

RESEARCH ARTICLE

A Spatio-Temporal Neural Network Learning System for City-Scale Carbon Storage Capacity Estimating

MOU CHAO^{1,2}, WEI MAIMAI¹, LIU HANZHANG¹, CHEN ZHIBO^{1,2}, AND CUI XIAOHUI^{1,2}¹School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China²Engineering Research Center for Forestry-Oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China

Corresponding author: Cui Xiaohui (cuixiaohui@bjfu.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant BLX202140, in part by the Beijing Forestry University "Double First-Class" Construction Project under Grant 450-GK112203008, and in part by the Beijing Forestry University National Training Program of Innovation and Entrepreneurship for undergraduates.


ABSTRACT Carbon storage capacity can be estimated to establish evaluation standards and statistics for carbon neutrality. Existing estimation methods including machine learning system have weakness modeling ability, and they are unable to deal with the complex topographies and temporal changes in vegetation in urban zones. However, a deep neural network has the potential ability to face such complex scenes because of its nonlinear fitting properties which has been widely used in industry. In this study, a novel and powerful neural network learning system named CiSL-NPP is proposed, to integrate multi-source data as an accurate, efficient, simple, long-term, city-scale carbon storage capacity measurement method. We use an unsupervised generation model spatio-temporal Masked AutoEncoder (st-MAE) and lightweight RNN model to efficiently obtain high-quality data that contain time-series information such as seasons; this minimizes any undue impact of meteorological, temporal, and spatial factors on the measurement owing to the multi-source, multi-modal nature of the data. In the MAE, we add seasonal coding to make it time-series sensitive; we also embed road network information to accurately perceive the complex topography of the city. Results for 16 cities in China and Europe revealed that the proposed method shows: 1) Higher-quality generated data (MSE is 0.13-0.29); 2) Accurate coverage of time series and complex geographical features; 3) Satisfaction of estimation demands with only 316 RMB cost; 4) Capability to evolve a long-term trend of urban vegetation carbon storage capacity in 4.2 days; and 5) Easily interpretable results which could, in practice, provide sound guidance for urban planning and decision-making processes.

INDEX TERMS Neural network learning system, unsupervised generation model, spatio-temporal masked autoencoder, carbon storage capacity measurement.

I. INTRODUCTION

More than 90% of human economic activity takes place in cities [1], which directly contribute at least 85% of global carbon emissions [2] and are regarded as the world's main battlefields of carbon neutrality [3]. Currently, carbon mitigation and carbon sink increment are the primary approaches to achieving urban carbon neutrality. The carbon mitigation approach, a complex systematic undertaking, is difficult to

implement [4]. Further, strict policies towards emissions may have consequences such as carbon leakage [5], [6]. The carbon sink increment approach shows more potential to achieve urban carbon neutrality. For example, studies indicate that the carbon sink of urban vegetation can offset more than 22.45% of carbon emissions [7], already achieving carbon neutrality in some areas [8]. In city, limited to a fixed-size space, the effective way to increase carbon sink is to use green resources (e.g., municipal parks) rationally according to their vegetation carbon storage capacity. Net Primary Productivity (NPP), a crucial metric reflecting the storage effect of vegetation,

The associate editor coordinating the review of this manuscript and approving it for publication was Jon Atli Benediktsson .

is the key role for increasing carbon storage capacity in the city [9], [10], [11].

However, estimating NPP accurately is a challenging problem [12], [13]. Compared to the large-scale and coarse particle forest estimation environment, urban vegetation is influenced by urban planning and meteorology; it shows irregular distribution and obvious periodic growth changes [14], [15]. City managers have to continuously monitor changes in city carbon sinks over extended periods of time [16], [17]. It is very difficult to accurately estimate NPP in a low-cost, efficient, and long-term manner at the city scale, which is also the big challenge facing in this work.

Existing approaches to estimating NPP can be roughly divided into traditional methods and simple artificial intelligence (AI) methods, neither of which can be directly applied to the urban environment. Traditional approaches including the plot inventory method, model inversion method (e.g., remote-sensing-based light utility efficiency (LUE) model), and flux observation method focus on forest ecology and do provide accurate results under large-scale estimation conditions. However, it is hard to use traditional methods for city-scale NPP estimation due to: 1) the requirement of long-term field observation which is costly as well as time- and labor-intensive [18], [19]; 2) both social and natural activities in cities significantly affect meteorology [20] which increasing the difficulty of obtaining high-quality remote sensing data in urban environments; 3) the complex topography of a given city also restricts the deployment of data acquisition sensors over sufficiently large scales [21]; 4) compared to forests, urban vegetation undergoes rapid changes in distribution and growth as cities develop, it cannot satisfy the requirements for long-term, continuous regulation [22], [23]. On the other hand, as neural networks have been widely used in industry fields [24], [25], simple neural networks (i.e., simple AI model) were initially applied to large-scale NPP estimation in efforts to improve computational efficiency and reduce the necessary cost, manpower, and time. Despite AI's powerful space-time computing capability, this approach is quite controversial in carbon neutrality [26].

Nowadays, the fusion methods have been used to estimate NPP at the city scale, and it is promising to be the most efficient method since solving the accuracy problem. One of them is the NPP function which proposed by Mngadi et al. [27] according to the modified MOD17 model. NPP function combines the advantages of the traditional methods and neural networks, and it is a fine-grained and accurate estimation method for region-level carbon sinks in cities. This approach simplifies the calculation by combining the deployment of sensors and remote sensing, which minimizes computational cost while accurately estimating small-range urban areas over certain time periods. However, the NPP function remains costly and inefficient when obtaining high-quality data. And it cannot perceive the city's complex spatial and temporal changes, so that it is not applicable for long-term, city-scale NPP estimation. Hence, we answer the research

questions in this work: 1) Is there a simple, low-cost, efficient and effective way to estimate carbon storage capacity in the city scale for long periods of time? 2) Can the method processes spatial and temporal information which hinder the carbon storage capacity estimation in city? 3) Is the carbon storage capacity estimating method "green"?

To achieve the low-cost, efficient, accurate, and long-term estimation of vegetation carbon storage capacity at the city scale, we propose a fused multi-source and multi-modal city-scale NPP estimation method, City-Scale Long-term NPP (CiSL-NPP), which has spatio-temporal perception ability by reconstructing high-quality data with spatio-temporal information. Based on the NPP function, from the perspective of data acquisition, we combine a LUE model and powerful deep neural networks model to fuse multi-source data (e.g., Photosynthetically Active Radiation, PAR, represents the portion of solar radiation that can be used by green plants for photosynthesis and is used to calculate the actual energy absorbed by vegetation) and urban road network data. CiSL-NPP adopts the efficient and simple calculation framework of the NPP function. An unsupervised MAE model is used in CiSL-NPP to generate remote sensing data with spatio-temporal information, which avoids the undue influence of meteorology or urban geographical environment complexities in obtaining high-quality remote sensing data. The proposed approach can perceive seasonal time series over urban complex topography by embedding temporal and urban road network information. Additionally, an RNN is trained to generate sensor data (e.g., PAR), further reducing data acquisition costs. According to experiments on data sets from 16 cities in China and Europe, we find that our approach not only estimates urban vegetation carbon storage efficiently, accurately, and easily at the city scale, but also has relatively low computational consumption, economic cost, and carbon expenditure cost.

The contributions of this work are highlighted as follows.

(1) A novel neural network learning system named CiSL-NPP is established to estimate carbon storage capacity at the city scale and it has the powerful modeling ability. By fusing remote sensing, road network, climate, and other multi-source data, this approach not only allows for simple, low-cost, accurate, and efficient calculation, but also can monitor urban carbon storage capacity for long periods of time. It can also dynamically evolve urban carbon storage capacity for further analysis.

(2) The st-MAE is introduced in the CiSL-NPP provides the ability of processing spatial and temporal data. In the application of city-scale carbon storage capacity estimating, the spatio-temporal ability makes it able to: a) combine macro climate regulation factors by embedding seasonal data and incorporating agricultural growth rules as time code, reflecting the evolution of NPP over the given time scale; b) incorporate data from the urban road network, allowing it to better define and differentiate urban vegetation areas. In this way, the accuracy of vegetation area reconstruction

is improved while enabling the analysis of complex urban topographies.

(3) With the benefits of transfer learning, the CiSL-NPP can be completed quickly and with less consumed resources. The model can be trained in 4.2 days for 316 RMB¹ and 20.724 CO₂ equivalent emissions (i.e., the metric for measuring carbon pollution introduced in [28], the lower, the more environmentally friendly). Thus, the CiSL-NPP can be used for actual decision-making regarding urban carbon neutrality.

The remainder of this manuscript is organized as the following: Section II introduces the background regarding carbon storage capacity estimating at city scale and the challenges in this field. Section III introduces the details of the proposed CiSL-NPP method. Section IV introduces the exhaustive experiments in this work to show the advantages of our method, such as quantitative analysis, qualitative analysis and case study. Section V introduces the results of experiments and discusses them comprehensively. Finally, Section VI concludes this paper and gives directions for future trends.

II. BACKGROUND

Through our exhaustive investigation, we noticed that although cities play a very important role in carbon neutrality, the existing carbon sequestration calculation methods are not well applied to the urban scale, which aroused our interest and led to the this work.

A. CARBON NEUTRALIZATION IN CITY

As mentioned above, cities contribute 85% of global industrial direct carbon emissions [2]. Due to continuous urbanization [2], cities have become key places for carbon neutrality in many countries [29], [30], [31]. Carbon mitigation and carbon sink increment are the two paths most often taken toward achieving urban carbon neutrality. Carbon mitigation is difficult and costly to implement as there is no unified standard for carbon emissions calculation [32], [33], and because many links and stakeholders are involved in this process [34], [35], [36]. For example, emissions reduction requires cities to transform their roles via cooperation among city-level governments, enterprises, and institutions at substantial costs of both time and financial resources [4]. Meanwhile, strict policies on carbon emissions have serious consequences including carbon leakage in developing countries [29], [37]. Carbon sink increment is a more promising approach. As carbon storage capacity in the urban area directly originates from vegetation [38], carbon sink increment with vegetation may be an efficient technique for achieving carbon neutrality in cities. Previous studies indicate that the increased carbon storage of vegetation during urbanization process can offset more than 22.45% of a city's carbon emissions [7].

In cities, the rational allocation of green resources is one of the most effective and direct ways to increase carbon sinks. Azaria et al. [39] found that urban green space distribution

and CO₂ storage have a 0.79 correlation coefficient. Gong and Luo [40] suggest that the carbon storage effect of urban vegetation can be increased four-fold by transforming simple green space into compound space. As urban greening rates increase year-by-year [41], an effective green resource allocation strategy is needed.

To develop a green resource allocation strategy, an efficient, accurate and convenient estimation of carbon storage capacity serves as a crucial foundation [30], [42]. NPP is one of the most significant metrics to quantify vegetation carbon storage capacity [9], [10], [11], and its accurate estimation has been a contentious and challenging issue [13], [43]. The approaches to estimating NPP include, as mentioned above, traditional methods and simple AI methods. Traditional methods mainly focus on the forest environment and provide accurate estimation results under large-scale estimation conditions. Sample plot inventory [44], model inversion [45], and flux observation [46] are included in this category. Simple AI methods have been extensively applied for NPP estimation over large scales as well [47], [48], [49], [50]. However, simple AI models yield quite different results from traditional methods, so their validity is questionable; simple AI models are more controversial in terms of NPP estimation [26]. Researchers have found that NPP varies greatly on different time scales [15] and that most estimation methods are greatly affected by climate change [51]. In addition, due to the various types of urban vegetation and complicated urban topographies [52], [53], accurately estimating NPP at the city scale using any of these tools is highly challenging.

Below, we introduce common NPP estimation methods and the efforts of previous researchers in urban NPP estimation.

B. EXISTING METHOD

Traditional methods and simple AI methods comprise the two main categories of exiting carbon storage estimation. Most NPP estimation processes are a combination or improvement upon methods in these two categories.

1) TRADITIONAL METHODS

There are three main types of traditional carbon storage estimation methods, namely, sample inventory, model inversion, and flux observation.

Sample inventory method. Sample (plot) inventory is a classical forest carbon storage estimation method that is operated primarily by measuring biomass [54]. Data for the target area is first acquired including forest type, average tree height, and diameter at breast height. These data are then used to calculate or model the carbon sink reserves [44], [55], [56]. This approach is simple and relatively accurate, but is limited by area conditions and may damage the forest under analysis. More importantly, it does not reflect the spatial and temporal characteristics of vegetation. The sample inventory method cannot be directly applied to city-scale NPP estimation.

Model inversion method. Model inversion calculates carbon storage by establishing a data model, for example, based

¹RMB is the legal tender of the People's Republic of China.

on remote sensing data. Forest carbon storage can be calculated through optical remote sensing data, synthetic aperture radar satellite data (SAR), and laser radar data (LiDAR). The LUE model is the most commonly used, which simulates the final NPP based on the absorption of photosynthetically active radiation (FAR) by vegetation and the actual photosynthetic utilization rate of vegetation [57]. Model inversion requires costly data acquisition, has strict data quality requirements, and cannot perceive spatio-temporal information, so it cannot be directly applied to urban scenes. Compared to large-scale forest scenes, urban climate variability [58] has a greater impact on remote sensing data [59].

Flux observation method. The flux observation system uses observation stations established over a large scale to observe the exchange flux of greenhouse gases in the atmosphere and ecosystem. The Eddy Covariance (EC) technique [60] is mainly used to measure carbon dioxide. The Pan-European Research Infrastructure (ICOS) established an observation system for the entire European continent and quantified the photosynthesis, respiration, and evaporation of vegetation via EC to calculate the carbon storage of vegetation [46]. This method requires numerous instruments and equipment for long-term detection, which is time-consuming, expensive, and may not properly reveal timely temporal and spatial changes. It does not provide efficient or low-cost carbon storage estimation [61] and thus is not suitable for NPP estimation at the city scale.

2) SIMPLE AI METHODS

AI can simulate human behavior in computers by learning, performing analysis, and making decisions; today, it is an indispensable technology ubiquitous throughout human society [62]. Many scholars have applied AI methods for NPP estimation. Most have used simple models such as Random Forest (RF) [63], Support Vector Machine (SVR) [64], and multi-layer perceptron (MLP) [65]. For example, Moosavi et al. [66] estimated regional NPP with an ANN model. Based on multispectral images, Peng et al. [67] applied RF and SVR models to estimate regional NPP at the canopy scale. Although simple AI shows potential for NPP estimation, the latest research by Wang et al. [26] shows that it does not accurately reflect climate and temperature changes when calculating carbon storage capacity. In addition, the results of NPP as estimated by simple AI models differ significantly from those of traditional methods and have not been widely recognized. It may be necessary to use more powerful deep neural networks [68], [69].

C. CITY-SCALE NPP ESTIMATION

A great deal of previous work has been done on NPP calculations, but studies on the urban scale are still lacking.

1) NPP FUNCTION

In contrast to natural environments like forests, cities have infrastructures such as parks, buildings, highways, and

vegetation areas like green belts and artificial ecological forests. This variety, as well as the anthropological effects (e.g., urban planning) upon these spaces, create significant changes in city-scale vegetation distribution [70], [71]. Currently, research on the estimation of the carbon storage capacity of urban vegetation is still in the preliminary stage. Most studies have applied common carbon storage estimation methods (Section.B) and their combinations, such as traditional methods and simple AI methods, which lack specific approaches for city-scale NPP estimation. In addition, urban management requires making scientific decisions by monitoring changes in urban carbon stocks over long time periods [16]. Consequently, city-scale NPP estimation is rendered ineffective by variable meteorological conditions, complex topographies, rapid seasonal changes in vegetation, difficulty in accurate estimation, and continuous regulation.

Existing NPP estimation methods cannot be applied in cities due to several limitations. Plot inventory, for example, is too expensive and incapable of predicting future spatial and temporal vegetation characteristics to apply. The LUE model (model inversion) is widely used, however, it cannot perceive temporal and spatial distribution characteristics due to its strict requirements for data quality and the extent to which it is influenced by climate and geomorphological conditions. The flux observation method requires a flux observation system, which is overly expensive to construct, and does not provide sufficiently fine-scale estimation for cities. Simple AI models have insufficient modeling ability and do not yield sufficiently accurate prediction results. A scientific, cost-effective, and readily applicable estimation method is urgently required [72], [73].

Mngadi et al. [27] proposed the NPP function, a fine-grained accurate carbon sink estimation method for urban areas, based on a modified LUE model (MOD17) in an effort to resolve the above problems. The NPP function first selects random sample points in the research area, then vegetation parameters (PAR data) and satellite remote sensing image data are collected through sensors.

The NPP function is simple, low in cost, and provides accurate NPP estimations in small urban areas. However, this method has limitations when applied to larger city-scale NPP estimation. First, it does not efficiently provide high-quality data as it requires the deployment of numerous sensors. The remote sensing data of most cities are affected by cloud cover and other meteorological factors. Monitoring cannot be conducted over certain time scales, and the accuracy of the results decreases significantly due to the inability to perceive urban space-time information. For example, in remote sensing images of Beijing, China, only 24% of days in the past five years have seen less than 5% average cloud cover (Fig.1(a)). As shown in Fig.1(b), due to the large administrative area and complex landforms in cities, regional meteorological conditions significantly vary despite the relatively little cloud cover, markedly reducing the available data and directly affecting the accuracy of evaluations. Although the NPP function still cannot be directly used to estimate NPP

at the city scale, it still shows remarkable potential for urban application due to its other advantages.

2) NEURAL NETWORK SYSTEM

Some scholars have argued the feasibility and broad application prospects of deep learning models to estimate NPP in large-scale forest scenarios [74], but no empirical research on urban application has been conducted. The application of deep learning for urban NPP estimation is limited by remote sensing data availability, time series perception ability, and climate perception ability.

Unsupervised learning [75] is a potential solution to the difficult and expensive training necessary for data acquisition. MAE, a recently developed unsupervised learning model, has been widely recognized for its powerful data reconstruction ability and application prospect [76], [77]. Few have attempted to apply MAE for urban NPP estimation and existing MAE models cannot solve time series problems.

RNN is generally considered to have time-series processing capability and is widely used in time-series prediction tasks. It is derived from a series of classical models [78]. However, RNN cannot be applied to city-scale NPP estimation directly as it does not reflect the influence of climatic or other factors.

III. METHOD

A. CiSL-NPP FRAMEWORK

We developed the novel CiSL-NPP estimation method for city-scale NPP estimation by using neural networks. As shown in Fig.2(a), traditional methods can only estimate NPP momentarily by collecting data at certain time points. Conversely, CiSL-NPP can accurately generate data with time series by introducing a powerful deep learning model to realize long-term NPP estimation by data generation (as shown in Fig.2(b)). CiSL-NPP generates urban vegetation characteristic data accurately and evolves the dynamic changes in urban carbon storage capacity through a spatio-temporal Masked Auto Encoder (st-MAE) model. In st-MAE, a seasonal temporal coding mechanism and spatial mask based on OpenStreetMap (OSM)² road network information are introduced to manage the vegetation growth cycle and urban topography. An RNN applicable to regression tasks is trained to generate the required sensor data (i.e., PAR). Fig.3 shows the calculation framework of CiSL-NPP including data alignment, data generation, and NPP calculation steps, which are discussed in further detail below.

Our loss function computes the Gaussian mean squared error (GMSE) between the reconstructed and original images in the pixel space. The optimization target of the proposed neural network system is described in

$$\min L = \frac{1}{2} \sum_{i=1}^n (\text{norm}(x_i) - \text{norm}(y_i)). \quad (1)$$

where x_i, y_i is the pix of an image.

²<https://wiki.openstreetmap.org>

The pseudo-code for the training part of the core model is shown in Algorithm 1.

Algorithm 1 Train

Input: epochs, batchsize, train_dataset, model_arguments

Output: weights

```

1: for i in epochs do
2:   for train_data in dataloader(train_dataset, batchsize)
   do weights = train_model(train_data, model_arguments)
3:   end for
4:   return weights
5: end for
6: function dataloader(dataset, batchsize)
7:   iterator = "extract batchsize data from the dataset"
8:   return iterator
9: end function
10: function train_model(data, arguments)
11:   weights = "train the model and return the weights"
12:   return weights
13: end function

```

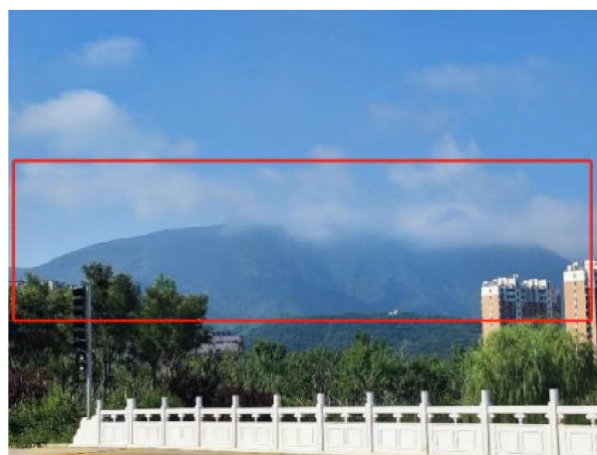
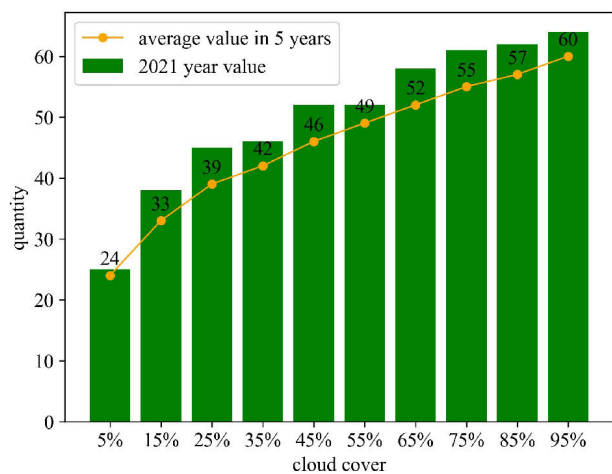
B. DATA ALIGNMENT

"Data alignment" refers here to aligning original data by cutting, splicing, matching, and other operations for input in the generation phase and subsequent processing by deep generation models (e.g., st-MAE). We conducted alignment for remote sensing data, OSM road network data, PAR data, and multi-source data matching.

Remote sensing data alignment. The interpretability of carbon neutral planning processes and results is vital to city managers [79]. NPP estimations are a fundamental basis for urban carbon neutralization. Therefore, the interpretability of NPP estimation results is crucial. Accordingly, we cut the original band data into an $H \times W$ data matrix based on the urban administrative region. We then concatted it in the third dimension in the order of 4-3-2 bands and 5-4-3 bands. The newly obtained RGB and false color $H \times W \times 3$ images were used as the original input of the generated data. The merged images were cut into several 224×224 images as remote sensing input data, yielding two types of remote sensing samples with $224 \times 224 \times 3$ in size. The remote sensing data was named by the data acquisition date to facilitate the introduction of time information during the data generation phase.

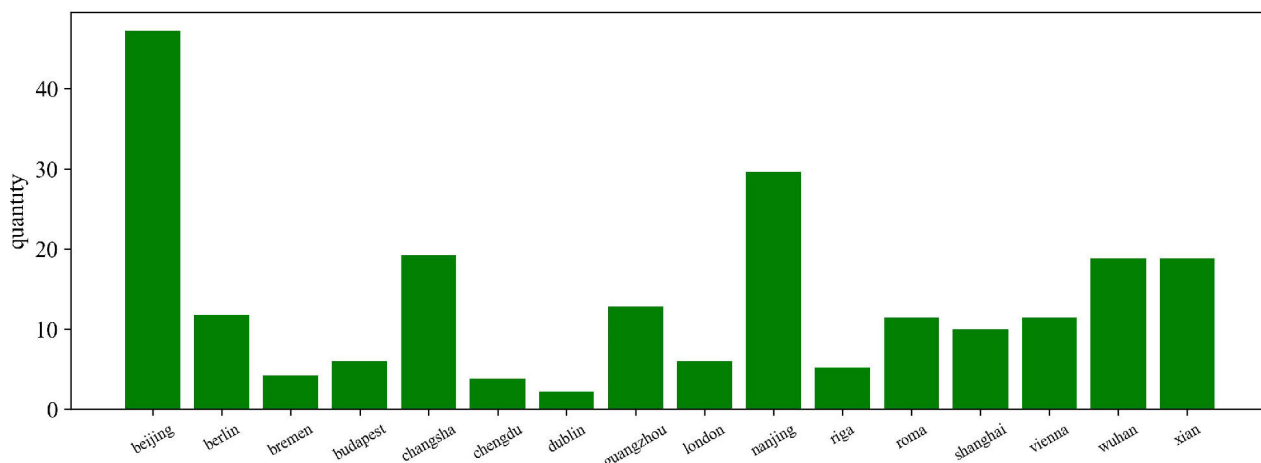
OSM road network data alignment. Similar to the remote sensing data alignment operation, the OSM road network information was first processed according to the $H \times W$ city size to obtain original data, then cut into a 224×224 input size. The order of cut OSM road network data we kept consistent with the remote sensing input data.

PAR data alignment. Due to the different resolution ratio of PAR and remote sensing data, the gathered data was first cut and aligned based on the latitude and longitude of the target $H \times W$ city size. The data were then merged according to their mean value on the time dimension and formed into an



(a)

(b)



(c)

FIGURE 1. Main limiting factors of NPP function: (a) Remote sensing image cloud cover statistics; (b) City is clear and cloudless while forest (suburb) is foggy;(c) Average annual availability of remote sensing data for each city (5 year period).

$H \times W \times t$ PAR data matrix, where the t value represents the time granularity (e.g., year, quarter, month).

Multi-source data matching. Due to the different resolution ratio of multi-source data, we matched data on the space dimension. The WGS84 coordinate system was selected to align the H and W values of the target city in various datasets to obtain spatially consistent data. The time-series information contained in CiSL-NPP allowed us to select quarterly granular PAR data and OSM road network data of the corresponding year to match the time dimension consistently. The effect of time granularity on NPP estimation was also evaluated in subsequent ablation experiments.

C. DATA GENERATION

“Data generation” refers here to generating the calculation of NPP input data at time t . The st-MAE model generates fPAR with spatio-temporal and climate information. The RNN model generates PAR data.

1) ST-MAE

st-MAE structure. As shown in Fig.3(b), the Encoder-Decoder framework of MAE was applied in this research. According to the NPP function calculation, data of 2,3,4 and 5 bands was input to the CiSL-NPP when calculating the NPP at t . Both false color and RGB images were generated with two respective st-MAE models to improve the interpretability of the CiSL-NPP method, as our goal was securing results for direct guidance in urban planning and decision-making. Additionally, we added timing embedding and an OSM road network mask to introduce the spatio-temporal information of urban vegetation into st-MAE.

Timing embedding. Four seasons were chosen as the basis for time coding so that CiSL-NPP would imply global climate change information. For each image, we assign a number according to the season in which it was taken, that is the spring, summer, autumn and winter correspond to 0, 1, 2 and 3 respectively. The numbers Patch (i.e., session) were

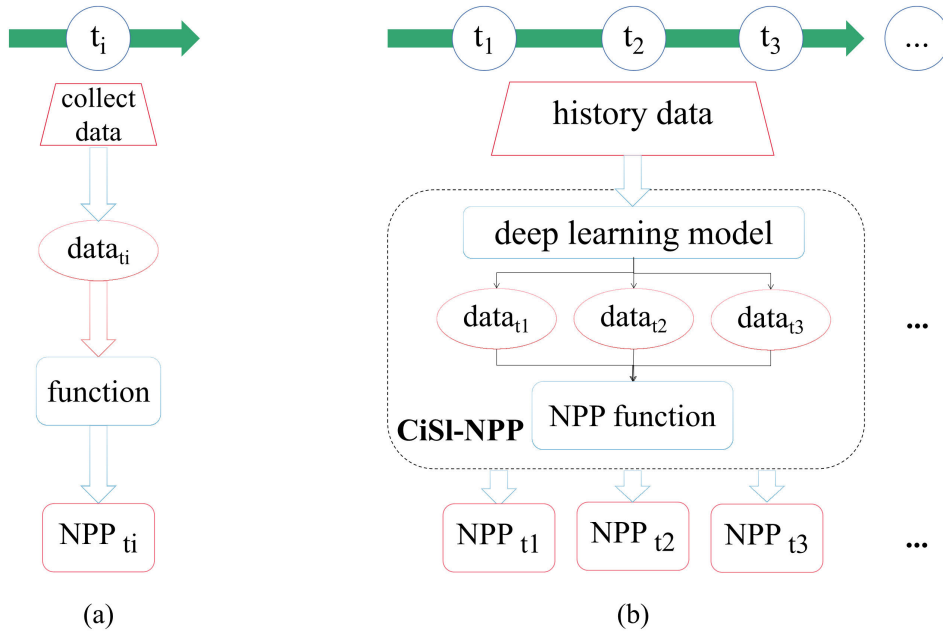


FIGURE 2. Comparison of CiSL-NPP and traditional methods: (a) Traditional NPP methods; (b) CiSL-NPP method.

projected to the same size as the position Embedding in MAE according to (2):

$$\text{Session embedding} = \text{LinearProjection} \left(\text{Patch}^{(\text{session})} \right) \quad (2)$$

where LinearProjection (·) is a learnable linear projection. Finally, this was added to each patch in st-MAE. OSM road network mask. We developed a new random masking mechanism by embedding road network information into the patch. First, the OSM road network data and image data were matched by geographic coordinates. Then, using urban geographic information from the OSM, each Patch(t) was categorized into either a vegetation area or an urban facility area according to (3); thus, we determined whether Patch(t) is an OSM token.

$$I_{\text{patch}} = \begin{cases} 1, & \Phi_i \geq \alpha \\ 0, & \Phi_i < \alpha \end{cases} \quad (3)$$

where Φ_i is the vegetation coverage of patches calculated by the OSM road network data. The decision threshold α was set to 0.4.

We also developed a new OSM road network mask mechanism. Considering the contribution of vegetation to NPP, we attempted to obtain more vegetation information through st-MAE by selecting as much urban vegetation as possible when establishing the OSM road network mask mechanism, as shown in (4).

Mask Patch

$$= \begin{cases} \text{OSM tokens} (\varepsilon), & \text{if rate (OSM tokens)} \geq \varepsilon \\ \text{OSM tokens}(1) + \text{ROP}, & \text{if rate (OSM tokens)} < \varepsilon \end{cases} \quad (4)$$

where OSM token (·) represents whether the masking OSM token is in the vegetation area or not, ROP represents “random other patches”.

Fig. 6(a) shows the distribution of urban vegetation screened from OSM. The city mask obtained from the OSM road network mask mechanism is shown in Fig. 6(b). The black pixel area in the figure represents the vegetation area and the white represents the urban facility area.

2) RNN

Due to the limited PAR data, we implemented a lightweight RNN autoregressive model to predict PAR at time t to prevent overfitting [80]. RNN outputs the data at each moment to produce a PAR prediction result of $H \times W \times t$.

3) NPP CALCULATION

CiSL-NPP was designed with the calculation paradigm of the NPP function which is calculated by (5)-(8). The PRI (Photochemical Reflectance Index) in Equation (5) is one of the indicators of the spectral reflectance of plant leaves, while APAR (Absorbed Photosynthetically Active Radiation) is the photosynthetically active radiation actually absorbed by plants. Previous studies have found that these two parameters can be used to directly calculate the NPP [27]. APAR can be calculated by equations (6) and (7), while the calculation of PRI is given in equation (8). Cloud-free images are screened and atmospheric correction is processed with satellite remote sensing data.

$$\text{NPP} = 0.5139(\text{PRI} \times \text{APAR}) - 1.9818 \quad (5)$$

$$\text{APAR} = f\text{fAR} \times \Sigma\text{PAR} \quad (6)$$

$$\text{fPAR} = (1.24 \times \text{NDVI}) - 0.168 \quad (7)$$

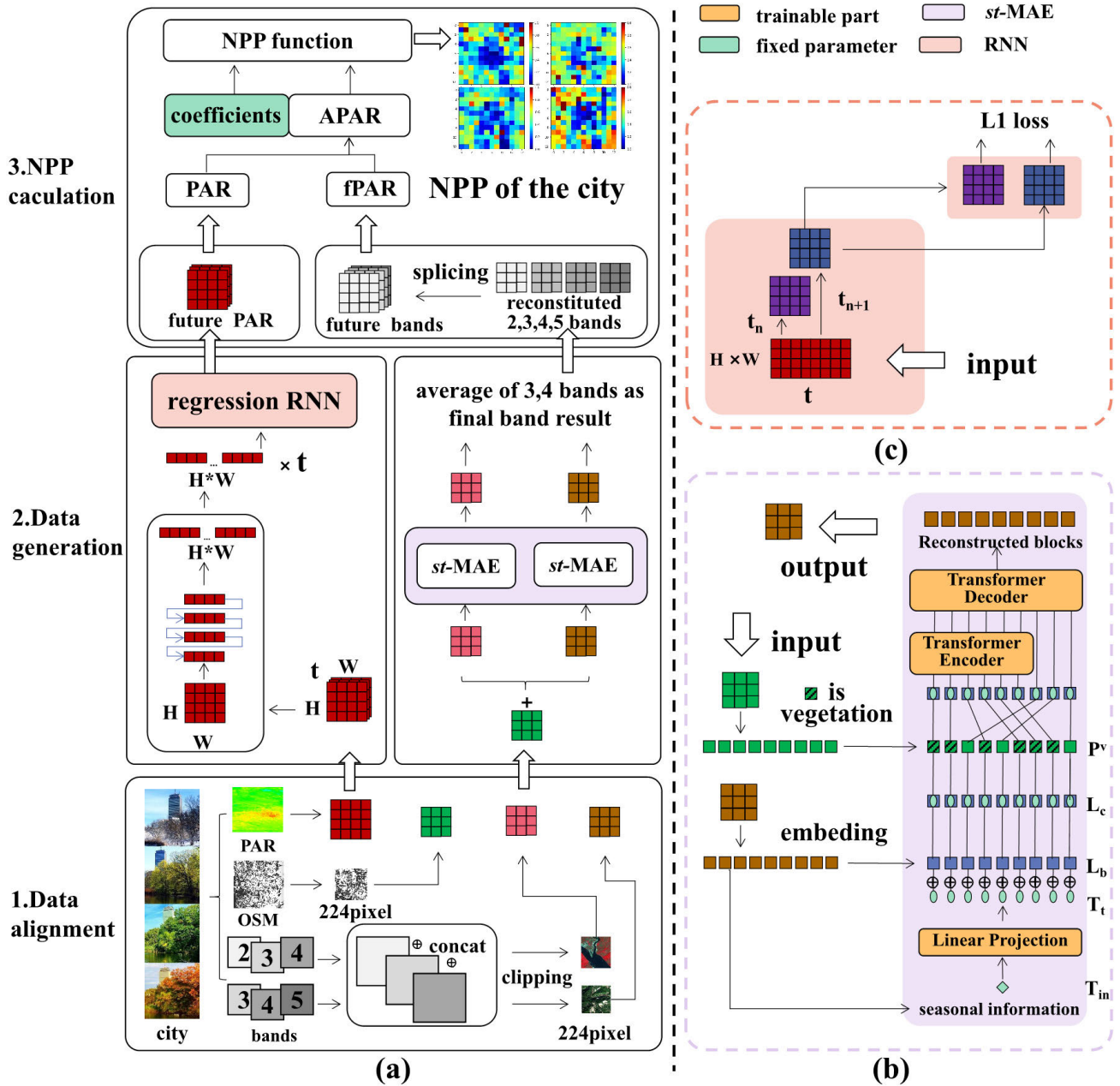


FIGURE 3. CiSL-NPP framework: (a) Overview of CiSL-NPP in bottom-up order, including data alignment, data generation, and NPP estimation; (b) RNN and st-MAE process. Note: Original remote sensing, OSM road network, and PAR data are cut, spliced, and matched as a suitable structure for deep learning models RNN and st-MAE. The RNN predicts time-series PAR data by sampling at each moment, fitting samples at the next moment, and propagating back through L1 loss. The st-MAE model generates time series data by embedding vegetation temporal features and urban topography with remote sensing and OSM road network data. Reconstructed data is input to the NPP function to calculate the city-scale NPP.

At time t , their relationship can be expressed by the following function:

$$NPP^t = f(PAR^t, NDVI^t, PRI^t) \quad (8)$$

where

$$NDVI^t = \left(\frac{NIR-Red}{NIR+Red} \right)$$

$$PRI^t = \left[0.53 \times \left(\frac{r_{band3}-r_{band2}}{r_{band3}+r_{band2}} \right) + 1 \right] / 2.$$

First, all data required for NPP calculation at time t is made available through data generation. Urban NPP estimation results can then be obtained on the time dimension via the process shown in Fig.2.

IV. EXPERIMENTS

A. DATA COLLECTION

Remote sensing data, an OSM road network, PAR, and NPP products (ground truth) were collected to calculate NPP in

this study. Since PAR data is only available until 2019, we only collect all data up to 2019. It is worth noting that non-up-to-date data will not affect the evaluation of the CiSL-NPP. Because of the existence of the generation mechanism, once the model is proven its effective, we can generate the latest data.

1) REMOTE SENSING DATA

Remote sensing data including 16-bit data of bands 2, 3, 4, and 5 of the Landsat 8 satellite OLI land imager (30 m/pixel accuracy) was obtained through the US Geological Survey (USGS) official website.³ In this website, we select the dataset of *Landsat 8-9 OLI/TIRS C2 L1* and then obtain the 2, 3, 4, and 5 bands of remote sensing images covering the study area. If the city is segmented in multiple remote sensing images, we match the images and crop it to the specified area. Finally, the unlabeled remote sensing data from 2013 to 2021 for eight large and medium-sized cities in China and eight in Europe were collected, totaling 300 GB. As shown in Table 1, the vegetation growth (i.e., false color image) and city conditions (i.e., RGB visible image) were visually observed by the 4-3-2 and 5-4-3 band combination, respectively. High-quality remote sensing data is the key to successful NPP estimation through deep learning [81], [82]. Unfortunately, our raw data contained significant amounts of cloud noise; there are few high-quality data with less than 5% clouds, as shown in Fig. 1(c). This restricts the application of the NPP function with remote sensing data at the city scale.

2) OSM ROAD NETWORK DATA

OSM road network data provides the latest and historical urban information of road and railway networks, buildings, water bodies, land use profiles, and more reflecting the complex topography and historical changes in a city [83]. OSM has historical integrity of more than 83% and has been widely applied in earth science, environmental science, and other study domains.

The OSM file records the point, line, and surface distribution information of each land type. We collected over 20 GB of OSM data from 16 cities between 2013 and 2021 for the purposes of this study. On the OSM website, we can directly download the overall ground data for the desired area and then use the code we wrote to extract the specified types for subsequent studies (for example, we extracted all the green space types to determine if each area is a vegetated area). The precise land type distribution information of the target areas was filtered for the proposed CiSL-NPP method. As an example, Fig. 4 shows the leisure land type in Beijing.

3) PAR DATA

PAR refers to the spectral component of solar radiation that is utilized for plant photosynthesis. PAR is generally acknowledged as a crucial model variable in NPP estimation [84], as it reflects the growth of plants.

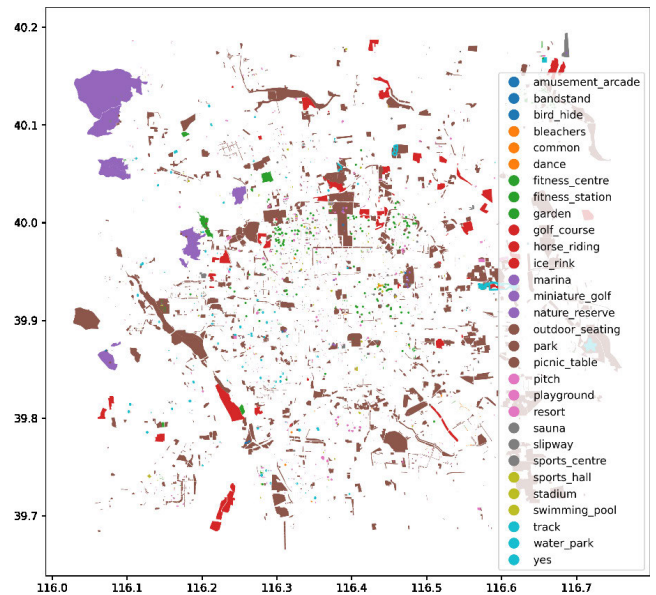


FIGURE 4. Leisure-type land of Beijing.

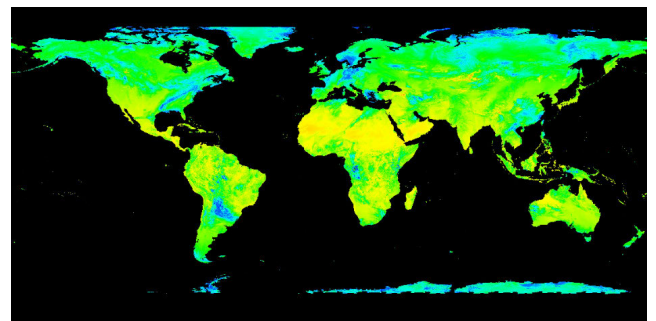


FIGURE 5. Sample of global PAR data.

The PAR global daily dataset used in this study was retrieved from Beijing Normal University⁴ based on MODIS and AVHRR (remote sensing products with full names Moderate-resolution Imaging Spectroradiometer and Advanced Very High Resolution Radiometer, respectively). We obtain PAR data of the whole year regarding different years from these products. MODIS data was estimated by a hybrid algorithm and AVHRR data was estimated by an improved look-up table algorithm. The resolution of the PAR dataset is 5 km. A total of 40 GB of PAR data between 2013 and 2019 were gathered. An example can be seen in Fig. 5.

4) NPP GROUND TRUTH


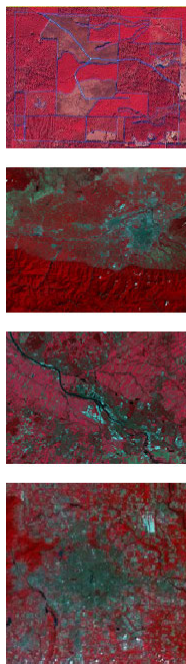

To evaluate the accuracy of the CiSL-NPP estimation method, the NPP product data⁵ which was generated based on the improved vegetation productivity estimation model is gathered as the ground truth. The NPP product is reliable and

³<https://earthexplorer.usgs.gov/>

⁴<http://www.geodata.cn/>

⁵<http://www.geodata.cn/>

TABLE 1. Description of remote sensing data.

| General information | Attribute | Row data example (Band2~5) | False color image example | RGB image example |
|---------------------|-------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| | | | | |
| Source | Landsat8 | | | |
| Years | 2013-2021 | | | |
| Size | >300GB | | | |
| Band | 2, 3, 4, 5 | | | |
| Cities | Beijing, Shanghai, Guangzhou, Nanjing, Changsha, Xi'an, Zhengzhou, Shenyang, Budapest, Berlin, London, Rome, Bremen, Dublin, Riga, Vienna |  |  |  |

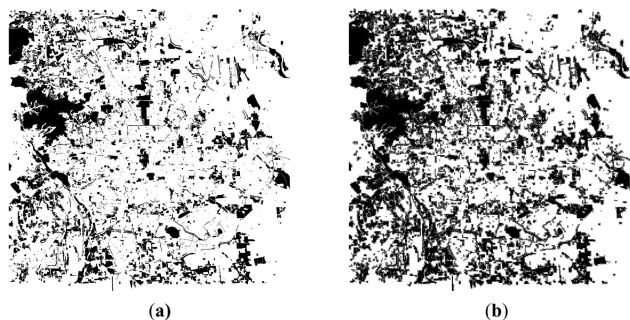


FIGURE 6. Examples of OSM road network mask mechanism.

accurate so it has been widely used [85], [86], [87]. The NPP product is projected in geographic latitude/longitude, with a spatial accuracy of 5km and a temporal resolution of 8 days. We collected about 3GB of PAR data between 2013 and 2018. Notably, the temporal and spatial accuracy of remote sensing data, road network data and PAR data are different from that of NPP production. Therefore, data alignment is required in estimation and evaluation.

B. EXPERIMENT SETTING

We conducted three experiments to validate CiSL-NPP: Quantitative analysis, qualitative analysis, and case study.

In the quantitative analysis process, we ran an st-MAE ablation experiment and RNN regression accuracy experiment to verify the accuracy of CiSL-NPP data and NPP estimation. In the qualitative analysis process, the cost of the CiSL-NPP method was analyzed in terms of the economy of such a large-scale model as well as its carbon emissions. In the case study, we tested the effectiveness of CiSL-NPP on examples of Beijing and Budapest from both qualitative and quantitative perspectives. We visualized an optimal model and part of the reconstructed data, then calculated long-term NPP changes in cities using CiSL-NPP and compared the results against those of traditional machine learning.

C. QUANTITATIVE ANALYSIS

1) ST-MAE ABLATION EXPERIMENT

Training settings. The batch size of st-MAE was set to 32, the learning rate was initialized to 1e-3, and the AdamW optimizer [88] was applied. The MAE pre-training model was used to perform transfer learning on the data sets of 16 cities and 1,000 epochs were trained after freezing some encoders.

Ablation experiments. Different ablation experiments were designed to compare the effects of the st-MAE model. First, to account for the use of transfer learning, the reconstruction effects of MAE-base and MAE-large of different scales

of MAE pre-training models were compared. We did not consider MAE-huge for its excessive cost and high carbon emissions. Next, frozen encoders with different layers were compared to observe the function of transfer learning. We also compared different linear projection timing-coding layers of 2, 4, 8 and 12 are compared. The reconstruction error MSE of st-MAE before and after adding timing coding was compared as well. We then compared reconstruction errors of the different mask rates 0.55, 0.65, 0.75, 0.85, and 0.95. Finally, we compared the OSM road network mask versus random mask mechanisms.

In the quantitative analysis section, our experiments are done on the validation set. Our validation set is set to the last year of data from the training set, which has no intersection with the training set.

2) RNN REGRESSION EXPERIMENT

Model structure. We adopted an RNN structure model with one layer of hidden nodes. The MLP layer uses the ReLu activation function and the loss function is L1 [89].

Training. The batch size during RNN training was set to 6, the step size to 1, the learning rate initially to 0.1, and the learning rate as halved per 100 epochs. The random initialization weight and tanh activation function were selected. A total of 1,000 epochs of RNN were trained.

PAR time granularity. To mitigate the shortcoming of RNN’s information loss over long time series, PAR data were sampled at different granularities of 15 days, 30 days (one month), 60 days (2 months), and 90 days (one quarter) for comparison.

D. QUALITATIVE ANALYSIS

1) ECONOMY

The data required by CiSL-NPP is not charged, so we analyzed its economy mainly by the cost of model training. The deep learning models used in CiSL-NPP are large scale and consume a large amount of computing power during training. The economy was calculated with rented cloud GPU computing power.

2) CARBON EMISSIONS

The high carbon emissions of large-scale deep learning models have received a great deal of research attention [90], [91]. The carbon emissions of CiSL-NPP were analyzed according to its large-scale depth models. We estimated CiSL-NPP’s carbon emissions by calculating the GFLOPs required to operate it. We calculated the conversion coefficients of GFLOPs and CO2 equivalent emissions (CO2EE) of computational models as per the mean value as follows [28]:

$$1GFLOPS = 5.22 \times 10^{-9}CO2EE \quad (9)$$

E. CASE STUDY

Visualizing st-MAE reconstruction of remote sensing images. According our quantitative experiments, the optimal visual reconstruction results of the st-MAE model were selected to

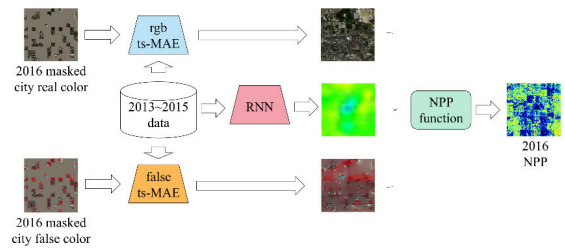


FIGURE 7. Framework of proposed long-term NPP forecasting method.

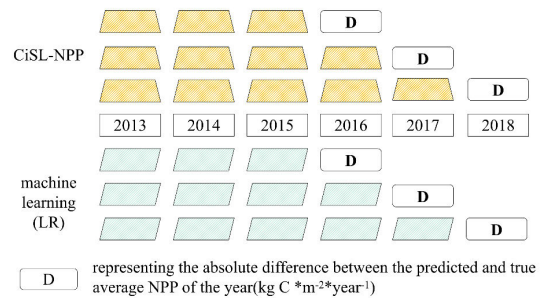


FIGURE 8. Comparison of CiSL-NPP with machine learning method.

directly verify the effect of CiSL-NPP. Two capital cities, Beijing and Budapest, were exemplified to visualize the RGB and false color images reconstructed by remote sensing images in spring, summer, autumn, and winter of 2018.

1) VISUALIZATION RESULTS

A space adjacent to vegetation areas and buildings in each city was randomly selected to directly visualize the reconstruction effect before and after adding the OSM road network mask. The NPP estimation results were observed in the form of a heat map.

2) CITY NPP EVOLUTION

To illustrate the effect of long-term NPP estimation, we divided the collected data into a training set and test set incrementally to, respectively, train and test CiSL-NPP. For example, the data from 2013 to 2015 were used to train the model, then the trained model was operated to estimate NPP from 2016 to 2018. The framework we used is shown in Fig. 7, where BJ4 and BD4 denote the models trained with data from 2013 to 2015 in Beijing and Budapest, respectively, and so on.

City NPP generation evolution. By searching for areas characterized by vegetation change during urban construction, we observed the evolution of synchronously generated NPP to determine the effectiveness of CiSL-NPP. Four specific areas that changed from green spaces to buildings were selected in Beijing and Budapest at random.

3) COMPARISON WITH MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

The advantages of CiSL-NPP were verified through comparing it to linear regression [92], [93], deep belief network (DBN) [94] and MultiLayer Perceptron (MLP) [95].

As shown in Fig. 8, the NPP of the following year was predicted by training different scale data sets of 3, 4, 5, and 6 years.

V. RESULTS AND DISCUSSION

A. QUANTITATIVE ANALYSIS

1) ST-MAE ABLATION RESULTS

Model size. As shown in Fig.9(a), st-MAE shows the best effect (0.222) when operated with MAE-large model parameters. Nevertheless, the reconstruction error MSE with the MAE-base model parameters only increased by 0.01 compared with the former one. We recommend st-MAE with the MAE-based model as per a workable trade-off between accuracy and economy; we used these techniques for subsequent experiments.

Encoder frozen layers. In the pre-trained model, the encoder layer has learned a large number of image features, and we consider directly adopting the features it has learned by freezing all encoder layers and gradually reducing the number of frozen layers to explore their impact. The baseline of the MAE-base encoder in this research is 12 layers. Fig.9(b) shows the MSE results of different cities, where excessive encoder layers decreased the effect. However, the effect was significantly improved by 8% when reducing 12 freezing layers to 8 layers. The minimum MSE was 0.219 with four freezing layers. Though CiSL-NPP requires retraining on some datasets, we find that it is unnecessary to create such a computational burden.

Timing coding projection layers. Since temporal encoding is very important for timing tasks, we paid much attention to the information transfer of temporal encoding and experimented with different effects of projection layers from few to many. As shown in Fig.9(c) and Table 2, the minimum reconstruction error MSE is 0.212 when there are four layers of temporal coding. The effect was improved when there were fewer layers, which highlights the low cost of st-MAE. The optimal settings for timing coding projection decreased the MSE ($0.212 < 0.222$, as shown in Fig.9(c)) suggesting that timing coding is effective for NPP estimation. Temporal vs. non-temporal embedding. As shown in Fig.9(d), the st-MAE results show a 0.98% increase after adding temporal information. We also compared the results before and after adding temporal information in different cities. Fig.9(d) shows where the effect with temporal coding was improved in most cities, at a maximum increase of 1.89%. The effect of temporal embedding differs in different cities as well, possibly due to seasonal information. Different cities have different dominant vegetation, so growth patterns may vary by season. Additionally, climate information was embedded in the CiSL-NPP model by temporal coding.

Mask rate. Since we added some coding, we need to consider whether a mask rate of 0.75 is still the best choice for the model, and we tested the effect of increasing or decreasing the mask rate on the results, using 0.75 as a starting point. Fig.9(e) shows that reconstruction works best with a

TABLE 2. Comparison of CiSL-NPP and traditional methods.

| Architecture | Ablation variable | Value | Mean MSE | Improvement | |
|--------------|-----------------------------|------------------|----------|-------------|-------|
| st-MAE-base | Number of frozen layers | 12 | 0.251 | - | |
| | | 8 | 0.232 | - | |
| | | 4 | 0.219 | - | |
| | | 2 | 0.222 | - | |
| | Number of projection layers | 12 | 0.218 | - | |
| | | 8 | 0.222 | - | |
| | | 4 | 0.212 | - | |
| | | 2 | 0.213 | - | |
| | Masking mechanism | Random | 0.229 | - | - |
| | | OSM road network | 0.222 | 0.222 | 3.20% |

TABLE 3. Results of RNN regression.

| Structure | Time granularity | PAR Average Value | L1 loss | Loss ratio | Improvement |
|-----------|------------------|-------------------|---------|------------|-------------|
| RNN | 15 | 6721.1 | 1015.3 | 15.10% | - |
| | 30 | 6972.8 | 953.5 | 13.60% | 9.90% |
| | 60 | 7110.8 | 317.5 | 4.50% | 70.20% |
| | 90 | 7175.5 | 231.8 | 3.20% | 78.81% |

0.75 mask rate; the effect is worse when increasing the rate beyond this point and is not improved by further reducing the rate. We recommend a 0.75 mask rate.

OSM road network vs random mask. Compared with the random mask, the st-MAE reconstruction error was reduced by 3.2% using the OSM road network mask (Fig.9(f)). We also find that the road network mask performs best in rapidly developing cities. For example, the OSM road network mask showed improvements of 11.75% and 6.88% to new first-tier cities Changsha and Nanjing, respectively.

2) RNN REGRESSION RESULTS

Our comparative results are shown in Table 3. When the time granularity was increased from 15 days to 30 days, 60 days, or 90 days, the loss decreased by 9.9%, 70.2%, and 78.81%, respectively. We find that the effect is best when the time granularity is one season.

B. QUALITATIVE ANALYSIS

Economic cost analysis. RMB is the most commonly used currency in China and we use it to compare economic expenses. We collected the majority of the popular cloud computing power rental platforms in China. The most cost-effective among them is 1.58 RMB per hour, which we selected for our purposes here. Based on the calculation time required for different models, we evaluated the economy of the platform in a step-wise process. First, as each epoch in the ts-MAE model costs about 6 min, the 1,000 epochs cost $1000 \times 6 / 60 \times 1.58 = 158$ RMB and reconstruction of RGB and false color model cost $2 \times 158 = 316$ RMB. Second, the RNN model takes about 2 min and the cost is negligible. Compared to the cost of transportation, equipment, and personnel as-calculated in a field study, CiSL-NPP shows significant economic advantages.

Model carbon emission analysis. CiSL-NPP contains a light RNN, which requires less calculation compared to st-MAE. Therefore, the RNN's carbon emissions are negligible. Table 4 shows a comparison of st-MAE with the classic VGG16 and vit-base models for carbon emissions. The CO2 emission of st-MAE-base is 39.2% lower than VGG16,

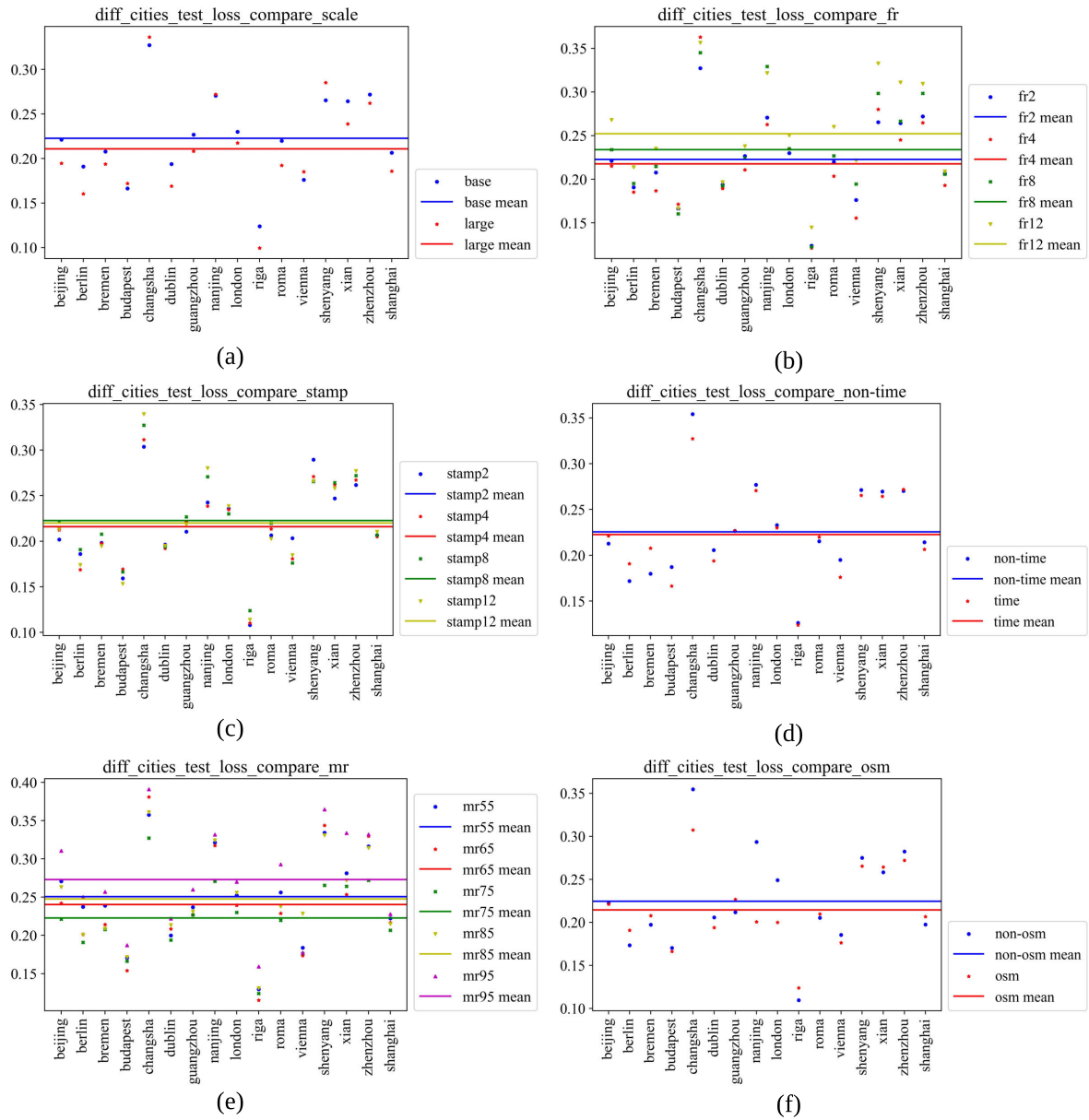


FIGURE 9. Ablation results: (a) Different scales of st-MAE; (b) Different numbers of frozen layers in encoder; (c) Different temporal encoding layers; (d) Temporal vs. non-temporal ablation results; (e) Effect of different mask rates; (f) OSM road network mask and random mask.

TABLE 4. Carbon emissions estimation.

| Model | Input Resolution | Params(M) | GFLOPs | CO2 equivalent emissions |
|--------------|------------------|-----------|--------|--------------------------|
| st-MAE-base | 224x224 | 26.87 | 18.84 | 20.724 |
| st-MAE-large | 224x224 | 27.46 | 40.64 | 44.704 |
| vgg16 | 224x224 | 138.36 | 31 | 34.1 |
| Vit-base | 224x224 | 86.57 | 33.72 | 37.092 |

suggesting that CiSL-NPP is environmentally friendly. However, the st-MAE-large increases by 115.7% compared to the st-MAE-base and is 31.1% higher than that of VGG16. With similar structures and the same scale parameter settings, st-MAE-base decreases 44.5% of CO2 emissions while st-MAE-large increases them by 20.5%. Hence, we did not

continue to use the st-MAE-large model in this study. The results suggest that CiSL-NPP based on st-MAE-base does not introduce excessive carbon emissions at a certain level of precision and in fact emits less carbon than other deep learning models.

Time overhead. In subsection III-A, we give the pseudo-code of the algorithm, where the core matrix operation process is in the Self-Attention part of the train_model method. The operation formula of the Self-Attention part is shown in Equation (10).

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (10)$$

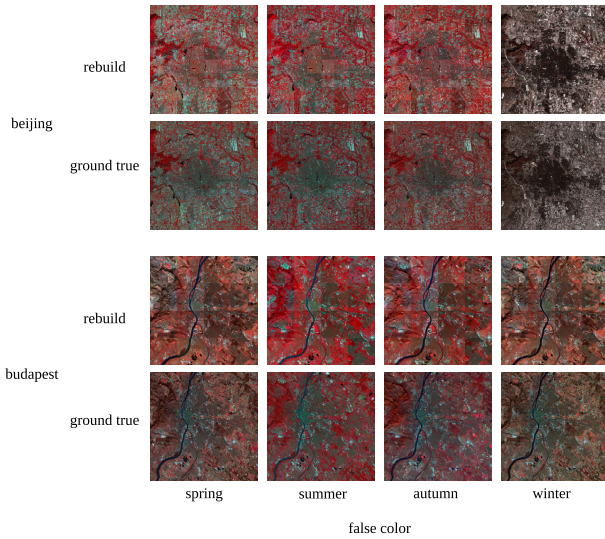


FIGURE 10. Comparison of remote sensing reconstruction effects in spring, summer, autumn, and winter in Beijing and Budapest.

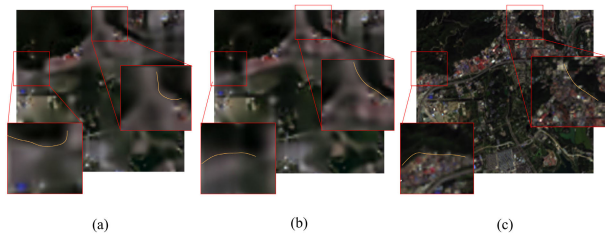


FIGURE 11. Visualization of the road network mask effect: (a) Without OSM road network mask; (b) With OSM road network mask; (c) Ground truth.

Among them, the similarity calculation QK^T means $n \times d$ with $d \times n$ operation to get $n \times n$ matrix with complexity $O(n^2d)$, and softmax is calculated for each row with complexity $O(n^2)$ for n rows, so they get $O(n^2d)$ after weighted sum. In the economic overhead above we mentioned that the main time overhead lies in the training of the st-MAE, i.e., 100 hours, plus the time for data processing (in 2 hours), for a total of $102/24 = 4.2$ days. However, the general traditional approach requires travel to the field to collect data, with significant time spent on commuting and equipment deployment (in Budapest, for example, we need up to two days of time overhead just for the round trip from Beijing, which grows exponentially as the study targets become more numerous, in addition to the cost of transportation involved).

C. CASE ANALYSIS

1) ANALYSIS OF VISUALIZATION RESULTS

Reconstruction results of st-MAE remote sensing image. Fig.10 directly illustrates that the reconstruction results of false color are very close to the ground truth. And it indicates that the st-MAE model can reconstruct reliable remote sensing images.

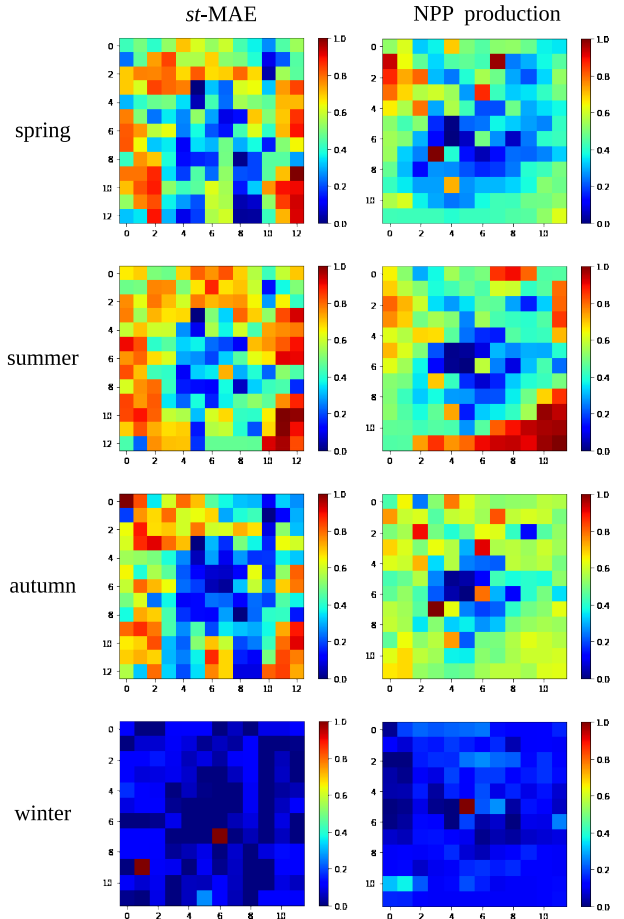


FIGURE 12. Comparison of reconstruction results with NPP production visualization.

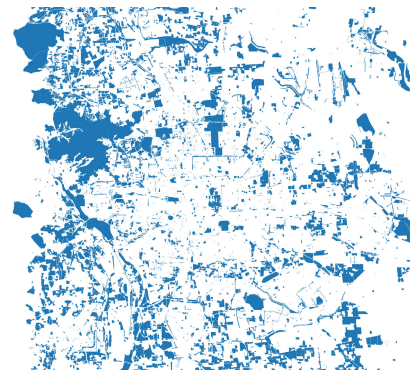


FIGURE 13. Visualization of vegetation distribution from OSM in Beijing.

OSM road network mask effect. As shown in Fig. 11(a), the model reconstruction results lost the edge shape (Fig.11(c)) at the junction of vegetation and non-vegetation areas without the OSM road network information. As shown in Fig. 11(b), The real edge shape was retained after adding the road network information, which implicitly improved the reconstruction effect of the st-MAE model.

NPP reconstruction results. As shown in Fig.12, the results of CiSL-NPP and NPP production have similar overall dis-

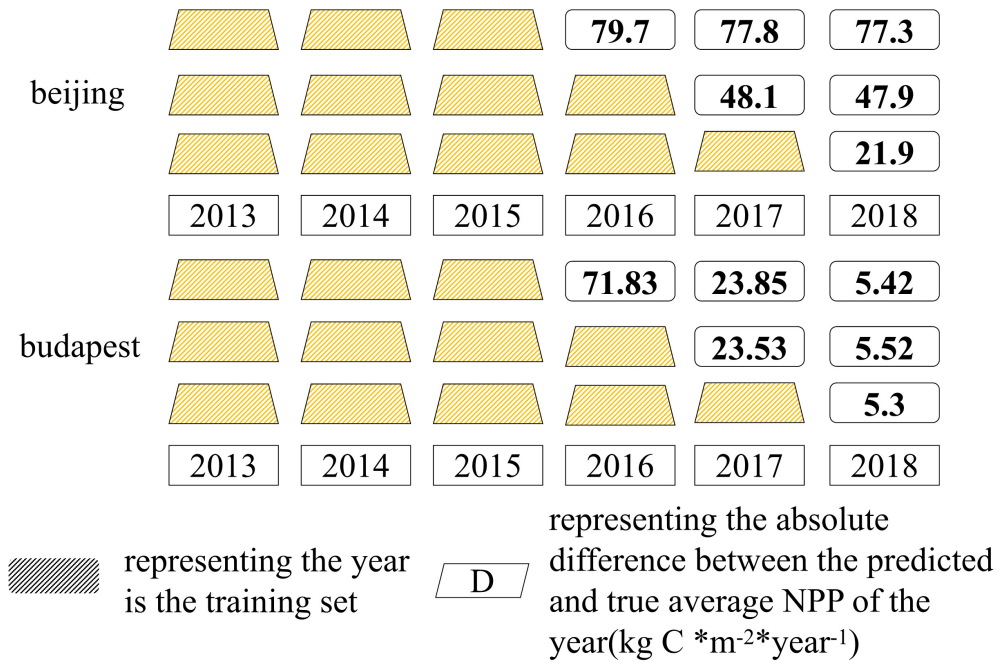


FIGURE 14. CiSL-NPP results of long-term simulation.

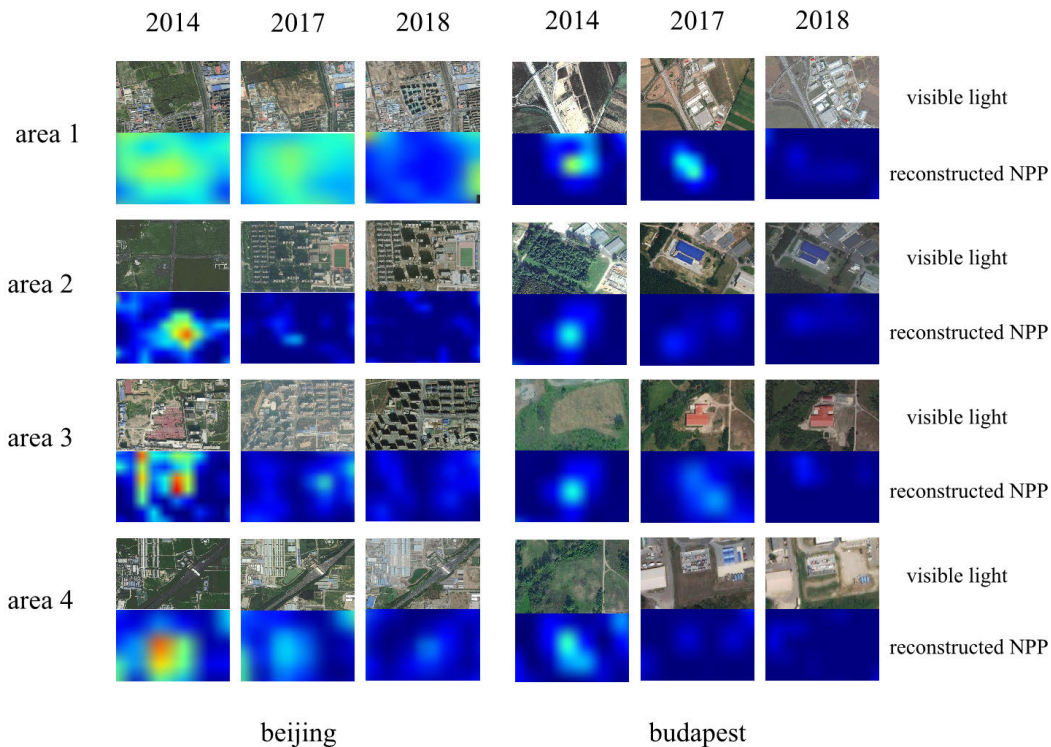


FIGURE 15. CiSL-NPP long-term simulation results on four different areas in Beijing and Budapest selected at random. Note that Upper parts show visible light representative of building construction. Lower parts show how CiSL-NPP simulates urban NPP evolution through time series and OSM road network masks.

tributions with increasing NPP values from the center to the surrounding areas – there are more plants in the center of cities than their outskirts. Fig.13 shows the visualization results of the OSM vegetation distribution. With a scattered

vegetation distribution, the CiSL-NPP reconstruction results are spatially more well-refined compared to the NPP production. Moreover, the results of CiSL-NPP vary by season. This suggests that time embedding works very well.

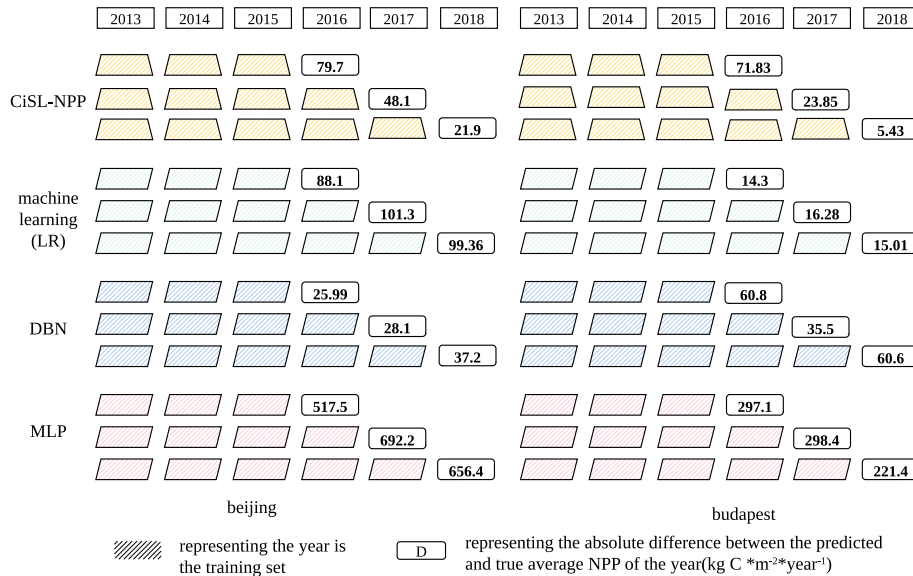


FIGURE 16. Comparison of CiSL-NPP with machine learning methods.

Taken together, our results demonstrate the strong interpretability of CiSL-NPP and its practical applicability in actual urban planning and decision-making processes.

2) ANALYSIS OF NPP EVOLUTION RESULTS

Long-term NPP prediction results. Fig. 14 shows the absolute NPP error in 2018. We find that the BJ4 model and the BJ5 model decreased by 38.03% and 71.67% over the BJ3 model, respectively. The BD5 model was 2% lower than the BD3 model, whereas the BD4 model was instead 1.85% higher than the BD3 model. This suggests that only a certain amount of data is required to allow the model to learn seasonal change information through temporal coding.

NPP generation evolution results. As shown in Fig. 15, in areas with significant changes in urban construction (e.g., transformation from green spaces to buildings), the estimations of CiSL-NPP directly reflect vegetation changes. The results also reflect the carbon sink reduction caused by the reduction of green space during urban construction.

3) COMPARISON WITH MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

As shown in Fig. 16, in the case of Beijing and Budapest in 2018, the absolute error of CiSL-NPP was reduced by 77.96%, 63.82%, 41.12% and 289.72% compared with simple machine learning models, DBN and MLP respectively. Overall, CiSL-NPP shows higher accuracy. As mentioned in Section IV-A, the proposed CiSL-NPP method has a small error over the three-year period which suggesting that it could work over a relatively long period of time as its generation mechanism.

VI. CONCLUSION

Effective city-scale NPP estimation, as a pathway towards carbon neutrality, is a highly challenging process. In this

study, to solve this problem, a spatio-temporal neural network learning system named CiSL-NPP was proposed. By fusing multi-source data including remote sensing, PAR, and road networks, the st-MAE with time embedding and OSM road network mask was used in the framework of CiSL-NPP. The detailed experiments such as ablation experiments, qualitative experiments and case study were carry out explaining the effectiveness of CiSL-NPP. Our method has a higher estimation efficiency, more effectiveness, lower carbon consumption and cost than other NPP estimation methods. The CiSL-NPP efficiently achieves long-term, city-scale NPP estimation with high spatial resolution. Compared with traditional estimation methods, by introducing the neural networks learning system, the proposed method shows higher accuracy in the city-scale NPP estimation field, and we can foresee that it will perform better when applied in the real scenes. CiSL-NPP also shows strong interpretability, giving it remarkable potential for practical city-planning applications. However, the NPP calculation does not use different parameters according to different cities, and in the future we plan to adjust the parameters to be adaptive in nature and extend this neural network system to more cities.

REFERENCES

- [1] H. Youn, L. M. A. Bettencourt, J. Lobo, D. Strumsky, H. Samaniego, and G. B. West, "Scaling and universality in urban economic diversification," *J. Roy. Soc. Interface*, vol. 13, no. 114, Jan. 2016, Art. no. 20150937.
- [2] K. Liu, Z. Ni, M. Ren, and X. Zhang, "Spatial differences and influential factors of urban carbon emissions in China under the target of carbon neutrality," *Int. J. Environ. Res. Public Health*, vol. 19, no. 11, p. 6427, May 2022.
- [3] W. Sha, Y. Chen, J. Wu, and Z. Wang, "Will polycentric cities cause more CO₂ emissions? A case study of 232 Chinese cities," *J. Environ. Sci.*, vol. 96, pp. 33–43, Oct. 2020.
- [4] A. Huovila, H. Siikavirta, C. A. Rozado, J. Rökman, P. Tuominen, S. Paiho, Å. Hedman, and P. Ylén, "Carbon-neutral cities: Critical review of theory and practice," *J. Cleaner Prod.*, vol. 341, Mar. 2022, Art. no. 130912.

- [5] R. Kondo, Y. Kinoshita, and T. Yamada, "Green procurement decisions with carbon leakage by global suppliers and order quantities under different carbon tax," *Sustainability*, vol. 11, no. 13, p. 3710, Jul. 2019.
- [6] S. Yu, X. Yuan, X. Yao, and M. Lei, "Carbon leakage and low-carbon performance: Heterogeneity of responsibility perspectives," *Energy Policy*, vol. 165, Jun. 2022, Art. no. 112958.
- [7] W. Y. Chen, "The role of urban green infrastructure in offsetting carbon emissions in 35 major Chinese cities: A nationwide estimate," *Cities*, vol. 44, pp. 112–120, Apr. 2015.
- [8] D. J. Nowak and D. E. Crane, "Carbon storage and sequestration by urban trees in the USA," *Environ. Pollut.*, vol. 116, no. 3, pp. 381–389, Mar. 2002.
- [9] P. F. Chen, "Monthly NPP dataset covering China's terrestrial ecosystems at north of 18°N (1985–2015)," *J. Global Change Data Discovery*, vol. 3, no. 1, pp. 34–41, 2019.
- [10] X. Guan, H. Shen, W. Gan, G. Yang, L. Wang, X. Li, and L. Zhang, "A 33-year NPP monitoring study in Southwest China by the fusion of multi-source remote sensing and station data," *Remote Sens.*, vol. 9, no. 10, p. 1082, Oct. 2017.
- [11] T. N. Kaye, K. L. Pendergrass, K. Finley, and J. B. Kauffman, "The effect of fire on the population viability of an endangered prairie plant," *Ecol. Appl.*, vol. 11, no. 5, pp. 1366–1380, Oct. 2001.
- [12] H. Lin, H. Ma, C. Nyandwi, Q. Feng, and T. Liang, "A new net primary productivity model and new management strategy of grassland classification based on CSCS in China," *Rangeland J.*, vol. 43, no. 6, pp. 285–295, Apr. 2021.
- [13] W. Ge, L. Deng, F. Wang, and J. Han, "Quantifying the contributions of human activities and climate change to vegetation net primary productivity dynamics in China from 2001 to 2016," *Sci. Total Environ.*, vol. 773, Jun. 2021, Art. no. 145648.
- [14] B. Guo, M. Lu, Y. Fan, H. Wu, Y. Yang, and C. Wang, "A novel remote sensing monitoring index of salinization based on three-dimensional feature space model and its application in the yellow river delta of China," *Geomatics, Natural Hazards Risk*, vol. 14, no. 1, pp. 95–116, Dec. 2023.
- [15] Y. Kai, T. Yichen, Y. Chao, Z. Feifei, Y. Quanzhi, and H. Lizhong, "NPP spatial and temporal pattern of vegetation in Beijing and its factor explanation based on CASA model," *Remote Sens. Natural Resour.*, vol. 27, no. 1, pp. 133–139, 2014.
- [16] E. A. Mohareb, H. L. MacLean, and C. A. Kennedy, "Greenhouse gas emissions from waste management—Assessment of quantification methods," *J. Air Waste Manag. Assoc.*, vol. 61, no. 5, pp. 480–493, 2011.
- [17] B. Crawford, A. Christen, and I. McKendry, "Diurnal course of carbon dioxide mixing ratios in the urban boundary layer in response to surface emissions," *J. Appl. Meteorol. Climatol.*, vol. 55, no. 3, pp. 507–529, Mar. 2016.
- [18] N. F. Sonti, P. M. Groffman, D. J. Nowak, J. G. Henning, M. L. Avolio, and E. J. Rosi, "Urban net primary production: Concepts, field methods, and Baltimore, Maryland, USA case study," *Ecol. Appl.*, vol. 32, no. 4, p. e2562, Jun. 2022.
- [19] X. Wei, J. Yang, P. Luo, L. Lin, K. Lin, and J. Guan, "Assessment of the variation and influencing factors of vegetation NPP and carbon sink capacity under different natural conditions," *Ecol. Indicators*, vol. 138, May 2022, Art. no. 108834.
- [20] S. Grimmond, "Urbanization and global environmental change: Local effects of urban warming," *Geograph. J.*, vol. 173, no. 1, pp. 83–88, Mar. 2007.
- [21] S. Ghimire, R. C. Deo, N. J. Downs, and N. Raj, "Global solar radiation prediction by ANN integrated with European centre for medium range weather forecast fields in solar rich cities of Queensland Australia," *J. Cleaner Prod.*, vol. 216, pp. 288–310, Apr. 2019.
- [22] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610–2623, 2010.
- [23] Y. Yan, C. Wu, and Y. Wen, "Determining the impacts of climate change and urban expansion on net primary productivity using the spatio-temporal fusion of remote sensing data," *Ecol. Indicators*, vol. 127, Aug. 2021, Art. no. 107737.
- [24] L. Jin, S. Liang, X. Luo, and M. Zhou, "Distributed and time-delayed-winner-take-all network for competitive coordination of multiple robots," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 641–652, Jan. 2023.
- [25] L. Jin, Y. Liufu, H. Lu, and Z. Zhang, "Saturation-allowed neural dynamics applied to perturbed time-dependent system of linear equations and robots," *IEEE Trans. Ind. Electron.*, vol. 68, no. 10, pp. 9844–9854, Oct. 2021.
- [26] K. Wang, A. Bastos, P. Ciais, X. Wang, C. Rödenbeck, P. Gentile, F. Chevallier, V. W. Humphrey, C. Huntingford, M. O'Sullivan, S. I. Seneviratne, S. Storch, and S. Piao, "Regional and seasonal partitioning of water and temperature controls on global land carbon uptake variability," *Nature Commun.*, vol. 13, no. 1, p. 3469, Jun. 2022.
- [27] M. Mngadi, J. Odindi, O. Mutanga, and M. Sibanda, "Estimating above-ground net primary productivity of reforested trees in an urban landscape using biophysical variables and remotely sensed data," *Sci. Total Environ.*, vol. 802, Jan. 2022, Art. no. 149958.
- [28] D. Patterson, J. Gonzalez, Q. Le, C. Liang, L.-M. Munguia, D. Rothchild, D. So, M. Texier, and J. Dean, "Carbon emissions and large neural network training," 2021, *arXiv:2104.10350*.
- [29] Y. Yu and N. Zhang, "Low-carbon city pilot and carbon emission efficiency: Quasi-experimental evidence from China," *Energy Econ.*, vol. 96, Apr. 2021, Art. no. 105125.
- [30] X. Du, L. Shen, Y. Ren, and C. Meng, "A dimensional perspective-based analysis on the practice of low carbon city in China," *Environ. Impact Assessment Rev.*, vol. 95, Jul. 2022, Art. no. 106768.
- [31] Z. Liu, Z. Deng, G. He, H. Wang, X. Zhang, J. Lin, Y. Qi, and X. Liang, "Challenges and opportunities for carbon neutrality in China," *Nature Rev. Earth Environ.*, vol. 3, no. 2, pp. 141–155, Dec. 2021.
- [32] S. Azarkamand, G. Ferré, and R. M. Darbra, "Calculating the carbon footprint in ports by using a standardized tool," *Sci. Total Environ.*, vol. 734, Sep. 2020, Art. no. 139407.
- [33] T. Liu, Q. Wang, and B. Su, "A review of carbon labeling: Standards, implementation, and impact," *Renew. Sustain. Energy Rev.*, vol. 53, pp. 68–79, Jan. 2016.
- [34] F. Nevens and C. Roorda, "A climate of change: A transition approach for climate neutrality in the city of Ghent (Belgium)," *Sustain. Cities Soc.*, vol. 10, pp. 112–121, Feb. 2014.
- [35] H. Vandevyvere and F. Nevens, "Lost in transition or geared for the S-curve? An analysis of Flemish transition trajectories with a focus on energy use and buildings," *Sustainability*, vol. 7, no. 3, pp. 2415–2436, Feb. 2015.
- [36] S. Jetoo, "Stakeholder engagement for inclusive climate governance: The case of the city of Turku," *Sustainability*, vol. 11, no. 21, p. 6080, Nov. 2019.
- [37] L. Qin, M. Y. Malik, K. Latif, Z. Khan, A. W. Siddiqui, and S. Ali, "The salience of carbon leakage for climate action planning: Evidence from the next eleven countries," *Sustain. Prod. Consumption*, vol. 27, pp. 1064–1076, Jul. 2021.
- [38] F. Ferrini, A. Fini, J. Mori, and A. Gori, "Role of vegetation as a mitigating factor in the urban context," *Sustainability*, vol. 12, no. 10, p. 4247, May 2020.
- [39] L. Azaria, "Carbon sequestration capability analysis of urban green space using geospatial data," in *Proc. ES Web Conf.*, vol. 73, 2018, Art. no. 03009.
- [40] Y. Gong and X. Luo, "Experimental study on the carbon sequestration benefit in urban residential green space based on urban ecological carrying capacity," *Sustainability*, vol. 14, no. 13, p. 7780, Jun. 2022.
- [41] X. Liu, S. Wang, P. Wu, K. Feng, K. Hubacek, X. Li, and L. Sun, "Impacts of urban expansion on terrestrial carbon storage in China," *Environ. Sci. Technol.*, vol. 53, no. 12, pp. 6834–6844, Jun. 2019.
- [42] A. Zhang and R. Deng, "Spatial-temporal evolution and influencing factors of net carbon sink efficiency in Chinese cities under the background of carbon neutrality," *J. Cleaner Prod.*, vol. 365, Sep. 2022, Art. no. 132547.
- [43] H. Wang, G. Liu, Z. Li, P. Wang, and Z. Wang, "Assessing the driving forces in vegetation dynamics using net primary productivity as the indicator: A case study in Jinghe River basin in the Loess Plateau," *Forests*, vol. 9, no. 7, p. 374, Jun. 2018.
- [44] X. Tang, X. Zhao, Y. Bai, Z. Tang, W. Wang, Y. Zhao, H. Wan, Z. Xie, X. Shi, B. Wu, and G. Wang, "Carbon pools in China's terrestrial ecosystems: New estimates based on an intensive field survey," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 16, pp. 4021–4026, 2018.
- [45] X. Liu, F. Pei, Y. Wen, X. Li, S. Wang, C. Wu, Y. Cai, J. Wu, J. Chen, K. Feng, and J. Liu, "Global urban expansion offsets climate-driven increases in terrestrial net primary productivity," *Nature Commun.*, vol. 10, p. 5558, Dec. 2019.

- [46] C. Rebmann, M. Aubinet, H. Schmid, N. Arriga, M. Aurela, G. Burba, R. Clement, A. De Ligne, G. Fratini, B. Gielen, and J. Grace, "ICOS eddy covariance flux-station site setup: A review," *Int. Agrophys.*, vol. 32, no. 4, pp. 471–494, Dec. 2018.
- [47] B. Lee, N. Kim, E.-S. Kim, K. Jang, M. Kang, J.-H. Lim, J. Cho, and Y. Lee, "An artificial intelligence approach to predict gross primary productivity in the forests of South Korea using satellite remote sensing data," *Forests*, vol. 11, no. 9, p. 1000, Sep. 2020.
- [48] K. Valaskova, J. Olah, J. Popp, and G. Lazaroiu, "Virtual modeling and remote sensing technologies, spatial cognition and neural network algorithms, and visual analytics tools in urban geopolitics and digital twin cities," *Geopolitics, Hist. Int. Relations*, vol. 14, no. 2, pp. 9–24, 2022.
- [49] M. Bugaj, V. Machova, and S. Bratu, "The geopolitics of smart sustainable city governance and simulated 3D environments," *Geopolitics, Hist., Int. Relations*, vol. 14, no. 2, pp. 152–167, 2022.
- [50] M. Nagy and G. Lazaroiu, "Computer vision algorithms, remote sensing data fusion techniques, and mapping and navigation tools in the Industry 4.0-based Slovak automotive sector," *Mathematics*, vol. 10, no. 19, p. 3543, Sep. 2022.
- [51] D. P. M. Zaks, N. Ramankutty, C. C. Barford, and J. A. Foley, "From Miami to Madison: Investigating the relationship between climate and terrestrial net primary production," *Global Biogeochemical Cycles*, vol. 21, no. 3, pp. n/a–n/a, Sep. 2007.
- [52] F. Li, W. Zheng, Y. Wang, J. Liang, S. Xie, S. Guo, X. Li, and C. Yu, "Urban green space fragmentation and urbanization: A spatiotemporal perspective," *Forests*, vol. 10, no. 4, p. 333, Apr. 2019.
- [53] P. Kowe, O. Mutanga, and T. Dube, "Advancements in the remote sensing of landscape pattern of urban green spaces and vegetation fragmentation," *Int. J. Remote Sens.*, vol. 42, no. 10, pp. 3797–3832, May 2021.
- [54] S. Jia, "Study on carbon storage of forest vegetation and its economic value in Henan province based on continuous forest resources inventory," *Hubei. Agric. Sci.*, vol. 55, pp. 1612–1616, 2016.
- [55] J. M. O. Scurlock, K. Johnson, and R. J. Olson, "Estimating net primary productivity from grassland biomass dynamics measurements," *Global Change Biol.*, vol. 8, no. 8, pp. 736–753, Aug. 2002.
- [56] Y. Luo, X. Wang, Z. Ouyang, F. Lu, L. Feng, and J. Tao, "A review of biomass equations for China's tree species," *Earth Syst. Sci. Data*, vol. 12, no. 1, pp. 21–40, Jan. 2020.
- [57] H. Zhang, R. Sun, Z. Xiao, J. Wang, and M. Wang, "Global 500M spatial resolution gross and net primary productivity products based on an improved light use efficiency model from 2000–2019," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2022, pp. 5762–5765.
- [58] S. Kedia, S. P. Bhakare, A. K. Dwivedi, S. Islam, and A. Kaginalkar, "Estimates of change in surface meteorology and urban heat island over North-west India: Impact of urbanization," *Urban Climate*, vol. 36, Mar. 2021, Art. no. 100782.
- [59] L. Bruzzone and B. Demir, "A review of modern approaches to classification of remote sensing data," in *Land Use and Land Cover Mapping in Europe (Remote Sensing and Digital Image Processing)*, vol. 18, I. Manakos and M. Braun, Eds. Dordrecht, The Netherlands: Springer, Jan. 2014, pp. 127–143.
- [60] B. Zhou, Z. Liao, S. Chen, H. Jia, J. Zhu, and X. Fei, "Net primary productivity of forest ecosystems in the Southwest Karst region from the perspective of carbon neutralization," *Forests*, vol. 13, no. 9, p. 1367, Aug. 2022.
- [61] M. Kang and S. Cho, "Progress in water and energy flux studies in Asia: A review focused on eddy covariance measurements," *J. Agricult. Meteorol.*, vol. 77, no. 1, pp. 2–23, 2021.
- [62] C. Zhang and Y. Lu, "Study on artificial intelligence: The state of the art and future prospects," *J. Ind. Inf. Integr.*, vol. 23, Sep. 2021, Art. no. 100224.
- [63] B. Yu, F. Chen, and H. Chen, "NPP estimation using random forest and impact feature variable importance analysis," *J. Spatial Sci.*, vol. 64, no. 1, pp. 173–192, Jan. 2019.
- [64] F. Yang, K. Ichii, M. A. White, H. Hashimoto, A. R. Michaelis, P. Votava, A.-X. Zhu, A. Huete, S. W. Running, and R. R. Nemani, "Developing a continental-scale measure of gross primary production by combining MODIS and AmeriFlux data through support vector machine approach," *Remote Sens. Environ.*, vol. 110, no. 1, pp. 109–122, Sep. 2007.
- [65] M. Xu, C. Hu, R. G. Najjar, M. Herrmann, H. Briceno, B. B. Barnes, J. O. R. Johansson, and D. English, "Estimating estuarine primary production using satellite data and machine learning," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 110, Jun. 2022, Art. no. 102821.
- [66] V. Moosavi, H. Malekinezhad, and B. Shirmohammadi, "Fractional snow cover mapping from MODIS data using wavelet-artificial intelligence hybrid models," *J. Hydrol.*, vol. 511, pp. 160–170, Apr. 2014.
- [67] M. Peng, W. Han, C. Li, X. Yao, and G. Shao, "Modeling the daytime net primary productivity of maize at the canopy scale based on UAV multispectral imagery and machine learning," *J. Cleaner Prod.*, vol. 367, Sep. 2022, Art. no. 133041.
- [68] Y. Zhang and Y. Zhang, "New discrete-time models of zeroing neural network solving systems of time-variant linear and nonlinear inequalities," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 2, pp. 565–576, Feb. 2020.
- [69] Y. Shi, B. Qiu, D. Chen, J. Li, and Y. Zhang, "Proposing and validation of a new four-point finite-difference formula with manipulator application," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1323–1333, Apr. 2018.
- [70] J. Hurley, A. Saunders, A. Both, C. Sun, B. Boruff, J. Duncan, M. Amati, and P. Caccetta, "Urban vegetation cover change in Melbourne: 2014–2018," Centre Urban Res., Melbourne, VIC, Australia, Tech. Rep. V2.0, 2019.
- [71] J. Zhong, J. Liu, L. Jiao, X. Lian, Z. Xu, and Z. Zhou, "Assessing the comprehensive impacts of different urbanization process on vegetation net primary productivity in Wuhan, China, from 1990 to 2020," *Sustain. Cities Soc.*, vol. 75, Dec. 2021, Art. no. 103295.
- [72] C. Milesi, C. D. Elvidge, R. R. Nemani, and S. W. Running, "Assessing the impact of urban land development on net primary productivity in the southeastern United States," *Remote Sens. Environ.*, vol. 86, no. 3, pp. 401–410, Aug. 2003.
- [73] F. Pei, X. Li, X. Liu, S. Wang, and Z. He, "Assessing the differences in net primary productivity between pre- and post-urban land development in China," *Agricult. Forest Meteorol.*, vols. 171–172, pp. 174–186, Apr. 2013.
- [74] T. Yu, Y. Pang, R. Sun, and X. Niu, "Spatial downscaling of vegetation productivity in the forest from deep learning," *IEEE Access*, vol. 10, pp. 104449–104460, 2022.
- [75] X. Liu, C. Yoo, F. Xing, H. Oh, G. El Fakhri, J.-W. Kang, and J. Woo, "Deep unsupervised domain adaptation: A review of recent advances and perspectives," *APSIPA Trans. Signal Inf. Process.*, vol. 11, no. 1, pp. 1–48, 2022.
- [76] S. Cao, P. Xu, and D. A. Clifton, "How to understand masked autoencoders," 2022, *arXiv:2202.03670*.
- [77] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 16000–16009.
- [78] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306.
- [79] J. Kandt and M. Batty, "Smart cities, big data and urban policy: Towards urban analytics for the long run," *Cities*, vol. 109, Feb. 2021, Art. no. 102992.
- [80] Y. Kim, H.-G. Kim, Z. Li, and H.-J. Choi, "Avoiding overfitting in deep neural networks for clinical opinions generation from general blood test results," in *MEDINFO: Precision Healthcare Through Informatics*, vol. 245. Amsterdam, The Netherlands: IOS Press, 2017, p. 1274.
- [81] R. Houborg, J. B. Fisher, and A. K. Skidmore, "Advances in remote sensing of vegetation function and traits," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 43, pp. 1–6, Dec. 2015.
- [82] P. Liu, K. K. R. Choo, L. Wang, and F. Huang, "SVM or deep learning? A comparative study on remote sensing image classification," *Soft Comput.*, vol. 21, no. 23, pp. 7053–7065, Dec. 2017.
- [83] M. Minghini and F. Frassinelli, "OpenStreetMap history for intrinsic quality assessment: Is OSM up-to-date?" *Open Geospatial Data. Softw. Standards*, vol. 4, no. 1, pp. 1–17, Dec. 2019.
- [84] L. Wang, W. Gong, Y. Ma, and M. Zhang, "Modeling regional vegetation NPP variations and their relationships with climatic parameters in Wuhan, China," *Earth Interact.*, vol. 17, no. 4, pp. 1–20, May 2013.
- [85] T. Cui, Y. Wang, R. Sun, C. Qiao, W. Fan, G. Jiang, L. Hao, and L. Zhang, "Estimating vegetation primary production in the Heihe river basin of China with multi-source and multi-scale data," *PLoS ONE*, vol. 11, no. 4, Apr. 2016, Art. no. e0153971.
- [86] T. Yu, R. Sun, Z. Xiao, Q. Zhang, G. Liu, T. Cui, and J. Wang, "Estimation of global vegetation productivity from global land surface satellite data," *Remote Sens.*, vol. 10, no. 2, p. 327, Feb. 2018.

- [87] M. Wang, R. Sun, A. Zhu, and Z. Xiao, "Evaluation and comparison of light use efficiency and gross primary productivity using three different approaches," *Remote Sens.*, vol. 12, no. 6, p. 1003, Mar. 2020.
- [88] S. Bodapati, H. Bandrupally, R. N. Shaw, and A. Ghosh, "Comparison and analysis of RNN-LSTMs and CNNs for social reviews classification," in *Advances in Applications of Data-Driven Computing*. New York, NY, USA: Springer, 2021, pp. 49–59.
- [89] M. Carvalho, B. L. Saux, P. Trouve-Peloux, A. Almansa, and F. Champagnat, "On regression losses for deep depth estimation," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2915–2919.
- [90] M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov, "Training and operation of an integrated neuromorphic network based on metal-oxide memristors," *Nature*, vol. 521, pp. 61–64, Dec. 2015.
- [91] W. Wan, "A compute-in-memory chip based on resistive random-access memory," *Nature*, vol. 608, no. 7923, pp. 504–512, 2022.
- [92] M. Wang, G. Liu, R. Sun, and Z. Xiao, "Assessment of NPP dynamics and the responses to climate changes in China from 1982 to 2012," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 6602–6605.
- [93] L. Song, M. Li, H. Xu, Y. Guo, Z. Wang, Y. Li, X. Wu, L. Feng, J. Chen, X. Lu, Y. Xu, and T. Li, "Spatiotemporal variation and driving factors of vegetation net primary productivity in a typical Karst area in China from 2000 to 2010," *Ecol. Indicators*, vol. 132, Dec. 2021, Art. no. 108280.
- [94] A. Krizhevsky and G. Hinton, "Convolutional deep belief networks on CIFAR-10," *Unpublished Manuscript*, vol. 40, no. 7, pp. 1–9, 2010.
- [95] A. H. Fath, F. Madanifar, and M. Abbasi, "Implementation of multilayer perceptron (MLP) and radial basis function (RBF) neural networks to predict solution gas-oil ratio of crude oil systems," *Petroleum*, vol. 6, no. 1, pp. 80–91, Mar. 2020.



MOU CHAO received the Ph.D. degree from the School of Computer Science, Chongqing University, in 2018.

His current research interests include double carbon research, educational data mining, data mining, and model interpretability.



WEI MAIMAI received the bachelor's degree in forestry engineering from Beijing Forestry University, in 2021, and the bachelor's degree in data science and big data technology, in 2023.

His current research interests include double carbon research and big data.



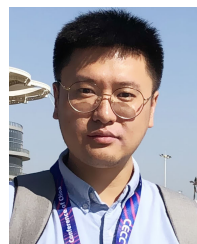
LIU HANZHANG is currently pursuing the bachelor's degree in engineering with Beijing Forestry University, Beijing, China.

His current research interests include double carbon research and big data.



CHEN ZHIBO received the Ph.D. degree.

He has been the Dean of the School of Information and the Head of the Computer Science and Technology Discipline, Beijing Forestry University, since 2013. He is currently a Professor and a Ph.D. Supervisor. His current research interests include double carbon research, smart forestry, and AI.



CUI XIAOHUI received the Ph.D. degree in computer application technology from Harbin Engineering University, in 2013.

His current research interests include double carbon research, smart forestry AI, and big data.

...