Cyanobacteria Bloom Prediction Research based on Multi-source Satellite Images

Meng Mu: 161301037@stu.njnu.edu.cn Yunmei Li: liyunmei@njnu.edu.cn

Nanjing Normal University

ABSTRACT

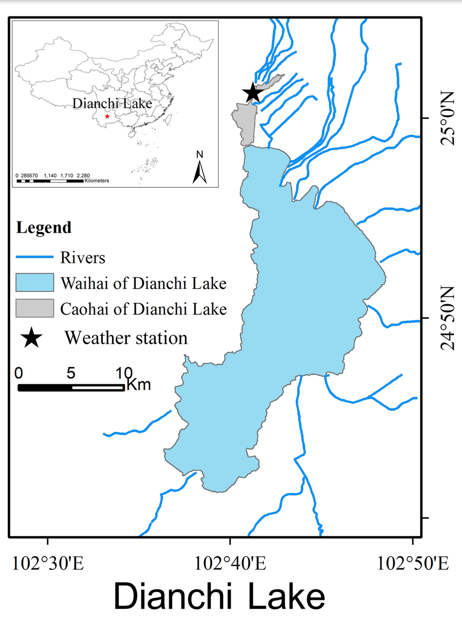
High-frequency and reliable data on cyanobacteria blooming over a long time is crucial to identify the outbreak mechanism of blooms and to forecast future trends. Owing to the heavy cloud cover of remote sensing observation in Dianchi Lake, HJ-1, GF-1 and Landsat-8 were used jointly to obtain the temporal-spatial distribution of bloom, which greatly increased the monitoring frequency. Due to meteorological factors are the potent catalysts for the occurrence and expansion of cyanobacteria bloom, this study establish Logistic regression model and Random forest (RF) model to predict the probability of cyanobacteria bloom based on meteorological factors. Several findings can be drawn (a): the accuracy is 0.71 for predicting the probability of bloom, and 0.73 for predicting the probability of significant bloom through Logistic model. (b) the forecast accuracy is 0.60 of RF model. These models demonstrated the feasibility of bloom forecasting the by meteorological data.

INTRODUCTION

- Cyanobacteria blooms have become a growing threat in freshwater systems.
- Bloom is closely related to meteorological factors.
- The combination of multi-source satellite sensors can greatly improve monitoring efficiency of blooms, especially in a cloudy and rainy lake.
- It is impossible to immediately curb the outbreak of blooms. In this situation, it is particularly important to predict the location and duration of the.

OBJECTIVE & STUDY AREA

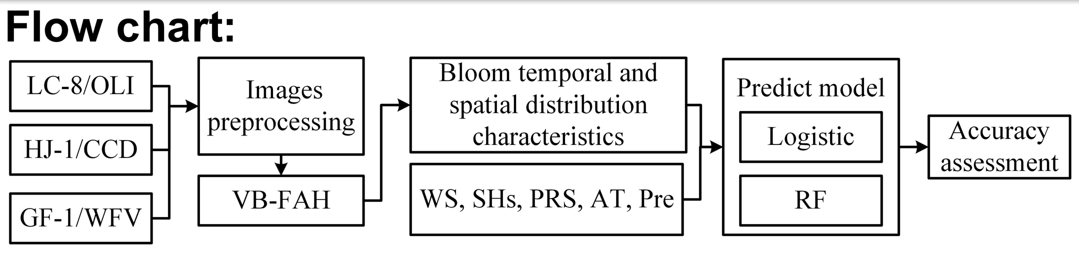
The objective of the study is to obtain the temporal and spatial distribution of cyanobacteria blooms in Dianchi Lake through multi-source satellite images. And then, correlated the responses of bloom dynamics to meteorological factors, to predict the the probability of blooms.



Dianchi Lake

METHODS

Flow chart:



- **Image preprocessing**

The images were processed in two steps: radiometric calibration and quick atmospheric correction in ENVI 5.3. The CCD and WFV images were geometrically corrected by referring to a fixed Landsat-8/OLI image using the nearest-neighbor approach, with RMSE within half of a pixel. The GF-1/WFV images were also resampled to 30 m spatial resolution.

- **VB-FAH**

VB-FAH = $(\rho_{NIR}-\rho_{Green})+(\rho_{Green}-\rho_{Red})*(\lambda_{NIR}-\lambda_{Green})/(2\lambda_{NIR}-\lambda_{Red}-\lambda_{Green})$

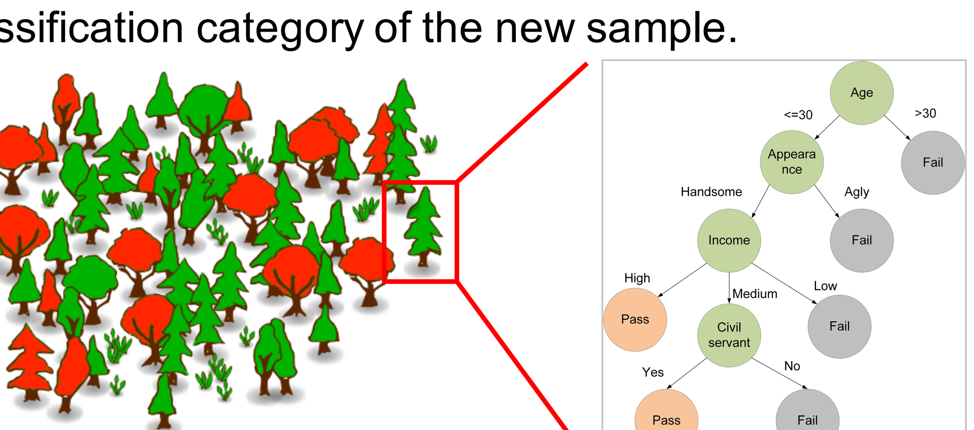
- **Logistic Model**

Hypothesis "1" represents the occurrence of bloom, "0" represents the absence of bloom, $p(y=1)$ represents the probability of bloom occurrence, and x_n represent influencing factors. Here:

$$\text{logit}(p) = \beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\ldots\ldots+\beta_nx_n$$

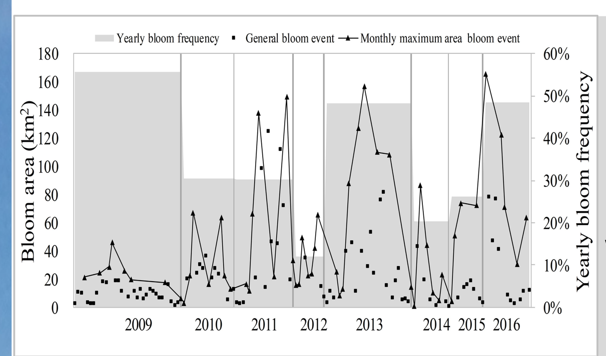
- **RF Model**

Create a forest in a random way and get the "rules". When a new input comes in, each tree in the forest votes, and the category with the most votes is the classification category of the new sample.



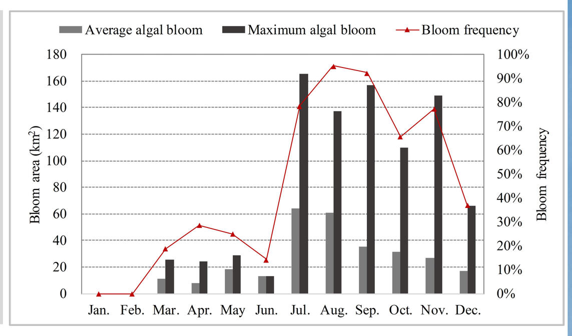
RESULTS

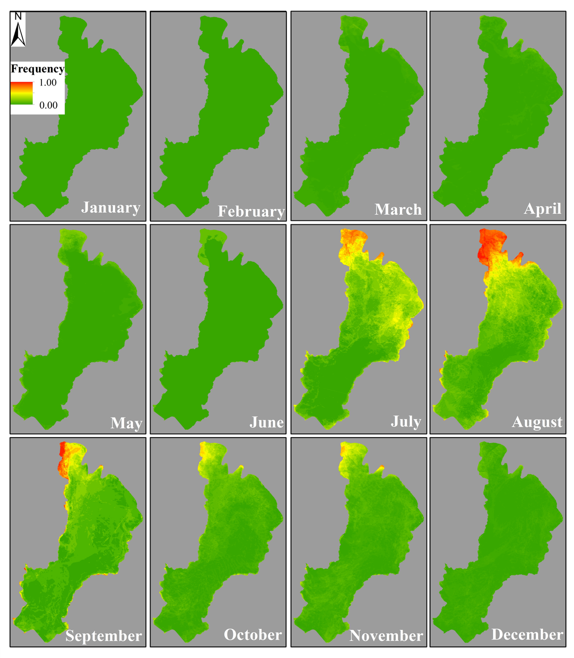
- **Temporal and spatial distribution of cyanobacteria blooms in Dianchi Lake**



Usually, the bloom began in March/April and ended in December, and the maximum bloom area was extended year by year. 2011, 2013 and 2016 were blooming year due to their larger bloom area and higher bloom frequency.

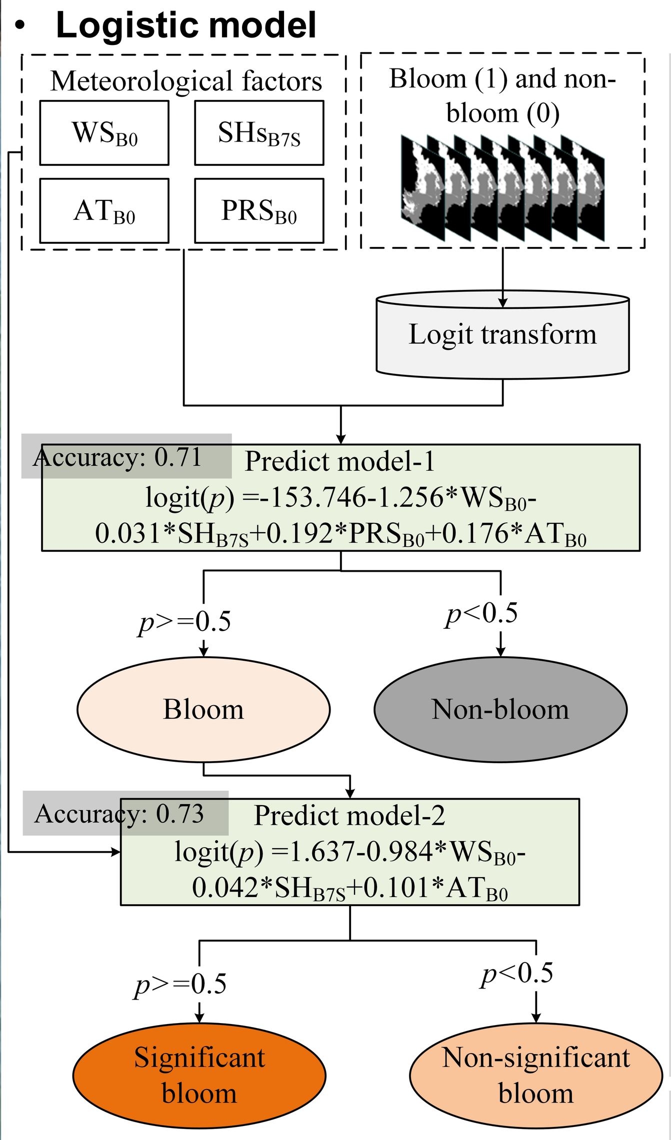
The monthly bloom frequency of Dianchi Lake followed a "three peaks" phenomenon, and no blooms were observed in January and February. From July, bloom frequency and size both increased significantly.





The northern lake experienced more blooms in all months (excluding January and February). Blooms first occurred in the northern region in March, then the coverage extended to the southern lake starting in July, and then to the entire lake. The high frequency of blooms lasted from July to November, then gradually decreased in December and completely disappeared in January.

- **Predict model of bloom probability in Dianchi Lake**
  - **Logistic model**

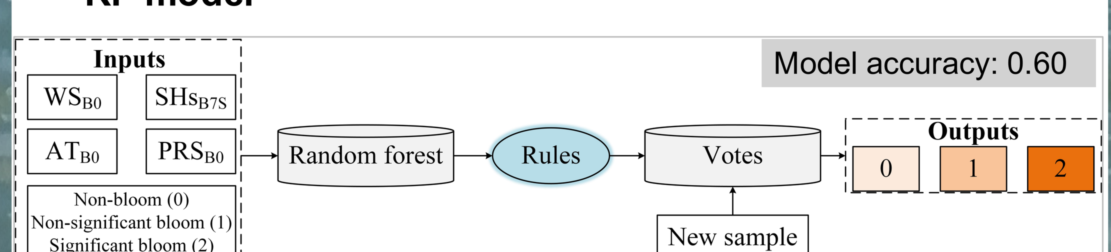


Modeling data set: 424 satellite images from 2009-2015; **Verification data set**: 30 images in 2016; **Independent variables**: $\mathbf{WS_{B0}}$: wind speed on the imagery day; $\mathbf{SHs_{B7S}}$: cumulative sunshine hours from the seven days before the imagery day; $\mathbf{AT_{B0}}$: air temperature on the imagery day; $\mathbf{PRS_{B0}}$: pressure on the imagery day; **Dependent variable of model-1:** the state of bloom (1) or non-bloom (0); **Dependent variable of model-2:** the state of significant bloom (1) or non-significant bloom (0);

Imagery day including bloom day and non-bloom day which determined by satellite images.

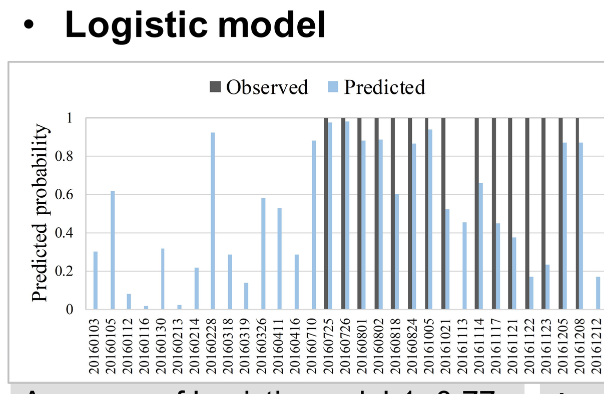
- **RF model**

Model accuracy: 0.60

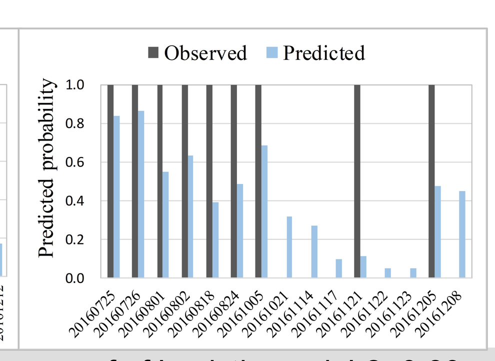


Firstly, five types of data are used as inputs, and put them into the RF to obtain "rules". Finally, new samples in validation data set learning the "rules", and the ranking is obtained by voting.

- **Accuracy assessment**
  - **Logistic model**

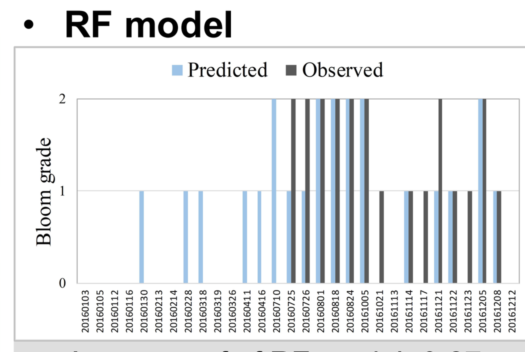


Accuracy of Logistic model-1: 0.77



Accuracy of of Logistic model-2: 0.80

- **RF model**

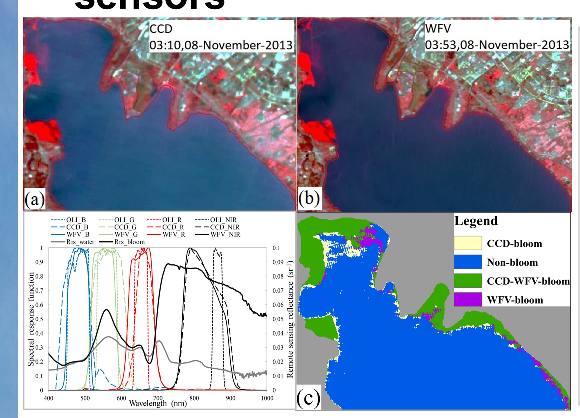


Accuracy of of RF model: 0.27

- In the first half of 2016, there was no valid images due to the heavy cloud.
- Logistic model required two steps to predict the bloom scale.
- The prediction accuracy of Logistic model is better than that of RF model.

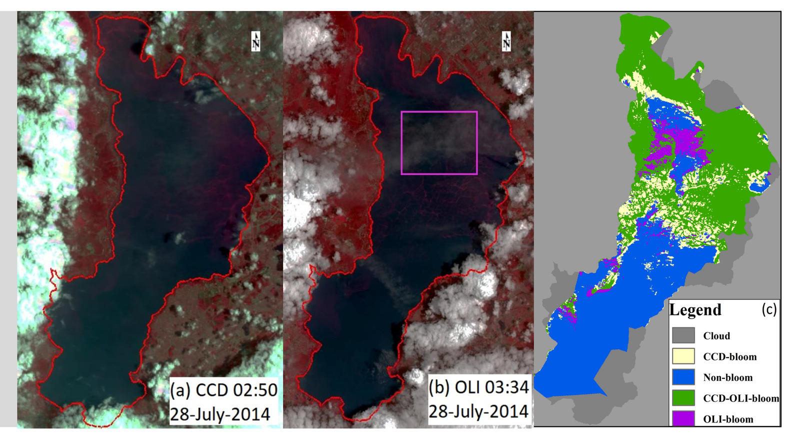
DISCUSSION

- **Consistency of bloom area retrieval by different sensors**

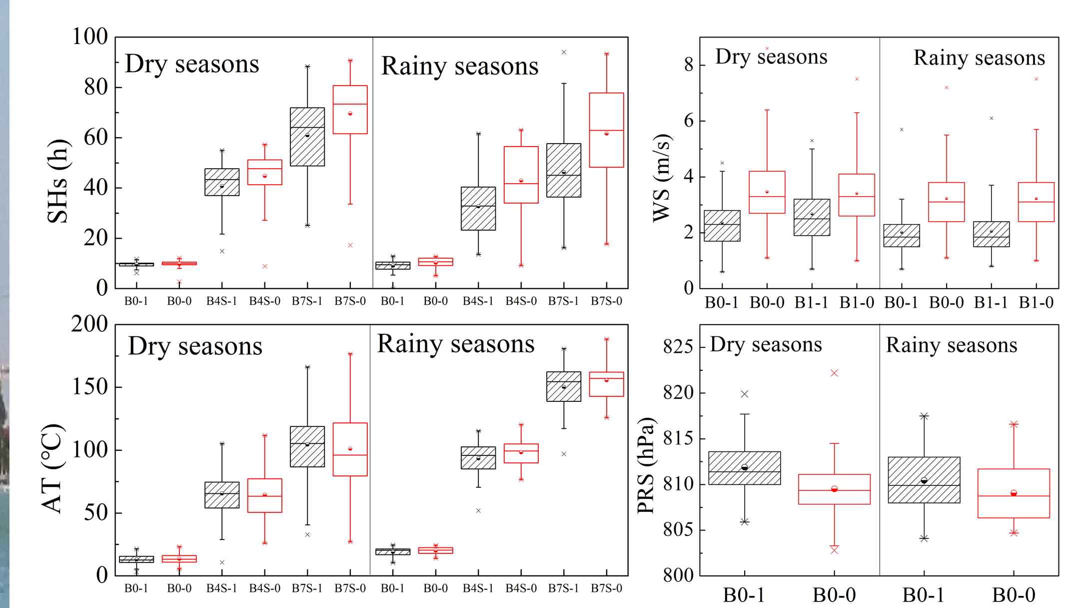


HJ-1/CCD and GF-1/WFV images were acquired on 8 November 2013, 43 minutes apart. The relative error was only 0.5094 km² between the two bloom areas.

HJ-1/CCD and Landsat-8/OLI images were acquired on 28 July 2014, 44 minutes apart. The discrepancy is mainly caused by the cloud.

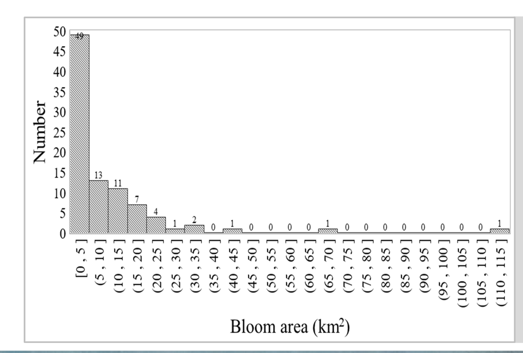


- **Response of bloom dynamic to meteorological factors**



SHs_{B7S} on bloom day and non-bloom day have a significant difference both in dry and rainy seasons; PRS_{B0} and AT_{B0} on bloom day and non-bloom day have a slight difference; WS_{B0} on bloom day and non-bloom day have a significant difference both in dry and rainy seasons.

- **Error analysis of Logistic predict model-1**



Most biased forecast events have small coverage area. This is because it may be in a recession, and the meteorological factors are no longer suitable for outbreak of bloom.

CONCLUSION

- To overcome the shortage of effective images in Dianchi Lake, three satellite sensors were used jointly, which greatly improves the monitoring frequency.
- The consistency analysis shows that LC-8/OLI, GF-1/WFV and HJ-1/CCD can be used in combination.
- The responses of bloom dynamics to meteorological factors were quantified based on the long-term bloom record, which shows that WS_{B0}, SHs_{B7S}, AT_{B0} and PRS_{B0} were highly related with the outbreak of blooms.
- Compared to RF model, Logistic model has higher precision to predict probability of blooms.
- The predict model demonstrated the feasibility of monitoring and forecasting the cyanobacteria blooms by meteorological data.

MAJOR REFERENCES

- Xing Q, Hu C (2016): Mapping macroalgal blooms in the Yellow Sea and East China Sea using HJ-1 and Landsat data: Application of a virtual baseline reflectance height technique. Remote Sensing of Environment 178, 113-126
- Shi K, Zhang Y, Zhou Y, Liu X, Zhu G, Qin B, Gao G (2017): Long-term MODIS observations of cyanobacterial dynamics in Lake Taihu: Responses to nutrient enrichment and meteorological factors. Sci Rep 7, 40326
- Chen, L., J. Zhang, X. Chen, M. Cai, L. I. Haijun & Y. U. Yongming (2016) Research on meteorological factors and Logistic prediction model of cyanobacterial blooms in Erhai Lake. Journal of Central China Normal University.
- Stumpf RP, Johnson LT, Wynne TT, Baker DB (2016): Forecasting annual cyanobacterial bloom biomass to inform management decisions in Lake Erie. Journal of Great Lakes Research 42, 1174-1183