High-Resolution Mapping of Gross Primary Production in Northeast China Using Landsat-8/9 and Sentinel-2 A/B

Xiaoyan Ma, Li Pan, and Haoming Xia

***Abstract*—Accurately estimating gross primary production (GPP) in terrestrial ecosystems is crucial for gaining a deeper understanding of the carbon cycle within the ecosystem and for predicting climate change. Although many GPP datasets are available, they often have low resolution, typically around 500 m or lower, which restricts their effectiveness in monitoring fragmented croplands and areas with high heterogeneity. In this study, we utilized optical satellite data from Landsat-8/9 and Sentinel-2A/B, along with meteorological data from ERA5, to generate a GPP dataset with a spatial resolution of 30 m for three provinces in Northeast China. This dataset was developed based on the Vegetation Photosynthesis Model and exhibited robust validation results when compared with SIF data and other existing GPP datasets. It provides a high-resolution GPP product that significantly enhances the precision of carbon cycle research in Northeast China. This research underscores the feasibility of producing high spatial resolution GPP products using Landsat-8/9 and Sentinel-2A/B optical satellite data. The resulting dataset offers a more refined GPP estimate for studies related to the terrestrial carbon cycle.**

***Index Terms*—Gross primary Production (GPP), image integration, mapping, vegetation photosynthesis model (VPM).**

I. Introduction

GROSS primary Production (GPP) refers to the amount of carbon dioxide (CO_2) from the atmosphere that is fixed into organic compounds by terrestrial plants through net photosynthesis, also known as total primary productivity. GPP determines the initial matter and energy entering terrestrial ecosystems and plays a key role in carbon balance, reflecting the ability of terrestrial ecosystems to offset anthropogenic CO_2 emissions [1]. Due to the susceptibility of photosynthesis to natural disasters such as droughts, floods, frosts, and hurricanes, accurate estimation of GPP can yield critical insights into ecosystem responses to extreme environmental events [2], [3], [4], [5]. Considering the pivotal role that Gross Primary Production (GPP) plays in ecosystem dynamics, coupled with the increasing imperative to comprehend the terrestrial biosphere's contribution to the carbon cycle, acquiring a comprehensive understanding of the spatiotemporal patterns of GPP is of paramount importance [6], [7], [8], [9], [10].

Since the 1980s, various methods for estimating GPP have been developed, which can be broadly categorized into four categories: the enzyme kinetic models [11], [12], the light use efficiency (LUE) models [13], [14], [15], machine learning techniques based on eddy covariance measurements [16], [17], [18] and methods using solar-induced chlorophyll fluorescence (SIF) [19], [20]. The inherent simplicity of the LUE model, coupled with its minimal input requirements, renders it a highly favored approach for remote sensing applications. LUE models, which are underpinned by satellite remote sensing data, offer the capability to deliver products with high temporal and spatial resolution. This characteristic renders them particularly suitable for large-scale GPP estimation, facilitating comprehensive environmental monitoring and ecosystem management [8].

The GPP estimated by LUE models is the product of LUE and the absorbed photosynthetically active radiation (APAR) by plants. Traditional LUE models typically assume a uniform LUE across a specific biological community, while APAR represents the energy absorbed by the entire canopy [14]. However, recent studies have shown that the photosynthetic capacity is not only affected by leaf quantity (leaf area index, LAI) but also by leaf quality (the photosynthetic rate of each leaf) [21], [22]. This is mainly related to the chlorophyll and nitrogen content of the leaves [23], [24], leading to variations in APAR even among plants with the same LAI. Recent research indicates that the fraction of photosynthetically active radiation absorbed by chlorophyll ($fPAR_{chl}$) can better capture the seasonal changes in vegetation photosynthetic capacity, significantly improving the seasonal representation of GPP [25], [26]. The vegetation photosynthesis model (VPM), based on this theory, has shown excellent results in large-scale GPP estimation [8], [27], [28], [29].

Manuscript received 21 March 2024; revised 14 June 2024; accepted 20 July 2024. Date of publication 23 July 2024; date of current version 5 August 2024. This work was supported in part by the National Natural Science Foundation Project of China under Grant 42271118 and Grant 32130066, in part by the Key Project of Research and Development of Ningxia, China under Grant 2022BEG03050, and in part by the Henan Provincial Department of Science and Technology Research Project under Grant 242102321160. *(Corresponding author: Haoming Xia.)*

Xiaoyan Ma is with the College of Geography and Environmental Science, Henan University, Kaifeng 475004, China (e-mail: xiaoyanma@vip.henu.edu.cn).

Li Pan is with the School of Biological Sciences, Center for Earth Observation and Modeling, University of Oklahoma, Norman, OK 73019 USA (e-mail: li.pan@ou.edu).

Haoming Xia is with the College of Geography and Environmental Science, the Key Laboratory of Geospatial Technology for the Middle and Lower Yellow River Regions (Henan University), Ministry of Education, Xinyang Academy of Ecological Research, Henan Dabieshan National Field Observation and Research Station of Forest Ecosystem, Henan University, Kaifeng 475004, China (e-mail: xiahm@vip.henu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3432581

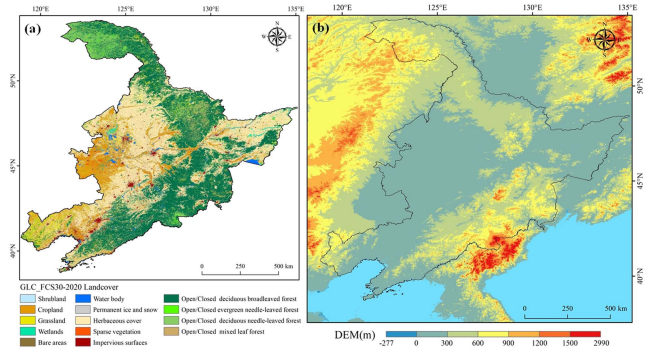

Fig. 1. (a) Land cover classification and (b) elevation.

The current mainstream GPP products have a spatial resolution of mostly 500 m [8], [14], [30], [31], which is quite limited for monitoring fragmented croplands (CRO) or areas with high heterogeneity. Compared to satellites like MODIS with moderate spatial resolution, the Landsat series of satellites offer a resolution of 30 m, providing more detailed spatial information. However, their 16-day revisit period limits their ability to distinguish crop phenology. The Sentinel-2 system, consisting of two satellites (Sentinel-2A and Sentinel-2B), has a revisit period of 5 days. Thus, the fusion of Landsat and Sentinel-2 data can yield an image dataset with high spatial and temporal resolution, leading to the generation of GPP products with a 30-m spatial resolution. This is extremely helpful for obtaining phenological information on crops and studying their growth conditions. Additionally, high-resolution data are required in environmental monitoring and precision agriculture management to obtain detailed information [6], [32], [33]. Therefore, there is an urgent need for high spatial resolution GPP products for accurate estimation of terrestrial plant GPP.

The objective of this study is to develop a 30-m spatial resolution GPP product for Northeast China using data from Landsat-8/9 and Sentinel-2 A/B satellites, employing the VPM model. This study also involves the comparative analysis and validation of the GPP product, assessing the efficacy of fused high-resolution satellite data in the estimation of GPP products. Some studies have shown that SIF contains information about APAR and LUE [34], [35], [36], which are closely related to the VPM model. Therefore, we compared the GPP_{VPM} dataset with the GOSIF product [see Fig. 5(a)] and also compared the APAR [see Fig. 5(d)] used in producing the GPP_{VPM} with the GOSIF product. The innovation of this study lies in the production of the first 30-m, 8-day GPP dataset for Northeast China. This dataset integrates Landsat-8/9 and Sentinel-2A/B data sources with ERA5 meteorological reanalysis data. Additionally, it utilizes land cover and crop distribution maps to generate high-precision C3/C4 pixel proportion layers, enhancing the estimation accuracy of the VPM model. Compared to existing coarse-resolution datasets, this dataset is more suitable for agricultural regions with surface spatial heterogeneity and fragmented landscapes. Furthermore, it serves as a model for the future production of high spatial resolution GPP products.

TABLE I
DATASETS USED IN THIS STUDY AND THEIR SPECIFICATIONS

Data source	Dataset	Derived variables	Spatial resolution	Temporal resolution	Data Source or Reference
Landsat-8	LC08/C02/T1_L2	EVI, LSWI	30 m	16-day	GEE platform
Landsat-9	LC09/C02/T1_L2		30 m	16-day	
Sentinel-2A/B	COPERNICUS/S2_SR		10 m	5-day	
ERA5	ECMWF Climate Reanalysis	Daytime temperature Daily mean PAR	0.1°	hourly	ECMWF
CLCD	Crop types distribution	Land cover type	30 m	annual	[45]
GLC_FCS30	Other types distribution		30 m	5-year	[46]
CDL	Maize distribution	C4 crop percentage	10 m	invariant	[50]
ISLSCP II	C4 vegetation percentage map	C4 grassland percentage	1° × 1°	invariant	[51]
GOSIF	GOSIF	Data comparison	0.05°	8-day	[31]
GOSIF GPP	GOSIF GPP		0.05°	8-day	
MODIS	MOD17A2H		500 m	8-day	GEE platform

II. MATERIALS AND METHODS

A. Study Area

The study area is located in Northeast China, extending from 38.5°N to 53.6°N and 118.6°E to 135.0°E. This region encompasses the provinces of Heilongjiang, Jilin, and Liaoning, collectively covering an expansive area of approximately 7 78 000 km^2. The predominant topography consists of plains and mountains, with altitudes ranging from 6 to 2667 m, and an average altitude of 428.35 m. The study area extends across different climate zones, transitioning from mid-temperate to cold regions from south to north. It experiences well-defined seasons with periods of rainfall and heat, typical of a temperate continental monsoon climate. The primary land cover types within this region include CRO, deciduous broadleaf forests (DBF), deciduous needleleaf forests (DNF), and evergreen needleleaf forests (ENF) [see Fig. 1].

B. Datasets and Preprocessing

1) Meteorological Data l Data: The meteorological data used in this study were obtained from the European Centre for medium-range weather forecasts (ECMWF) climate reanalysis dataset (ERA5), which includes hourly shortwave radiation and surface temperature at 2 m above ground (see Table I) [37]. The temporal resolution of the time series data in this study is 8 days. Therefore, we used the average temperature from 10 A.M. to 6 P.M. each day as the daytime temperature and then averaged

these temperatures over every 8 days to create 8-day daytime temperature data. The shortwave radiation (PAR) data required unit conversion, as described in the following formula [38]:

$$\mathrm{PAR}_{\mathrm{DT\_ERA5}} = \mathrm{SW}_{\mathrm{rad}} \times f_{\mathrm{PAR}} \times \beta \tag{1}$$

where $\mathrm{SW}_{\mathrm{rad}}$ (J m^{-2}) is downward shortwave radiation in the ERA5 dataset; $f_{\mathrm{PAR}} = 0.45$ is the fraction of SW to PAR [39], $\beta = 4.56$ μmol J^{-1} is the energy-quanta conversion factor [40].

2) Landsat-8/9 and Sentinel-2A/B Data: Landsat-8 and Landsat-9 were launched in February 2013 and September 2021, respectively. They are equipped with the operational land imager (OLI) and the OLI-2, the revisit cycle for a single Landsat satellite is 16 days, while the combined revisit period for both satellites is reduced to 8 days. Sentinel-2 consists of two satellites: Sentinel-2A, launched in June 2015, and Sentinel-2B, launched in March 2017. These satellites carry a multispectral instrument (MSI) with spatial resolutions of 10, 20, and 60 m, and the combined revisit period for both satellites has a revisit period of 5 days.

The surface reflectance (SR) data utilized inherently compensates for atmospheric effects, including scattering and absorption. This compensation yields more accurate and consistent representations of the Earth's surface, an essential factor for ensuring reliable temporal and spatial comparisons. Such comparisons are critical for long-term environmental monitoring and the detection of ecological changes [41]. Consequently, this study employed all available Landsat-8/9 and Sentinel-2A/B SR remote sensing data from the study area for the duration of the research period, constituting our initial image set. The datasets used in this research are "LANDSAT/LC08/C02/T1_L2," "LANDSAT/LC09/C02/T1_L2," and "COPERNICUS/S2_SR" in the Google Earth Engine (GEE) platform (see Table I).

Cloud and cloud shadow detection is key to processing optical remote sensing data. For Landsat-8/9 images, we used the FMask algorithm in GEE to detect and remove poor-quality observations in the images, reducing the impact of clouds and cloud shadows. This method identifies potential cloud pixels based on the physical properties of clouds and detects cloud shadows through the dimming effect in the near-infrared band. For Sentinel-2A/B image data, we used the quality band in the metadata to identify poor-quality observations in the images. The metadata identifies observations that include cirrus clouds and opaque clouds as poor-quality observations, and these are stored as NODATA in the image to generate a poor-quality observation mask. High-quality observation images are obtained by removing clouds and cloud shadows with a cloud mask.

Due to the differences in orbit, spatial, and spectral configurations of sensors, which can affect the physical measurement values and radiometric properties of images [42], each sensor's wavelength may be slightly different. Therefore, it is necessary to normalize remote sensing reflectance data to obtain comparable results. In this study, ordinary least squares regression coefficients were used to convert MSI bands to OLI band standards. The bands that need to be processed are ρ_{red} and ρ_{NIR}. The regression coefficients for the two MSI bands calculated in this study are 0.9103 and 0.9701 [43], [44].

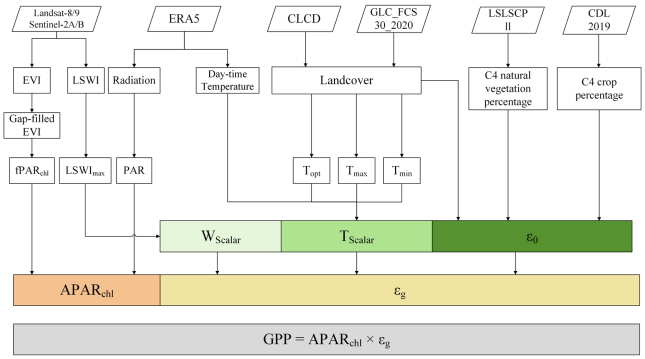

Fig. 2. Datasets and workflow of VPM to calculate GPP.

3) Land Cover Data: The VPM model incorporates different parameters tailored to various biotic communities. The land cover datasets used here are CLCD [45] and GLC_FCS30_2020 (see Table I) [46]. Because GLC_FCS30_2020 significantly underestimates CRO, and CLCD has a better recognition effect on CRO [47], we use CLCD to determine the distribution of CRO, and the rest is classified according to the classification results of GLC_FCS30_2020.

C. Methods

1) Reconstructed 30 m Enhanced Vegetation Index and Land Surface Water Index Data: Fig. 2 shows the workflow to produce the map of GPP. The enhanced vegetation index (EVI) [48] and the land surface water index (LSWI) [49] are the two main input parameters for the VPM. They are calculated based on the terrestrial reflectance data obtained from the fusion of Landsat-8/9 and Sentinel-2A/B satellites.

EVI has better sensitivity in the vegetation saturation area, eliminates soil background signals, and reduces atmospheric interference, better reflecting the spatial heterogeneity of vegetation. The calculation formula is as follows:

$$\mathrm{EVI} = 2.5 \times \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{red}}}{\rho_{\mathrm{NIR}} + 6 \times \rho_{\mathrm{red}} - 7.5 \times \rho_{\mathrm{blue}} + 1} \tag{2}$$

where ρ_{NIR}, ρ_{red}, and ρ_{blue} represent the near-infrared band reflectance values, the red band, and the blue band reflectance values, respectively. The blue band is used to correct the red band affected by the atmosphere. As the chlorophyll content in green vegetation increases, more red light is absorbed, more near-infrared light is scattered due to the arrangement of cell walls, and the introduction of the blue light band reduces the impact of aerosols on the red band.

LSWI is a vegetation moisture content index, and the calculation formula is as follows:

$$\mathrm{LSWI} = \frac{\rho_{\mathrm{NIR}} - \rho_{\mathrm{SWIR}}}{\rho_{\mathrm{NIR}} + \rho_{\mathrm{SWIR}}} \tag{3}$$

where ρ_{NIR} and ρ_{SWIR} represent the near-infrared band reflectance values and the shortwave infrared band reflectance values, respectively.

Because the lower temporal resolution of a single sensor may miss critical phenological changes in vegetation, multisource

sensor data are used to increase the number of available observation images during the research period. Considering the different temporal resolutions and overlapping shooting areas of the Landsat-8/9 and Sentinel-2A/B satellites, it is possible to observe more than twice at different times and areas on the same day, which may lead to abnormal observation values. In addition, different sensors can also cover the same area within a few days, and the images generated by this will also cause differences in EVI or LSWI. Therefore, it is necessary to generate time series with equal time intervals to reduce the impact of insufficient and uneven observations. In this article, the maximum value of the high-quality values of all images on each pixel within 8 days is calculated as the observation value of the 8-day composite image. Through this step, the temporal resolution of the dataset is reconstructed to 8 days.

2) C4 Percentage for CRO and Natural Vegetation: There are significant differences in solar energy utilization between C3/C4 plants [52]. C4 plants generally do not exhibit light saturation effects due to the absence of photorespiration. Many previous studies have shown that C3/C4 plants should be treated differently in LUE models [53]. Many models do not consider the differences between C3/C4 plants, which is one of the main reasons for the underestimation of GPP in CRO.

In this study, our primary focus is on maize as the representative of C4 crops, due to the relatively minor cultivation area and the complexities involved in accurately quantifying other C4 crops. We refer to the distribution of maize in other studies (see Table I) [50] to obtain the proportion of C4 crops. For natural vegetation, we use the C4 proportion in the International Satellite Land Surface Climatology Project (ISLSCP II) [51] C4 dataset for calculation (see Table I).

$$\text{Ratio}_{\text{C4}} = \frac{\text{C4 crop percentage}}{\text{Total crop percentage}} \tag{4}$$

where the C4 CRO is obtained from the CDL, and the total CRO is obtained from the proportion of CRO in CLCD. The C4 proportions for grasslands, sparse shrublands, wetlands, and natural vegetation are determined according to the values in the ISLSCP II C4 dataset.

3) VPM Model: The VPM model calculates GPP as the product of light absorption by chlorophyll of the vegetation (APAR_{chl}) and the efficiency (ε_g) that converts the absorbed energy to carbon fixed by plants through photosynthesis

$$\text{GPP} = \text{APAR}_{\text{chl}} \times \varepsilon_g \tag{5}$$

where APAR_{chl} is the product of photosynthetically active radiation (PAR) and ε_g is the efficiency of carbon fixation by photosynthesis

$$\text{APAR}_{\text{chl}} = \text{PAR} \times f\text{PAR}_{\text{chl}} \tag{6}$$

where PAR is obtained from ERA5 data, and $f\text{PAR}_{\text{chl}}$ is the fraction of PAR absorbed by chlorophyll, which is a linear function of EVI, the formulas are as follows [49]:

$$f\text{PAR}_{\text{chl}} = \text{EVI} - 0.1. \tag{7}$$

The ε_g is primarily influenced by temperature limitations (T_{scalar}) and water stress (W_{scalar}), and the formulas are as follows:

$$\varepsilon_g = \varepsilon_0 \times T_{\text{scalar}} \times W_{\text{scalar}} \tag{8}$$

where ε_0 and T_{scalar} are based on biomes. ε_0 values can be found in Table II, it is the maximum value of the efficiency that converts the absorbed energy to carbon fixed by plants through photosynthesis. For CRO and C4-containing natural vegetation, ε_0 is calculated based on the ratio of C3 to C4 plants using a weighted average. The formulas for T_{scalar} and W_{scalar} are as follows:

$$T_{\text{scalar}} = \frac{(T - T_{\max}) \times (T - T_{\min})}{(T - T_{\max}) \times (T - T_{\min}) - (T - T_{\text{opt}})^2} \tag{9}$$

$$W_{\text{scalar}} = \frac{1 + \text{LSWI}}{1 + \text{LSWI}_{\max}} \tag{10}$$

where $\text{LSWI}_{\max}$ is calculated based on the 90th percentile of annual LSWI. T, $T_{\max}$, $T_{\min}$, and T_{opt}, respectively, represent the daytime average temperature, the highest temperature for photosynthesis, the lowest temperature for photosynthesis, and the optimal temperature for photosynthesis. T is derived from ERA5 daytime average temperature, while the other three parameters are biome-specific and can be found in Table II.

TABLE II
BIOME SPECIFIC LOOKUP-TABLE

IGBP class	ε_0(g C/mol APAR)	$T_{\min}$(°C)	$T_{\max}$(°C)	T_{opt}(°C)
ENF	0.5250	-1	40	20
EBF	0.5250	-2	48	28
DNF	0.5250	-1	40	20
DBF	0.5250	-1	40	20
MF	0.5250	-1	48	19
CSH	0.5250	-1	48	25
OSH	0.5250	1	48	31
WSA	0.5250	-1	48	24
SAV	0.5250 (C3) 0.7875 (C4)	1	48	30
GRA	0.5250 (C3) 0.7875 (C4)	0	48	27
WET	0.5250 (C3) 0.7875 (C4)	-1	40	20
CRO	0.5250 (C3) 0.7875 (C4)	-1	48	30
URB	0.5250	0	48	27
CNV	0.5250 (C3) 0.7875 (C4)	0	48	27

This table is adopted from [27]

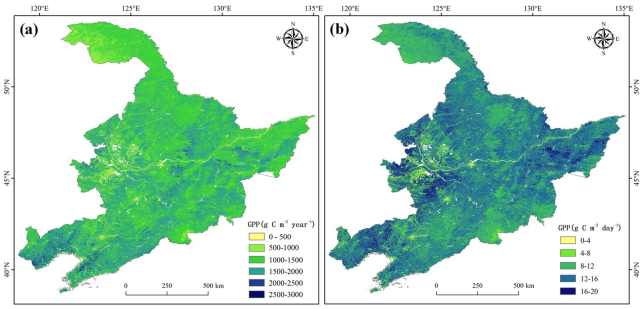

Fig. 3. (a) Total annual GPP and (b) maximum daily GPP for the study area in 2022.

III. Results

A. Annual Map of 30 m GPP Across Northeast China in 2022

Fig. 3 shows the annual and maximum daily GPP at a 30-m spatial resolution for the study area in 2022. The average annual GPP for the three northeastern provinces of China was calculated to be 1402.62 gC·m^{-2} day^{-1}. Notably, the highest annual GPP was predominantly observed in the southwestern and eastern mountainous areas [see Fig. 3(a)], where primarily characterized by DBF. Conversely, the lowest annual GPP was recorded in the western part of the study area, where sparse herbaceous vegetation predominates. In terms of daily GPP, the highest values were recorded in the southwestern and northeastern mountainous regions [see Fig. 3(b)], again largely covered by DBF. One reason for the inconsistent spatial distribution is the existence of fallow periods in CRO, leading to high daily GPP values but low annual GPP. This phenomenon is also evident in the frequency distribution plot in Section III-B. You can obtain the dataset online.[1]

B. Frequency Distribution of Annual Total GPP and Daily Maximum GPP in Northeast China

The frequency distribution of the annual total GPP and daily maximum GPP, as determined by GPP$_{\mathrm{VPM}}$, of the main land cover types can be found in Fig. 4, with the red line indicating the corresponding average value. The highest annual total GPP, mainly ranging between 1200 and 1800 gC·m^{-2} year^{-1} with an average of 1553.85 gC·m^{-2} year^{-1}, was observed in DBF. The highest daily maximum GPP, primarily ranging between 10 and 15 gC·m^{-2} day^{-1} with an average of 12.77 gC·m^{-2} day^{-1}, was observed in CRO.

C. Accuracy Assessment for the 2022 GPP Map

We conducted a comprehensive comparison of the GPP$_{\mathrm{VPM}}$ dataset against the GOSIF product, as illustrated in Fig. 5(a), and analyzed the APAR used in the GPP$_{\mathrm{VPM}}$'s production against the GOSIF product, detailed in Table I. Furthermore, we evaluated the GPP$_{\mathrm{VPM}}$ dataset alongside other GPP datasets, including MOD17A2H [see Fig. 5(c)] and GOSIF GPP [see

1[Online]. Available: https://www.scidb.cn/s/QJBzAz.

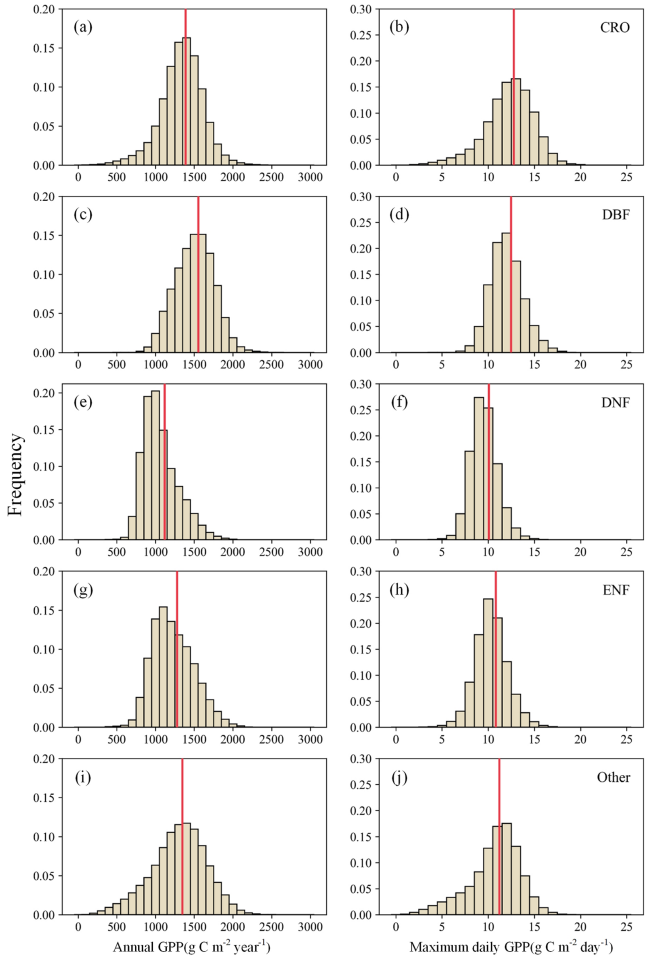

Fig. 4. (a), (c), (e), (g), (i) Frequency distribution graphs of annual total GPP and (b), (d), (f), (h), (j) daily maximum GPP for the main land cover types as determined by GPP$_{\mathrm{VPM}}$, with the red vertical line representing the average value.

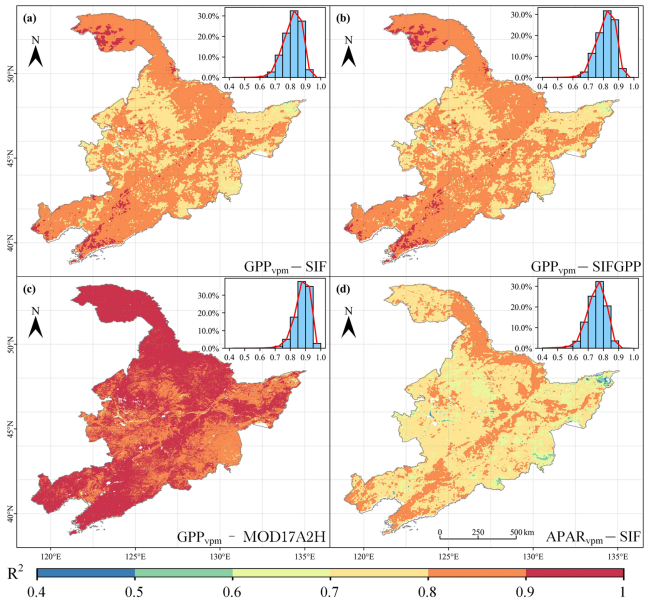

Fig. 5. Spatial distribution and frequency statistics of R^2.

Fig. 5(b)], through a correlation analysis involving 46 8-day composite datasets per pixel. This analysis calculated the coefficient of determination (R^2) for each pixel, resulting in a spatial distribution of R^2 values. The correlation outcomes reveal that the R^2 value, indicative of the dataset concordance, for GPP_{VPM} versus GOSIF hovers around 0.85. A similar R^2 value of approximately 0.85 is observed when comparing GPP_{VPM} with GOSIF GPP. The comparison between GPP_{VPM} and MOD17A2H GPP datasets yields a slightly higher R^2 of around 0.9, whereas the R^2 for APAR versus GOSIF is about 0.75. These R^2 values fall within an acceptable range, underscoring the satisfactory nature of our validation results.

IV. Discussion

A. Potential of Landsat-8/9 and Sentinel-2 A/B in Estimating GPP Products

In regions characterized by fragmented CRO or significant heterogeneity, GPP datasets derived from imagery with coarser spatial resolutions (>500 m) may struggle to accurately distinguish between different land cover types, introducing uncertainties into GPP assessments [54]. However, the reconstruction of continuous, high-resolution datasets (30 m), such as that from Landsat-8/9 and Sentinel-2 A/B, could generate reliable and high-resolution GPP data. Meanwhile, unlike commercially available optical data from platforms like WorldView, GeoEye, QuickBird, and IKONOS, which are costly and challenging to access, Landsat-8/9 and Sentinel-2 A/B provide high-resolution data free of charge [55].

This study focuses on the year 2022 to demonstrate the generation of GPP data for Northeast China. The abundance of images and the quality of observational data bolstered the scientific validity and accuracy of the GPP data generated using the VPM. The comparative and validation analyses conducted with other datasets yielded satisfactory results. Areas with lower R^2 values are distributed in the plains of the western and northeastern parts of the study area, where the main land cover type is CRO. The possible reason is that GPP_{VPM} dataset provides more detailed results in highly heterogeneous areas such as CRO compared to coarse-spatial resolution datasets. With the anticipated increase in freely available Landsat-9 satellite data, future research could extend GPP dataset production to cover broader areas and additional years. This expansion would facilitate the examination of temporal and spatial variations in plant GPP at a high spatial resolution over extended periods, offering valuable insights into ecosystem dynamics.

B. Improvements in Model Input Data

We tried to improve the VPM input data reported by VPM v2.0 [8] in three ways: first, VPM v2.0 uses NCEP Reanalysis II climate data. However, this study uses climate reanalysis data from the ECMWF, known as ERA5. Compared to NCEP reanalysis II, ERA5 has a higher spatial and temporal resolution, providing more detailed and refined meteorological data. Its spatial resolution in the horizontal direction is 0.25° (about 25 km), and the temporal resolution is hourly. This enables ERA5 to provide more refined geographic information and more accurate meteorological changes. Additionally, ERA5 uses more advanced physical parameterization schemes and improved modeling techniques, which help better simulate the behavior of the atmosphere and earth systems [37]. This makes the simulation of climate and meteorological processes by ERA5 more accurate. The comparative validation of the final GPP data also shows that the GPP products calculated using the ERA5 climate dataset perform well. Second, the land cover datasets used here are CLCD and GLC_FCS30_2020, they all have a spatial resolution of 30 m, which enables better differentiation of various biomes. Besides, CLCD has been proven to exhibit high accuracy in classifying CRO in China [47]. Finally, the prevailing assumption regarding C4 crops has been a simplistic division, presuming an equal representation of C3 and C4 crops [8]. In contrast, this study employs the CDL dataset specifically categorizing C4 crops [50]. All of these improvements are aimed at enhancing the accuracy of GPP estimation results.

C. Uncertainty

Although we have merged images from Landsat-8/9 and Sentinel-2A/B, the presence of clouds and cloud shadows makes it challenging to guarantee a satellite image for each pixel every 10 days [42]. In cases of missing images, we employ linear interpolation for filling, but this method only ensures the continuity of the data time series. Linear interpolation may lead to underestimation or overestimation of data at turning points, thereby causing GPP to be too large or too small, introducing certain uncertainties. Additionally, for the C4 crop, our primary focus is on maize, which may add uncertainties in the computation of other C4 crops. Moreover, the ISLSCP II C4 dataset is quite old and has a lower spatial resolution, adding uncertainties, yet it is the best estimate of C4 vegetation distribution currently available. Furthermore, the absence of ERA5 data in coastal areas [37], such as around the city of Dalian, results in the lack of GPP for some pixels in these regions.

V. Conclusion

Accurate estimation of GPP in terrestrial ecosystems is pivotal for enhancing our understanding of the ecosystem carbon cycle and aiding climate change predictions. This study leverages optical satellite data from Landsat-8/9 and Sentinel-2A/B, combined with ERA5 meteorological data, to generate a GPP dataset for Northeast China in 2022. Utilizing the VPM model, we achieved a 30-m spatial resolution, offering a refined and reliable assessment of GPP for the region's natural vegetation and CRO. Validation against SIF data and comparison with other GPP datasets confirmed the robustness of our results.

Our research underscores the potential of utilizing Landsat-8/9 and Sentinel-2A/B optical satellite data for crafting high spatial resolution GPP products. Such an approach is particularly beneficial for capturing GPP dynamics in areas characterized by high heterogeneity and fragmented CRO. This work sets a precedent for future high-resolution GPP product development

on a global scale. Looking ahead, the anticipated increase in Landsat-9 data availability will facilitate more comprehensive analyses of temporal and spatial variations in plant GPP, offering deeper insights into ecosystem processes over extended periods.

REFERENCES

[1] A. Á. Ballantyne, C. Á. Alden, J. Á. Miller, P. Á. Tans, and J. White, "Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years," *Nature*, vol. 488, no. 7409, pp. 70–72, 2012.

[2] P. Ciais et al., "Europe-wide reduction in primary productivity caused by the heat and drought in 2003," *Nature*, vol. 437, no. 7058, pp. 529–533, 2005.

[3] Y. Zhang et al., "Precipitation and carbon-water coupling jointly control the interannual variability of global land gross primary production," *Sci. Rep.*, vol. 6, no. 1, 2016, Art. no. 39748.

[4] A. Ahlström et al., "The dominant role of semi-arid ecosystems in the trend and variability of the land CO_2 sink," *Science*, vol. 348, no. 6237, pp. 895–899, 2015.

[5] Y. Zhang, X. Xiao, S. Zhou, P. Ciais, H. McCarthy, and Y. Luo, "Canopy and physiological controls of GPP during drought and heat wave," *Geophys. Res. Lett.*, vol. 43, no. 7, pp. 3325–3333, 2016.

[6] C. Beer et al., "Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate," *Science*, vol. 329, no. 5993, pp. 834–838, 2010.

[7] G. R. Yu et al., "Spatial patterns and climate drivers of carbon fluxes in terrestrial ecosystems of China," *Glob. Change Biol.*, vol. 19, no. 3, pp. 798–810, 2013.

[8] Y. Zhang et al., "A global moderate resolution dataset of gross primary production of vegetation for 2000–2016," *Sci. Data*, vol. 4, no. 1, pp. 1–13, 2017.

[9] B. D. Stocker, J. Zscheischler, T. F. Keenan, I. C. Prentice, S. I. Seneviratne, and J. Peñuelas, "Drought impacts on terrestrial primary production underestimated by satellite monitoring," *Nature Geosci.*, vol. 12, no. 4, pp. 264–270, 2019.

[10] L. Hu et al., "COS-derived GPP relationships with temperature and light help explain high-latitude atmospheric CO_2 seasonal cycle amplification," *Proc. Nat. Acad. Sci.*, vol. 118, no. 33, 2021, Art. no. e2103423118.

[11] G. D. Farquhar, S. V. von Caemmerer, and J. A. Berry, "A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species," *Planta*, vol. 149, pp. 78–90, 1980.

[12] D. De Pury and G. Farquhar, "Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models," *Plant, Cell Environ.*, vol. 20, no. 5, pp. 537–557, 1997.

[13] Y. Zhang et al., "Development of a coupled carbon and water model for estimating global gross primary productivity and evapotranspiration based on eddy flux and remote sensing data," *Agricultural Forest Meteorol.*, vol. 223, pp. 116–131, 2016.

[14] S. W. Running, R. R. Nemani, F. A. Heinsch, M. Zhao, M. Reeves, and H. Hashimoto, "A continuous satellite-derived measure of global terrestrial primary production," *Bioscience*, vol. 54, no. 6, pp. 547–560, 2004.

[15] W. Bi et al., "A global 0.05 dataset for gross primary production of sunlit and shaded vegetation canopies from 1992 to 2020," *Sci. Data*, vol. 9, no. 1, 2022, Art. no. 213.

[16] X.-J. Zhu et al., "Mapping Chinese annual gross primary productivity with eddy covariance measurements and machine learning," *Sci. Total Environ.*, vol. 857, 2023, Art. no. 159390.

[17] J. Xiao et al., "Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data," *Agricultural Forest Meteorol.*, vol. 148, no. 11, pp. 1827–1847, 2008.

[18] G. Tramontana et al., "Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms," *Biogeosciences*, vol. 13, no. 14, pp. 4291–4313, 2016.

[19] L. Guanter et al., "Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence," *Proc. Nat. Acad. Sci.*, vol. 111, no. 14, pp. E1327–E1333, 2014.

[20] R. Chen, L. Liu, X. Liu, and U. Rascher, "CMLR: A mechanistic global GPP dataset derived from TROPOMIS SIF observations," *J. Remote Sens.*, vol. 4, 2024, Art. no. 0127.

[21] J. Wu et al., "Leaf development and demography explain photosynthetic seasonality in Amazon evergreen forests," *Science*, vol. 351, no. 6276, pp. 972–976, 2016.

[22] J. Kattge, W. Knorr, T. Raddatz, and C. Wirth, "Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models," *Glob. Change Biol.*, vol. 15, no. 4, pp. 976–991, 2009.

[23] R. Houborg, A. Cescatti, M. Migliavacca, and W. Kustas, "Satellite retrievals of leaf chlorophyll and photosynthetic capacity for improved modeling of GPP," *Agricultural Forest Meteorol.*, vol. 177, pp. 10–23, 2013.

[24] G. Piñeiro, M. Oesterheld, and J. M. Paruelo, "Seasonal variation in aboveground production and radiation-use efficiency of temperate rangelands estimated through remote sensing," *Ecosystems*, vol. 9, pp. 357–373, 2006.

[25] Q. Zhang, E. M. Middleton, H. A. Margolis, G. G. Drolet, A. A. Barr, and T. A. Black, "Can a satellite-derived estimate of the fraction of PAR absorbed by chlorophyll (FAPARchl) improve predictions of light-use efficiency and ecosystem photosynthesis for a boreal aspen forest?," *Remote Sens. Environ.*, vol. 113, no. 4, pp. 880–888, 2009.

[26] Q. Zhang et al., "Estimation of crop gross primary production (GPP): FAPARchl versus MOD15A2 FPAR," *Remote Sens. Environ.*, vol. 153, pp. 1–6, 2014.

[27] Y. Zhang et al., "Consistency between sun-induced chlorophyll fluorescence and gross primary production of vegetation in North America," *Remote Sens. Environ.*, vol. 183, pp. 154–169, 2016.

[28] S. Zhou et al., "Dominant role of plant physiology in trend and variability of gross primary productivity in North America," *Sci. Rep.*, vol. 7, no. 1, 2017, Art. no. 41366.

[29] L. Qiao and H. Xia, "The impact of drought time scales and characteristics on gross primary productivity in China from 2001 to 2020," *Geo-Spatial Inf. Sci.*, pp. 1–19, 2024.

[30] Y. Zhang et al., "Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017," *Remote Sens. Environ.*, vol. 222, pp. 165–182, 2019.

[31] X. Li and J. Xiao, "Mapping photosynthesis solely from solar-induced chlorophyll fluorescence: A global, fine-resolution dataset of gross primary production derived from OCO-2," *Remote Sens.*, vol. 11, no. 21, 2019, Art. no. 2563.

[32] L. Yang, L. Wang, J. Huang, L. R. Mansaray, and R. Mijiti, "Monitoring policy-driven crop area adjustments in northeast China using Landsat-8 imagery," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 82, 2019, Art. no. 101892.

[33] Y. Guo, H. Xia, X. Zhao, L. Qiao, Q. Du, and Y. Qin, "Early-season mapping of winter wheat and garlic in Huaihe basin using Sentinel-1/2 and Landsat-7/8 imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8809–8817, 2023.

[34] L. Guanter et al., "Using field spectroscopy to assess the potential of statistical approaches for the retrieval of sun-induced chlorophyll fluorescence from ground and space," *Remote Sens. Environ.*, vol. 133, pp. 52–61, 2013.

[35] M. Rossini et al., "Red and far red sun-induced chlorophyll fluorescence as a measure of plant photosynthesis," *Geophys. Res. Lett.*, vol. 42, no. 6, pp. 1632–1639, 2015.

[36] X. Yang et al., "Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a temperate deciduous forest," *Geophys. Res. Lett.*, vol. 42, no. 8, pp. 2977–2987, 2015.

[37] J. Muñoz-Sabater et al., "ERA5-Land: A state-of-the-art global reanalysis dataset for land applications," *Earth Syst. Sci. Data*, vol. 13, no. 9, pp. 4349–4383, 2021.

[38] Y. Ryu, C. Jiang, H. Kobayashi, and M. Detto, "MODIS-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000," *Remote Sens. Environ.*, vol. 204, pp. 812–825, 2018.

[39] D. Meek, J. Hatfield, T. Howell, S. Idso, and R. Reginato, "A generalized relationship between photosynthetically active radiation and solar radiation 1," *Agron. J.*, vol. 76, no. 6, pp. 939–945, 1984.

[40] D. G. Dye, "Spectral composition and quanta-to-energy ratio of diffuse photosynthetically active radiation under diverse cloud conditions," *J. Geophysical Res., Atmos.*, vol. 109, no. D10203, 2004.

[41] Z. Jin et al., "Smallholder maize area and yield mapping at national scales with Google Earth engine," *Remote Sens. Environ.*, vol. 228, pp. 115–128, Jul. 2019, doi: 10.1016/j.rse.2019.04.016.

[42] E. Mandanici and G. Bitelli, "Preliminary comparison of Sentinel-2 and Landsat 8 imagery for a combined use," *Remote Sens.*, vol. 8, no. 12, 2016, Art. no. 1014. [Online]. Available: https://www.mdpi.com/2072-4292/8/12/1014

[43] D. P. Roy et al., "Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity," *Remote Sens. Environ.*, vol. 185, pp. 57–70, 2016.

[44] H. K. Zhang et al., "Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences," *Remote Sens. Environ.*, vol. 215, pp. 482–494, Sep. 2018, doi: 10.1016/j.rse.2018.04.031.
[45] J. Yang and X. Huang, "The 30 m annual land cover dataset and its dynamics in China from 1990 to 2019," *Earth System Sci. Data*, vol. 13, no. 8, pp. 3907–3925, 2021.
[46] X. Zhang, L. Liu, X. Chen, Y. Gao, S. Xie, and J. Mi, "GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery," *Earth Syst. Sci. Data*, vol. 13, no. 6, pp. 2753–2776, 2021, doi: 10.5194/essd-13-2753-2021.
[47] C. Zhang, J. Dong, and Q. Ge, "Quantifying the accuracies of six 30-m cropland datasets over China: A comparison and evaluation analysis," *Comput. Electron. Agriculture*, vol. 197, 2022, Art. no. 106946.
[48] A. Huete, K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira, "Overview of the radiometric and biophysical performance of the MODIS vegetation indices," *Remote Sens. Environ.*, vol. 83, no. 1, pp. 195–213, Nov. 2002, doi: 10.1016/S0034-4257(02)00096-2.
[49] X. Xiao et al., "Observation of flooding and rice transplanting of paddy rice fields at the site to landscape scales in China using VEGETATION sensor data," *Int. J. Remote Sens.*, vol. 23, no. 15, pp. 3009–3022, 2002.
[50] N. You et al., "The 10-m crop type maps in Northeast China during 2017–2019," *Sci. Data*, vol. 8, no. 1, 2021, Art. no. 41.
[51] C. J. Still, J. A. Berry, G. J. Collatz, and R. S. DeFries, "Global distribution of C3 and C4 vegetation: Carbon cycle implications," *Glob. Biogeochem. Cycles*, vol. 17, no. 1, pp. 6-1–6-14, 2003.
[52] C. Jin et al., "Effects of in-situ and reanalysis climate data on estimation of cropland gross primary production using the Vegetation Photosynthesis Model," *Agricultural Forest Meteorol.*, vol. 213, pp. 240–250, 2015.
[53] H. Yan et al., "Improved global simulations of gross primary product based on a new definition of water stress factor and a separate treatment of C3 and C4 plants," *Ecol. Model.*, vol. 297, pp. 42–59, 2015.
[54] Y. Zheng et al., "Improved estimate of global gross primary production for reproducing its long-term variation, 1982–2017," *Earth Syst. Sci. Data*, vol. 12, no. 4, pp. 2725–2746, 2020, doi: 10.5194/essd-12-2725-2020.
[55] R. Li, H. Xia, X. Zhao, and Y. Guo, "Mapping evergreen forests using new phenology index, time series Sentinel-1/2 and Google Earth Engine," *Ecol. Indicators*, vol. 149, 2023, Art. no. 110157.

Xiaoyan Ma received the B.S. degree, in 2021, in geographic information science from the Henan University, Kaifeng, China, where he is currently working toward the master's degree in civil engineering and water conservancy.

His research interests include agricultural remote sensing, vegetation phenology, and climate change.

Li Pan received the M.S. degree in civil engineering and water conservancy from the Henan University, Kaifeng, China, in 2022. He is currently working toward the Ph.D. degree in center for earth observation and modeling with University of Oklahoma, Norman, OK, USA.

His research interests focuses on quantitative remote sensing, with special focuses on vegetation phenology and climate change.

Haoming Xia received the Ph.D. degree in physical geography from the Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, China, in 2016.

He is currently an Associate Professor with the College of Geography and Environmental Science, Henan University, Zhengzhou, China. His research interests include quantitative remote sensing, with special focuses on land cover mapping, vegetation phenology, and climate change.