Front. Earth Sci. 2025, 19(1): 41−53
https://doi.org/10.1007/s11707-022-1052-8

RESEARCH ARTICLE

Long-term evolution of winter habitats in Poyang Lake derived from satellite imagery using machine learning methods

Siqin QI[1], Jiayi PAN (✉)[1,4], Adam T. DEVLIN[2,3,4]

1 School of Geography and Environment, Jiangxi Normal University, Nanchang 330022, China
2 Cooperative Institute for Marine and Atmospheric Research, School of Ocean and Earth Science and Technology, University of Hawaii at Mānoa, Honolulu, HI 96822, USA
3 Department of Oceanography, University of Hawaii at Mānoa, Honolulu, HI 96822, USA
4 Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong, China

Abstract Poyang Lake is a freshwater lake in China which is a vital winter habitat for many kinds of wildlife and a critical component of the regional ecology. Here, we use Landsat satellite imagery to systematically assess habitat characteristic changes from 1990 to 2021. Four machine learning methods including random forest (RF), gradient boosting tree (GBT), support vector machine (SVM) and classification and regression trees (CART) are analyzed by comparing the overall accuracy and Kappa coefficients. The results show that the accuracy of random forest is higher than that of the other three machine learning methods. The long-term characteristics of Poyang Lake winter habitat types are extracted from Landsat satellite images using the RF method. These results show that the mudflat area was larger than water surface and sand. After 2012, more mudflat area had been converted into grassland, which is related to the early onset of the winter dry season in the Poyang Lake area. The habitats were scattered and fragmented from 1990 to 1998; after 1997−1998, however, the degree of landscape patch density and interference decreased, indicating a decreased impact of human-related interference and natural factors on the evolution of habitats in and around Poyang Lake.

Keywords Poyang Lake, habitats, machine learning, landscape index

1 Introduction

Poyang Lake is the largest freshwater lake in China. The water level of the lake exhibits a significant seasonal variation (Mu et al., 2020) due to the influence of the East Asian monsoon. During the wet season (April to September) the water rises and the areal extent of Poyang Lake reaches up to 3000 km^2 (Feng et al., 2012). During the dry season (October to March) the water level declines significantly, and parts of the lake floor emerge, leading to a number of mutually unconnected dish-lakes, as the overall lake water surface shrinks to 1000 km^2 (Zhou et al., 2018). With this decline in water level comes increased areas of wetlands and other land types, constituting important habitats for wintering migratory birds such as white cranes as well as other wildlife (Kanai et al., 2002; Mu et al., 2020), thus rendering Poyang Lake rich in biodiversity.

The habitat conditions are affected by a combination of natural factors and human activities (Li et al., 2020a). Extensively developed land use may induce significant burdens on ecological environments (Naveh, 1994; Li et al., 2020b). The landscape pattern is an important parameter for quantifying habitat heath conditions and is mainly defined by landscape heterogeneity and diversity (Li et al., 2021). Degradation of habitats can affect landscape patterns and lead to landscape fragmentation (Khadka et al., 2021). Landscape pattern analysis can be used to describe and explain the process of associated habitat (land) type variations (Zhang et al., 2003; Wang and Ogawa, 2017).

Satellite images acquired from remote sensing methods are an important data source to study land type and habitat characteristics. Zhao et al. (2003) classified satellite images using the maximum likelihood method, grouping land covers into five types: water body, lacustrine vegetation, floodplain, cropland, and open land.

Received November 21, 2022; accepted January 29, 2023

E-mail: panj@cuhk.edu.hk

Li et al. (2022) used the nearest neighbor method to classify the Poyang Lake habitat areas into five types: forest, grassland, cropland, construction land, and unused land. These studies mainly used unsupervised classification algorithms.

Various machine learning methodologies and algorithms including the support vector machine (SVM), classification and regression trees (CART), gradient boosting regression tree (GTB), and random forest (RF) methods have been increasingly applied to analyze land use changes from different remotely sensed data sources. These machine learning algorithms have been shown to perform better in land type classifications than conventional supervised/unsupervised methods. Pal (2008) used three classifiers for land cover classification and found that the classification accuracy of the support vector machine (SVM) method is better than maximum likelihood classifier (MLC) and neural network (NN) methods. Shao and Lunetta (2012) compared the support vector machine with neural networks and CART algorithms, showing that SVM was superior to other classifiers, especially with small training sample sizes. Jia et al. (2014) utilized different classifiers for land cover classification, demonstrating that the support vector machine (SVM) method can achieve better classification accuracy than the traditional maximum likelihood classifier (MLC) method.

Recently accelerating climate change and anthropogenic factors have significantly affected Poyang Lake habitats, degraded its ecological function (Zhang et al., 2012; Chen et al., 2018; Yao et al., 2018; Zhou et al., 2019), and endangered the wintering wildlife in Poyang Lake (Tang et al., 2016; Xia et al., 2016). Rapid urbanization has led to land use change and habitat degradation in the Poyang Lake Eco-Economic Zone (PLEEZ) from 1988 to 2018 based on the InVEST (Integrated Valuation of Ecosystem Services and Tradeoffs) model analysis (Zhang et al., 2020). The Poyang Lake water area slightly decreased due to changes in temperature, evaporation, and rainfall, as well as anthropogenic factors such as the Three Gorges Project (Liu et al., 2021). It has been suggested that the long-term lake shrinkage might favor the growth of wetlands via an alteration of habitat conditions in the Poyang Lake area (Tang et al., 2016). A vegetation-based IBI (V-IBI) method was tested to assess the ecological status of the Poyang Lake wetland and classify its overall health conditions. These results showed that development of appropriate conservation measures is vital for wetland restoration and conservation (Yang et al., 2018).

Thus, investigation of the winter habitat characteristics and associated landscape patterns are critical to understanding variations of the Poyang Lake ecological system. Here, machine learning methods are applied to obtain Poyang Lake winter habitat types and landscape pattern indexes from Landsat satellite images, and the accuracy of the various methods are compared. The long-term variations of habitat types and landscape pattern indices are analyzed to reveal their responses to different factors. The remainder of the paper is organized as follows. Section 2 describes the study area and data, Section 3 gives the methodology used in this study, Section 4 presents the results, discussions are in Section 5, and conclusions are in Section 6.

2 Study area and data

2.1 Study area

Poyang Lake is located south of the middle reaches of the Yangtze River (28°22′−29°45′N and 115°47′−116°45′E). Water inflow to Poyang Lake come from five rivers, the Raohe, Xinjiang, Fuhe, Ganjiang, and Xiushui, and local precipitation is also a significant water source for the lake during the wet season. The Poyang Lake water is discharged into the Yangtze River through a relatively narrow channel on its north side between Xingzi and Hukou (Fig. 1). The study area is shown in Fig. 1, showing the maximum water surface area of Poyang Lake and the surrounding areas which consist of farmlands and other areas. These regions constitute the

Fig. 1 The study area covering Poyang Lake and surrounding regions.

major winter habitats of the Poyang Lake area, and they are selected to reveal long-term variations of the habitat characteristics.

2.2 Data

We here employ 18 cloud-free level L2A TM and OLI images of Landsat 5 and 8 to extract the Poyang Lake winter habitat types from 1990 to 2021, listed in Table 1. Landsat 8 OLI has eight bands with a spatial resolution of 30 m and a panchromatic band of 15 m spatial resolution. Landsat 5 TM consists of six bands with a spatial resolution of 30 m and one band with a spatial resolution of 120 m. The repeat cycle of Landsat satellite observations is 16 days.

3 Methodology

3.1 Habitat indices

Several land type indices are calculated for assisting machine learning classifications. These indices include the normalized difference vegetation index (NDVI), improved normalized difference water index (MNDWI), normalized building index (NDBI), and enhanced vegetation index (EVI), which can be used to identify basic types of land surfaces. The NDVI index shows vegetation distribution and reflects vegetation growth status, given by (Kaufman and Tanre,1992),

Table 1 Landsat images used in this study. All Landsat images were obtained from the GEE (Google Earth Engine) platform

Date	Sensor	Resolution/m
23 Jan 1990	Landsat 5 TM	30
17 Dec 1993	Landsat 5 TM	30
07 Dec 1995	Landsat 5 TM	30
10 Jan 1997	Landsat 5 TM	30
18 Dec 1999	Landsat 5 TM	30
08 Jan 2002	Landsat 5 TM	30
15 Dec 2004	Landsat 5 TM	30
21 Dec 2006	Landsat 5 TM	30
10 Dec 2008	Landsat 5 TM	30
11 Jan 2009	Landsat 5 TM	30
14 Jan 2010	Landsat 5 TM	30
24 Dec 2013	Landsat 8 OLI	30
13 Feb 2015	Landsat 8 OLI	30
16 Dec 2016	Landsat 8 OLI	30
19 Dec 2017	Landsat 8 OLI	30
05 Feb 2018	Landsat 8 OLI	30
23 Jan 2019	Landsat 8 OLI	30
12 Jan 2021	Landsat 8 OLI	30

$$\mathrm{NDVI} = \frac{NIR - \mathrm{Red}}{NIR + \mathrm{Red}}, \tag{1}$$

where *NIR* represents the near-infrared band value (near-infrared B4 in Landsat 5 and near-infrared B5 in Landsat 8), and Red denotes the red band value (B3 in Landsat 5 and B4 in Landsat 8). The MNDWI index can be used to identify water surface information in satellite images (Xu, 2006) as

$$\mathrm{MNDWI} = \frac{(P(\mathrm{Green}) - p(SWIR))}{(p(\mathrm{Green}) + p(SWIR))}, \tag{2}$$

where Green is the green band value (band B1 in Landsat 5 and band B2 in Landsat 8), and *SWIR* is the short-wave infrared band value (the short wave infrared B5 in Landsat 5 and the short wave infrared B6 in Landsat 8). *P* is a constant. The NDBI index can determine building land surface (Zha et al., 2003), given by

$$\mathrm{NDBI} = \frac{SWIR - NIR}{SWIR + NIR}, \tag{3}$$

where *SWIR* is the short-wave infrared band value (the short wave infrared B5 in Landsat 5 and the short wave infrared B6 in Landsat 8), and *NIR* is the near-infrared band value (near-infrared B4 in Landsat 5 and near-infrared B5 in Landsat 8). The EVI index inherits the advantages of the NDVI index, and improves the issues caused by vegetation saturation, incomplete atmospheric impact corrections, and soil background. Thus, EVI enhances the sensitivity of vegetation in the high biotope and reduces soil background and atmospheric influences. The EVI index is important for monitoring of vegetation changes with high sensitivity and superiority (Liu and Huete, 1995):

$$\mathrm{EVI} = 2.5 \times \frac{(NIR - Red)}{(NIR + 6 \times \mathrm{Red} - 7.5 \times \mathrm{Blue} + 1)}, \tag{4}$$

where Red represents red band B3 in Landsat 5 and red band B4 in Landsat 8, and Blue repents band B1 in Landsat 5 and B2 in Landsat 8. *NIR* is the near-infrared B4 in Landsat 5 and near-infrared B5 in Landsat 8.

The quality of sample points selected in satellite images is directly related to the accuracy of habitat information extracted, so typical and pure representative pixels should be selected as the sample points. We thus uniformly select sample points from the Landsat images acquired in the GEE. These sample points are randomly arranged and divided into training samples and test samples according to a ratio of 7:3. Finally, RF, SVM, GTB, and CART are used for classification. The confusion matrix, overall accuracy, and Kappa coefficient are obtained.

3.2 Machine learning algorithms

We use four machine learning algorithms: random forest (RF), support vector machine (SVM), gradient boosting

tree (GBT), and classification and regression tree (CART), to classify and analyze the Poyang Lake habitat types.

3.2.1 Random forest

The random forest method is a machine learning ensemble algorithm that combines multiple weak classifiers, such as classification and regression tree, and final classifications are based on their voting results. Samples and features of each weak classifier are selected randomly from the total samples with full features to ensure the random forest classification will be low-noise and properly fit. Each weak classifier may choose an optimal part among randomly assigned features as a node to control the subtree partitioning, enhancing the model generalization ability. Because each weak classification term is independent and integrated with the others, the random forest can adaptably deal with nonlinear and multi-feature data. In general, the random forest model is simple to implement, fast to train, has high precision, and has significant advantages in large-scale processing of data. Two parameters need to be set in the random forest establishment step: the number of classification trees and the number of predictor variables used for each node segmentation. Although many classification trees will bring computational complexity, they will also improve the classification results. On the GEE platform, we set the number of classification trees as 20, the bag fraction as 0.8, and the other parameters use the default settings.

3.2.2 Gradient boosting tree

The gradient boosting tree method is an iterative decision approach that consists of multiple decision trees. The conclusions of all trees are summed up to make final results, that is, the value of the current model is used as an approximate value of the residual of the lifting tree model by using a negative gradient of a loss function. It has the flexibility to handle various types of data, including continuous and discrete values. Compared with the SVM method, the prediction accuracy can also be relatively high in the case of a relatively short parameter adjustment time. When loss functions, such as the Huber and Quantile functions, are added to the algorithm, the robustness to outliers reaches a maximum. To obtain the best classification results, we set the number of classification trees as 20.

3.2.3 Support vector machine

The support vector machine method is a binary classification model which maps points in low-dimensional to higher-dimensional space, makes them linearly separable, and then use linear partitioning to determine the classification boundary. Its purpose is to find a hyperplane to segment the samples. This segmentation aims to maximize the interval and turn it into a convex quadratic problem. When the training samples are not linearly separable, a nonlinear support vector machine is used which maximizes the kernel function and soft interval. Linear classifiers use hyperplane type boundaries, and nonlinear classifiers use hypersurfaces. The support vector machine has more advantages than neural networks in the face of no local minimum problems. Moreover, SVM can deal with high-dimensional data sets well and has a strong generalization ability.

3.2.4 Classification & regression tree

The classification and regression tree method is a classical decision method dealing with classification or regression tasks involving continuous data. It consists of feature selection, tree generation, and pruning with a binary segmentation technology. Meanwhile, it divides the current sample set into two sub-sample groups and makes each non-leaf node generated with only two branches. The classification and regression tree categorizes each characteristic in a binary manner (including label characteristics and the continuity characteristics). If the label features from three properties can be one of the two attributes, the other property falls into the same category. Since the classification and regression tree generation algorithms allow the generated decision trees to grow naturally as large as possible, it is easy to cause overfitting. Therefore, the pruning strategy can be adopted to avoid overfitting.

3.3 Landscape pattern index

The habitat type landscape pattern index is a simple quantitative index that summarizes habitat pattern information and reflects aspects of its structural composition and spatial configuration. Landscape fragmentation and landscape interference indices may reflect the habitat pattern evolution in the study area. The landscape division and patch density indices are utilized in this study to reveal the spatial structure of habitat types, and these indexes can indicate landscape fragmentation and interference. The patch density (PD) is one of the indicators unveiling habitat fragmentation, and the patch number of a specific type accounts for the proportion of the whole landscape area. It shows that the greater the patch density, the higher the degree of fragmentation of the landscape. Landscape division degree refers to the degree of separation of different patches in a landscape type. The aggregation index (AI) reflects the aggregation and dispersion of patches in the landscape and examines

the connectivity between patches of each landscape type. The smaller the value, the more discrete the landscape. The landscape interference index represents the ability of landscape to resist external disturbance and self-recovery. The greater the intensity of human and natural interference, the stronger the sensitivity of the whole land landscape ecosystem and the greater the risk of landscape ecology.

We select the landscape patch density index (P_i), landscape division index (D_i), and landscape splitting index (F_i) to construct the landscape interference index (G_i) for Poyang Lake habitats (Zang et al., 2005), given by:

$$G_i = W_1 \times P_i + W_2 \times D_i + W_3 \times F_i, \quad (5)$$

where W_1, W_2, and W_3 are the weighting coefficients of PD, SPLIT, and DIVISION, respectively. Following previous approaches (Zang et al., 2005; Jing et al., 2008; Li and Li, 2008), the weighting coefficients of PD, SPLIT, and DIVISION are assigned as 0.5, 0.3, and 0.2, respectively.

The raster attribute table is constructed for the habitat type classification result map and imported into the Fragstats software (V4.2.1). Then, a rectangular window of 200 m × 200 m is established using a moving window module. The landscape index of the window is calculated in the window center. The landscape index of the raster is calculated by moving the window from the upper left of the image raster to the lower right, covering the whole image. The obtained landscape indexes are normalized as dimensionless values in ArcGIS.

3.4 Image preprocessing

The Poyang Lake winter habitat area is classified into six types: water body, grassland, mudflat, sandy land, farmland, and other. The NDVI, MNDWI, NDBI, and EVI indices are used to extract and distinguish the habitat characteristics. MNDWI is introduced to distinguish water surface from other habitat types. NDVI and EVI are used to separate the areas of mudflat and grassland where image pixels are easily mixed. Simultaneously, NDBI is employed to extract construction area. Then, the four machine learning algorithms describe above are used for further habitat classifications.

4 Results

4.1 Comparison of the machine learning methods

The resultant habitat classification accuracies and Kappa coefficients are given in Table 2. All classification accuracies of each method are above 98%, and the Kappa coefficients exceed 0.98. the random forest (RF) algorithm had an overall accuracy of 99.6% and a Kappa coefficient of 0.994, which was the best classification performance (Table 2).

Table 2 Overall accuracies and Kappa coefficients of the four machine learning algorithms for the Landsat image of 12 January 2021

Year	RF	GTB	SVM	CART
Overall accuracy	0.9960	0.9952	0.9930	0.9934
Kappa coefficient	0.9941	0.9930	0.9896	0.9902

Table 3 Confusion matrices of four machine learning algorithms for the Landsat image of 12 January 2021

Algorithms	Land type	Water	Grassland	Mudflat	Sand	Farmland	Other
RF	Water	8636	0	0	0	0	2
	Grassland	0	3448	10	0	1	2
	Mudflat	0	0	6635	1	2	3
	Sand	0	0	0	1295	0	2
	Farmland	0	2	8	7	50	1
	Other	4	5	8	0	3	176
GBT	Water	8635	0	0	0	0	3
	Grassland	0	3443	16	0	1	1
	Mudflat	2	1	6632	0	0	7
	Sand	0	0	27	1270	0	0
	Farmland	0	2	8	6	51	1
	Other	5	4	11	0	3	173
SVM	Water	8635	0	1	0	0	2
	Grassland	0	3449	9	0	1	2
	Mudflat	0	0	6634	5	2	0
	Sand	0	0	60	1236	1	0
	Farmland	0	0	12	0	56	0
	Other	20	7	18	0	3	148
CART	Water	8636	0	0	0	0	2
	Grassland	0	3445	11	0	1	4
	Mudflat	0	3	6598	17	11	12
	Sand	0	0	30	1267	0	0
	Farmland	0	2	11	6	49	0
	Other	4	4	11	0	6	171

The confusion matrices for the four machine learning algorithms are listed in Table 3, showing that the RF method has small classification errors, and the pixels of water surface and sand are less mixed than other types. For the RF method, the misclassification errors of grassland and mudflat are larger than those of water surface and sand. The reason for this is that, generally, mudflats and grassland are spatially connected. Although there are some classification errors for grassland, farmland, and mudflats, RF performs better than the other machine learning methods in the classification of habitat types in the study areas.

4.2 Long-term variability of habitat types in Poyang Lake

We use the RF method to classify the winter habitat area of Poyang Lake for the Landsat images from 1990 to 2021 based on our preliminary comparisons in the previous step. The classification results, overall accuracy, and Kappa coefficient of the random forest classification of the Landsat TM and OLI images are shown in Table 4, which indicates that most of the overall accuracies and Kappa coefficients of the RF method are ~99%, although in 2006, the overall accuracy and Kappa coefficient are 98.78% and 0.9828, respectively. These results suggest that the random forest method is satisfactory for analyzing winter habitat characteristics in the Poyang Lake area.

Figure 2 shows the percentage of Poyang Lake winter habitat types from 1990 to 2021, and the areal changes from 1990 to 2021 are illustrated in Fig. 3. The water surface decreased in winter; the mudflat was the primary habitat type in this period. In 1990, however, the proportion of water surface was larger than that of the mudflat area. Sand, farmland, and other areas accounted for a small proportion of the total area below 20% in most years, except that in 1997 when they accounted for as much as 40% of the total area. Among the six habitat types, the water surface area experienced the most significant change, with the largest area in 2002 and the smallest in 1993, though this was not related to the water level difference (discussed below). In most years, the sand area was small; however, it increased rapidly from 1995 to 1999 due to anthropogenic factors. In addition to the water surface, the most significant habitat types in Poyang Lake are mudflats and grasslands.

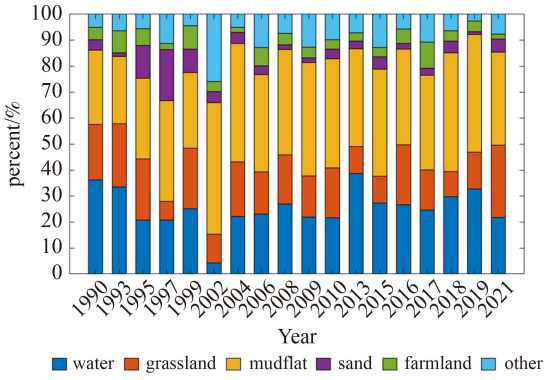

Fig. 2 Percentage of Poyang Lake winter habitat types from 1990 to 2021.

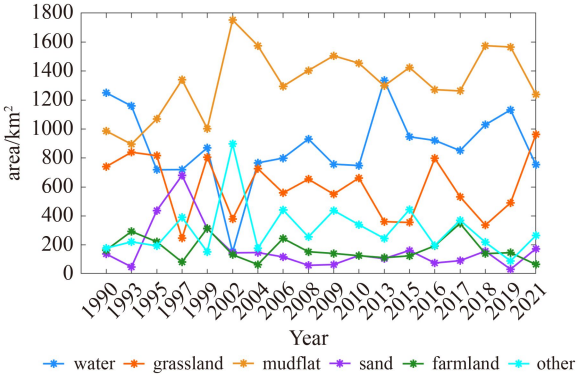

Fig. 3 Changes in Poyang Lake winter habitat area from 1990 to 2021.

Table 4 Overall accuracy and Kappa coefficient for random forest classification of Landsat images from 1990 to 2021

Date	Overall accuracy	Kappa coefficient
23 Jan 1990	0.9958	0.9938
17 Dec 1993	0.9965	0.9951
17 Dec 1995	0.9954	0.9935
10 Jan 1997	0.9974	0.9960
18 Dec 1999	0.9950	0.9932
08 Jan 2002	0.9940	0.9915
15 Dec 2004	0.9949	0.9928
21 Dec 2006	0.9878	0.9828
10 Dec 2008	0.9970	0.9954
11 Jan 2009	0.9979	0.9967
14 Jan 2010	0.9956	0.9936
24 Dec 2013	0.9972	0.9947
13 Dec 2015	0.9962	0.9935
16 Dec 2016	0.9974	0.9957
19 Dec 2017	0.9957	0.9929
05 Feb 2018	0.9896	0.9829
23 Jan 2019	0.9958	0.9923
12 Jan 2021	0.9960	0.9941

Figure 4 illustrates classification thematic maps from 1990 to 2021, which reveal spatial variations of the classified winter habitat types of Poyang Lake. The grasslands generally appeared around mudflats; however, in 2009 and 2015, the mudflats and grasslands were scattered, and the grasslands and mudflats alternatively appeared in the thematic maps (Fig. 4). The sand area was primarily observed near the main channel of Poyang Lake (especially in 1997) due to sand mining activities. The farmland around Poyang Lake was mainly located in the south-east of the study area, with a decline in area from 1998 to 2004. This may be partly due to a local policy implementation of "farmland returning to lake".

The Three Gorges Dam (TGD) in the upstream Yangtze River can modify hydrological conditions in Poyang Lake, and also can affect Poyang Lake habitats. We compared the Poyang Lake habitat types for the pre-dam and the post-dam periods. Here, the pre-dam period is selected as 1990−2002, and the post-dam period is 2004−2021. Tables 5 and 6 list the landscape transfer matrix of the Poyang Lake winter habitats for 1999−2020

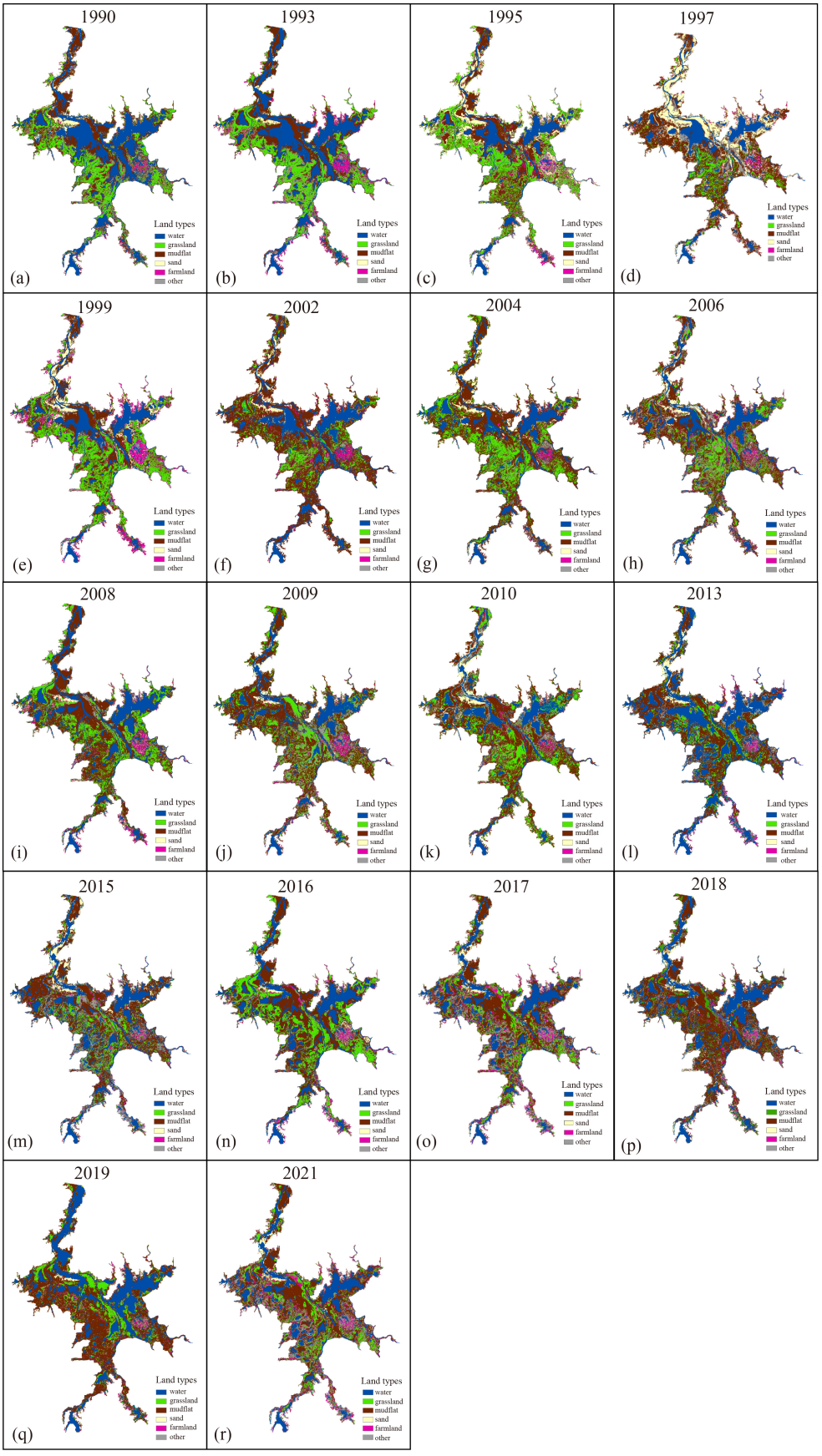

Fig. 4 Poyang Lake habit types obtained from Landsat images with the random forest method.

Table 5 Land type transfer matrix from 1990 to 2002

Area conversion/km^2	Water	Grassland	Mudflat	Sand	Farmland	Other	Total (1990)
Water	783.14	65.69	317.09	21.09	12.40	50.72	1250.13
Grassland	4.99	220.30	453.53	2.83	32.47	26.01	740.13
Mudflat	69.04	54.01	708.30	56.44	59.90	38.48	986.17
Sand	17.04	2.24	59.65	49.39	5.75	3.37	137.44
Farmland	2.75	20.44	105.65	1.11	28.60	5.67	164.22
Other	19.88	16.11	107.55	1.55	11.47	21.05	177.66
Total (2002)	896.84	378.84	1751.77	132.41	150.59	145.30	3455.75

Table 6 Land type transfer matrix from 2004 to 2021

Area conversion/km^2	Water	Grassland	Mudflat	Sand	Farmland	Other	Total (2004)
Water	460.00	23.27	201.65	15.94	1.41	64.67	766.94
Grassland	96.43	402.40	163.93	14.82	4.62	44.48	726.68
Mudflat	161.92	456.77	702.85	90.43	36.82	123.41	1572.20
Sand	16.48	17.44	66.62	37.32	2.86	7.18	147.90
Farmland	1.12	11.29	30.13	4.63	12.86	5.41	65.44
Other	17.00	50.24	72.48	10.58	6.74	19.56	176.59
Total (2021)	752.95	961.41	1237.65	173.72	65.31	264.71	3455.75

and 2004−2021, respectively, calculated based on Markov transfer matrices that can quantify both the conversion of different land types and the conversion rate (Feng et al., 2010).

Table 5 indicates that in the pre-dam period from 1990 to 2002, mudflats increased substantially. A large amount of the water area (as high as ~317.09 km^2) was converted to mudflats; ~453.53 km^2 grassland areas was also converted to mudflats. The sand and farmland areas converted into mudflats were also relatively large. The mudflat area reached its maximum of 1751.77 km^2 in 2002. In the post-dam period from 2004 to 2021 (Table 6), more mudflat was converted into grassland; in 2004, 456.77 km^2 of mudflat was transformed into grassland, accounting for 29.05% of the total mudflat area. Overall, the mudflat area decreased in the post-dam period as compared with that in the pre-dam period. The construction of the TGD has led to earlier winter dry seasons in Poyang Lake, with the result that the wet mudflat dries up and is converted into grassland. This situation may bring negative effects to the habitats which provide aquatic foods for wintering birds in the Poyang Lake area.

4.3 Evolution of habitat landscape patterns

Figure 6 shows the AI (Fig. 5(a)), PD (Fig. 5(b)), and landscape interference indices (Fig. 5(c)) for 1990 to 2021. From 1990 to 1997, landscape patch density fluctuated year by year, indicating that the distribution of all the winter habitat types in the Poyang Lake was highly scattered. A decreased landscape aggregation index leads to excessive fragmentation and a decreased connectivity between habitat patches. The landscape interference index increased, reflecting the fact that the habitat areas were continuously disturbed. From 1997 to 2021, both the landscape patch density and landscape interference indices decreased, and the landscape aggregation index increased. Thus, the connectivity of land types has been enhanced, leading to congregation of each habitat type.

These habitat landscape indexes suggest that there was a significant shift in 1997−1998. Prior to and later than 1997−1998, the landscape of the habitats in Poyang Lake exhibited reversed evolution characteristics. Figure 3 shows that after 1997, the sand and farmland areas in the Poyang Lake area reduced. Sand and farmland areas are not the major components of the Poyang Lake habitat system, but the correlation of their areal changes with the landscape indices should be considered. A decrease in farmland area may be related to the implementation of a local policy known as “farmland returning to lake” which began in 1997−1998. This policy helped to reduce the interference of human activities on Poyang Lake. In 1998, the Yangtze River Basin, including the Poyang Lake area, experienced a major flood event that could be a possible reason for the landscape index change observed in our study area.

Figure 6 shows the spatial patterns of the landscape patch density, landscape aggregation index, and landscape interference index for 1997, 1999, and 2021. For each year considered, the AI (Figs. 6(a), 6(b), and 6(c)),

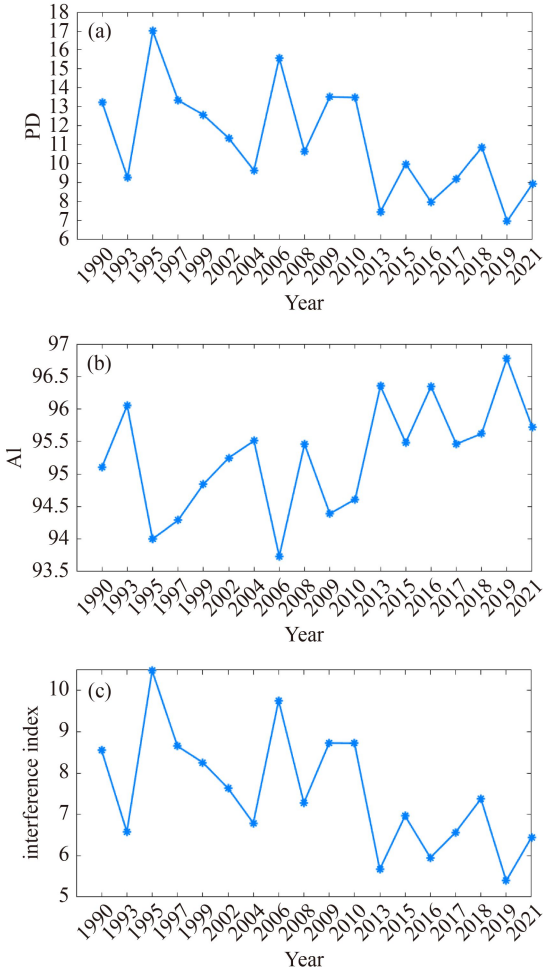

Fig. 5 The PD (a), AI (b), and landscape interference index (c) from 1990 to 2021, respectively.

PD (Figs. 6(d), 6(e), and 6(f)), and landscape interference indices (Figs. 6(g), 6(h), and 6(i)) had similar patterns. In the south-east Poyang Lake area near Lianhu, a low landscape aggregation index, high landscape patch density, and high landscape interference index were observed, suggesting that in these areas, the winter habitats are more vulnerable to external forcing, such as human activities, than in other areas.

By comparing Figs. 6(a, d, g), Figs. 6(b, e, h), and Figs. 6(c, f, i), we find that the areas with high degree of landscape aggregation index increased, and the areas with high degrees of landscape patch density and interference index decreased from 1997 to 2021. The evolutions of these landscape parameters reflect an improvement in the habitat connectivity, suggesting that the integrity of the habitats has been restored since 1997−1998.

The period of 1990−1998 was generally typified by fragmentation and interference of the Poyang Lake land types. From 1998 to the present, the landscape fragmentation and interference declined, which may benefit habitat restoration for certain kinds of wildlife.

4.4 Effects of water levels

All the satellite images used were taken during the low-water winter dry season. However, there was also a year-to-year water level variability, which makes it necessary to examine whether the difference of the water level in different years can affect the habitat classification results. Figure 7(a) shows water levels at Xingzi, Duchang, Tangyin, and Kangshan stations in Poyang Lake, the water area, and the mean water level of these stations on the satellite image acquisition dates in the years from 1990 to 2021. The water levels at Duchang and Xingzi experienced large variations with a maximum level difference of ~4.0 m and ~3.72 m, respectively. The water level variation ranges were small at Kangshan and Tangyin stations. Xingzi and Duchang are both located in a narrow channel of the lake, and therefore, the water levels may change relatively significantly compared with those at Kangshan and Tangyin. Figure 7(a) shows that the water levels at Kangshan and Tangyin were stable for the satellite acquisition dates from 1990 to 2021. In addition, Figure 7(a) shows that the smallest water was observed in 2002, while the lowest water level occurred in 2015 at Xingzi and Duchang. This suggests that the long-term variation of the water area was not consistent with that of water level, and further implying that the water area variations from the classification results in this study were not likely caused by water level changes. Figure 7(b) gives a scatter plot of the water area versus the mean water level, showing that for most cases, the water area was unrelated to water level as the mean water level was lower than 11.4 m. Therefore, although water levels were different for the different satellite acquisition time periods, these differences in the water level do not seem to influence the long-term variations of our habitat classification results.

The grassland, mudflat, sand, and farmland habitat areas regressed versus the water level are displayed in Fig. 8. The correlations of grassland, mudflat, sand, and farmland areas with water level are low and likely insignificant, and the R^2 between the areas of the grassland, mudflat, sand, and farmland and the water level are 0.06, 0.03, 0.01, and 0.07, respectively. This suggests that water level is not a major factor influencing the variations of the different habitats in Poyang Lake, at least for our study cases.

5 Discussion

The water area of Poyang Lake has significant seasonal and interannual variations. We next consider the maximum water area during the study period to better reveal winter habitat changes free of the water surface variation. This maximum water surface is obtained by integrating the total water surface via counting pixels of water in each of the satellite images. Next, the maximum

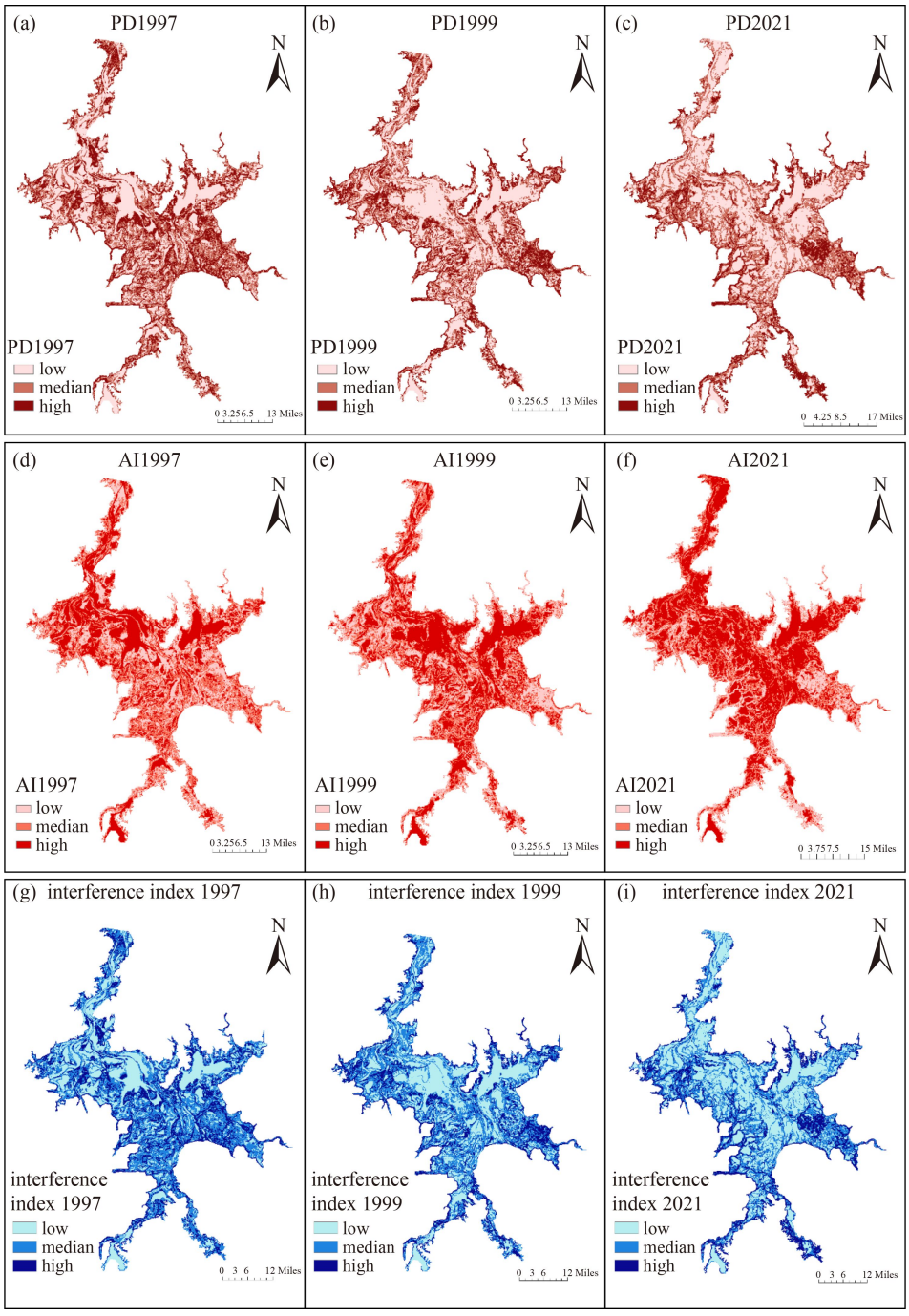

Fig. 6 PD, AI and landscape interference indices for different Poyang Lake land types.

water surface is excluded from the satellite images, so that the rest of the area in the satellite images does not contain any water surface. Figure 9 shows areas of the different habitat types (excluding maximum water surface) from 1990 to 2021 and indicates that mudflats and grasslands were the main habitat types, consistent with the results of Fig. 3. The evolution of grassland, sand, and farmland areas were almost same as those for the water surface-included results (shown in Fig. 3), and the general trend of the mudflat evolution was consistent with that in Fig. 3, except for 2010. This may be because the mudflat and water body are frequently connected and excluding the water area may affect mudflat area. However, the general evolution trend is same for both water surface-included and -excluded cases.

6 Conclusions

We here applied machine learning methods of random forest (RF), support vector machine (SVM), gradient boosting tree (GBT), and classification and regression

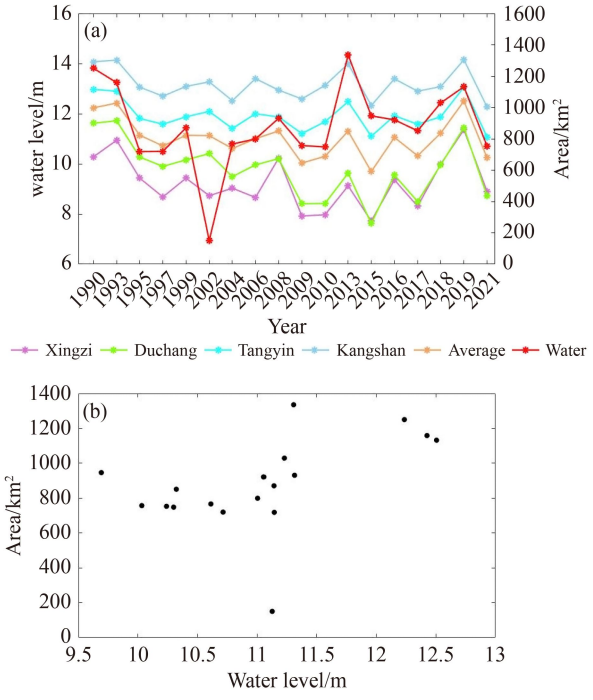

Fig. 7 Water levels at Xingzi, Duchang, Tangyin, and Kangshan stations in Poyang Lake, the water area, and the mean water level of these stations (a) and the water area versus water levels (b) on the satellite image acquisition dates for 1990 to 2021.

tree (CART) to classify and analyze winter habitat types and landscape characteristics in the Poyang Lake area using Landsat satellite imagery. Preliminary test results showed that the RF method gave the best results in overall accuracy and Kappa coefficient. Therefore, the RF method is applied to cloud-free Landsat satellite images over Poyang Lake from 1990 to 2021, showing that for most images, the overall accuracy and Kappa coefficient are higher than 0.99, suggesting that the RF machine learning method can well capture habitat types and landscape characteristics with high accuracy.

The classification results indicate that mudflats, water surface, and grasslands were the three leading winter habitat types of Poyang Lake. Generally, the mudflat area was larger than the other types. The sand and farmland areas were only a small portion of the total area (less than 20%) in most years. Further analysis reveals that after 2012, the mudflat area decreased, being converted into grassland, which could be caused by an early onset dry winter season and is likely related to the construction of the Three Gorge Dam. The derived landscape indices reveal that the habitat types were significantly scattered and fragmented prior to 1997−1998. After 1997−1998, however, the landscape fragmentation and interference indices decreased, and decreased human interference and other natural factors are likely responsible for the observed habitat landscape evolution.

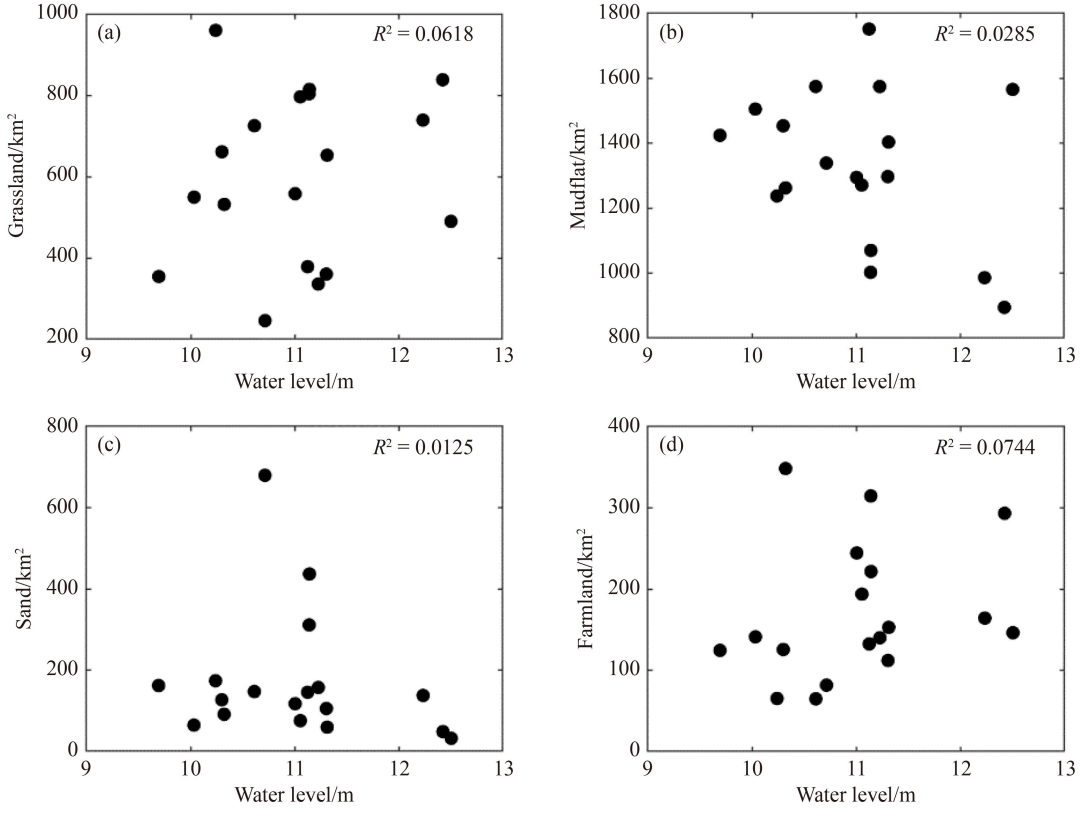

Fig. 8 The habitat areas of grassland (a), mudflat (b), sand (c), and farmland (d) versus the water level on the satellite image acquisition dates in the years from 1990 to 2021.

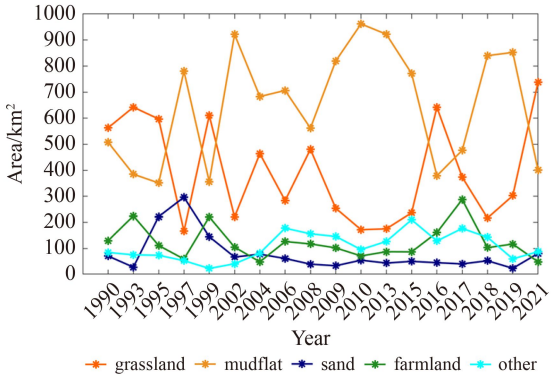

Fig. 9 Winter habitat areas excluding the maximum water surface.

Acknowledgments This study is supported by the National Basic Research Program of China (No. 2021YFB3900400)

References

Chen B, Chen L, Huang B, Michishita R, Xu B (2018). Dynamic monitoring of the Poyang Lake wetland by integrating Landsat and MODIS observations. ISPRS J Photogramm Remote Sens, 139: 75–87

Feng L, Hu C, Chen X, Cai X, Tian L, Gan W (2012). Assessment of inundation changes of Poyang Lake using MODIS observations between 2000 and 2010. Remote Sens Environ, 121: 80–92

Feng Y, Luo G, Zhou D, Han Q, Lu L, Xu W, Zhu L, Yin C, Dai L, Li Y (2010). Effects of land use change on landscape pattern of a typical arid watershed in the recent 50 years: a case study on Manas River Watershed in Xinjiang. Acta Ecol Sin, 30(16): 4295–4350

Jia K, Liang S, Wei X, Yao Y, Su Y, Jiang B, Wang X (2014). Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. Remote Sens (Basel), 6(11): 11518–11532

Jing Y, Zhang S, Li Y (2008). Ecological risk analysis of rural-urban ecotone based on landscape structure. Chinese J Eco, 27(2): 229–234 (in Chinese)

Kanai Y, Ueta M, Germogenov N, Nagendran M, Mita N, Higuchi H (2002). Migration routes and important resting areas of Siberian cranes (*Grus leucogeranus*) between northeastern Siberia and China as revealed by satellite tracking. Biol Conserv, 106(3): 339–346

Kaufman Y J, Tanre D (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. IEEE Trans Geosci Remote Sens, 30(2): 261–270

Khadka S, Gyawali B R, Shrestha T B, Cristan R, Banerjee S B, Antonious G, Poudel H P (2021). Exploring relationships among landownership, landscape diversity, and ecological productivity in Kentucky. Land Use Policy, 111: 105723

Li H, Fang C, Xia Y, Liu Z, Wang W (2022). Multi-scenario simulation of production-living-ecological space in the Poyang Lake area based on remote sensing and RF-Markov-FLUS model. Remote Sens (Basel), 14(12): 2830

Li J, Zhou K, Dong H, Xie B (2020a). Cultivated land change, driving forces and its impact on landscape pattern changes in the Dongting Lake Basin. Int J Environ Res Public Health, 17(21): 7988

Li X, Li J (2018). Analysis on region landscape ecological risk based on GIS - a case study along the lower reaches of the Weihe River. Arid Zone Res, 25(6): 899–903 (in Chinese)

Li J, Zhou K, Dong H, Xie B (2020b). Cultivated land change, driving forces and its impact on landscape pattern changes in the Dongting Lake Basin. Int J Environ Res Public Health, 17(21): 7988

Li P, Zuo D, Xu Z, Zhang C, Han Y, Sun W, Pang B, Ban C, Kan G, Yang H (2021). Dynamic changes of land use/cover and landscape pattern in a typical alpine river basin of the Qinghai-Tibet Plateau, China. Land Degrad Dev, 32(15): 4327–4339

Li Y, Sun X, Zhu X, Cao H (2010). An early warning method of landscape ecological security in rapid urbanizing coastal areas and its application in Xiamen, China. Ecol Modell, 221(19): 2251–2260

Liu H, Huete A (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. IEEE Trans Geosci Remote Sens, 33(2): 457–465

Liu H, Yuan H, Wang S, Zheng L, Liao M (2021). Spatiotemporal dynamics of water body changes and their influencing factors in the seasonal lakes of the Poyang Lake region. Water, 13(11): 1539

Mu S, Li B, Yao J, Yang G, Wan R, Xu X (2020). Monitoring the spatio-temporal dynamics of the wetland vegetation in Poyang Lake by Landsat and MODIS observations. Sci Total Environ, 725: 138096

Naveh Z (1994). From biodiversity to ecodiversity: a landscape-ecology approach to conservation and restoration. Restor Ecol, 2(3): 180–189

Pal M (2008). Ensemble of support vector machines for land cover classification. Int J Remote Sens, 29(10): 3043–3049

Shao Y, Lunetta R S (2012). Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS J Photogramm Remote Sens, 70: 78–87

Tang X, Li H, Xu X, Yang G, Liu G, Li X, Chen D (2016). Changing land use and its impact on the habitat suitability for wintering Anseriformes in China's Poyang Lake region. Sci Total Environ, 557-558: 296–306

Wang J, Ogawa S (2017). Analysis of dynamic changes in land use based on landscape metrics in Nagasaki, Japan. J Appl Remote Sens, 11(1): 016022

Xia S, Liu Y, Wang Y, Chen B, Jia Y, Liu G, Yu X, Wen L (2016). Wintering waterbirds in a large river floodplain: hydrological connectivity is the key for reconciling development and conservation. Sci Total Environ, 573: 645–660

Xu H (2006). Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. Int J Remote Sens, 27(14): 3025–3033

Yang W, You Q, Fang N, Xu L, Zhou Y, Wu N, Ni C, Liu Y, Liu G, Yang T, Wang Y (2018). Assessment of wetland health status of Poyang Lake using vegetation-based indices of biotic integrity. Ecol Indic, 90: 79–89

Yao J, Zhang Q, Ye X, Zhang D, Bai P (2018). Quantifying the impact of bathymetric changes on the hydrological regimes in a large floodplain lake: Poyang Lake. J Hydrol (Amst), 561: 711–723

Zang S, Liang X, Zhang S (2005). GIS- based analysis of ecological

risk on land use in Daqing City. J Nat Disast, 14(4): 141–145 (in Chinese)

Zha Y, Ni S, Yang S (2003). An effect approach to automatically extract urban land-use from TM imagery. J Remote Sens, 7(1): 37–40+82 (in Chinese)

Zhang Q, Fu B, Chen L (2003). Several problems about landscape pattern change research. Sci Geograph Sin, 3: 264-270 (in Chinese)

Zhang Q, Li L, Wang Y, Werner A D, Xin P, Jiang T, Barry D A (2012). Has the Three-Gorges Dam made the Poyang Lake wetlands wetter and drier? Geophys Res Lett, 39(20): 2012GL053431

Zhang T, Gao Y, Li C, Xie Z, Chang Y, Zhang B (2020). How human activity has changed the regional habitat quality in an eco-economic zone: evidence from Poyang Lake eco-economic zone, China. Int J Environ Res Public Health, 17(17): 6253

Zhao S, Fang J, Ji W, Tang Z (2003). Lake restoration from impoldering: impact of land conversion on riparian landscape in Honghu Lake area, Central Yangtze. Agric Ecosyst Environ, 95(1): 111–118

Zhou H, Luo Z, Tangdamrongsub N, Zhou Z, He L, Xu C, Li Q, Wu Y (2018). Identifying flood events over the Poyang Lake Basin using multiple satellite remote sensing observations, hydrological models and in situ data. Remote Sens (Basel), 10(5): 713

Zhou Y, Ma J, Zhang Y, Li J, Feng L, Zhang Y, Shi K, Brookes J D, Jeppesen E (2019). Influence of the Three Gorges Reservoir on the shrinkage of China's two largest freshwater lakes. Global Planet Change, 177: 45–55