Extracting Information on Wetlands in Northeast China Based on Time Series FY3/MERSI Data

Rui Feng, Jinwen Wu, Wenying Yu, Ruipeng Ji, and Yushu Zhang*
Institute of Atmospheric Environment, China Meteorological Administration, 110166, Shenyang, China

Abstract. Using Fengyun Medium Resolution Spectral Imager (FY3/MERSI) data from 2011 to 2018, this paper analyzed the characteristics of original spectra and the normalized difference vegetation index (NDVI) of lakes, reservoirs, rivers, paddy fields and swamps in Northeast China, A large-scale wetland extraction model was built using the decision tree classification method, and accuracy was verified in typical sites using China High-resolution Earth Observation System (CHEOS-1) data. The results showed that there were differences in the spectral curves and NDVI of different wetland types, especially the swamps, paddy fields and other landforms that showed a growth cycle with more significant differences in spectral curves and index curves than those of lakes, reservoirs and rivers. The total area of wetlands in Northeast China was 79,123.4 km$^2$, including 5262.3 km$^2$ of lake and reservoir wetlands, 6514.7 km$^2$ of river wetlands, 15284.7 km$^2$ of swamp wetlands and 52061.7 km$^2$ of paddy fields. The accuracy verification showed that the area accuracy of wetlands was over 91.8%, the overall classification accuracy of wetlands was over 84.8% and the Kappa coefficient was over 0.6446.

1 Introduction

Wetlands are among the most biodiverse ecological landscapes in nature, forming major ecosystems worldwide[1,2]. At present, global environmental change is accelerating, with unprecedented modifications to land use and land cover[3-5], which inevitably will be reflected in changes in the structure and function of wetland landscape elements. As human activities affecting wetlands expand, tracking changes in wetlands is becoming more important[6].

Remote sensing data with wide and frequent repeat coverage and large amounts of information are easy to obtain, meaning they have been widely used in wetland classification research[7]. Using medium- and high-resolution data from satellites such as Landsat, Satellite Pour l’Observation de la Terre (SPOT), and China High-resolution Earth Observation System (CHEOS-1), monitoring and classification of small-scale wetlands have been conducted with methods such as unsupervised classification and maximum likelihood, support vector machines and neural network models[8,9]. Frohn et al.[10] added time series remote sensing data analysis to a model built for extraction of forest and swamp wetland classes, with a classification accuracy of up to 92.8%.
The decision tree classification method is a way to obtain classification rules by combining discriminant knowledge and image spectral features to generate the logical discriminant. The CART (classification and regression tree) algorithm has high computational efficiency and a clear structure. Davranche et al.[11] used the decision tree classification method to extract reed wetland distribution in the Camargue region of the Rhone Delta in southern France based on SPOT-5 time series images, with a the cross-verification accuracy up to 98.7%. The decision tree method was also used to extract river and lake wetlands based on Landsat data, with extraction accuracy ranging from 6% to 10% higher than that of the maximum likelihood method[12].

Low-resolution satellite data have a low spatial resolution, but their wide monitoring coverage, convenient access and abundant spectral information, especially the high revisit cycle of twice a day, make research on wetlands at regional and even global scales possible[13]. Xu et al.[14] used the normalized difference vegetation index (NDVI) of Moderate Resolution Imaging Spectroradiometer (MODIS) data to study wetlands in the Ruoergai Plateau, and the extraction accuracy reached 82%. Chen et al.[15] used the enhanced vegetation index (EVI) of MODIS time series data to monitor the permanent and seasonal waters and swamps of Dongting Lake, and the classification accuracy reached 87.87%. This research shows that, although low-resolution satellite data have less spatial resolution, they are still feasible for a wide range of wetland monitoring. Using Fengyun Medium Resolution Spectral Imager (FY3/MERSI) data from 2011 to 2018, this paper analyzed the time series original spectra and NDVI of wetlands such as lakes, reservoirs, paddy fields and swamps in Northeast China to build a wetland extraction decision tree model. CHEOS-1 data were used to verify the accuracy. The research contributes data and methods for the Chinese Fengyun series satellites to carry out large-scale wetland extraction.

2 Materials and methods

2.1 Study area

The study area is in Northeast China, with a geographical range of 38.73°N-53.56°N and 118.85°E -135.04°E, including Liaoning, Jilin and Heilongjiang provinces (Figure 1). The total area is 787300 km2. The region has one of the largest wetland areas in China, accounting for approximately 48.3% of the total national wetlands[16]. Wetlands in the study area have some of the greatest impacts on human activities. Twelve wetlands in Northeast China, such as Zhalong, are included in The List of Wetlands of International Importance in China, accounting for 24.5% of the national total.

Figure 1. Location of study area.
2.2 Data sources

FY3/MERSI data were selected from 2011 to 2018, including FY3B/MERSI, FY3C/MERSI and FY3D/MERSI. To facilitate the extraction of wetland information, remote sensing data were collected from April to October, and cloud coverage was lower than 20%. FY3/MERSI data were obtained from the China Satellite Data Service Network, and the CHEOS-1 data came from the terrestrial observation satellite data service platform. FY3/MERSI data were processed in the "Northeast China Wetland Remote Sensing Monitoring and Evaluation Application Demonstration System", CHEOS-1 data were processed in the ENVI software.

2.3 Research methods

This paper used the decision tree method based on the CART algorithm to extract the wetland classes and analyze the spectral characteristics of FY3/MERSI's lakes, reservoirs, swamps, paddy fields and other wetlands in the study area and the curve changes of NDVI time series. Extraction parameters and rules for different wetlands were developed in combination with the crop growth periods to identify wetlands in Northeast China.

2.4 Sample selection

A wetland classification system was established based on the characteristics of wetlands in the study area, including landform types such as lake and reservoirs, paddy fields, rivers, swamp wetlands and others. A key step was the selection of samples to establish the decision tree classification rules, which had a direct impact on the classification accuracy. Therefore, for selection of each sample, at least 300 points were selected from each landform using existing geographic background data and visual interpretation of CHEOS-1 satellite data with uniform distribution as the standard.

2.5 NDVI

The NDVI was used to analyze time series changes to lakes and reservoirs, rivers, paddy fields and swamp wetlands. The NDVI was calculated using the following formula[17]:

\[ NDVI = \frac{\rho_{\text{nir}} - \rho_{\text{red}}}{\rho_{\text{nir}} + \rho_{\text{red}}} \]  

where: \( NDVI \) is the normalized difference vegetation index; \( \rho_{\text{nir}} \) is reflectivity of the near infrared band, and \( \rho_{\text{red}} \) is reflectivity of the red band. In FY3/MERSI, 3-channel and 4-channel were used for calculation.

2.6 Decision tree classification methods

The CART algorithm is a decision tree algorithm proposed by Breiman. It is characterized by taking full advantage of the binary tree form, segmenting the sample with a segmentation rule, and repeating different segmentation rules cyclically until the data cannot be further segmented. The method is divided into two parts: tree growth and tree pruning. The process of tree growth classifies the subsets using training samples through discrimination at the tree nodes. Each time of discrimination is the process of a subset
division, until the tree growth phase is completed. The tree growth process uses the Gini coefficient[18], where the value is the corresponding error rate when the node's category identifier is arbitrarily chosen. The formula is as follows:

\[
\text{Gini Index} = 1 - \sum_{j=1}^{J} p_j^2(j)
\]  

(2)

where: \( p(j) \) is the probability of the node belonging to the j-type sample number of the total number of samples, and J is the number of sample categories. The function of tree pruning is to prevent the data from being over fitted by tree growth, and it uses the cross verification method to generate the optimal decision tree with the smallest error.

3 Results

3.1 Analysis of the spectral characteristics of wetlands

The spectral reflection characteristics of lakes and reservoirs, as well as paddy fields, rivers and swamps are shown in Figure 2. In mid-May, the spectral reflectance of these four types of wetlands was the same, occurring within the blue band and the green band. The spectral reflectance of paddy fields was higher than that of the other three wetlands, although this difference was only approximately 2%. The spectral reflectance of lakes and reservoirs in the near-infrared and mid-infrared bands was less than 5%, which is significantly lower than that of the other three wetlands. Typically, their spectral reflectance decreased continuously with increasing wavelength. The spectral curves of paddy fields, rivers and swamps were very similar, showing double peaks in the near-infrared and mid-infrared regions. The peak in the mid-infrared band was higher than the peak in the near-infrared band, and its reflectivity dropped sharply by approximately 6% near 940 nm (18 band).

In mid-August, the spectral curves of lakes and reservoirs, as well as paddy fields, rivers and swamps were very consistent throughout the visible light range. The curves for paddy fields and swamps were almost coincident. Likewise, the curves for rivers and lakes/reservoirs were almost coincident. In the near-infrared, rivers had a weak peak, but there were strong peaks for paddy fields and swamps, showing a difference of approximately 8%–10%. Using near-infrared data for both mid-May and mid-August, allowed lakes/reservoirs and rivers to be distinguished, although paddy fields and swamps could not be distinguished using these data. Further analysis of NDVI time series is needed to distinguish between these two classes.

Figure 2. Reflectance spectral curves for various types of wetlands: (a) mid-May; and (b) mid-August.
3.2 Analysis of change of the NDVI time series

The time series of NDVI values for lakes and reservoirs, as well as rivers, paddy fields and swamps from May to October are shown in Figure 3. During the growing season, the NDVI values for rivers, paddy fields and swamps were greater than 0, while the NDVI values for lakes and reservoirs were all below 0. The NDVI values for rivers did not change significantly with time, but fluctuated within an interval of 0.2. The NDVI values formed single-peaked curves for paddy fields and swamps, with maximum values in early August. The difference between paddy fields and swamps was that the NDVI values of the paddy fields began to increase rapidly in late June, yielding a steep slope. In contrast, the NDVI values of swamps increased slowly from late May reaching their highest values in early August, showing no quick change. Therefore, it was possible to distinguish between paddy fields and swamps using the spectral difference between early August and late May.

Figure 3. Normalized difference vegetation index (NDVI) time series for various types of wetlands from May to October.

3.3 Decision tree classification rules

In this paper, we selected a certain number of samples from the different landform types, such as lakes, reservoirs, paddy fields, swamp wetlands and rivers, to create classified training samples for each category. The growing season change data of time series NDVI was used as representative data for separability analysis of samples. The separability coefficient was greater than 1.9767, which means a good separability. The training samples were used to establish a decision tree model, generating 17 leaf nodes, as shown in Figure 4.
Figure 4. Classification rules of the CART algorithm decision tree.

3.4 Classification results

The above decision tree classification rules were used to extract lakes and reservoirs, rivers and swamp wetlands and paddy fields in Northeast China(Figure 5). The paddy fields in the study area were a one crop per annum type, which were distributed in the plain or valley areas such as the Liaohe Plain and the Sanjiang Plain. The water classes included lakes, reservoirs and rivers throughout Northeast China. The Songhua River in the north is wide, giving clear extraction results. Swamp wetlands were mainly distributed in the estuary of the Liaohe River, northwestern Jilin, and southern Heilongjiang. The total area of wetlands extracted by satellite remote sensing data was 79123.4km², as shown in Table 1. Overall, the distribution area and range of wetlands in the map of wetland extraction were consistent with the description of the wetlands, which indicates that the wetland extraction using this method is effective.

Figure 5. Wetland extraction for Northeast China in FY3/MERSI data.

Table 1. Wetland area extraction results in Northeast China.

| Province   | Swamp wetlands | Paddy      | Lake and Reservoir | River  | Total    |
|------------|----------------|------------|--------------------|--------|----------|
| Heilongjiang| 13401.7        | 36586.4    | 3039.2             | 5973.7 | 59001.0  |
| Jilin      | 922.6          | 8997.4     | 1270.5             | 329.6  | 11520.1  |
| Liaoning   | 960.4          | 6477.9     | 952.6              | 211.4  | 8602.3   |
3.5 Accuracy verification

The verification of wetlands extent in FY3/MERSI data mainly focused on analysis from two aspects: the spatial match and the accuracy of area extraction. As Northeast China has a north-south span covering 16° of latitude and an east-west span covering 20° of longitude, the verification of the whole area using high-resolution satellite data was not possible. We selected typical test areas for accuracy verification with CHEOS-1 remote sensing data.

Both CHEOS-1 data and FY3/MERSI data used the WGS-84 coordinate system and UTM projection. FY3/MERSI data were resampled to 16 m and the 2-cropping data were removed using ENVI software to ensure the consistency of area. CHEOS-1 data wetland extraction also used the decision tree method and GIS data for comparison, to improve extraction accuracy.

In this paper, two swamps and one lake were chosen to verify the wetland extraction results from FY3/MERSI data. The verification area of swamps included four types of wetlands, which had large areas of relatively pure wetland and fragmented areas that were mixed with other landform types for different degrees of purity. After consideration, we chose the single-cropping CHEOS-1 data for Panjin City in Liaoning Province and the single-cropping CHEOS-1 data for Zhalong in Heilongjiang Province as the verification data. The overall classification accuracy of the Panjin wetlands was 84.8%, with a Kappa coefficient of 0.7048 (Figure 6).

![Figure 6. Panjin wetland comparison of CHEOS-1 data and FY3/MERSI data: (a) CHEOS-1; (b) FY3/MERSI.](image)

The overall classification accuracy of the Zhalong wetland was 93.4%, with a Kappa coefficient of 0.6446 (Figure 7). In the Zhalong wetland, water covered 103.9km² and 113.2km² and the swamp area was 980.0km² and 1060.7km², for FY3/MERSI data and CHEOS-1 data, respectively. The area extraction accuracy for the Zhalong wetlands was over 91.8%.

![Figure 7. Zhalong wetland comparison of CHEOS-1 data and FY3/MERSI data: (a) CHEOS-1; (b) FY3/MERSI.](image)

The overall classification accuracy for Wolong Lake was 91.9% (Figure 8), and the area of Wolong Lake extracted by FY3/MERSI data was 57.43 km², compared with 57.97 km².
from the CHEOS-1 data. The area extraction accuracy for Wolong Lake reached 99%. This was related to the rectangular water surface and clear boundary of Wolong Lake.

Figure 8. Wolong Lake comparison of CHEOS-1 data and FY3/MERSI data: (a)CHEOS-1; (b) FY3/MERSI; (c) Comparison image for CHEOS-1 and FY3/MERSI.

4 Conclusions

In this paper, time series FY3/MERSI data were used to analyze the spectral information of wetland landforms. Combined with the characteristics of crop growth and development, decision-tree method was used to extract the wetlands of Northeast China. The main conclusions were as follows:

There were differences in the spectral curves for different types of wetlands, especially for the swamp wetlands and paddy fields with growth cycles. Lake reservoirs and river wetlands could be distinguished using near-infrared data in mid-May and mid-August.

There were clear change characteristics in the time series for the NDVI. It was possible to distinguish paddy fields and swamps through NDVI differences between early August and late May.

The lakes and reservoirs, rivers and swamp wetlands and paddy fields in Northeast China were extracted using the decision tree classification model. The total area of wetlands was 79,123.4km². The single-crop CHEOS-1 data for Panjin City in Liaoning Province were used for verification. The area accuracy of wetlands was over 91.8%, the overall classification accuracy of wetlands was over 84.8% and the Kappa coefficient was over 0.6446.

Supported by the Department of Science & Technology of Liaoning Province (2018108004), and State Administration of Science. Technology and Industry for National Defense, PRC (70-Y40G09-9001-18/20).

References

1. X. Mou, X. Liu, B. Yan, B. Cui, Wetland Sci, 13, 19(2015).
2. P. Wang, R. Wan, G. Yang, Wetland Sci, 15, 114(2017).
3. D. Han, Y. Yang, Y. Yang, K. Li, Acta Ecol. Sin., 32, 1293(2012).
4. Z. Szantoi, F.J. Escobedo, A. Abd-Elrahman, L. Pearlstine, B. Dewitt, S. Smith, Environ Monit Assess, 187, 262(2015).
5. M. Gracz, P.H. Glaser, Wetl Ecol Manag, 25, 87(2017).
6. V. Klemas, Int J Remote Sens, 34, 6286(2013).
7. R. Chopra, V.K. Verma, P.K. Sharma, Int J Remote Sens, 22, 89(2001).
8. R. Martin, L. Brabyn, C. Beard, J Spat Sci, 59, 269(2014).
9. J.C. Xiao, W.X. Ou, H.Y. Fu, Acta Ecol. Sin., 33, 7496(2013).
10. R.C. Frohn, E.D. Amico, C. Lane, B. Autrey, J. Rhodus, H.X. Liu, Wetlands, 32, 289(2012).
11. A. Davranche, G. Lefebvre, B. Poulin, 114, 552(2010).
12. L. Zhang, R. Ruan, Geo. & Spatial Info. Tech., 38, 87(2015).
13. M. Zhang, Y.N. Zeng, Y.S. Zhu, J REMOTE SENS, 21, 479(2017).
14. H. Xu, Z. He, S. Dan, B. Dan, C. Yang, J Glaciol Geocryol, 29, 450(2007).
15. Y. Chen, Z. Niu, S. Hu, H. Zhang, J. Hydrol. Eng., 47, 1093(2016).
16. Y.Y. Chen, J.Z. Ma, X.T. Liu, China wetlands encyclopedia, 10(2009).
17. J.A. Schell, Nasa Special Publication, 351, 309(1973).
18. W. Yu, X. Liu, D. Sun, W. Jiang, W. Qu, Trans. Chin. Soc. Agric. Eng., 32, 212(2016).