Land Use Simulation of Shanghai Based on Multi-source Data Integration

Yundan Bai[1,2,3,a], Fayun Li[1,2,*] and Weiyu Yu[1,2]

[1]School of Ecological Technology and Engineering, Shanghai Institute of Technology, Shanghai, 201418, China
[2]Department Institute of Beautiful China and Ecological Civilization, University Think Tank of Shanghai Municipality, Shanghai, 201418, China
[3]Institute of Oriental Language and Culture, Shanghai Vocational College of Business and Foreign Languages, Shanghai, 201399, China

[a]email: baiyd@sicfl.edu.cn
*Corresponding author's e-mail: lifayun@sit.edu.cn

Abstract. With the continuous improvement of big data mining technology, massive and multi-source land use spatio-temporal big data began to emerge. Big data thinking plays an increasingly important role in ecological monitoring, smart city construction, public safety and support for major decisions. Taking Shanghai as an example, this paper integrates ESA CCI, GlobeLand 30 and MODIS data by voting method, and inputs MCE-CA-Markov model to obtain more accurate prediction results. This study provides a new simulation thinking, which can provide ideas for the prediction of Shanghai and other similar cities.

Keywords: multi-source data; land use simulation; integration.

1 Introduction

With the continuous progress of satellite technology and the increasing demand for production, high-resolution land use data emerge in endlessly. Massive real-time land use data are widely used in urban land use/cover classification[1,2], forest vegetation monitoring[3,4], urban impervious surface detection[5] and urban pollution detection[6]. Land use data is the basis of land use simulation, which can provide necessary information support for land use simulation. Land use classification has always been a hot topic in the field of remote sensing. Land use data classification is closely related to many factors such as sensors, image preprocessing and classification algorithms.

The classifiers of different algorithms have different sensitivities to different ground objects, and the probability of different classifiers correctly identifying the same ground object is also different, which has certain subjectivity. There are complementary advantages between different types of classifiers. Because each classifier has its own limitations, Gincinto and Roli compared the application of different classifiers in remote sensing data classification and found that ' no single classifier is omnipotent '[7]. Therefore, in addition to developing advanced classifiers with machine

P. Xiang et al. (eds.), *Proceedings of the 2023 5th International Conference on Hydraulic, Civil and Construction Engineering (HCCE 2023)*, Atlantis Highlights in Engineering 26,
https://doi.org/10.2991/978-94-6463-398-6_65

learning, artificial intelligence technology, or algorithms that improve existing classifiers, integrating existing classifiers is also an effective method.

The voting method is also an application of the integration idea. It follows the principle of ' minority obeys majority '. The classification combination rule holds that the group 's decision-making judgment is better than the individual 's decision-making judgment, and classifies the categories with the same classification type of the most classifiers as the categories of the image pixels to be classified[8]. The voting method can integrate the complementary information between classifiers and reduce the error of a single classifier by linearly integrating the classification results of multiple classifiers[9]. Therefore, the results selected by voting are often more accurate than the prediction results of a single classifier.

2 Methods

2.1 Voting integration strategy

The most commonly used combination strategy in ensemble learning is the voting method. For the classification problem, each weak classifier gives its own prediction results, and then the final result of the combination is obtained by the voting method[10]. In this paper, the absolute majority voting method is used to fuse three land use data sets based on different algorithms. If the classification results of two or more base classifiers are the same, then the classification result is considered to be the final category[11]. If the classification results of the three data sets are not the same, it is considered that there is no classification result and it is divided into other land use. The voting strategy is shown in Figure 1:

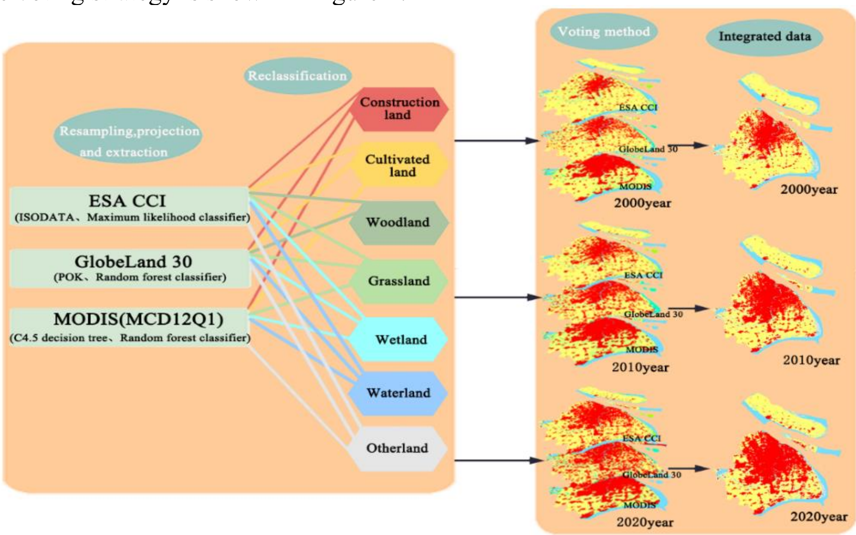


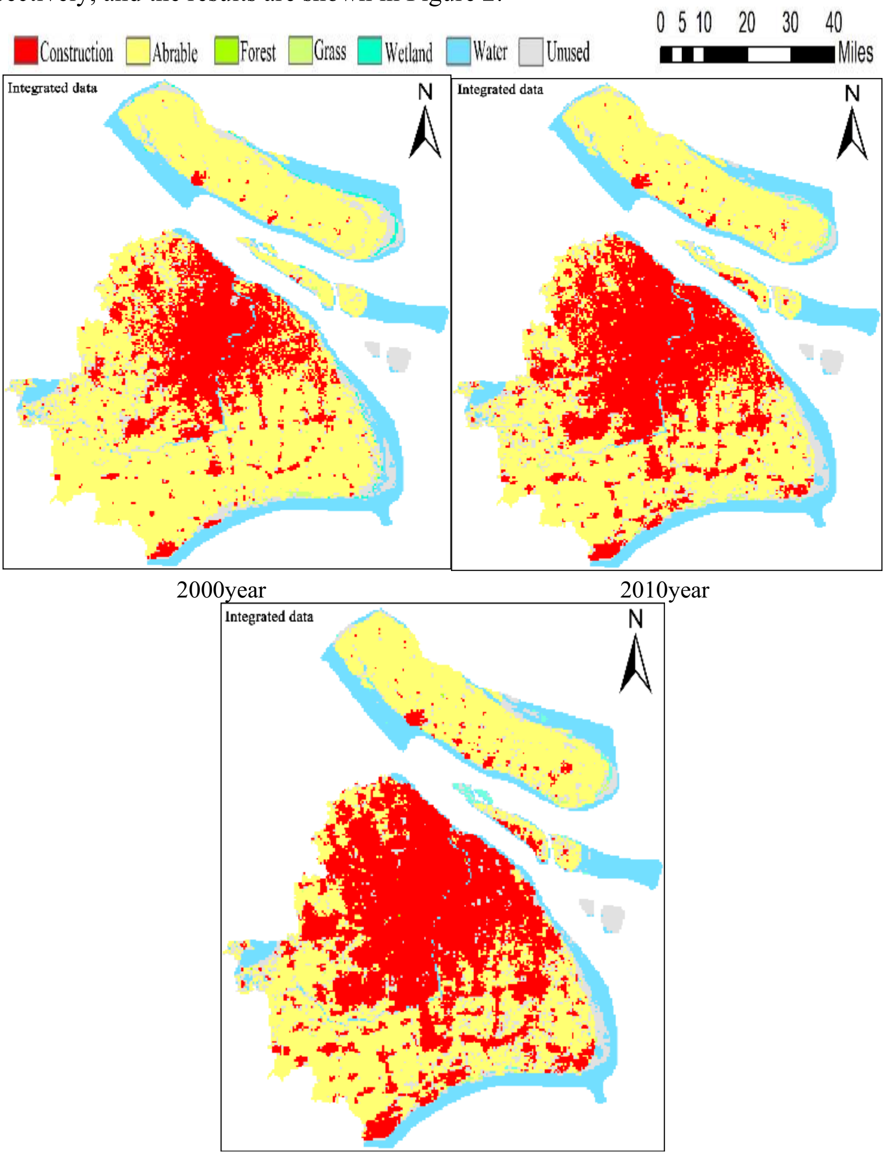
Fig. 1. Voting strategy flow chart

In RStudio, based on the voting strategy, the ifelse function is used to fuse the land use data of the three datasets into new data. The specific code is as follows:

Algorithm A:raster data 1 B:raster data 2 C:raster data 3 JC:Integrated data

```
A<-raster("A.tif")#import raster data1
A<-as.data.frame(A, xy=T) #convert the raster data1 into a list
B<-raster("B.tif")#import raster data2
B <-as.data.frame(B, xy=T) #convert the raster data2 into a list
C<-raster("C.tif")#import raster data3
C <-as.data.frame(C, xy=T) #convert the raster data into a list
JC <-cbind(A,B[,3]) #combines the third column of raster data1 and data2
JC <-cbind(JC,C[,3]) #merge JC and the third column of data3
Colnames(JC) [3:5] <-c ("A","B","C") #JC's column 3-5 was renamed A,B,and C.
JC$JC<-
        ifelse((JC$A==1 & JC$B==1 & JC$C==1)#if A,B,C three columns are type1
          |(JC$A==1 & JC$B==1 & JC$C!=1)#or if both columns A and B are type1
|(JC$A!=1 & JC$B==1 & JC$C==1)#or if both columns B and C are type1
|(JC$A==1 & JC$B!=1 & JC$C==1)#or if both columns A and C are type1
,1,#It is considered that the sample is type 1, otherwise it is classified as else for re-voting.
        (ifelse((JC$A==2 & JC$B==2 & JC$C==2)#if A,B,C three columns are type2
|(JC$A==2 & JC$B==2 & JC$C!=2)#or if both columns A and B are type2
|(JC$A!=2 & JC$B==2 & JC$C==2)#or if both columns B and C are type2
|(JC$A==2 & JC$B!=2 & JC$C==2)#or if both columns A and C are type2
,2,#It is considered that the sample is type 2, otherwise it is classified as else for re-voting.
(ifelse((JC$A==3 & JC$B==3 & JC$C==3)#if A,B,C three columns are type3
|(JC$A==3 & JC$B==3 & JC$C!=3)#or if both columns A and B are type3
|(JC$A!=3 & JC$B==3 & JC$C==3)#or if both columns B and C are type3
          |(JC$A==3 & JC$B!=3 & JC$C==3)#or if both columns A and C are type3
,3,#It is considered that the sample is type 3, otherwise it is classified as else for re-voting.
(ifelse((JC$A==4 & JC$B==4 & JC$C==4)#if A,B,C three columns are type4
|(JC$A==4 & JC$B==4 & JC$C!=4)#or if both columns A and B are type4
|(JC$A!=4 & JC$B==4 & JC$C ==4)#or if both columns B and C are type4
          |(JC$A==4 & JC$B!=4 & JC$C ==4)#or if both columns A and C are type4
,4,#It is considered that the sample is type 4, otherwise it is classified as else for re-voting.
(ifelse((JC$A==5 & JC$B==5 & JC$C==5)#if A,B,C three columns are type5
|(JC$A==5 & JC$B==5 & JC$C!=5)#or if both columns A and B are type5
|(JC$A!=5 & JC$B==5 & JC$C==5)#or if both columns B and C are type5
          |(JC$A==5 & JC$B!=5 & JC$C==5)#or if both columns A and C are type5
,5,#It is considered that the sample is type 5, otherwise it is classified as else for re-voting.
(ifelse((JC$A==6 & JC$B==6 & JC$C==6)#if A,B,C three columns are type6
|(JC$A==6 & JC$B==6 & JC$C!=6)#or if both columns A and B are type6
|(JC$A!=6 & JC$B==6 & JC$C==6)#or if both columns B and C are type6
          |(JC$A==6 & JC$B!=6 & JC$C==6)#or if both columns A and C are type6
,6,#It is considered that the sample is type 6, otherwise it is classified as else for re-voting.
(ifelse((JC$A==7 & JC$B==7 & JC$C==7)#if A,B,C three columns are type7
|(JC$A==7 & JC$B==7 & JC$C!=7)#or if both columns A and B are type7
|(JC$A!=7 & JC$B==7 & JC$C==7)#or if both columns B and C are type7
          |(JC$A==7 & JC$B!=7 & JC$C==7)#or if both columns A and C are type7
,7,#It is considered that the sample is type 7, otherwise it is classified as else for re-voting.
(ifelse((JC$A==0 & JC$B==0 & JC$C==0)#if A,B,C three columns are type0
|(JC$A==0 & JC$B==0 & JC$C!=0)#or if both columns A and B are type0
|(JC$A!=0 & JC$B==0 & JC$C==0)#or if both columns B and C are type0
          |(JC$A==0 & JC$B!=0 & JC$C==0)#or if both columns A and C are type0
,0,8))))))))))))))))#It is considered to be type 0, otherwise it is classified as type8.
          JC<-JC[,-c(3:5)]#delete the A、B、C three columns of the integrated data
JC<-rasterFromXYZ(JC)#rasterized integrated data
writeRater(JC,"JC.tif",overwrite=TRUE)#export integrated data
```

Perform the above code to integrate the three data sets in 2000,2010 and 2020 respectively, and the results are shown in Figure 2:



2000year 2010year

2020year

Fig. 2. LUCC integrated data

2.2 Accuracy comparison

In order to pursue the optimal model performance and maximize the integration of information from multiple LUCC data products[12], this study uses the MCE-CA-

Markov model[13] based on the voting method to simulate the future land use of Shanghai in 2030, as shown in Figure 3.

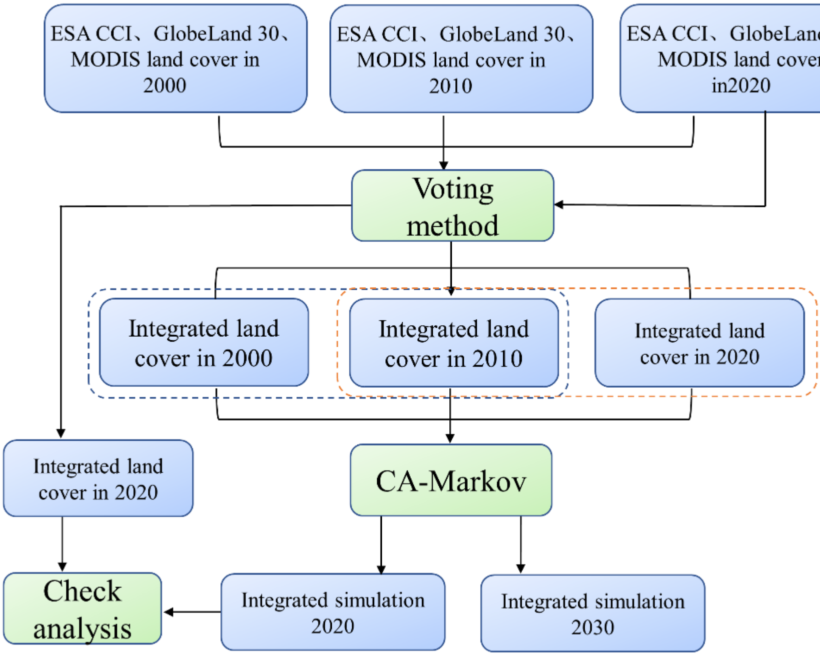


Fig. 3. LUCC integrated data

3 Results

3.1 LUCC simulation of Shanghai in 2020

Based on the three integrated land use data of 2000 and 2010, the land use change in 2020 is simulated. As shown in Figure 4, the simulated 2020 land use data is compared with the 2020 data integrated with the original data of the three data sets, and the Kappa value is 0.96. Comparing the simulation results (Fig.4) with the integrated data, the difference in construction land area is 605.00 km^2, the difference in cultivated land area is 495.00 km^2, the difference in water area is 26.00 km^2, and the difference in other land area is 91.50 km^2. Overall, the increase and decrease of construction land and cultivated land area showed a ' shift ' state.

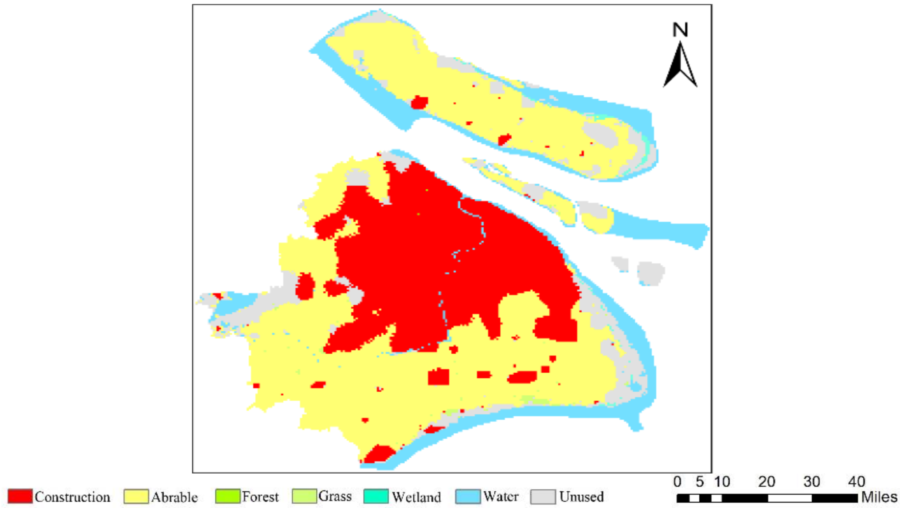


Fig. 4. Simulation results for 2020

3.2 LUCC simulation of Shanghai in 2030

The land use conversion scenario of Shanghai from 2020 to 2030 is set as a continuation of the model from 2010 to 2020, and the LUCC transition probability matrix and suitability atlas from 2010 to 2020 are introduced under different voting strategies. It is predicted that in 2030, the proportion of construction land area is 45.81 %, the proportion of cultivated land area is 28.37 %, the proportion of forest land area is 0.12 %, the proportion of grassland area is 0.32 %, the proportion of wetland area is 1.35 %, the proportion of water area is 13.05 %, and the proportion of other land area is 10.98 %, as shown in Fig.5. Among them, cultivated land is mainly converted into construction land, wetland and other land, and the fastest change in area is forest land and wetland.

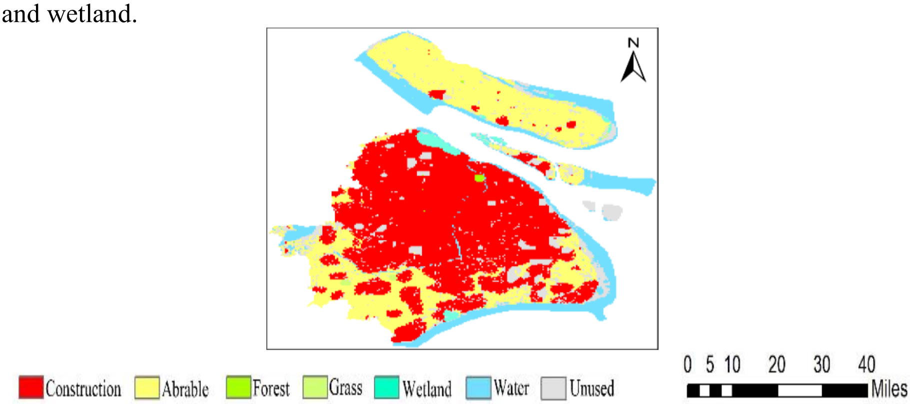


Fig. 5. Simulation results for 2030

4 Conclusions

According to the development model of 2010-2020[14], it is expected that by 2030, the area of construction land in Shanghai will still be in a state of expansion, the area of cultivated land and grassland will decrease, the area of forest land and grassland will increase, the area of water area will remain unchanged, and the area of other land will decrease. Therefore, Shanghai should strictly control the scale of construction land, optimize the internal structure of construction land, fully consider the integrity of new projects and surrounding building layout, and establish a land use model with economic, social and ecological benefits.

References

1. Du Peijun, Xia Junshi, Xue Chaohui, Tan Kun, Su Hongjun, Bao Rui. Advances in hyperspectral remote sensing image classification [J]. Journal of Remote Sensing, 2016,20 (02) : 236-256.
2. Mansour S, Al-Belushi M, Al-Awadhi T. Monitoring land use and land cover changes in the mountainous cities of Oman using GIS and CA-Markov modelling techniques[J]. Land use policy, 2020, 91: 104414.
3. Chen Qiao, Chen Yongfu, Ju Hongbo. Study on the monitoring method of vegetation change in Tianbao District based on 3S technology [J].Forestry Science Research, 2013,26 (06) : 736-743.
4. Prodhan F A, Zhang J, Sharma T P P, et al. Projection of future drought and its impact on simulated crop yield over South Asia using ensemble machine learning approach[J]. Science of The Total Environment, 2022, 807: 151029.
5. Duan Pan, Zhang Fei, Liu Changjiang. Impervious surface extraction and spatial difference analysis of typical cities in Xinjiang based on Sentinel-2A/B[J]. Journal of Remote Sensing, 2022,26 (07): 1469-1482.
6. Xia Junshi. Ensemble learning based hyperspectral remote sensing image classification[D]. China University of Mining and Technology, 2013.
7. Xia Junshi, Du Peijun, Zhang Wei. Key technologies and system implementation of remote sensing image multi-classifier integration [J]. Technology News, 2011,29(21): 22-26.
8. Djan'na H K, Diekkrüger B, Näschen K, et al. Impact of the accuracy of land cover data sets on the accuracy of land cover change scenarios in the Mono River Basin, Togo, West Africa[J]. 2019.
9. Chen Feng, Xia Jiaoxiong. Research on scholar interest mining based on fusion text topic model [J]. Computer Age, 2022 (09): 31-35.
10. Fang Binhao, Huang Hong, Zhou Ziyun. Research on intrusion detection methods based on multiple feature selection and mixed sampling [J]. Journal of Sichuan University of Light Chemical Technology (Natural Science Edition), 2022,35 (04): 85-93.
11. Lei Haochuan. Research on remote sensing image classification based on multi-classifier integration [D]. China University of Geosciences (Beijing), 2018.
12. Li C, Yu W, Dzodzomenyo M, et al. Growing spatial overlap between dam-related flooding, cropland and domestic water points: A water–energy–food nexus management challenge in Malawi and Ghana[J]. Frontiers in Water, 2021, 3: 730370.

13. Cunha E R. da, Santos CAG, Silva RM, da, Bacani VM, Pott A (2021) Future scenarios based on a CA-Markov land use and land cover simulation model for a tropical humid basin in the Cerrado/Atlantic forest ecotone of Brazil[J]. Land Use Policy, 2020, 101.
14. Quan Ruisong. Prediction of land use change and its hydrological effects in Shanghai based on scenario simulation [J]. Journal of Natural Resources, 2018, 33 (09): 1552-1562.