The threshold determination methods of water body information extraction using GF-1 satellite image

R P Ji\textsuperscript{1,2,*}, W Y Yu\textsuperscript{1,2}, R Feng\textsuperscript{1,2}, J W Wu\textsuperscript{1,2} and Y S Zhang\textsuperscript{1,2,*}

\textsuperscript{1} Institute of Atmospheric Environment, China Meteorological Administration, Shenyang 110166, China
\textsuperscript{2} Key Laboratory of Agrometeorological Disasters, Liaoning Province, Shenyang 110166, China
\textsuperscript{*} jiruipeng@163.com

* The corresponding author: Y S Zhang, Email: yushuzhang@126.com

Abstract: The GF-1 satellite is the first satellite of China High-resolution Earth Observation System (referred to short as CHEOS) with the advantages of high spatial resolution, high temporal resolution and 4 multi-spectral channels, which can quickly extract inland water body information. At present, the methods of water body information extraction by remote sensing mainly include single-band threshold method, spectral relation method, normalized difference vegetation index (NDVI) method, normalized difference water body index (NDWI) method and shade water body index (SWI) method. The results of water body information extraction will vary with different methods; especially, the thresholds determination of each method is an important factor in the accuracy of water body information extraction. In this paper, the water body information of Qinghai Lake, Hongze Lake, and Hulun Lake were extracted by NDWI index, and the optimal thresholds of water body information extraction were determined by using Jeffries-Matusita (J-M) distance method and density slice method was in the range of -0.061~0.093. Finally, the accuracy of the extraction results was analyzed, and the overall accuracy is more than 88%.

1. Introduction
The water body is a major part of the surface hydrosphere, which is the natural water bodies bordered by relatively stable land, including rivers, lakes, seas, glaciers, snowfields, reservoirs, ponds, etc. Inland water body participates in the water cycle of the biosphere, being closely related to the survival of the earth’s organisms while enabling sustainable development of human societies. As a key indicator for evaluating the ecological environment, inland water bodies are of great significance for the maintenance, restoration, and development of the ecological environment. Due to the affection of climate changing and human’s activities, China's water resources are uneven in time and spatial distribution, resulting in a prominent contradiction between water supply and demand of some regions. Therefore, scientifically and accurately monitoring and evaluating water body change has important theoretical and practical significance for the evaluation of ecological environment quality.

Remote sensing technology is considered to be the most economical, rapid and effective methods of water change monitoring. In particular, CHEOS satellites enrich the data sources of water information extraction, which has become one of the major means to improve the accuracy of water body information extraction. Gf-1 satellite is equipped with two PMS and four WFV cameras, with the
researches on remote sensing extraction of water body information focus on the comparison of methods or indices [1-3], but rarely in threshold determination of each index and its accuracy. However, the threshold determination is the focus of water extraction research [4-5]. Therefore, based on the WFV data of GF-1 satellite, NDWI was calculated to extract the water bodies’ information of Qinghai Lake, Hongze Lake, and Hulun Lake, and the optimal thresholds of NDWI were determined by J-M distance method and density slice method combined with manual visual interpretation in this paper [6].

### Table 1. Technical parameters of GF-1 satellite payload

| satellite payload | Band order | band       | Spectral range/μm | Spatial resolution/m | Revisit time/d |
|-------------------|------------|------------|-------------------|----------------------|---------------|
| PMS (2-m resolution) | PAN 1     | Panchromatic | 0.45~0.90         | 2                     | 4             |
|                   | 2         | Blue       | 0.45~0.52         |                       |               |
|                   | 3         | Green      | 0.52~0.59         |                       |               |
|                   | 4         | Red        | 0.63~0.69         | 8                     | 4             |
| WFV (16-m resolution) | 1         | Blue       | 0.45~0.52         | 16                    | 2             |
|                   | 2         | Green      | 0.52~0.59         |                       |               |
|                   | 3         | Red        | 0.63~0.69         |                       |               |
|                   | 4         | Near infrared | 0.77~0.89     |                       |               |

### 2. Materials and methods

#### 2.1. Study areas and data resources

Considering the representativeness of different regions and water body types, three lakes were selected as study areas, including Qinghai Lake in Qinghai Province, Hongze Lake in Jiangsu Province and Hulun Lake in Inner Mongolia region. Gf-1 satellite data used in this paper were WFV multispectral images with a spatial resolution of 16m. The study area and data information were shown in table 2.

| Study area    | Province   | Sensor | Imaging time   | Data resource                                      |
|---------------|------------|--------|----------------|----------------------------------------------------|
| Qinghai Lake  | Qinghai    | WFV2   | 2016-7-15      | Land Observation Satellite Data Service Platform of China Center for Resources |
|               |            | WFV3   | 2017-8-13      |                                                    |
| Hongze Lake   | Jiangsu    | WFV1   | 2016-7-28      | Satellite Data and Application                      |
|               |            | WFV3   | 2018-7-20      | (http://218.247.138.119:7777/DSSPlatfor m/index.html) |
| Hulun Lake    | Inner Mongolia | WFV3 | 2017-8-27      |                                                    |
|               |            | WFV3   | 2018-8-23      |                                                    |

#### 2.2. Data processing

Being all 1A level data products, the GF-1 images in the paper were preprocessed by radiometric calibration, atmospheric correction, geometric correction, solar elevation correction, etc. The data processing was completed in ENVI5.3 SP1 software.

#### 2.3. Threshold determination methods

**2.3.1 Jeffries-Matusita (J-M) distance method.** The basic principle is to assume that the object characteristic values of water and non-water body samples accord with a normal distribution, and the separability is used to evaluate the association between two types in a certain feature, so as to select the classification feature. J-M distance method is used to compute the separability, and Gaussian
mixture model is used to calculate the optimal threshold between water and the non-water body samples for a certain feature.

The value of J-M distance is in the range 0 to 2, which represents the samples separability. When JM=2, it indicates that the two types of features are completely separated under the selected classification features; when the JM value is smaller, it indicates that the separation is poor and there will be a large number of misclassified objects. The formulas are as follows [7]:

\[ J = 2(1-e^{-B}) \]  

\[ B = \frac{1}{8}(m_1 - m_2)^2 \frac{2}{\sigma_1^2 + \sigma_2^2} + \frac{1}{2} \ln \left( \frac{\sigma_1^2 + \sigma_2^2}{2\sigma_1 \sigma_2} \right) \]

Where, \( J \) is the J-M distance, \( B \) is Bhattacharyya distance, \( m_i \) and \( m_j \) are the object characteristic mean values of water and non-water body samples, \( \sigma_i \) and \( \sigma_j \) are the object standard deviations of water and non-water body samples.

The optimal threshold \( T \) between two types in a certain feature is calculated based on the Gaussian mixture model as follows:

\[ T = m_2 \sigma_1^2 - m_1 \sigma_2^2 + \sigma_1 \sigma_2 \sqrt{2(m_1 - m_2)^2 + 2A(\sigma_1^2 - \sigma_2^2)} \]

\[ A = \ln \left( \frac{\sigma_2^2}{m_1} \times \frac{\sigma_1^2}{m_2} \right) \]

When the J-M distance value is less than 1.7, \( T \) calculated with formulas (3) might be not accurate. Marpu proposed a solution and adjustment criteria as follows:

\[
\begin{cases} 
1.25 \geq |J| \geq 0.5 & T = m_2 \\
1.75 > |J| > 1.25 & T = \frac{T + m_2}{2} \\
|J| \geq 1.75 & T = T
\end{cases}
\]

2.3.2 Density slice method. Density slice is to highlight the hue of a certain density level (or corresponding object), the hue density of the image is divided into several levels, which pixels above a certain gray level into one area, and pixels lower than a certain gray level into another area. Different colors indicated different density levels, and the valley value between the two peaks was selected as the range of threshold.

2.4 Normalized difference water body index (NDWI)

The method is suitable for water body information extraction in flat terrain areas. If NDWI≥T, the object was water body:

\[ NDWI = \frac{B_{green} - B_{nir}}{B_{green} + B_{nir}} \]

Where, NDWI is normalized difference water body index, \( B_{green} \) and \( B_{nir} \) are reflectivities of the green band and near-infrared band, \( T \) is the experience threshold calculated with formula (1)~(5).

3. Results

Water and non-water bodies of Hongze Lake, Qinghai Lake, and Hulun Lake were sampled separately by using the J-M distance method, and the characteristic variables of NDWI were calculated. The NDWI thresholds of 6 temporal times in three lakes were determined according to the determination rule of \( T \) and the density slice method. The results were shown in table 3 and table 4.
Table 3. NDWI thresholds calculated with J-M method

| Date      | NDWI characteristic values | Hongze Lake samples | Qinghai Lake samples | Hulun Lake samples |
|-----------|-----------------------------|---------------------|----------------------|--------------------|
|           |                             | Water body          | Non-water body       | J-M                | Threshold          |
| 2016/7/28 | Mean value                 | 0.285415            | -0.447314            | 2.00               | 0.04468            |
|           | Standard Deviation         | 0.055081            | 0.127762             |                    |                    |
| 2018/7/20 | Mean value                 | 0.156209            | -0.480746            | 1.99               | 0.015858           |
|           | Standard Deviation         | 0.034705            | 0.131662             |                    |                    |

| Date      | NDWI characteristic values | Water body          | Non-water body       | J-M                | Threshold          |
|-----------|-----------------------------|---------------------|----------------------|--------------------|
|           | Mean value                 | 0.531983            | -0.467056            | 2.00               | 0.176931           |
|           | Standard Deviation         | 0.074188            | 0.137084             |                    |                    |
| 2017/8/13 | Mean value                 | 0.417058            | -0.542514            | 1.97               | -0.01854           |
|           | Standard Deviation         | 0.15115             | 0.181527             |                    |                    |

| Date      | NDWI characteristic values | Water body          | Non-water body       | J-M                | Threshold          |
|-----------|-----------------------------|---------------------|----------------------|--------------------|
|           | Mean value                 | 0.154898            | -0.467629            | 2.00               | -0.12608           |
|           | Standard Deviation         | 0.06376             | 0.078992             |                    |                    |
| 2017/8/27 | Mean value                 | 0.372972            | -0.539701            | 2.00               | 0.006651           |
|           | Standard Deviation         | 0.083788            | 0.132594             |                    |                    |

Table 4. NDWI thresholds calculated with density slice method

| Date      | Area          | Threshold |
|-----------|---------------|-----------|
| 2016/7/28 | Hongze Lake   | 0.0       |
| 2018/7/20 |                | 0.012934  |
| 2016/7/15 | Qinghai Lake  | 0.126394  |
| 2017/8/13 |                | 0.0       |
| 2017/8/27 |                | -0.05     |
| 2018/8/23 | Hulun Lake    | 0.0       |

According to the determined threshold values, the water body information of the three Lakes was extracted respectively (Figure 1), and the relative errors were calculated compared with the actual area. The results are shown in Table 5.
4. Discussion
The basic principle of extracting inland water body information with satellite remote sensing technology is according to the spectral characteristics. Within the visible light range, water reflectance is the highest in the blue and green band; and as the wavelength increases, water reflectance gradually decreases; and in the near-infrared band, water reflectance is very low and almost completely be absorbed.

This paper does not focus on the water body information extraction method but discusses how to determine the threshold of the normalized difference water body index for extracting water body information. The study is aiming to remind researchers to pay attention to the affection of the threshold determination method.

5. Conclusion
Based on formula (1) to formula (5), J-M distance values could completely separate the water and non-water body features were all above 1.97. In J-M distances method, the extraction threshold of Hongze Lake was 0.04468, 0.015858, Qinghai Lake’s was 0.176931, -0.01854, and Hulun Lake’s was -0.12608, 0.006651. In the density slice method, Hongze Lake’s was 0, 0.012934, Qinghai Lake’s was 0.126394, 0, and Hulun Lake’s was -0.05, 0. The difference in the relative errors between the two methods was only 0.9%, which explained that the threshold values determined by the two methods were closer.

Water bodies affected by various factors have different thresholds. The threshold determination with JM and density slice methods could avoid relying on repeated experiments. According to three times standard deviation method [8], it is recommended to the range of NDWI is 0.216 ~ 0.247, two times standard deviation method recommended to 0.139 ~ 0.170, and one times standard deviation method recommended to 0.061~ 0.093. The threshold ranges are the optimal threshold.
Acknowledgments
This study was financially supported by China Meteorological Standardization Project of CMA (QX/T-2018-58) and State Administration of Science Technology and Industry for National Defense (70-Y40G09-9001-18/20).

References
[1] Duan Q Y, Meng L K, Fan Z W, Hu W G, Xie W J 2015 Remote Sens. Land & Resources 27 79- 84
[2] Zha L, Gong H L, Hu Z W, Du H Y 2015 J. Capital Normal Univ. (Natural Sci. Ed.) 36 85-89
[3] Chen W Q, Ding J L, Li Y H, Niu Z Y 2015 Resources Sci. 37 1166-72
[4] Xi X Y, Shen N, Li X J 2009 Computer Eng. & Design 30 993-6
[5] Jian X, Chen H, Xing ZH Y, Fang T, Yin L Y 2014 Sci. Tech. Engrg. 14 267-274
[6] Wang A L, Liu J, Wang C Y, Wang R Y 2017 J. Shandong Agri. Univ. (Natural Sci. Ed.) 48 70-74
[7] Jiang C M 2016 Diss. (Fujian: Fujian Normal University) p 2-3
[8] Huang E H, Pan D L, Li SH J, He X Q 2006 J. Mar. Sci. 24 91-97