Multistep Prediction of Land Cover From Dense Time Series Remote Sensing Images With Temporal Convolutional Networks

Jining Yan, Xiaodao Chen, Yunliang Chen, and Dong Liang

Abstract—**Time series prediction (TSP) of land use/land cover (LULC) is an important scientific issue, but forecasting LULC changes at lead times of multiple time steps at fine time scales remains problematic. Especially in the context of current rapid economic and social development, the traditional one-step prediction models with a five-year or ten-year cycle cannot meet the application needs of land management departments. Temporal convolutional networks (TCNs) outperform other traditional TSP approaches. Therefore, we have proposed a pixel-level multistep TSP (pMTSP) approach that employs TCNs to carry out multistep prediction of land cover from dense time series remote sensing images, making up for the shortcomings of low accuracy, coarse time granularity, and labor-consuming of the current LULC prediction approaches. The results of comparative experiments with seasonal-trend decomposition procedure based on LOcally wEighted regreSsion Smoother and autoregression (STL-AR), seasonal autoregressive integrated moving average , and dynamic harmonics regression using single enhanced vegetation index time series, as well as the comparative experiment with the cellular automata-Markov model using real moderate resolution imaging spectroradiometer image time series, showed that the pMTSP can accurately extrapolate the change trend of the time series in fine-scale and obtain highly consistent prediction results with actual data, performing better than the other four contrasting algorithms in 23-step LULC prediction. The pMTSP can be used for multistep, fine-time-scale, and long time-series land cover prediction, which is of great guiding significance for the sustainable development and utilization of land resources.**

Index Terms—**Dense time series, land use/land cover (LULC), multistep prediction, pixel-level, temporal convolutional networks (TCNs).**

I. Introduction

TIME series prediction (TSP) of land use/land cover (LULC) [1], [2] is an important scientific issue that can help better understand the land use process, as well as changing laws and trends [3], [4], so as to support land use planning and decision-making for land use planners and resource managers in the future [5]. Scientists from all over the world have paid close attention to the subject of land cover change prediction, proposed many excellent prediction algorithms and models, and achieved many successful results. For example, on the pixel level, some classical regression and prediction algorithms, such as seasonal-trend decomposition procedure based on LOcally wEighted regreSsion Smoother (LOESS) and autoregression (STL-AR), seasonal autoregressive integrated moving average (SARIMA) [6], dynamic harmonics regression (DHR), and robust iteratively reweighted least squares (RIRLS) [7], [8], have been used to fit and predict the surface reflectance (SR) [9], leaf area index (LAI) [10], and other indexes of LULC. In addition, some postclassification algorithms, such as the cellular automata (CA)-Markov model [11] and the artificial neural network-based cellular automaton (ANN-CA) model [12], have been used to predict future land use situations based on historical classification results.

However, as for the previously used pixel-level LULC prediction algorithms, the prediction accuracy is generally not high, might be due to the defects of the model itself (for example, because STL often fails to extract the seasonality component accurately when seasonality shift and fluctuation exist [13], all STL-based time-series models cannot get accurate forecasting results), or not suitable for long time series forecasting [14] [such as the ARIMA-based algorithms, including autoregressive integrated moving average (ARIMA), SARIMA, etc.], or less generally applicable for nonlinear and nonstationary data sets [15] (such as the Fourier-decomposition-based algorithms, including DHR and RIRLS, etc.). As for the postclassification methods, the input data of the models is the LULC classification results of each time point, which may often be imprecise and ignore temporal dependencies that can be derived from remote-sensing time series [16], [17]. If the classification results of multiple time phases are input into the prediction model at the same time, it is easy to produce error accumulation, which makes the final prediction result less accurate. In addition, it is particularly time-consuming and laborious to make land use classifications at each time point [18]. Hence, the time granularity of the experimental data used in the previous studies was relatively coarse, and the forecast results were often at five-year or ten-year intervals. In the context of the current rapid economic and

Manuscript received June 1, 2020; revised August 5, 2020; accepted August 28, 2020. Date of publication September 1, 2020; date of current version September 16, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 41801363, and in part by the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) under Grant CUG170689. *(Corresponding author: Yunliang Chen.)*

Jining Yan, Xiaodao Chen, and Yunliang Chen are with the School of Computer Science and Hubei Key Laboratory of Intelligent Geo-Information Processing, China University of Geosciences, Wuhan 430074, China (e-mail: yanjn@cug.edu.cn; chenxd@cug.edu.cn; chenyl@cug.edu.cn).

Dong Liang is with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute and the University of Chinese Academy of Sciences, Chinese Academy of Sciences, Beijing 100049, China (e-mail: liangdong@radi.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3020839

social development, urban land use changes are more frequent. If we continue to use five or ten years as the forecast unit, the method will not be able to meet the application needs of land management departments [19]. In addition, most of the previous research work has focused on predicting the results at the next time step, which is called a one-step prediction, but cannot predict multiple time series into the future [20]. That is to say, LULC prediction results for future fine time scales are not available, preventing land use planning at finer timescales. Therefore, choosing a novel time series analysis model and using dense time series land cover satellite data to make multistep land cover predictions at fine time scale will be of great significance for urban land use planning and environmental protection [21].

Temporal convolutional network (TCN), one of the members of the convolutional neural network (CNN) [22], [23] family, can be used to fit and predict long time series with high precision and high efficiency due to its simple structure and novel network. Besides, it outperforms canonical recurrent networks such as long short-term memory networks [24] while demonstrating longer effective memory and has been successfully used in handwritten text recognition [25], action segmentation and detection [26], and so on. Hence, in this article, we propose a pixel-level multistep TSP (pMTSP) approach that employs TCNs to carry out multiple time steps prediction of land cover from dense time series remote sensing images on the pixel-level. Fully considering the inherent periodic characteristics of land cover time series, the pMTSP divides the long time series remote sensing images into time sequences with annual cycle, and then realizes the high-precision sequence-to-sequence prediction with the help of the TCN's flexible receptive field and temporal convolution features. The pMTSP is capable of predicting the entire subsequence in one-shot manner, overcoming the error transmission and accumulation problems of the recursive one-step forecasting method which uses isolated time points to make direct multistep forecasting. In addition, the pMTSP makes pixel-level prediction from remote sensing images, effectively avoiding the problems of error accumulation, coarse time granularity, and labor-consuming caused by the postclassification prediction approaches.

The remainder of this article is organized as follows. In the next section, we summarize the related work and put forward the problems and challenges faced by multistep LULC prediction. Section III introduces in detail the background knowledge and implementation process of the pMTSP approach. Section IV describes the experiments and result analysis as well as comparison and validation. Finally, in Section V, we provide a summary and conclude the article.

II. Related Work

According to the review results of the state-of-the-art land cover prediction, the popularly used land cover prediction methods can be classified into two types— pixel-level prediction methods and postclassification prediction methods.

A. Pixel-Level Prediction Methods

Pixel-level prediction methods are those that forecast LULC in the future using past time series remote sensing pixels. The popularly used methods include STL-AR, SARIMA [6], DHR, and RIRLS [7], [8], and so on.

1) STL-AR: The STL model is able to decompose a series into trend, seasonal, and residual components based on a LOESS [27]. It cannot be directly used to forecast future values without the help of the autoregression (AR) model. STL and AR have been used to model and predict the 2007 LAI values using six years (2001–2006) of the Moderate Resolution Imaging Spectroradiometer (MODIS) LAI product [28]. Since STL decomposition can isolate noise components in time series, the STL-AR model is more sensitive to noise in the data. However, the STL assumes that trend and seasonal components are smooth and slowly changing [13], which reduces the final prediction accuracy to some extent.
2) SARIMA: The SARIMA model is formed by including additional seasonal terms in the ARIMA model, which is a class of model that captures a suite of different standard temporal structures in time series data. Based on the assumption that the time series is a set of stochastic variables that depend on the time t, SARIMA can interpret the whole time series according to some rules or mathematical methods. However, one of the limitations of the SARIMA model is the stationarity of a time series, and it is often difficult for a time series to meet this modeling requirement in practice. Consequently, it is necessary to transfer a nonstationary time series into a stationary one by differencing before the SARIMA prediction is applied [28]. In addition, the SARIMA model is a typical short-term forecasting approach, and it is not suitable for long time series analysis [14].

   The SARIMA model has been used to model MODIS LAI time series, from 2001 to 2006, and predict 2007 LAI values. The comparative results showed that the SARIMA model gave better prediction results than STL-AR.
3) DHR and RIRLS: The DHR and RIRLS models both use the Fourier harmonic component to fit the previous time series and predict future values. The DHR model has an advantage of expressing seasonal or periodic components, so it is suitable for analyzing the time series with remarkable seasonal variations. The three components of the time series-trend, season, and residual-can be fitted by the Fourier harmonic components. But unlike the ordinary Fourier analysis, the harmonic coefficients of each component vary with time change, which also reflects the dynamic characters of the model. Benefitting from these multiscale sine and cosine components, the DHR model could fit the time series as much as possible to obtain a relatively smooth result. This model has been used to analyze the LAI time series products, and the annual prediction results showed that it was very effective in predicting the short-term LAI on a pixel basis [10].

   According to the data quality of the time-series remote sensing images, along with a long-term trend component, the RIRLS model can also be transformed into three other typical types—a simple model with only four coefficients, an advanced model with six coefficients, and full model with eight coefficients [9]. The simple model has been successfully applied to the continuous change detection and classification algorithm of land cover for a Landsat scene located in New England, USA, but it had problems when used for places where intra-annual changes occurred [7]. The similar idea of the advanced model has been employed in the continuous monitoring of forest disturbance algorithm for one Landsat scene located between Georgia and South Carolina, USA [29]. Furthermore, the full model has been successfully used for detecting forest disturbance from a satellite image time series [30].

Due to the DHR and RIRLS models, both being capable of modeling the seasonality of the data, the prediction values will not be influenced by vegetation phenology and sun angle differences [9]. However, due to the fact that the classical Fourier analysis is less generally applicable for nonlinear and nonstationary data sets [15], the accuracy of extrapolated predictions for surface reflectivity time series is not high.

Therefore, the conclusion is that we should develop a better forecasting model to obtain more accurate land use prediction results in advance.

B. Postclassification Prediction Methods

Postclassification prediction methods utilize classified land use data of each previous period to predict future land cover. Due to the need to perform land use classification for each time point, the efficiency of these methods is very low. Therefore, the most common strategy adopted with these methods is to use fixed time interval land use classification results, such as one year, five years, or ten years, to make advance predictions using Markov and its improved models. For example, López *et al.* [31]successfully predicted land cover and land use change in the urban fringe for the next 20 years using Markov chains and regression analyses; Yirsaw *et al.* [11] employed a CA-Markov model to predict future LULC changes for the year 2020, based on the mapping results of the years 1990, 2000, and 2010, and validations with the actual data showed an overall satisfactory result; Saputra and Lee [12] applied an ANN-CA model to predict LULC changes in 2050 and 2070 in North Sumatra, Indonesia, and the comparison between the predicted and the real LULC maps for 2010 illustrated high agreement.

However, one of the disadvantages of this strategy is coarse time granularity, leading to the fact that it cannot provide land use change details at fine time scales. If we use the postclassification land use time series to make multistep predictions, the classification error of each time point will accumulate and lead to less accurate forecasting results. In addition, this strategy ignores temporal dependencies and change periodicity that can be derived from remote-sensing time series, which directly affect the final prediction accuracy.

Through the above review of two typical commonly used LULC prediction algorithms, the main issues can be identified as the following.

1) Low prediction accuracy, not only for pixel-level prediction methods but also for postclassification prediction methods.
2) Previous work mainly focused on the prediction of the next time point, but little research focused on the prediction of future multistep time series.
3) Coarse time granularity, which is very obvious for the post-classification prediction methods and is not be able to meet the current application needs of land management departments.
4) Laborious and time-consuming nature, which are the most obvious disadvantages of postclassification prediction methods. Therefore, we propose the pMTSP approach employing TCNs to predict multiple-time-step land cover, from dense time series remote sensing images on the pixel-level, making up for the shortcomings of low accuracy, coarse time granularity, and labor-consuming of current TSP approaches.

III. Multistep Prediction of Land Cover

The overall technical solution of the multistep prediction of land cover from dense time series remote sensing images with TCNs mainly includes the following three steps:

1) remote sensing pixel series preprocessing;
2) TCN-based multistep pixel series prediction; and
3) performance evaluation and optimization.

A. Data Preprocessing

Resulting chiefly from varying atmospheric conditions and sunsensor-surface viewing geometries [32], the remote sensing pixel series may be affected by prevalent noise. Hence, the direct method is to suppress outliers using high-pass, low-pass, band-pass, or band-stop filters to reduce the influence of low-quality data [33]. Considering the noise characteristics of the time-series remote sensing images, the Whittaker filter was adopted because it can provide a better fit to the raw time series data than the Fourier analysis, asymmetric Gaussian model, or double logistic model [34]. In addition, in order to realize the multistep prediction, the long time-series remote sensing images should be divided into several pixel vectors, and the length of each vector is the same as the annual change cycle of the original time series.

B. TCN-Based Multistep TSP

1) Background Knowledge of the TCN: There are two important components of the TCN: one is the 1-D fully convolutional network (1-D FCN) architecture [26], and the other is the causal convolutions. Therefore, the TCN can be simply expressed as TCN = 1D FCN + causal convolutions.

1) 1-D FCN: In FCN, all hidden layers are all convolutional layers, which is different from using a fully connected layer to obtain a fixed-length feature vector after the convolutional layer in classic CNN [35]. The FCN can accept input images of any size, and its output is the same size as the input images, thanks to upsampling after the last convolutional layer, i.e., deconvolution [26]. If the input images become 1-D series data, then the FCN becomes 1-D FCN, which can take a sequence of any length and map it to an output sequence of the same length.

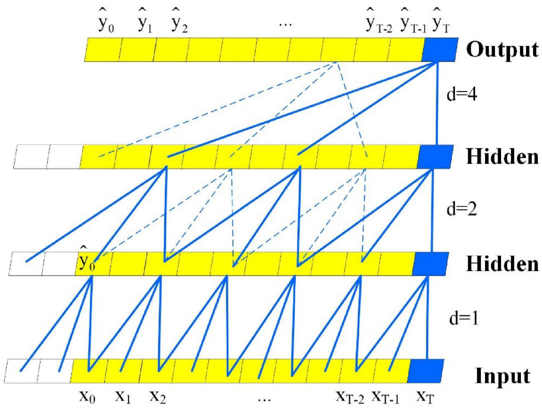


Fig. 1. Dilated causal convolution with dilation factors $d = 1, 2, 4$ and filter size $k = 3$.

2) Causal Convolutions: Due to the input image size of CNN needs to be fixed, it is not suitable for dealing with sequence problems [26], [36], and so causal convolution is applied to deal with these issues. Yet a simple causal convolution is challenged by sequence tasks with long histories, the dilated convolution, which enables an exponentially large receptive field [37], to be employed.

Specifically, the dilated convolution has a dilation rate parameter apart from the size of the convolution kernel, which is mainly used to indicate the size of the dilation, and this is its main difference from the simple causal convolution [38]. For a 1-D sequence input $X \in R^n$ and a filter $\mathrm{f}: \{0, \ldots, k-1\} \rightarrow R$, the dilated convolution operation F on element s of the sequence is defined as the following:

$$\mathrm{F(s)} = (X *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot X_{s-d \cdot i} \tag{1}$$

where d is the dilation factor, k is the filter size, and $\mathrm{s} - d \cdot i$ accounts for the direction of the past. Then, the architectural elements in a TCN can be illustrated in Fig. 1. When $d = 1$, the dilated convolution reduces to a regular convolution; if the filter size k is chosen to be larger and the dilation factor d increases, the receptive field of the TCN will be increased. Thus, dilated convolution can be used to deal with long sequence problems.

2) TCN-Based Multistep Prediction: Given a time series $Q = q_1, q_2, \cdots, q_{n*T}$, where T is the period and $n * T$ represents the total length of the time series. Our goal is to predict the next multistep values of Q. In general, as for time series forecasting, is to predict the observation at the next time step, called one-step prediction, as only one time step is to be predicted. If we want to predict the next T-step values, the common way is to split the multistep prediction problem into several one-step prediction subproblems [39], as follows:

$$\begin{aligned} q_{n*T+1} &= \mathrm{model}_1(q_{(n-1)*T+1} + q_{(n-1)*T+2} + \cdots + q_{n*T}) \\ q_{n*T+2} &= \mathrm{model}_2(q_{(n-1)*T+2} + \cdots + q_{n*T} + q_{n*T+1}) \\ &\cdots\cdots \\ q_{(n+1)*T} &= \mathrm{model}_T(q_{n*T} + \cdots + q_{(n+1)*T-2} + q_{(n+1)*T-1}). \end{aligned} \tag{2}$$

However, this method does not take into account the inherent periodic characteristics in the time series, and the last prediction result in the time series is based on its previous prediction result, which easily causes error transmission and accumulation, resulting in the accuracy of the prediction result gradually decrease as the prediction step length increase. Therefore, we adopted the multiple output strategy, which involved developing one model that was capable of predicting the entire forecast sequence in a one-shot manner, as the following:

$$\begin{aligned} &[q_{n*T+1}, q_{n*T+2}, \ldots, q_{(n+1)*T}] \\ &= model(q_{(n-1)*T+1}, q_{(n-1)*T+2}, \ldots, q_{n*T}). \end{aligned} \tag{3}$$

In fact, this prediction method is to divide the time series into several vectors according to the period T, and the length of each vector is T. The input and output of the prediction model are vectors, not individual sequence points. But the "model" in (6) must be trained on the previous values to get accurate multistep prediction results. Therefore, the first step of the pMTSP is to transform the time series Q into several vectors with length of T. Then, each vector was put into the TCN model for iterative training. Once the model training was complete, it could be used to predict the next multistep values. Finally, the predicted vector would be reverse-transformed into a time series, which was the multi-step prediction results. The overall technical solution is as shown in Fig. 2.

In addition, in order to improve the accuracy of multistep prediction results as much as possible, the TCN construction process needed to consider the characteristics of the time series. In this study, the initial filter size k was set equal to the period T of the time series, and the model input parameter size must also be fit for the time-series vector, so as to consider the data of one period as a whole.

C. Performance Evaluation and Optimization

In order to evaluate the forecasting performance of the pMTSP, as well as reverse optimize parameters of the TCN model, we chose the following two methods.

1) Time Dimension Evaluation: The time dimension evaluation is mainly used to compare the actual remote sensing pixel series with the multistep prediction results using two metrics—one is the root-mean-square error (RMSE), and the other is the Pearson correlation coefficient (PCC) [40]. Their calculation formulas are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{m} (a_i - p_i)^2}{m}} \tag{4}$$

$$\mathrm{PCC} = \frac{\sum_{i=1}^{m} (a_i - \overline{a}) \cdot (p_i - \overline{p})}{\sqrt{\sum_{i=1}^{m} (a_i - \overline{a})^2} \cdot \sqrt{\sum_{i=1}^{m} (p_i - \overline{p})^2}} \tag{5}$$

where a represents the actual time series value, $\overline{a}$ represents the mean value of the actual value, p represents the prediction results, $\overline{p}$ represents the mean value of the prediction results, and m represents the length of the time series data. If the PCC value is small or the RMSE value is large, the "epochs" parameters, as well as other parameters, can be considered to adjust if necessary.

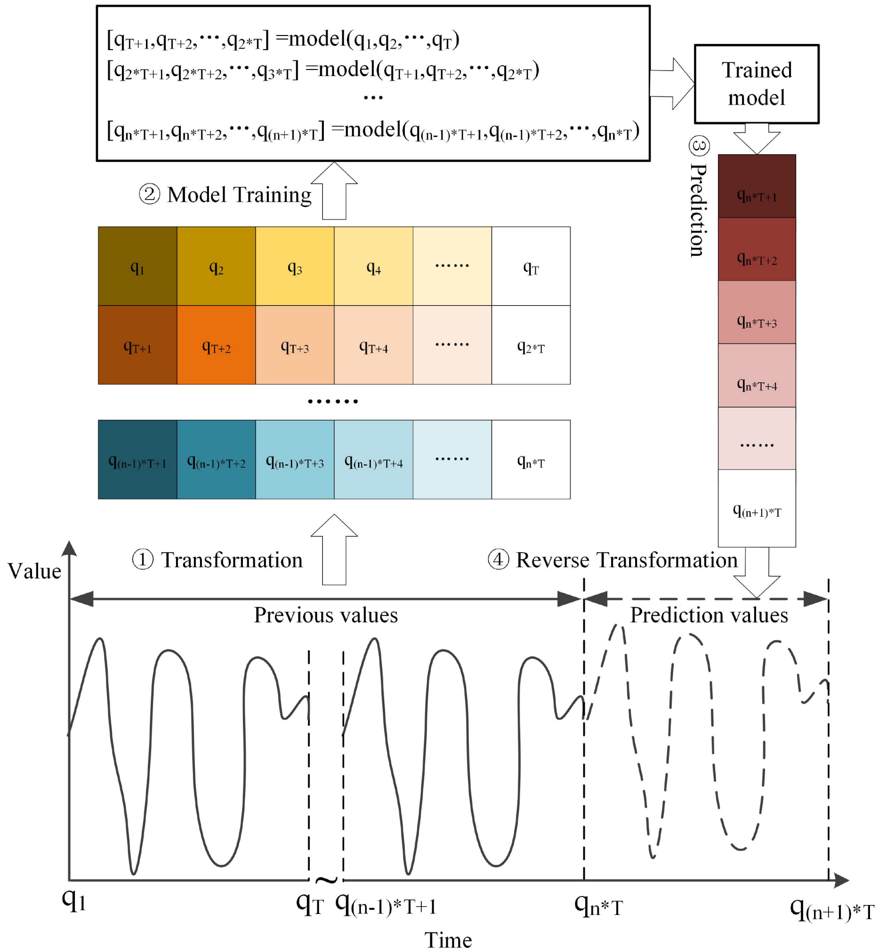


Fig. 2. Overall technical solution of the pMTSP.

2) Spatial Dimension Evaluation: Spatial dimension evaluation is used to compare the difference between the actual and the multistep prediction results of the same time point. The peak-signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) two well-known objective image quality metrics were adopted [41]. The PSNR is the ratio of the energy of the peak signal to the average energy of the noise, usually expressed in decibels (dBs). The calculation formula of PSNR is as follows:

$$\mathrm{PSNR}(aI, pI) = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE(aI, pI)}\right) \quad (6)$$

where

$$\mathrm{MSE}(aI, pI) = \frac{1}{\mathrm{M} \cdot \mathrm{N}} \sum_{i=1}^{M} \sum_{j=1}^{N} (aI_{ij} - pI_{ij})^2 \quad (7)$$

where aI and pI denote the actual and predicted grey-level images, MAX_I is the maximum gray value, generally assigned 255 (8-b grey-level image), and i, j denote the width and length of the image. The larger the PSNR value, the smaller the difference between the predicted and the actual images.

SSIM is a full-reference image quality evaluation index that measures image similarity from three aspects: luminance (L), contrast (C), and structural (S). The calculation formula of SSIM is as follows:

$$\mathrm{SSIM}(aI, pI) = \mathrm{L}(aI, pI) \cdot \mathrm{C}(aI, pI) \cdot \mathrm{S}(aI, pI) \quad (8)$$

where

$$\begin{cases} \mathrm{L(aI, pI)} = \frac{2 \cdot \mu_{\mathrm{aI}} \cdot \mu_{\mathrm{pI}}}{\mu_{\mathrm{aI}}^2 + \mu_{\mathrm{pI}}^2} \\ \mathrm{C}(aI, pI) = \frac{2 \cdot \sigma_{\mathrm{aI}} \cdot \sigma_{\mathrm{pI}}}{\sigma_{aI}^2 + \sigma_{pI}^2} \\ \mathrm{S}(aI, pI) = \frac{2 \cdot \sigma_{\mathrm{aIpI}}}{\sigma_{aI} + \sigma_{pI}} \end{cases} \quad (9)$$

where μ_{aI} and μ_{pI} denote the mean values of actual and predicted images, σ_{aI} and σ_{pI} denote the standard deviation of actual and predicted images, and σ_{aIpI} is the covariance of both images. The value range of SSIM is $0 - 1$. The larger the SSIM value, the more similar between the predicted and the actual images.

In addition, since the proposed pMTSP approach would be compared with the typical postclassification prediction method, the K-means [42] unsupervised classification method was introduced to perform land use classification for the actual data and the prediction results and then to

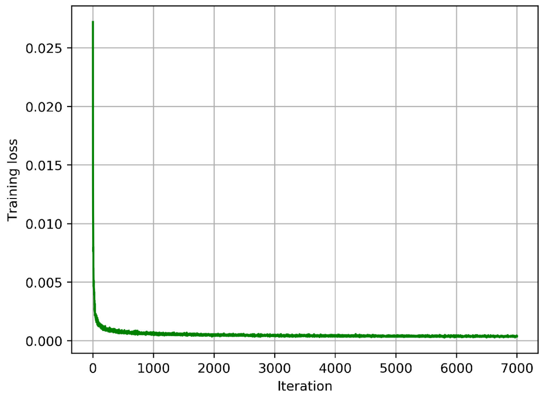


Fig. 3. Trend curve of the training loss with the training times.

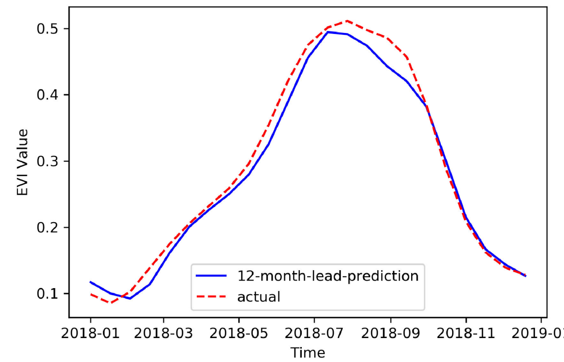


Fig. 4. Predicted and actual EVI curves using the pMTSP method.

evaluate the classification accuracy equivalent to the actual land use results.

IV. Validation and Results

Validation of the pMTSP was conducted using a typical index product of LULC-enhanced vegetation index (EVI), which can demonstrate sharper growing season peaks and exhibits greater sensitivity to canopy structure differences than other vegetation indexes such as the normalized difference vegetation index [33], [43]. In order to make comparisons with those benchmark prediction methods, the experimental data adopted: 1) a single randomly selected EVI time series, measuring prediction accuracy in time dimension; and 2) the real MODIS image time series, verifying its ability to solve practical LULC-prediction problems.

A. Multistep Prediction in Single EVI Time Series

1) Experimental Data: The experimental data used were a randomly selected EVI time series from an actual MODIS Terra 16-day composite data (MOD13Q1), which was collected from January 1, 2001 to December 19, 2018, a total of 18 years (414 time points), in Wuhan, China. The advance prediction step was set to 12 months. That is, using the previous 12-m EVI time series (23 time points) to predict the next 12-m EVI time series (23 time points). In our validation experiments, we chose 17 years of the dataset (from 2001 to 2018, 391 time points) for model training and the last year of the dataset (2018, 23 time points) for validation.

2) TCN-Based TSP: The TCN is a typical TSP algorithm based on iterative optimization. Hence, the determination of the iteration number is a very important issue. In our experiment, we set the initial maximum number of training sessions to 7000 during the model training process and compared the trend of the training loss with the training times to select a “relatively suitable number.” Here, the “relatively suitable number” means that the iteration number cannot be so large as to make the model overfit and time-consuming or so small as to result in a low prediction accuracy [44]. The trend curve of the training loss with the training times was as shown in Fig. 3.

As can be seen from Fig. 3, the loss value gradually decreased and stabilized after 5000 training sessions with the number of training rising. Therefore, we had reason to believe that the 5000 may be the “relatively suitable number” of training sessions. The comparison chart of the 12-m-lead-prediction results and the actual time series are as shown in Fig. 4.

As can be seen from Fig. 4, the 12-m-lead-prediction time series closely fit the actual time series. Except for mid-July to mid-September, the prediction results of other time intervals were almost identical to the actual values. This may be due to the abrupt change of EVI time series in mid-July leading to the parameters of the TCN model being unable to adjust in time. However, even during the periods of deviation, the maximum relative deviation error was not more than 5%. The overall 0.01915 RMSE value and 0.9960 PCC value also strongly demonstrated the high multistep prediction accuracy of TCNs.

3) Comparative Experiments: In order to measure prediction accuracy in the time dimension, we chose three frequently used pixel-level prediction methods, STL-AR, SARIMA, and DHR, to carry out comparative experiments. All of the comparative experiments used the same 16-d MODIS EVI time series with the pMTSP test. All experiments used 17 years of the dataset (from 2001 to 2018, 391 time points) for model training and the last year of the dataset (2018, 23 time points) for validation.

1) STL-AR: The STL model was used to decompose the EVI time series into trend, season, and residual subcomponents, then perform TSP on each subcomponent using the AR model; prediction results of each subcomponent were combined to get the final results. The results of the 23-step prediction using STL-AR are shown in Fig. 5.

   As can be seen in Fig. 5, the 12-m-lead-prediction curve deviates far from the actual curve, and the maximum relative deviation error from July to September reached about 20%. In the actual EVI time series, the two months are in the turning range of the EVI value from the prosperity to the decline. That is to say, the STL-AR model cannot quickly capture the changing trend, leading to huge errors in timing transition intervals. In addition, although the prediction series has similar growth or declining trends to the actual value overall, it cannot perceive the change details, eventually leading to the predicted land cover results not being consistent with the actual situation. The

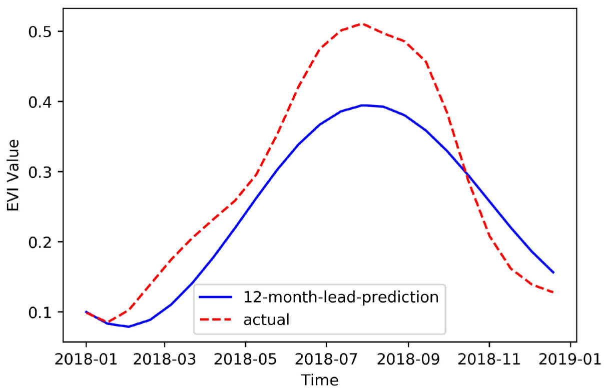


Fig. 5. Predicted and actual curves using the STL-AR method.

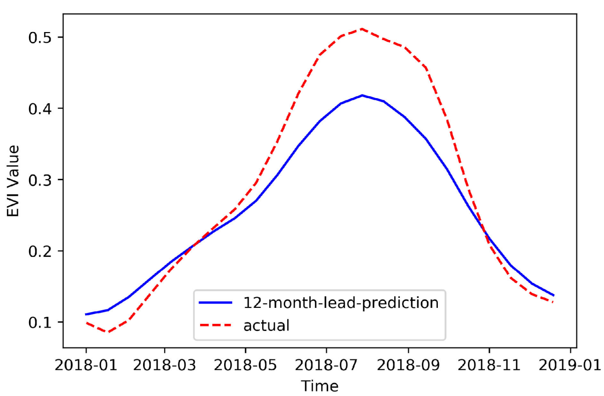


Fig. 6. Predicted and actual EVI curves using the SARIMA method.

0.0686 RMSE value and 0.9536 PCC value also reflect the deficiency of STL-AR model.

2) SARIMA: The SARIMA model was mainly used to analyze univariate time series with trend and seasonal elements through setting seasonal autoregressive order P, seasonal difference order D, seasonal moving average order Q, and time step m of a single seasonal cycle. In this comparative experiment, these four parameters were set to 1, 1, 1, and 23, respectively. The results of the 23-step prediction using SARIMA are shown in Fig. 6.

As can be seen from Fig. 6, the 12-m-lead-prediction curve has the same downward and upward trends as the actual value. The 0.9967 PCC value also illustrates the high degree of fit. However, around August, the maximum relative deviation error between the predicted and actual values still reaches about 20%, which shows that the SARIMA model is weak in predicting the time-varying intervals. The 0.0554 RMSE value also reflects the deficiency of the SARIMA model.

3) DHR: The DHR model uses the dynamic Fourier harmonic component to fit the previous time series and predict the future values, and its dynamically changing sine and cosine components can be used to fit historical time series with high accuracy. However, the DHR model has weak extrapolated prediction ability, so the prediction accuracy

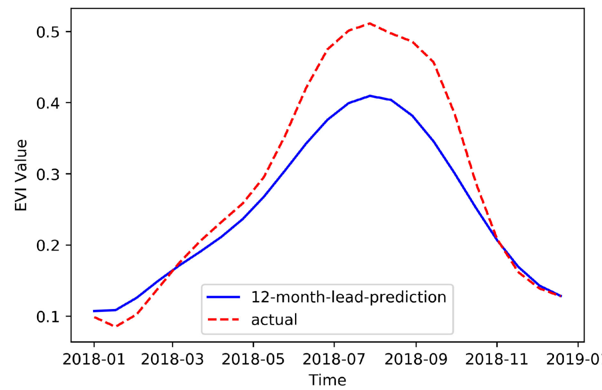


Fig. 7. Predicted and actual EVI curves using the DHR method.

TABLE I
RMSE, PCC, AND TOTAL RUNNING TIME OF pMTSP, STL-AR, SARIMA, AND DHR

Method	RMSE	PCC	Total running time (seconds)
pMTSP	0.01915	0.9960	1018
STL-AR	0.0686	0.9536	20
SARIMA	0.0554	0.9967	23
DHR	0.0598	0.9965	827

for the future is not high. The results of 23-step prediction using DHR are shown in Fig. 7.

As can be seen from Fig. 7, the 12-m-lead-prediction curve has the same change trend as the actual value. But the maximum relative deviation error between the predicted and actual values still reaches about 20%, which may lead to a wrong prediction type of LULC. The 0.0598 RMSE value and 0.9965 PCC value also reflect the deficiency of DHR model.

4) Discussion: From the above experimental results, it can be concluded that: 1) STL-AR, SARIMA, and DHR can accurately extrapolate the change trend of the time series, but the relative deviation error between the predicted and actual values during the growth trend change intervals is very high. The pMTSP results can be a close fit to the actual time series, except for some small deviation errors in the turning points of the series. 2) Through comparing the RMSE and PCC results between the predicted and the actual values, it can be concluded that the pMTSP has the best results, followed by SARIMA and DHR, and the STL-AR is the worst. However, in terms of total running time, the STL-AR has the highest working efficiency, and the pMTSP has the longest running time, taking 1018 s for 5000 iterations of the TCN model training (see Table I). In summary, the pMTSP can accurately extrapolate the change trend of the time series and obtain highly consistent prediction results, performing better than STL-AR, SARIMA, and DHR overall in multi-step TSP.

B. Multistep Prediction in Real MODIS Image Time Series

1) Experimental Data: The experimental data used were a real 18-year EVI time series of MOD13Q1, collected from January 1, 2001 to December 19, 2018, in Wuhan, China. In the nearly past 20 years, Wuhan has experienced rapid urbanization development such as urban expansion and urban transformation, which is suitable as a research object for multistep prediction of land cover. Just like the parameter setting of the single EVI time series experiment, the prediction step was also set to 12 months, the first 17 years of the dataset (from 2001 to 2018, 391 time points) was used for model training, and the last year of the dataset (2018, 23 time points) was used as the ground truth for validation. Notably, the Whittaker filter was used to suppress outliers, reduce the influence of low-quality data, and filter the noise of each pixel series.

2) TCN-Based Land Cover Prediction: According to the network building experience in the experiment of single EVI time series, the number of training sessions was set to 5000 during the TCN model training process. In addition, because the LULC prediction experiment was performed on the pixel-level, the training and prediction process of each pixel was independent. Hence, the entire calculation process could be accelerated using multigraphics processing units (GPUs) in parallel [45]. Therefore, we successfully applied for two computing nodes from China's Tianhe-2 supercomputing clusters, and each node has four Nvidia Tesla K80 GPUs. The experimental MODIS EVI time series was divided into eight subcomponents in the spatial dimension, and these eight subcomponents were parallel performed with land cover prediction based on a batch job mechanism [46]. The final 23-step prediction results are shown in Fig. 8. For simplicity, only four of the prediction results, which were evenly distributed in the four seasons of the year 2018, are shown for comparison with the actual images. In addition, the histograms for and differences between the actual and prediction images were also calculated.

As can be seen in Fig. 8, from the analysis of image texture features [47], the TCN prediction results have almost the same texture features as the actual data, which is also supported by the fact that the four different images do not reflect obvious texture features. Therefore, we believed that the 12-m-lead-TCN-prediction method could accurately predict the overall land cover situation of the next 12 months, which also reflected the advantages of the pMTSP; from the analysis of image grayscale characteristics, the prediction results of the two points of 20 180 306 and 20 181 219 were almost the same as the actual data of those time points, and the 20 180 610 and 20 180 914 prediction results were slightly different from the actual data. This may be due to the fact that the EVI value is in the transition period from June to September, and the parameters of the TCN model cannot be adjusted in time. However, it can be seen from the four difference histogram images that most of the differences between the actual and predicted values were between -0.1 and 0.1, which strongly supported the high accuracy of the TCN's prediction results.

In addition, in order to quantify the prediction accuracy overall, the RMSE, PCC, PSNR, and SSIM results between the actual and TCN-predicted values were calculated (see Fig. 9).

As can be seen from Fig. 9, except for a few small areas, the RMSE values in most areas are less than 0.1 or close to 0, the PCC values are larger than 0.9 or close to 1, the PSNR values at 23 time points are greater than 115 dBs, and the SSIM values at 23 time points are larger than 0.97 or close to 1, which strongly illustrate that the results obtained by the pMTSP are very slightly different from the actual values, both in spatial and time dimensions. In other words, the 23 MODIS EVI images predicted by TCNs for the next 12 months are highly similar to the actual obtained images.

3) Comparative Experiments With CA-Markov: In order to further evaluate the prediction accuracy of the pMTSP in the spatial dimension, we chose the commonly used postclassification prediction method CA-Markov to carry out comparative experiments. Because the input of CA-Markov model was land use classification results, it would have been too time-consuming to perform land use classification for all of the 414 EVI images. Hence, we only chose four typical time points to compare. The details are as follows.

1) Taking into account the rapid development of Wuhan city, we chose to employ the land use classification results of 2016 and 2017 to predict the land use situation for 2018. Therefore, eight MODIS EVI images collected on March 6, June 10, September 14h, and December 19 in 2016 and 2017 were selected.
2) In order to maximize the fine-scale prediction of CA-Markov, we performed K-means unsupervised classification on the EVI images in 2016 and 2017, with the classification number set to 10. It should be noted that the classification number set to 10 was only to finely compare the accuracy of CA-Markov and pMTSP. The numbers 1 to 10 were not given clear names for land use types, but generally represented grassland, forest, impervious, water and their subtypes. The CA-Markov model we used was integrated in TerrSet software, which is an integrated geographic information system and remote sensing software developed by Clark Labs at Clark University for the analysis and display of digital geospatial information [48], [49].
3) The actual and TCN prediction results for March 6, June 10, September 14, and December 19 in 2018 were also performed K-means unsupervised classification, with the classification number set to 10.

The final prediction results by CA-Markov, the K-means classification results of actual data, and the K-means classification results of TCN prediction values were as shown in Fig. 10.

As can be seen from Fig. 10, the unsupervised classified results of the actual data and TCN prediction values were almost the same in most areas, except for a few minor differences. However, the prediction results of CA-Markov model were obviously different from the actual classified results, especially for the prediction of image texture information. For quantitative evaluation, we randomly selected cross-validation samples in the actual classification results and evaluated the relative overall classification accuracy (OCA) and relative kappa coefficient (Kappa) of the TCN-predicted classification results and the CA-Markov prediction results [50]. The word "relative" here means that we assumed that the cross-validation samples selected from the actual classification results were completely correct, and all the accuracy evaluation results were based on this assumption (see Table II).

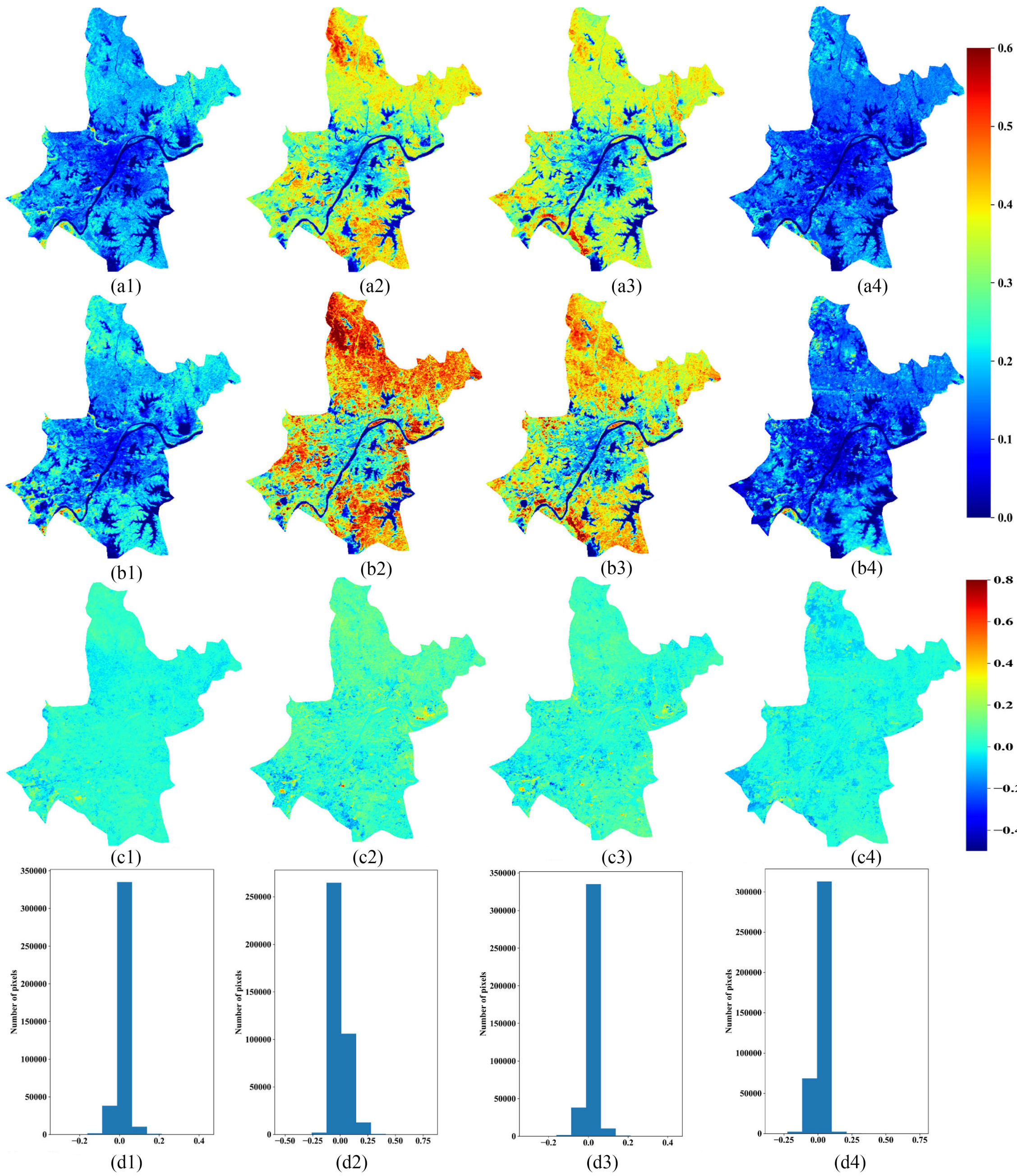
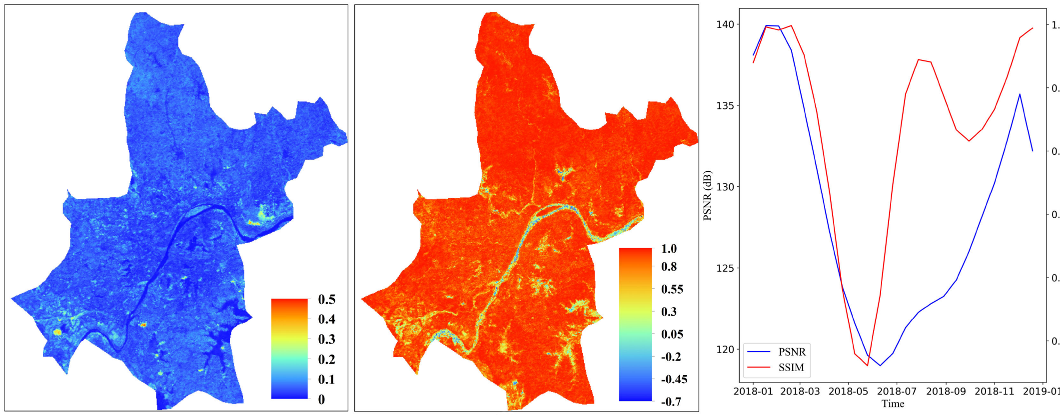


Fig. 8. Comparison of the pMTSP results with the actual images. Images a1, a2, a3, and a4 are the TCN prediction results; b1, b2, b3, and b4 are the actual values; c1, c2, c3, and c4 are the differences between the actual and prediction images; and d1, d2, d3, and d4 are the histograms of each of the difference results. In order to display results intuitively, the TCN prediction, actual, and difference images are all given false colors. The same color bar is used for the actual and prediction values, and the difference results adopt a separate color bar.

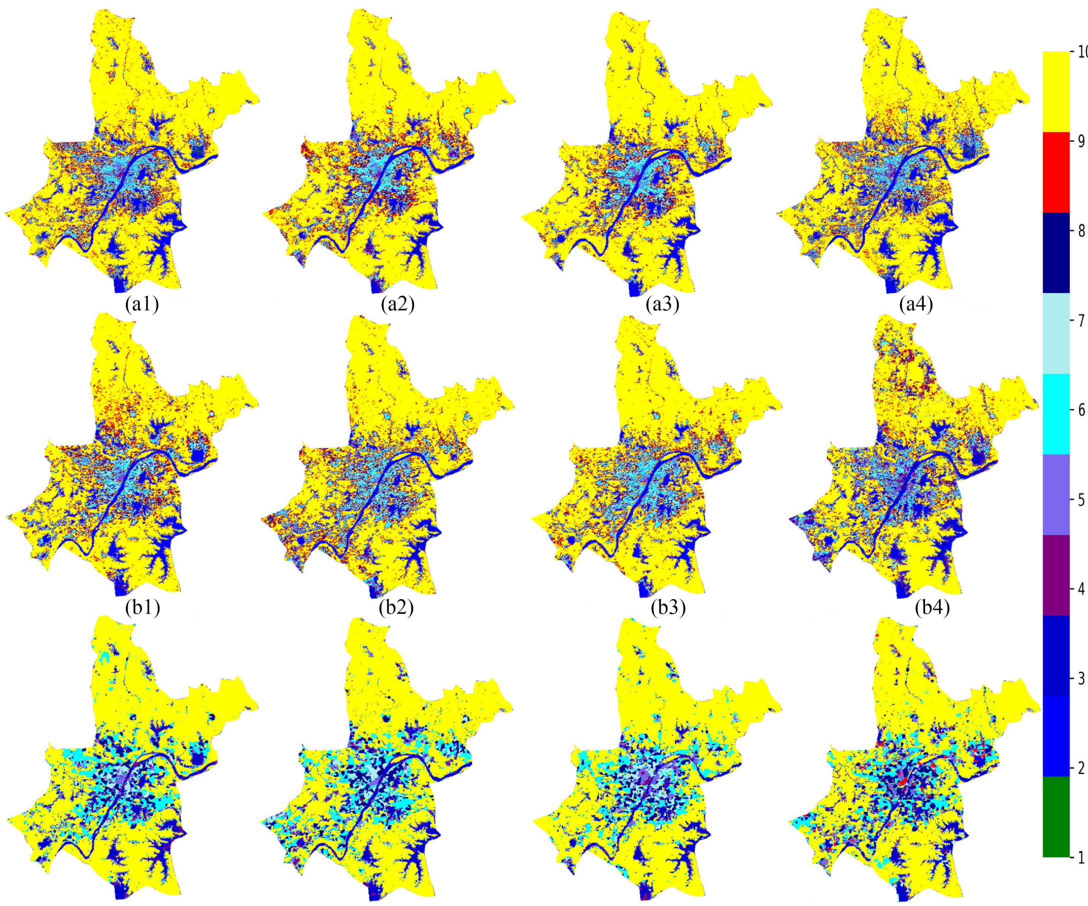
TABLE II
ACCURACY COMPARISON OF THE POST-pMTSP CLASSIFICATION AND CA-MARKOV PREDICTION RESULTS

post-pMTSP classification results			CA-Markov prediction results		
Time point	Relative OCA	Relative Kappa	Time point	Relative OCA	Relative Kappa
20180306	84.7101%	0.8338	20180306	75.6669%	0.7147
20180610	82.2047%	0.8013	20180610	74.5693%	0.7993
20180914	82.6819%	0.8073	20180914	70.6229%	0.7784
20181219	82.8490%	0.8083	20181219	77.4565%	0.7386



(a) (b) (c)

Fig. 9. RMSE, PCC, PSNR, and SSIM results between the actual and TCN-predicted values. (a) RMSE. (b) PCC. (c) PSNR & SSIM.



(a1) (a2) (a3) (a4)

(b1) (b2) (b3) (b4)

(c1) (c2) (c3) (c4)

Fig. 10. K-means unsupervised classification results of the actual and TCN prediction values, as well as the prediction results by CA-Markov. The numbers 1 to 10 refer to the classification labels, and different categories display different colors on the classified images. (a1) TCN Prediction Result(20180306). (a2) TCN Prediction Result(20180610). (a3) TCN Prediction Result(20180914). (a4) TCN Prediction Result(20181219). (b1) Actual Result(20180306). (b2) Actual Result(20180610). (b3) Actual Result(20180914). (b4) Actual Result(20181219). (c1) CA-Markov Result(20180306). (c2) CA-Markov Result(20180610). (c3) CA-Markov Result(20180914). (c4) CA-Markov Result(20181219).

As can be seen from Table II, all of the relative OCA and Kappa values of the post-pMTSP classification results were better than those of the CA-Markov. In other words, the agreement between the TCN-predicted classification results and the actual LULC types was higher than that of the CA-Markov model. Therefore, the conclusion was that the pMTSP method outperformed the postclassification CA-Markov model in LULC prediction.

C. Discussion

Through testing by single EVI time series and real MODIS image time series, the main characteristics of pMTSP could be obtained.

The advantages of this multistep LULC prediction method are as follows.

1) The pMTSP, using the TCN deep learning algorithm to perform multistep land cover prediction, can accurately extrapolate the change trend of the time series and obtain highly consistent prediction results with actual data.
2) The pMTSP performs land cover prediction from dense time series remote sensing images on the pixel-level and can effectively solve the problems of low accuracy, coarse time granularity, and labor-consuming in the postclassification prediction methods.
3) Through training the historical data of the entire time series and then predicting the change trend in the future, the pMTSP is possible to fully consider temporal dependencies and change periodicity that can be derived from remote-sensing time series, to improve the prediction accuracy.
4) Based on the idea of sequence-to-sequence prediction, the pMTSP is capable of predicting the entire time sequence in one-shot manner, overcoming the error transmission and accumulation problems of the recursive one-step forecasting method which uses individual sequence points to make direct multistep forecasting.

The pMTSP also has the following limitations.

1) The pMTSP results can be a very close fit with the actual time series, but there are still some small deviation errors in the turning points of the series.
2) Compared with the postclassification methods, the pMTSP can greatly save manpower. But the pixel-level LULC multistep prediction is still very computationally demanding. Though eight GPUs of China's Tianhe-2 supercomputing clusters were adopted, the 18-year EVI time series (a total of 162 736 pixels with 414 time points) 23-step prediction in Wuhan city still took more than 24 days [51], [52].
3) Since the pMTSP method works on the pixel-level, it is more suitable for performing multistep land cover prediction from a slightly lower spatial resolution remote sensing images. For LULC prediction using submeter level remote sensing data, the pixel-level prediction method may cause more noise pollution for the final results [53].

V. Conclusion

In order to solve the problems faced by the current LULC prediction issues that include low prediction efficiency, coarse time granularity, and labor-consuming, we propose the pMTSP approach. This approach employs a simple structure and novel deep learning network, i.e., the TCN, to perform multistep LULC prediction. The results of comparative experiments with STL-AR, SARIMA, and DHR using single EVI time series, as well as the comparative experiment with CA-Markov model using real MODIS image time series, showed that the pMTSP can accurately extrapolate the change trend of the time series on a fine scale and obtain highly consistent prediction results with the actual data, performing better than the other four contrasting algorithms in multistep LULC prediction. In addition, the pMTSP performs LULC prediction from dense time series remote sensing images on the pixel-level and can obtain fine time scale LULC prediction results that are able to meet the application needs of land management departments in the context of current rapid economic and social development.

However, its computationally demanding nature is the most prominent problem faced by the pMTSP. The time series of each pixel needs to go through two stages of model training and prediction, but the iterative training of the TCN model is very computationally demanding. In practical application scenarios, a remote sensing image often contains millions, or even tens of millions, of pixels, and the workload of pixel-level multistep prediction is very heavy. In future work, we plan to use an unsupervised classification algorithm to cluster the time series with similar change characteristics, then choose one of the time series curves of the same category for model training, and finally use the transfer learning method [54] to transfer the model parameters of the trained time series to the pixel series that are not involved in training, reducing model training times so as to improve the working efficiency of the entire model.

Appendix

Following Abbreviations Are Used in This Manuscript

TSP	Time series prediction
LULC	Land use/land cover
TCN	Temporal convolutional network
LOESS	LOcally wEighted regreSsion Smoother
STL-AR	Seasonal-trend decomposition procedure based on LOESS and autoregression
SARIMA	Seasonal autoregressive integrated moving average
DHR	Dynamic harmonics regression
CA	Cellular automata
pMTSP	Pixel-level multistep time series prediction

References

[1] F. E. Fassnacht *et al.*, "Importance of sample size, data type and prediction method for remote sensing-based estimations of aboveground forest biomass," *Remote Sens. Environ.*, vol. 154, pp. 102–114, 2014.
[2] Z. Hao, A. AghaKouchak, N. Nakhjiri, and A. Farahmand, "Global integrated drought monitoring and prediction system," *Scientific Data*, vol. 1, 2014, Art. no. 140001.
[3] B. C. Pijanowski, D. G. Brown, B. A. Shellito, and G. A. Manik, "Using neural networks and GIS to forecast land use changes: A land transformation model," *Comput., Environ. Urban Syst.*, vol. 26, no. 6, pp. 553–575, 2002.
[4] F. Gerber, R. de Jong, M. E. Schaepman, G. Schaepman-Strub, and R. Furrer, "Predicting missing values in spatio-temporal remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2841–2853, May 2018.
[5] D. Guan, H. Li, T. Inohae, W. Su, T. Nagaie, and K. Hokao, "Modeling urban land use change by the integration of cellular automaton and markov model," *Ecological Model.*, vol. 222, no. 20/22, pp. 3761–3772, 2011.
[6] Z. Xiao, S. Liang, J. Wang, B. Jiang, and X. Li, "Real-time retrieval of leaf area index from modis time series data," *Remote Sens. Environ.*, vol. 115, no. 1, pp. 97–106, 2011.
[7] Z. Zhu and C. E. Woodcock, "Continuous change detection and classification of land cover using all available Landsat data," *Remote Sens. Environ.*, vol. 144, pp. 152–171, 2014.

[8] P. Fu and Q. Weng, "A time series analysis of urbanization induced land use and land cover change and its impact on land surface temperature with Landsat imagery," *Remote Sens. Environ.*, vol. 175, pp. 205–214, 2016.

[9] Z. Zhu, C. E. Woodcock, C. Holden, and Z. Yang, "Generating synthetic Landsat images based on all available Landsat data: Predicting Landsat surface reflectance at any given time," *Remote Sens. Environ.*, vol. 162, pp. 67–83, 2015.

[10] B. Jiang, S. Liang, J. Wang, and Z. Xiao, "Analysis and prediction of MODIS LAI time series with dynamic harmonic regression model," *J. Remote Sens.*, vol. 14, no. 1, pp. 13–32, 2010.

[11] E. Yirsaw, W. Wu, X. Shi, H. Temesgen, and B. Bekele, "Land use/land cover change modeling and the prediction of subsequent changes in ecosystem service values in a coastal area of China, the Su-Xi-Chang region," *Sustainability*, vol. 9, no. 7, 2017, Art. no. 1204.

[12] M. H. Saputra and H. S. Lee, "Prediction of land use and land cover changes for North Sumatra, Indonesia, using an artificial-neural-network-based cellular automaton," *Sustainability*, vol. 11, no. 11, 2019, Art. no. 3024.

[13] Q. Wen, J. Gao, X. Song, L. Sun, H. Xu, and S. Zhu, "RobustSTL: A robust seasonal-trend decomposition algorithm for long time series," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, pp. 5409–5416.

[14] A. A. Adebiyi, A. O. Adewumi, and C. K. Ayo, "Comparison of ARIMA and artificial neural networks models for stock price prediction," *J. Appl. Math.*, vol. 2014, 2014, Art. no. 614342.

[15] P. Singh, S. D. Joshi, R. K. Patney, and K. Saha, "The fourier decomposition method for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. A: Math., Phys. Eng. Sci.*, vol. 473, no. 2199, 2017, Art. no. 20160871.

[16] R. Flamary, M. Fauvel, M. Dalla Mura, and S. Valero, "Analysis of multitemporal classification techniques for forecasting image time series," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 953–957, May 2015.

[17] J. Inglada, A. Vincent, M. Arias, B. Tardy, D. Morin, and I. Rodes, "Operational high resolution land cover map production at the country scale using satellite image time series," *Remote Sens.*, vol. 9, no. 1, 2017, Art. no. 95.

[18] X. Li, W. Chen, X. Cheng, and L. Wang, "A comparison of machine learning algorithms for mapping of complex surface-mined and agricultural landscapes using ZiYuan-3 stereo satellite imagery," *Remote Sens.*, vol. 8, no. 6, 2016, Art. no. 514.

[19] F. Yulianto, T. Maulana, and M. R. Khomarudin, "Analysis of the dynamics of land use change and its prediction based on the integration of remotely sensed data and CA-Markov model, in the upstream Citarum Watershed, West Java, Indonesia," *Int. J. Digit. Earth*, vol. 12, no. 10, pp. 1151–1176, 2019.

[20] A. Venkatraman, M. Hebert, and J. A. Bagnell, "Improving multi-step prediction of learned time series models," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 3024–3030.

[21] A. M. Hersperger *et al.*, "Urban land-use change: The role of strategic spatial planning," *Global Environ. Change*, vol. 51, pp. 32–42, 2018.

[22] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[23] N. Wang, F. Chen, B. Yu, and Y. Qin, "Segmentation of large-scale remotely sensed images on a spark platform: A strategy for handling massive image tiles with the MapReduce model," *ISPRS J. Photogrammetry Remote Sens.*, vol. 162, pp. 137–147, 2020.

[24] W.-S. Hu, H.-C. Li, L. Pan, W. Li, R. Tao, and Q. Du, "Spatial-spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4237–4250, Jun. 2020.

[25] J. Gan, W. Wang, and K. Lu, "In-air handwritten chinese text recognition with temporal convolutional recurrent network," *Pattern Recognit.*, vol. 97, 2020, Art. no. 107025.

[26] C. Lea, M. D. Flynn, R. Vidal, A. Reiter, and G. D. Hager, "Temporal convolutional networks for action segmentation and detection," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 156–165.

[27] J. Verbesselt, R. Hyndman, G. Newnham, and D. Culvenor, "Detecting trend and seasonal changes in satellite image time series," *Remote Sens. Environ.*, vol. 114, no. 1, pp. 106–115, 2010.

[28] B. Jiang, S. Liang, J. Wang, and Z. Xiao, "Modeling modis lai time series using three statistical methods," *Remote Sens. Environ.*, vol. 114, no. 7, pp. 1432–1444, 2010.

[29] Z. Zhu and C. E. Woodcock, "Object-based cloud and cloud shadow detection in Landsat imagery," *Remote Sens. Environ.*, vol. 118, pp. 83–94, 2012.

[30] J. Verbesselt, A. Zeileis, and M. Herold, "Near real-time disturbance detection using satellite image time series," *Remote Sens. Environ.*, vol. 123, pp. 98–108, 2012.

[31] E. López, G. Bocco, M. Mendoza, and E. Duhau, "Predicting land-cover and land-use change in the urban fringe: A case in Morelia City, Mexico," *Landscape Urban Planning*, vol. 55, no. 4, pp. 271–285, 2001.

[32] J. N. Hird and G. J. McDermid, "Noise reduction of NDVI time series: An empirical comparison of selected techniques," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 248–258, 2009.

[33] I. Nitze, B. Barrett, and F. Cawkwell, "Temporal optimisation of image acquisition for land cover classification with random forest and modis time-series," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 34, pp. 136–146, 2015.

[34] P. M. Atkinson, C. Jeganathan, J. Dash, and C. Atzberger, "Inter-comparison of four models for smoothing satellite sensor time-series data to estimate vegetation phenology," *Remote Sens. Environ.*, vol. 123, pp. 400–417, 2012.

[35] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Fully convolutional neural networks for remote sensing image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 5071–5074.

[36] F. Schuiki, M. Schaffner, F. K. Gürkaynak, and L. Benini, "A scalable near-memory architecture for training deep neural networks on large in-memory datasets," *IEEE Trans. Comput.*, vol. 68, no. 4, pp. 484–497, Apr. 2018.

[37] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, *arXiv:1511.07122*.

[38] J. Zhang, H. Wu, W. Su, X. Wang, M. Yang, and J. Wu, "A new approach for classification of epilepsy eeg signals based on temporal convolutional neural networks," in *Proc. 11th Int. Symp. Comput. Intell. Des.*, 2018, vol. 2, pp. 80–84.

[39] A. Galicia, R. Talavera-Llames, A. Troncoso, I. Koprinska, and F. Martínez-Álvarez, "Multi-step forecasting for big data time series based on ensemble learning," *Knowl.-Based Syst.*, vol. 163, pp. 830–841, 2019.

[40] K. D. Singh, D. Ramakrishnan, and L. Mansinha, "Relevance of transformation techniques in rapid endmember identification and spectral unmixing: A hypespectral remote sensing perspective," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 4066–4069.

[41] Q. Poterek, P. Herrault, G. Skupinski, and D. Sheeren, "Deep learning for automatic colorization of legacy grayscale aerial photographs," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2899–2915, May 28, 2020.

[42] R. Kuçak, E. Özdemir, and S. Erol, "The segmentation of point clouds with k-means and ANN (artifical neural network)," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 595–598, 2017.

[43] J. Yan, L. Wang, W. Song, Y. Chen, X. Chen, and Z. Deng, "A time-series classification approach based on change detection for rapid land cover mapping," *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 249–262, 2019.

[44] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.

[45] A. Agathos, J. Li, D. Petcu, and A. Plaza, "Multi-GPU implementation of the minimum volume simplex analysis algorithm for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2281–2296, Jun. 2014.

[46] W. Xue *et al.*, "Ultra-scalable CPU-MIC acceleration of mesoscale atmospheric modeling on Tianhe-2," *IEEE Trans. Comput.*, vol. 64, no. 8, pp. 2382–2393, Aug. 2014.

[47] J. Yuan, D. Wang, and R. Li, "Remote sensing image segmentation by combining spectral and texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 16–24, Jan. 2013.

[48] H. Khawaldah, I. Farhan, and N. Alzboun, "Simulation and prediction of land use and land cover change using GIS, remote sensing and CA-Markov model," *Global J. Environ. Sci. Manag.*, vol. 6, pp. 215–232, 2019.

[49] Clark Labs, Worcester, MA, USA, *Terrset Geospatial Monitoring and Modeling Software*. Accessed: Feb. 15, 2020. [Online]. Available: https://clarklabs.org/terrset/.

[50] B. Huang, B. Zhao, and Y. Song, "Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery," *Remote Sens. Environ.*, vol. 214, pp. 73–86, 2018.

[51] L. Wang, Y. Ma, A. Y. Zomaya, R. Ranjan, and D. Chen, "A parallel file system with application-aware data layout policies for massive remote sensing image processing in digital earth," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 6, pp. 1497–1508, Jun. 2015.

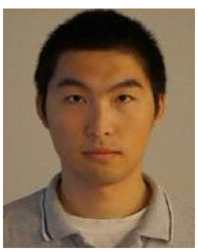
[52] J. Liu, Y. Xue, K. Ren, J. Song, C. Windmill, and P. Merritt, "High-performance time-series quantitative retrieval from satellite images on a GPU cluster," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 2810–2821, Aug. 2019.

[53] W. Han, R. Feng, L. Wang, and Y. Cheng, "A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 23–43, 2018.

[54] M. Wurm, T. Stark, X. X. Zhu, M. Weigand, and H. Taubenböck, "Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks," *ISPRS J. Photogrammetry Remote Sens.*, vol. 150, pp. 59–69, 2019.

Jining Yan received the Ph.D. degree in signal and information processing from the University of Chinese Academy of Sciences, Beijing, China.

He is an Associate Professor with the School of Computer Science, China University of Geoscience, Wuhan, China. His research interests include remote sensing data processing, time-series analysis and change detection, cloud computing in remote sensing, and applied oceanography.



Xiaodao Chen received the B.Eng. degree in telecommunication from the Wuhan University of Technology, Wuhan, China, in 2006, the M.Sc. degree in electrical engineering from and the Ph.D. degree in computer engineering from Michigan Technological University, Houghton, MI, USA, in 2009 and 2012, respectively.

He is currently an Associate Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. His research interests include design automation for petroleum system, high performance computing, and optimization.

Yunliang Chen received the B.Sc. and M.Eng. degrees from the China University of Geosciences, Wuhan, China, and the Ph.D. degree from Huazhong University of Science and Technology, Wuhan, China.

He is currently an Associate Professor with the School of Computer Science, China University of Geosciences. His research interests included computer network engineering, cloud computing, and IoT.

Dong Liang received the B.Sc. degree in applied mathematics from the University of Hull, Hull, U.K., in 2006, and the M.Sc. degree in applied mathematics from Mälardalen University, Västers, Sweden, in 2009. He is currently pursuing the Ph.D. degree in cartography and GIS from the University of Chinese Academy of Sciences, Beijing, China, with a focus on big Earth data and polar remote sensing.