>79.001</td>
<td>significant</td>
</tr>
<tr>
<td>NH₃-N</td>
<td>38.933</td>
<td>significant</td>
</tr>
</tbody>
</table>

From spatial variation of DO (positive indicator) and COD (negative indicator) concentration mean value of 13 sites from 2000 to 2008 (Fig. 4), it can be clearly seen that COD is at about the same level at 1-4 site and its concentration is close to sameness after passing site 5, which means that pollution load gets tremendously intensified after sample 4. Dilution from the abouchement of tributary to pollutants, and diffusion and degradation that main stream itself brings to the pollutants, is not enough to lessen the COD concentration in the water; however, at site 11, 12, 13, concentration of DO tends to pick up. DO content is an important indicator to decide whether the water quality is good or not. [4] Although at these sites, COD concentration has not been able to degrade much, yet it can be seen that reproduction and growth of aquatic life in the water body has already started to recover. The acceleration of oxygenolysis speed of aquatic organism and intensification of self-purifying capacity of water body can better help the degradation of the pollutants downstream.

Fig. 4 The concentration variation comparison between DO and COD

4.2.3. Result of KDE based on spatially classified samples

After KDE to samples of each cluster area, the conclusion reaches that from Fig. 5, concentration of COD, BOD₅ and NH₃-N is commonly low and on the whole there’s little rangeability of concentration change, which means a comparatively light pollution load and a smooth change. Although there’s a large wave range of monitoring value of NH₃-N concentration in Cluster 1 area, its concentration is mainly centralized between 0 and 2.5 mg/L. From the concentration probability density of four monitoring indicators, Cluster 1 area is significantly different from Cluster 2 and 3 area, while there’s fewer difference between Cluster 2 and 3 area, which agrees with the clustering analysis result.
4.3. Seasonal variation of water quality

4.3.1. Time division basis

Some research shows that there’s an asymmetrical rainfall distribution of Weihe River within a year, mostly from June to September, with precipitation reaching 57% of that in a total year. [5] According to monthly data on flows of Weihe River main stream from 1998 to 2005 obtained from hydrologic monitor site in Xianyang, the average monthly flow of dry, normal and rainy season is respectively 30.56 m$^3$/s, 32.22 m$^3$/s, 141.51 m$^3$/s. Flow sees Fig. 6.
It can be seen that flow fluctuation of Weihe River presents certain cycle along with season change. And because flow is closely related with precipitation[6], and there’s a significant seasonal difference of flow, there’s a need to discuss water quality indicator of rainy, normal and dry season respectively.

4.3.2. Result of KDE based on seasonally classified samples

Observe and study overall condition of water quality based on all monitoring samples of 13 sites of Weihe River from 2000 to 2008, and employ kernel density estimation. (See Fig. 7)

![Fig. 7 The probability density of water quality indicators in different seasons](image)

It can be seen that concentration probability density function of four monitoring indicators of normal season is similar to dry season, and is significantly different from rainy season.

5. Water quality assessment based on KDE samples

Considering water quality of most monitor sites of the main stream in the area of Shaanxi of Weihe River is far from the water function standard, classification assessment has no actual meaning to the reach. Here a concept of inappropriate degree to water body function is introduced to precisely quantize the degree that water quality fails to satisfy the water functional requirement. Because on one hand, the damage that water quality deterioration brings to water body function lies on the concentration of pollutants; on the other hand, it depends on the duration of pollution, single pollution index [7][8] (or named as quotient value method, also as exceeding times with regard to negative indicator, meaning the ratio of current pollutant concentration to prescribed limit) is now often employed to weigh the harm brought by concentration:

\[ I_m = C_m / C^c \]  \hspace{1cm} (1)
where: $C_m$ means concentration of the $m$th pollutant, and $C'_m$ serves as the prescribed limit of concentration that functional zone puts to the $m$th pollutant;

while DO does not belong to the category of pollutants, and the degree of contamination can refer to the following equation:

$$I_m = \frac{C'_m}{C_{DO}}$$

(2)

where: $C_{DO}$ means concentration of DO and $C'_m$ serves as the prescribed limit that functional zone put to DO.

According to probability density from kernel density estimation, the exposure time relating to some concentration could be estimated to get $E_m$, single factor influence of the $m$th water quality indicator in the area, together with the pollutant concentration said above:

$$E_m = \int_0^{\infty} I_m(x) \times P_m(x) dx$$

(3)

where: $I_m(x)$ means single pollution index at the time of the $m$th water quality indicator concentration being $x$. $P_m(x)$ is the corresponding probability with indicator concentration being $x$.

Inappropriate degree represents adverse impact of current water quality on water body function, and is an accumulative sum of the influence of all the indicators:

$$E = E_1 + E_2 + \ldots + E_n$$

(4)

where $n$ means the total number of the monitoring indicators.

The result shows in Fig. 8:
From Fig. 8, water quality of Weihe River main stream in the rainy season is the best and Cluster 1 area has the best water quality condition. In the area of Cluster 1, there’s no significant change of inappropriate degree between three seasons and water quality of the rainy season is slightly better than the other two; in the area of Cluster 2, water quality gets the worst in normal season, mainly because of low quantity of DO in the water body and comparatively high quantity of NH$_3$-N; in the area of Cluster 3, water quality turns the worst in dry season, mainly because of lack of discharge of stream. Compared with the standard brought by corresponding functional zone, it can be seen that the most principal indicator that affects water quality of the area is NH$_3$-N, then BOD$_5$ and COD, DO the least.

6. Discussion

6.1. Analysis of water quality variation

The result of PCA-CA of water quality shows that water quality indicators have distinctively different characteristics with spatial change from the upper stream downward, and also much similarities with administrative division: four sites of Cluster one area are situated in Baoji which has smaller population than Xi’an and Xianyang and fewer industrial plants; six sites of Cluster two area are in Xi’an and Xianyang, both with dense population and heavy industrial pollution; three sites of Cluster three area are in Weinan, of which the agriculture is relatively developed while the industry is not, and the pollutants are mostly from non-point sources.

Considering the time, the seasonal change of water quality of the main stream of Weihe River shows a big influence from runoff. High probability value and small variation coefficient of four monitoring
indicators shows an improvement of each indicators in rainy season when there’s a large runoff of the main stream. Because discharge from the industrial and living source has nothing to do with the seasonal change, but is related with industrial production cycle and population respectively, according to the monitoring data of nine years, the increase of runoff in rainy season has distinctively diluted the pollutants. Although pollution of agricultural plane source will be elevated with an increase of rainfall, the overall result is an improvement of water quality of the main stream of Weihe River.

6.2. Analysis of the assessment result

Cluster 1 area has little seasonal change, with main variation on BOD and COD\textsuperscript{Me}. When rainy season comes, the increased runoff should have effectively diluted the pollutants in the water, but in this area water quality has not much changed. The failure of the living waste to get effectively treated since long time ago because of sparse residence of the rural population results in a heavy pollution to the rural environment, therefore, there’s a need to properly deal with the living waste from the area along the bank to avoid the domestic sewage from directly entering the main stream.

Cluster 2 area has the worst index value within a year except that of COD. Especially a big decrease of the amount of DO results in a bad lack of oxygen in the water, meanwhile, the appropriate degree reaches the highest. In rainy season, water quality gets better, especially the negative effect of COD and BOD\textsubscript{5} on water function reduced almost a half, which shows that in this area main pollutants come from the point source, and can be improved a lot when the runoff increases.

In Cluster 3 area, except DO, the negative effect on water function of the other three indexes (BOD\textsubscript{5}, COD, NH\textsubscript{3}-N) changes inversely with the magnitude of the runoff. In dry season, the inappropriate degree of Cluster 2 area and Cluster 3 area are nearly the same, which shows that Weihe River has the poorest capability of self-purification during dry season.

7. Conclusion

According to the data from 2000 to 2008, there’s a distinctive difference of the water quality characteristic of the main stream of Weihe River in administrative divisions: water quality turns the best in the upstream Baoji, then the worst in Xi’an and Xianyang, and better then in Weinan. Xianyang and Xi’an with a dense population and developed industry reach the highest inappropriate degree in normal season, mainly because of a low amount of DO and a high amount of NH\textsubscript{3}-N in the water; areas in Baoji with a small population and Weinan with poorly developed industry reach the highest inappropriate degree in dry season, mainly because of a large amount of COD and BOD. In rainy season, the main stream of Weihe River in Shanxi area reaches the lowest inappropriate degree.

On the whole, the water quality indicators with their influence on the area in a descending order are: NH\textsubscript{3}-N>BOD\textsubscript{5}>COD>DO.

Acknowledgements

This study was supported by the National Key Research Program on Water Pollution Control and Remediation (2009ZX07212-002-01).
References


