Spatiotemporal Prediction of Algae Blooms for Water Quality Management

Youliang Chen, Hamed Karimian

chenyouliang@jxust.edu.cn

hamedk@pku.edu.cn

2024-06-04
Outline

1. Research background
2. Multi-source data
3. Methods
4. Results
5. Early Warning system for Cyanobacterial Blooms
6. Conclusions and Future work
In recent years, frequent outbreaks of cyanobacterial blooms in water bodies have significantly impacted the ecological environment and residents' lives. The main cause of cyanobacteria blooms is eutrophication.

The process where water bodies, such as lakes, rivers, and coastal areas, become enriched with nutrients (e.g., Nitrates), leading to excessive growth of algae and other aquatic plants.

The harmful substances in cyanobacterial blooms can enter water sources, posing dangers to both human and animal health.
The outbreak of cyanobacterial blooms is particularly severe in China's Taihu Lake, Dianchi Lake, and Chaohu Lake. According to the Ecological Environment Bulletin issued by the Ministry of Ecology and Environment in 2022, nearly 25% of China's lakes are moderately eutrophic, and 4.3% are severely eutrophic.
1. Research background

2. Multi-source data

3. Methods

4. Results

5. Early Warning system for Cyanobacterial Blooms

6. Conclusions and Future work
The Huancheng River in Zhejiang Province, is an important source of Taihu Lake, which is one of the five largest freshwater lakes in China. The Huancheng river is located in the subtropical monsoon climate zone, with a mild and humid climate and four distinct seasons,
Water quality data and pre-processing

- In order to comprehensively collect the chlorophyll-a concentration in the water quality of the study area, a group of water quality monitoring stations were deployed in the Ring Road River at regular intervals, and each group included three monitoring points, a total of 33 monitoring points were deployed. Water quality data were collected using sensor AP-7000 from 2020-05, to 2020-09 respectively, 14 indicators such as:

<table>
<thead>
<tr>
<th>Ammonia</th>
<th>Ammonium</th>
<th>Nitrate</th>
<th>pH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlorophyll</td>
<td>Temperature</td>
<td>Turbidity</td>
<td>pressure</td>
</tr>
</tbody>
</table>
This study utilized the image data of GF-1 and Sentinel-2. The preprocessing of GF-1 image data mainly includes radiometric calibration, atmospheric correction, resampling (2m spatial resolution), and land-water separation. Since the image data acquired by Sentinel-2 is at L2A level, there is no need for radiometric calibration and atmospheric correction. The preprocessing of Sentinel-2 image data mainly includes Pansharpening (10m spatial resolution) and land-water separation.
Outline

1. Research background
2. Multi-source data
3. Methods
4. Results
5. Early Warning system for Cyanobacterial Blooms
6. Conclusions and Future work
In this study, random forest technique was utilized to predict chl-a concentrations.
chlorophyll-a concentration inversion band combinations

Based on the remote sensing images of GF-1 and Sentinel-2, the reflectance spectral features of the monitoring stations were extracted separately, and the spectral reflectance features were plotted with wavelength as the X-axis and reflectance as the Y-axis.

Therefore, the two bands corresponding between the reflection peak and the absorption valley are selected to construct a series of RS index to monitor chl-a.

Reflectance spectral characteristics of monitoring stations based on remotely sensed image data
In order to determine the correlation between chlorophyll-a concentration and band inversion models, correlation analysis was performed.

Correlation between GF-1 band combination and measured chlorophyll-a concentration

<table>
<thead>
<tr>
<th>Band combination</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>(b4-b3)/(b4+b3)</td>
<td>0.529**</td>
</tr>
<tr>
<td>1/(b4-b3)</td>
<td>-0.213</td>
</tr>
<tr>
<td>1/(b4+b3)</td>
<td>-0.338</td>
</tr>
<tr>
<td>b4/b3</td>
<td>0.532**</td>
</tr>
</tbody>
</table>

b3=680nm, b4=810nm

As can be seen from the results, among the GF-1 band combinations, the b4/b3 combination showed the highest correlation between reflectance and measured chlorophyll-a concentration, with an R of 0.532.
### Correlation between Sentinel-2 band combination and measured chlorophyll-a concentration

<table>
<thead>
<tr>
<th>Band combination</th>
<th>Correlation coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>((b_3-b_2)/(b_3+b_2))</td>
<td>0.051</td>
</tr>
<tr>
<td>(1/(b_3-b_2))</td>
<td>-0.031</td>
</tr>
<tr>
<td>(1/(b_3+b_2))</td>
<td>-0.032</td>
</tr>
<tr>
<td>(b_3/b_2)</td>
<td>0.056</td>
</tr>
<tr>
<td>((b_5-b_4)/(b_5+b_4))</td>
<td>0.131</td>
</tr>
<tr>
<td>(1/(b_5-b_4))</td>
<td>-0.041</td>
</tr>
<tr>
<td>(1/(b_5+b_4))</td>
<td>-0.040</td>
</tr>
<tr>
<td>(b_5/b_4)</td>
<td>0.114</td>
</tr>
<tr>
<td>((b_7-b_6)/(b_7+b_6))</td>
<td>0.016</td>
</tr>
<tr>
<td>(1/(b_7-b_6))</td>
<td>-0.173</td>
</tr>
<tr>
<td>(1/(b_7+b_6))</td>
<td>-0.156</td>
</tr>
<tr>
<td>(b_7/b_6)</td>
<td>0.021</td>
</tr>
<tr>
<td>((b_9-b_8A)/(b_9+b_8A))</td>
<td>(0.326^{**})</td>
</tr>
<tr>
<td>(1/(b_9-b_8A))</td>
<td>0.237</td>
</tr>
<tr>
<td>(1/(b_9+b_8A))</td>
<td>-0.280*</td>
</tr>
<tr>
<td>(b_9/b_8)</td>
<td>0.309*</td>
</tr>
</tbody>
</table>

As analyzed, the reflectance of the \((b_9-b_8A)/(b_9+b_8A)\) band combination has the highest correlation with the measured chlorophyll-a concentration, with an \(R\) of 0.326.
Outline

1. Research background
2. Multi-source data
3. Methods
4. Results
5. Early Warning system for Cyanobacterial Blooms
6. Conclusions and Future work
Evaluation of inversion model

To test the feasibility of the inversion model, the validation experiment was performed on 20% of the data for each satellite.

Random forest inversion model accuracy
chlorophyll-a concentration inversion results

On May 31 and July 22, the water quality condition of the river was good with low cyanobacterial blooms occurring. In the rest of the time period, the water quality of this river was poor with different degrees of cyanobacterial bloom phenomenon, among which, on May 13, the chlorophyll-a concentration was 15ug/L~20ug/L, and the degree of occurrence of cyanobacterial bloom was serious. Overall, it reflects the concentration of chlorophyll-a in the water quality of the Circular River during this time period.

Inversion results of chlorophyll-a concentration based on the random forest model
Comparison of inversion results based on APPEL index

The APPEL index (Approach by elimination) was used to compare the performance of our approach with the approach by Alem. et al. 2012.

\[ APPEL = R_{rs}(\lambda_{\text{mir}}) - [(R_{rs}(\lambda_{\text{blue}}) - R_{rs}(\lambda_{\text{mir}}))^2 + (R_{rs}(\lambda_{\text{red}}) - R_{rs}(\lambda_{\text{mir}}))] \]

Fitting results of APPEL inversion model

<table>
<thead>
<tr>
<th>Sensor type</th>
<th>Equation expressions</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>GF-1</td>
<td>( y = 9\times10^{-6}x + 7.429 )</td>
<td>0.001</td>
</tr>
<tr>
<td>Sentinel-2</td>
<td>( y = -8\times10^{-6}x + 11.642 )</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Relation between APPEL model and measured chlorophyll-a concentration
The APPEL inversion model was applied to the GF-1 remote sensing data. On May 31, the inversion results of the APPEL model were chlorophyll-a concentrations ranging from 5 ug/L to 10 ug/L. The inversion results of the Random Forest showed that the chlorophyll-a concentration of the river on that day had both areas greater than 10 ug/L and areas less than 10 ug/L, and there were multiple intervals of chlorophyll-a concentration value results. Combined with the measured data, the inversion results of the Random Forest model were closer to the chlorophyll-a concentration situation. On July 22, the inversion results of the APEEEL model were the same as those of May 31, which manifested the chlorophyll-a concentration of 5 ug/L~10 ug/L. The Random Forest model inversion results reflected that there were areas with low values of chlorophyll-a concentration, and there were also some rivers with areas with chlorophyll-a concentration greater than 10 ug/L, which showed a slight hydrophobic phenomenon.
APPEL inversion model was applied to Sentinel-2 remote sensing data. In the inversion results on May 13, the inversion results of the APPEL model manifested that the chlorophyll-a concentration ranged from 10ug/L to 15ug/L, and the inversion results of the random forest model appeared the area of greater than 15ug/L. On the 1st of August, the inversion results of the APPEL model still manifested that the chlorophyll-a concentration was in the range of 10ug/L to 15ug/L, and the inversion results of the random Forest model inversion results manifested an overall chlorophyll-a concentration below 10ug/L, with fewer areas greater than 10ug/L.

Comparison of inversion results based on APPEL and RF model
Comparison of results based on evaluation indicators

To test the accuracy of the inversion of the random forest model, the remotely sensed reflectance corresponding to 20% of the measured chlorophyll-a concentration data was substituted into the inversion model as the independent variable, and the inversion values obtained by the inversion model were compared with the measured values to evaluate the accuracy of the model. The results show that the accuracy of the random forest inversion model has been greatly improved and the model performance is stable.

Accuracy evaluation results of each inversion model
The most influencing factors

• variable importance:

Through usage of variable importance, one can interpret and compare the importance of predictors in a model.

\[
VI = \sum_{T=1}^{T} \frac{\Delta R_{mT}}{T}
\]

<table>
<thead>
<tr>
<th>Predictor</th>
<th>VI (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chl-a</td>
<td>100</td>
</tr>
<tr>
<td>water temperature</td>
<td>42</td>
</tr>
<tr>
<td>oxygen content</td>
<td>24</td>
</tr>
<tr>
<td>ammonia</td>
<td>23</td>
</tr>
<tr>
<td>Sentinel,GF-1</td>
<td>22</td>
</tr>
<tr>
<td>Salinity</td>
<td>22</td>
</tr>
</tbody>
</table>
1. Research background
2. Multi-source data
3. Methods
4. Results
5. Early Warning system for Cyanobacterial Blooms
6. Conclusions and Future work
Early Warning and Application of Cyanobacterial Blooms

Based on our proposed approach, we developed a Chl-a concentration prediction system which has been applied in the monitoring project of a cyanobacterial bloom outbreak in Huzhou City, Zhejiang Province, China. The system was able to analyze the data collected by sensors, generate analysis reports, and predict the cyanobacterial bloom outbreaks in 48 hours in advance. After one year of operation, the predicted cyanobacterial blooms were consistent with the actual situation, indicating that the model can provide scientific reference and decision support for preventing and managing urban inland river water pollution.
Chl-a concentration prediction system

Cyanobacteria remote dynamic monitoring and early warning system

Log in to your account: CYL
Real name: Cyl
User Status: Normal
Department:

Xintang Port Cyanobacteria situation

Level: Level 3

Analyze reports
106. Weekly Report2022.11.28.docx 11-28
103. Weekly Report2021.11.7 .docx 11-7
100. Weekly Report2021.10.17.docx 10-17
95. Weekly Report2022.9.5.docx 9-5
94. Weekly Report2022.8.29.docx 8-29
93. Weekly Report 2022.8.22.docx 8-22

Chl-a concentration prediction system
### 系统界面

**表格内容**

<table>
<thead>
<tr>
<th>编号</th>
<th>项目编号</th>
<th>监测点号</th>
<th>传感器号</th>
<th>监测站号</th>
<th>类别</th>
<th>监测因子</th>
<th>值（单位）</th>
<th>日期/时间</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200001</td>
<td>200001-70-82</td>
<td>7号站-1</td>
<td>010001</td>
<td>7号站</td>
<td>AP7000-1</td>
<td>温度(Temp)</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>200001</td>
<td>200001-70-82</td>
<td>7号站-1</td>
<td>010001</td>
<td>7号站</td>
<td>AP7000-1</td>
<td>气压(Baro)</td>
<td>143.40</td>
</tr>
<tr>
<td>3</td>
<td>200001</td>
<td>200001-70-82</td>
<td>7号站-1</td>
<td>010001</td>
<td>7号站</td>
<td>AP7000-1</td>
<td>pH</td>
<td>6.62</td>
</tr>
<tr>
<td>4</td>
<td>200001</td>
<td>200001-70-82</td>
<td>7号站-1</td>
<td>010001</td>
<td>7号站</td>
<td>AP7000-1</td>
<td>DO</td>
<td>80.20</td>
</tr>
<tr>
<td>5</td>
<td>200001</td>
<td>200001-70-82</td>
<td>7号站-1</td>
<td>010001</td>
<td>7号站</td>
<td>AP7000-1</td>
<td>ORP</td>
<td>915</td>
</tr>
<tr>
<td>6</td>
<td>200001</td>
<td>200001-70-82</td>
<td>7号站-1</td>
<td>010001</td>
<td>7号站</td>
<td>AP7000-1</td>
<td>EC</td>
<td>1158</td>
</tr>
<tr>
<td>7</td>
<td>200001</td>
<td>200001-70-82</td>
<td>7号站-1</td>
<td>010001</td>
<td>7号站</td>
<td>AP7000-1</td>
<td>DO-Sat</td>
<td>2.4320</td>
</tr>
<tr>
<td>8</td>
<td>200001</td>
<td>200001-70-82</td>
<td>7号站-1</td>
<td>010001</td>
<td>7号站</td>
<td>AP7000-1</td>
<td>SSO</td>
<td>10.00</td>
</tr>
<tr>
<td>9</td>
<td>200001</td>
<td>200001-70-82</td>
<td>7号站-1</td>
<td>010001</td>
<td>7号站</td>
<td>AP7000-1</td>
<td>Chlorophyll a</td>
<td>197</td>
</tr>
<tr>
<td>10</td>
<td>200001</td>
<td>200001-70-82</td>
<td>7号站-1</td>
<td>010001</td>
<td>7号站</td>
<td>AP7000-1</td>
<td>BGA-PC</td>
<td>6661</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**地图**

地图显示了地理位置信息。
Outline

1. Research background
2. Multi-source data
3. Methods
4. Results
5. Early Warning system for Cyanobacterial Blooms
6. Conclusions and Future work
Conclusions

- In this study, the measured water quality data from May 2020 to September 2020 were used as the data source, and GF-1 and Sentinel-2 remote sensing data were utilized to produce surface distribution of chlorophyll-a concentration and remote sensing imagery, and a random forest model was constructed to invert the chlorophyll-a concentration and monitor the outbreak of cyanobacterial blooms.

- Our results showed that the GF-1 and Sentinel-2 random forest inversion models were highly accurate with small errors, and the $R^2$ of the models were 0.93 and 0.87, respectively.
As a result of the RF variable importance indicator, the main influencing factors on chl-a concentration are chl-a_current, water temperature, oxygen content, and ammonia, Sentinel,GF-1 and Salinity.
Future work:

The study of chlorophyll-a concentration in the water body can make an analysis of the phenomenon of cyanobacterial bloom, and provide a theoretical basis for the prevention and treatment of water pollution in the city. Although this study has achieved certain results, there are still deficiencies that need to be further improved, mainly including the following aspects:

(1): The inversion model used in this study was limited by the time-series limitations of measured chlorophyll-a concentration data and remotely sensed data. Future work can utilize Envisat data, Zhuhai-1 and other medium and high-resolution data to construct long time-series data for chlorophyll-a concentration inversion.

(2): This study only took into account 14 influencing factors, such as ammonia nitrogen, ammonium salt and phycocyanin, and did not consider the influence of other influencing factors on chlorophyll-a concentration. Future studies can consider more influencing factors, such as land use, meteorological conditions, to continue to explore the mechanism of cyanobacterial bloom outbreaks.
Future work:

(3): Applying those models that consider spatial and temporal dependencies among data can improve the prediction accuracy.
References:

Thank you for your attention