Lake Ice Phenology Extraction using Machine Learning Methodology

Pengfei Xie¹, ², Yubao Qiu²*, XingXing Wang², Lijuan Shi² and Wenshan Liang²

¹. School of Artificial Intelligence, Hebei University of Technology, Tianjin 300400; ². Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094.

Correspondence: qiuyb@aircas.ac.cn

Abstract. Lake-ice phenology is one of the key cryosphere indicators. Based on changes in time series of daily lake ice coverage and the microwave brightness temperature, two machine-learning methods were selected for estimating the phenology of lake ice. Using MODIS lake ice coverage time series, a convolutional neural network (CNN) was used to automatically classify lakes as either having or not having regular freezing signatures. The classification results were compared to the test samples and classification accuracies of 91% and 100% were found for lakes that do and do not freeze annually, respectively. In order to extract the lake ice phenology from passive microwave brightness temperature data, support vector regression (SVR) was used: the freeze-up start (FUS) and break-up end (BUE) were extracted using time series of estimated emissivity. The lake ice phenology results obtained using SVR were compared with the reference results. The values of R² for FUS and BUE were 0.8928 and 0.8899, respectively. These results show that it is possible to use AI methods for the fast extraction of the phenology of lake ice and that these methods can be applied to lakes globally.

Keywords: Lake-ice phenology, Convolutional Neural Network (CNN), Support Vector Regression (SVR), MODIS, Passive Microwave Radiometer

1. Introduction

“Lake Ice Phenology” is a term used to describe the seasonal cycle of lake-ice cover, including the freezing period, thawing period and ice-cover duration [1]. Remote sensing monitoring of lake ice phenology mainly uses optical and passive microwave data. Optical remote sensing uses the difference between the reflectivity of ice and water to identify lake freezing and thawing. To extract the lake ice phenology, thresholds are set based on the average reflectivity of the whole lake [2] or the proportion of lake ice to water for the whole lake [5]. In this way, the monitoring of lake ice phenology in Alaska and European Alpine regions was realized. The resolution of MODIS data is between 250 m and 1000 m, and it can be used to extract ice phenology information for smaller lakes [6]. However, there are still many lakes for which phenological information cannot be extracted using MODIS. Some of these lakes are small, some are irregular (for example, the ratio of the circumference to the area is high), and some remain ice-free all year round due to the geology or regional climate. Cloud and polar night conditions also limit the ability of optical imagery to capture details of lake ice phenology in the
Arctic. How to efficiently and quickly judge the lakes that can extract the lake ice phenology with large area and long-time series is the premise of using MODIS data for lake ice phenology monitoring. The freezing-thawing state of lakes can also be monitored by passive microwave brightness temperature, which is mainly based on the variation principle of ice and water emissivity during freezing-thawing of lakes. The lake ice phenology parameters were obtained by setting the threshold of the extracted lake brightness temperature [7-10]. Although the spatial resolution of passive microwave data is relatively coarse, the data has a long time series (from 1978 to now), and is less affected by the environment such as cloud and polar night. The data of passive microwave remote sensing can provide a basis for the study of phenology monitoring and climatology of lake ice in the long time scale or large spatial scale. Similarly, due to the difference in shape, size and geographical location of the lakes, the brightness temperature of the lakes in the freezing/thawing period are quite different, so it is difficult to achieve the extraction of lake ice phenology in a large range of Eurasia continent with a unified threshold. Therefore, it is necessary to develop a rapid and uniform method to extract lake ice phenology for a wide range of lakes.

With the aim of solving these two problems, in this study the characteristics of MODIS lake ice coverage time series and passive microwave brightness temperature time series were analysed. A convolution neural network was used to realize the classification of lakes based on the MODIS lake ice coverage time series so as to identify the lakes where MODIS data cannot obtain lake ice phenology. At the same time, AMSE-E and AMSR-2 H-polarized 18.7 GHz microwave brightness temperature data were used to approximately calculate the emissivity of the lakes, and the lake ice phenology was extracted using support vector regression.

2. Methodology and Data Samples

2.1. Time Series Characteristics Analysis of MODIS Lake Ice Coverage and Passive Microwave Brightness Temperature for Lake

2.1.1. Time Series Characteristics Analysis for MODIS Lake Ice Coverage

The lake ice coverage time series extracted from MODIS optical data can be used to determine lake ice phenology. According to the existing classification results [6] for Tibetan plateau lakes, the lake ice coverage time series have different characteristics for lakes for which the lake ice phenology is difficult to extract (Figure 1) and it is hard to model and identify these lakes using traditional mathematical methods. Deep learning can autonomously extract the feature database of sample datasets using training samples to enable the identification of lakes.

![Figure 1. MODIS lake ice coverage time series for lakes which it is difficult to extract the lake ice](image-url)
2.1.2. Time Series Characteristics Analysis of Passive Microwave Brightness Temperature for Lakes

The composition and physical properties of an object determine the response of microwave radiation that interacts with it. The brightness temperature of the microwave radiation captured by a microwave radiometer is affected by the emissivity of the object, the observation angle and the polarization mode. In ideal conditions, it is assumed that the brightness temperature received by the sensor is the radiation brightness temperature of the object itself. Ignoring the influence of the roughness of the object's surface and other factors, the brightness temperature of the object's microwave radiation can be expressed by equation (1):

\[ T_b(f, \theta, p) = e(f, \theta, p)T \]  

where \( f \) is the frequency, \( e \) is emissivity of the object, which is one of the main physical parameters that affect microwave radiation. \( p \) is the polarization mode, which includes vertical polarization and horizontal polarization. \( \theta \) is the incidence angle of the satellite sensor. \( T \) and \( T_b \) are the temperature and the brightness temperature of the object, respectively.

When \( \theta = 55^\circ, T = 0^\circC, f = 18.7 \text{ GHz}, p = \text{H} \), under ideal conditions, the emissivity of water and ice were calculated to be 0.2726 and 0.7828, respectively. This difference in emissivity between water and ice can be used to distinguish between the freezing and thawing of lakes. When the ice–water phase change occurs in the lake, the brightness temperature of the lake also shows an obvious abrupt change (Figure 2). Using these abrupt change points, clear freezing and thawing dates (freeze-up start (FUS) and break-up end (BUE)) dates can be obtained. Due to the different geographical locations of lakes, the variations in brightness temperature caused by temperature differences are themselves very variable and there are no uniform rules. In order to reduce the influence of temperature change on the determination of the freeze/thaw status of lakes, we calculated a sample approximate emissivity time series for lakes by combining the temperature data with equation (1) and then extracted the lake ice phenology using a machine-learning regression model.

2.2. Machine-Learning Methods

2.2.1. Lakes Classification Using a Convolution Neural Network

A convolutional neural network (CNN) is a multi-layer neural network composed principally of an input layer, convolutional layer, pooling layer and output layer. The convolutional layer and pooling layer are hidden layers. The input layer is used to receive the original data (in this case, a total of 5,834 days of MODIS daily lake ice coverage covering the period from 2002 to 2018). The convolutional layer extracts sample features through the receptive field; the features obtained by the receptive field are independent of the translation, scaling and rotation of the sample data features[11], thus reducing the influence of noise. The pooling layer reduces the amount of data to be processed by using the principle of local images correlation, which, in this case, reduced the number of characteristics of the
lake ice coverage time series. Compared with the commonly used feed-forward neural network, the training efficiency of a CNN is higher. The output layer maps the extracted features to the final labels (types of lake). As a feature-learning method, CNN transforms the original data into more abstract expressions through a simple non-linear model, and does not need too much manual participation during the feature-learning process [12].

The structure of the convolutional neural network designed in this experiment is shown in Figure 3. The network was mainly composed of three convolution layers having 50, 250 and 1000 convolution kernels, together with a maximum pooling layer, a global maximum pooling layer and a fully-connected layer.

2.2.2. Support Vector Regression for Lake Ice Phenology Extraction

The Support Vector Machine (SVM) is a machine-learning method based on statistical learning theory [13]. The main features are that a kernel function is applied to transform a non-linear space problem into a linear space problem and problems that exist in traditional classification methods, such as over-learning dimension disaster local minimization in classical learning methods, can be effectively avoided by applying the principle of structural risk minimization [14]. In this study, the SVM had a good generalization ability if the samples of FUS and BUE were small. Support Vector Regression (SVR) can be carried out by extending the SVM from a classification problem to a regression problem [15]. SVR looks for a regression plane through the training sample set so that the distance between all the data and the plane is the smallest, thus obtaining the final regression model.

2.3. Preparation of Data Samples

2.3.1. Sample from the MODIS Lake Ice Coverage Time Series

In this study, a one-dimensional time series of daily lake ice coverage for 490 lakes on the Tibetan Plateau covering the period from 2002 to 2018 (a total of 5,834 days) [6] was used as the feature vector. In the cited article, the lakes are divided into four types. For type-I lakes, it is difficult to extract the freeze-thaw signal due to the influence of salt flats. For type-IV lakes, it is difficult to extract the lake ice phenology either because of the size and shape of the lakes or because the lakes do not freeze in winter. The characteristics of the time series of type-I and type-IV lakes are not clear and it is difficult to extract the lake ice phenology for these types. Therefore, these lakes are classified as belonging to the same category. In the case of type-II lakes, the four lake ice phenology parameters can be calculated. For type-III lakes, only the FUS and BUE parameters can be extracted. In order to have a balanced distribution within the sample set, 150 type-II lakes, 60 type-III lakes and 150 type-IV lakes were selected as training samples for random equalization; 50 type-II lakes, 30 type-III lakes and 50 type-IV lakes were selected as test samples for random equalization.

2.3.2. Data Samples for Time Series of Approximate Passive Microwave Emissivity

For the monitoring of lake freezing and thawing with passive microwave data (see 2.1.2), we selected the horizontally polarized passive microwave brightness temperature at 18.7 GHz from August 2002 to July 2017. Brightness temperatures for the period August 2002 to July 2010 were obtained from the
Calibrated Passive Microwave Daily EASE-Grid 2.0 (CETB), which has a resolution of $6.25 \times 6.25$ km. The data from August 2012 to July 2017 were based on the original AMSR-2 orbital brightness temperature released by the Japanese Aerospace Exploration Agency (http://global.jaxa.jp/); these data had a resolution of $22 \times 14$ km. We also examined a range of temperature variables derived from the ERA-Interim Reanalysis 2-m air temperature dataset at a 0.5-degree spatial resolution [16]. The nearest-neighbor method was used to extract the daily temperature and brightness temperature of lakes. The lake ice phenology for 70 lakes on the Tibetan Plateau and Mongolian Plateau was extracted for the period 2002 to 2017 using the brightness temperature time series. August 1 to July 31 of the next year was taken as a hydrological year. Ignoring the influence of atmospheric and internal ice temperature on the passive microwave observations, and using the temperature data for the lakes, equation (1) was used to obtain the approximate emissivity of the lakes at 18.7 GHz horizontal polarization for use as sample features (Figure 4). Due to the absence of passive microwave data from October 2011 to July 2012, the lake brightness temperature sequences from August 2010 to July 2012 were not used. Finally, 884 sets of FUS samples and 881 sets of BUE samples were obtained. Random non-repeat sampling was used to select 70% of the FUS samples and BUE samples as the training samples for the SVR, and the remaining 30% were used as the test set for the accuracy test.

![Figure 4. Approximate emissivity sample (Qinghai Lake) at 18.7 GHz H-polarization](image)

3. Results and Analysis

3.1. Classification of Lakes Based on MODIS observations

The convolution neural network was used to train the sample set; the variations in model accuracy and losses that occurred during the training process are shown in Figure 5. It can be seen that the accuracy of the model training converges rapidly. The accuracy of the test set reaches its highest value at a training epoch of 27. A detailed analysis of the classification results for the test set (see Table 1) showed that the precision of the classification predictions for type-II lakes was 92%, compared to 90% for type-III lakes and 100% for type-I and type-IV lakes. Due to the similar characteristics of lake ice coverage time series for type II and type III lakes during periods of complete melting and initial freezing, the classification of some of the type II and type III lakes is confused. The convolution neural network can realize the classification of a large number of lakes and remove the lakes that are difficult to extract lake ice phenology. This can provide a basis for extracting the lake ice phenology of lakes over large areas and for long time periods.
Figure 5. CNN model accuracy and losses for the training and test samples

| Table 1. Lake classification confusion matrix |
|-----------------------------------------------|
| Predicted class | Type II | Type III | Type I and Type IV | Total | Precision |
| Actual class    |         |          |                    |       |           |
| Type II         | 46      | 4        | 0                  | 50    | 0.92      |
| Type III        | 3       | 27       | 0                  | 30    | 0.9       |
| Type I and type IV | 0     | 0        | 50                 | 50    | 1         |

3.2. Lake-Ice Phenology Extraction Based on Passive Microwave Observations

Regression was conducted on the approximate emissivity test set that was extracted from the passive microwave data using the support vector regression method. Results for the lake ice phenology (FUS and BUE) were thus obtained, and these were compared with visually interpreted lake ice phenology (as shown in Figure 6).

Figure 6. Comparison of lake ice phenology derived from visual observations with that predicted by the SVR

Both the FUS and BUE passed the validity test. The $R^2$ for the FUS was 0.8928 and the RMSE was 7.6058; the $R^2$ for BUE was 0.8899 and the RMSE was 9.2254. This shows that the SVR method can effectively extract lake ice phenology. SVR is greatly affected by the amount of data and the quality of
samples. The sample data used in this study did contain some errors. Since the visual interpretation of the lake ice phenology was obtained from the lake brightness temperatures without any standardization of the linear interpolation, the missing passive microwave brightness temperature data because of the orbital scanning vacancy may have introduced some errors. In addition, there is no unified standard for lakes in different regions that allows lake ice phenology to be extracted by applying a simple threshold to time series of lake brightness temperatures: this will also produce some errors in the sample data.

4. Conclusion
In this study, the machine-learning methods were applied to extract the ice phenology of a large number of lakes using remote sensing data. Using the MODIS lake ice coverage time series, a CNN was used to realize the classification of lakes in cases both where the phenology could be extracted and where it could not. The classification precision for type-II and type-III lakes in the test set that had clear lake ice phenology information was 92% and 90%, respectively whereas the classification accuracy for type-I and type-IV lakes with difficult-to-extract lake ice phenology was 100%. This method thus provides a foundation for the rapid extraction of lake ice phenology based on MODIS lake-ice data. The approximate emissivity time series for the lakes were extracted from passive microwave data, and the support vector regression method was implemented to automatically extract the lake ice phenology. Using the test set, the R² values for FUS and BUE were found to be 0.8928 and 0.8899, respectively. This implies that this new method can be used to obtain the lake ice phenology for a large range of lakes.

Following this, using MODIS data, the lake ice coverage time series for lakes were extracted. The convolutional neural network was used to classify a large number of lakes in Eurasia, and the lake ice phenology was extracted. In order to avoid the influence of the polar night in the Arctic region, the approximate emissivity time series for the lakes were obtained based on passive microwave data, and lake ice phenology was extracted by SVR to supplement the data on the lake ice phenology for Eurasia. Finally, a lake ice phenology dataset for the entire Eurasian continent was obtained, which will provide data for studies of the three-pole climate change.

References
[1] Duguay C R, Bernier M, Gauthier Y, and Kouraev A 2014 Remote Sensing of the Cryosphere, Remote sensing of lake and river ice. chapter 12 pp 273-306.
[2] Latifovic R, and Pouliot D 2007 Remote Sensing of Environment, Analysis of climate change impacts on lake ice phenology in Canada using the historical satellite data record. 106(4), 492–507.
[3] Weber H, Riffler M, Nõges T, and Wunderle S 2016 Remote Sensing of Environment, Lake ice phenology from AVHRR data for European lakes: An automated two-step extraction method. 174, 329–340.
[4] Šmejkalová T, Edwards M E and Dash J 2016 Scientific Reports, Arctic lakes show strong decadal trend in earlier spring ice-out. 6(1).
[5] Reed B, Budde M, Spencer P, and Miller, A E 2009 Remote Sensing of Environment, Integration of MODIS-derived metrics to assess interannual variability in snowpack, lake ice, and NDVI in southwest Alaska. 113(7), 1443–1452.
[6] Qiu Y, Xie P, Leppäranta M, Wang X, Lemmetyinen J, Lin H, and Shi L 2019 Big Earth Data, MODIS-based Daily Lake Ice Extent and Coverage dataset for Tibetan Plateau. 3(2), 170–185.
[7] Kang K-K, Duguay C R, and Howell S E L 2012 The Cryosphere, Estimating ice phenology on large northern lakes from AMSR-E: algorithm development and application to Great Bear Lake and Great Slave Lake, Canada. 6(2), 235–254.
[8] Qiu Y, Guo H, Ruan Y, Fu X, Shi L and Tian B 2017 Science Data Bank, A dataset of microwave brightness temperature and freeze-thaw for medium-to-large lakes over the High Asia region 2002 – 2016. (2).
[9] Cai Y, Ke C-Q, and Duan Z 2017 Science of The Total Environment, Monitoring ice variations in Qinghai Lake from 1979 to 2016 using passive microwave remote sensing data. 607-608, 120–131.
[10] Du J, Kimball J S, Duguay C, Kim Y, and Watts J D 2017 The Cryosphere, Satellite microwave assessment of Northern Hemisphere lake ice phenology from 2002 to 2015. 11(1), 47–63.
[11] Huang J T, Li J, and Gong A Y 2015 IEEE International Conference on Acoustics. IEEE, An analysis of convolutional neural networks for speech recognition. 4989-4983.
[12] LeCun Y, Bengio Y and Hinton G 2015 Nature, Deep learning. 521(7553), 436–444.
[13] Vapnik V 1998 Statistical learning theory. chapter 10-11, pp.401-492.
[14] Ding S-F, Qi B-J, Tan H-Y 2011 Journal of the University of Electronic Science and Technology of China, An Overview on Theory and Algorithm of Support Vector Machines. 40(1),2-10.
[15] Smola A J, and Schölkopf B 2004 Statistics and Computing, A tutorial on support vector regression. 14(3), 199–222.
[16] D P Dee, S M Uppala, A J Simmons, P Berrisford, P Poli, S Kobayashi, U Andrae, M A Balmaseda, G Balsamo, P Bauer, P Bechtold, A C M Beljaars, L van de Berg, J Bidlot, N Bormann C Delsol, R Dragani M Fuentes, A J Geer, L Haimberger, S B Healy, H Hersbach, E V Holm, L Isaksen, P Källberg, M Köhler, M Matricardi, A P McNally, B M Monge - Sanz, J - J Morcrette, B - K Park, C Peubey, P de Rosnay, C Tavolato, J - N Thépaut, F Vitart 2011 Quarterly Journal of the Royal Meteorological Society, The ERA - Interim reanalysis: configuration and performance of the data assimilation system. 137(656): 553-597.

**Funding:** Supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant number XDA19070201), Multi-Parameters Arctic Environmental Observations and Information Services (Grant number 2017YFE0111700) and the International Cooperation Program of Chinese Academy of Sciences (Grant number 131211KYSB20170041).