CNSIF: A reconstructed monthly 500-m spatial resolution solar-induced chlorophyll fluorescence dataset in China

Kaiqi Du[1,2], Guilong Xiao[1,2], Jianxi Huang[1,2,3*], Xiaoyan Kang[4], Xuecao Li[1,2], Yelu Zeng[1,2], Quandi Niu[1,2],

5 Haixiang Guan[1,2], Jianjian Song[1,2]

[1] College of Land Science and Technology, China Agricultural University, Beijing 100083, China

[2] Key Laboratory of Remote Sensing for Agri-Hazards, Ministry of Agriculture and Rural Affairs, Beijing 100083, China

[3] Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu 611756, China

10 [4] Key Laboratory of Ecosystem Network Observation and Modeling, Institute of Geographic Sciences and Natural Resources
Research, Chinese Academy of Sciences, Beijing 100101, China

Correspondence to: Jianxi Huang (jxhuang@cau.edu.cn)

Abstract. Satellite-derived solar-induced chlorophyll fluorescence (SIF) offers valuable opportunities for monitoring large-
scale ecosystem functions. However, the inherent trade-off between satellite scan range and spatial resolution, along with

15 incomplete spatial coverage and irregular temporal sampling, limits its broader application. In this study, we developed a 500-
m spatial resolution monthly SIF dataset for the China region (CNSIF) from 2003 to 2022, using a data-driven deep learning
approach based on high-resolution apparent reflectance and thermal infrared data. The results indicate that CNSIF effectively
captures the spatial patterns of vegetation photosynthetic activity and exhibits a positive annual growth trend of 0.054.
Comparisons with tower-based observations validated the ability of CNSIF to track changes in photosynthetic intensity over

20 time across different ecosystems. Furthermore, the strong correlation ($R^2_{2016} = 0.768$, $R^2_{2020} = 0.743$; P<0.001) between
CNSIF and the MODIS monthly Gross Primary Production (GPP) product demonstrates its potential for estimating carbon
flux. CNSIF's higher-resolution estimation of photosynthetic activity offers a promising tool for monitoring vegetation
dynamics across China and estimating fragmented agricultural production. It enables the incorporation of ecosystem
fragmentation effects into earth observation and carbon cycle systems. The CNSIF dataset is available at

25 https://doi.org/10.6084/m9.figshare.27075145 (Du et al., 2024).

1 Introduction

 The photosynthesis proxy capability of solar-induced chlorophyll fluorescence provides us with unprecedented opportunities to examine vegetation dynamics at both sub-daily and interannual scales over large regions (Xiao et al., 2021; Du et al., 2024). This is crucial for studying net-zero carbon emissions, the carbon cycle, and ecosystem functions (Doughty et al., 2022).

5 However, the satellites used to generate SIF products are primarily designed to monitor greenhouse or trace gases in Earth's atmosphere. As a result, they typically feature low spatial resolution or non-continuous scanning patterns (Köhler et al., 2018). Currently, satellite-based SIF products can be broadly categorized into two types based on their swath coverage and spatial resolution. The first category includes SIF products from satellites with narrow swath widths and small footprints. For example, the Chinese Carbon Dioxide Observation Satellite Mission (TanSat) provides SIF products with a spatial resolution of 2 km $\times$

10 2 km (Du et al., 2018; Yao et al., 2021). The Orbiting Carbon Observatory-2 (OCO-2) generates SIF products with a spatial resolution of 1.3 km $\times$ 2.25 km (Frankenberg et al., 2014). Slightly different from OCO-2, the Orbiting Carbon Observatory-3 (OCO-3) serves as a replica of OCO-2, which provides SIF products with a spatial resolution of 1.6 km $\times$ 2.2 km since May 2019 (Eldering et al., 2019). Notably, the TROPOspheric Monitoring Instrument (TROPOMI) has a swath width of 2,600 km and a spatial resolution of 3.5 km $\times$ 5.5 km (Köhler et al., 2018). TROPOSIF stands out for its unique advantage of combining

15 a wide swath with relatively high spatial resolution. The Greenhouse Gases Observing Satellite (GOSAT) provides SIF products with a spatial resolution of 10.5 km $\times$ 10.5 km, but its observation frequency is relatively low (Joiner et al., 2011). The second category consists of SIF products derived from satellites with wide swath coverage but coarser spatial resolution. These include the Global Ozone Monitoring Experiment-2 (GOME-2) SIF product with a spatial resolution of 40km $\times$ 80km (Joiner et al., 2013) and the Scanning Imaging Absorption Spectrometer for Atmospheric CHartographY (SCIAMACHY) SIF

20 product with a spatial resolution of 30km $\times$ 240km (Köhler et al., 2015). These SIF products have garnered significant attention from researchers across various fields, including remote sensing, climate studies, Earth observation system science, and plant physiology (Mohammed et al., 2019; Xiao et al., 2019; Porcar-Castell et al., 2021). However, the inherently low spatial resolution of satellites, as well as narrow and incomplete spatial coverage and irregular temporal sampling, have prevented SIF's broader application (Xiao et al., 2021). There has been some recent work to improve

25 the resolution of existing SIF products through physically-based approaches. Köhler et al. (2018) proposed the first global

0.05 ° TROPOMI SIF observations (Caltech TROPOMI SIF) by relying on the Fraunhofer line principle and singular value decomposition. Zhang et al. (2018) used surface reflectance data and OCO-2 SIF to train a neural network, generating a 0.05 ° spatially continuous SIF (CSIF) product under clear-sky conditions. Li and Xiao (2019), utilizing discrete OCO-2 SIF observations and meteorological reanalysis data, developed a 0.05 ° global SIF dataset (GOSIF) through a data-driven approach.

5 Duveiller et al. (2020) established a semi-empirical relationship between SIF and MODIS explanatory variables using a light-use efficiency model, downscaling GOME-2 SIF from 0.5 ° to 0.05 °. Chen et al. (2022) employed machine learning to reconstruct a 0.05 ° TROPOMI SIF under clear-sky conditions (RTSIF), extending the temporal coverage of the dataset to 20 years. Gensheimer et al. (2022) developed a convolutional neural network called SIFnet to enhance the spatial resolution of TROPOSIF by leveraging coarse SIF observations and high-resolution auxiliary data carrying GPP-related information. Zou

10 et al. (2024) used a pseudo-invariant calibration method to correct the temporal degradation of the GOME-2A instrument, generating a 0.05 ° global GOME-2A SIF dataset (TCSIF) suitable for time trend analysis.

However, these SIF data are still coarse for many research areas. Although most of the downscaled SIF datasets have a spatial resolution of 0.05 °, there has been a persistent effort to estimate SIF at finer spatial scales due to the presence of fine-scale phenomena such as ecosystem fragmentation and vegetation cover diversification in the earth observation system and the

15 carbon cycle system (Brudvig et al., 2015). Zhang et al. (2020a) used a data-driven approach combining optical, SIF, and environmental data to predict maize yields, finding that coarse-resolution SIF products (0.05 °–1 °) could only capture a small portion of the spatial characteristics of crops. Turner et al. (2020) applied an oversampling method to obtain 500-m spatial resolution SIF data for California, demonstrating strong consistency with tower-based GPP measurements. Zhang et al. (2021c) further downscaled the 0.05 ° GOSIF product to obtain SIF estimates with a spatial resolution of 0.008 °. They combined it with

20 surface temperature data to monitor agricultural drought during the maize growing season. Kang et al. (2023) explored the spatiotemporal prediction-driving mechanism of SIF, downscaling the monthly GOSIF product to plot size, and successfully predicting cotton yields at the field scale one to two months before harvest.

Additionally, most large-scale vegetation photosynthesis studies have focused on North or South America (Turner et al., 2021; Zhang et al., 2023; Li et al., 2023). For example, Turner et al. (2021) reconstructed 500-m SIF estimates for the contiguous

25 United States, observing annual GPP anomalies driven by extreme climate events, again circumstantially demonstrating the

linear relationship between SIF and GPP. Zhang et al. (2023) constructed an afternoon photosynthetic inhibition index based on OCO-3 SIF, revealing that during the 2020 drought in the southwestern United States, vegetation sensitivity to vapor pressure deficit increased as the drought intensified. Their findings offer new insights into the mechanisms of plant mortality and how they affect the terrestrial carbon and water cycles. Li et al. (2023), using environmental satellite observations

5   combined with OCO-3 SIF, observed that photosynthetic responses to the 2020 U.S. heatwave varied across vegetation types and drought gradients, underscoring the need to consider the asymmetric increase in GPP when evaluating carbon-climate interactions under environmental stress. These studies have deepened our understanding of vegetation photosynthetic responses to environmental stress in specific ecosystems across the Americas. However, there is a pressing need to investigate photosynthetic changes across diverse vegetation types in a broader range of ecosystems. In particular, a more comprehensive

10   understanding of vegetation dynamics in the East Asian region is needed, covering a variety of ecosystems such as drylands in northern China, paddy fields in southern China, subtropical forests, alpine meadows on the Tibetan Plateau, and estuarine wetlands.

Advancing the development of satellite-based SIF datasets is essential for their global-scale application. However, despite the continuous production of satellite-based SIF products in recent years, no publicly available long-term, high-resolution SIF

15   dataset for the East Asian region has been available so far due to the limitations of the spatial resolution of SIF-related explanatory variables. This study aims to provide a monthly 500-m spatial resolution SIF dataset for the China region from 2003 to 2022, offering a comprehensive understanding of historical vegetation dynamics in the East Asian region. We applied a data-driven deep learning approach to develop a China regional high-resolution SIF dataset (CNSIF), based on high-resolution surface reflectance and thermal infrared data from the Landsat 7 / 8 and Sentinel-2 satellites. The performance of

20   CNSIF was rigorously evaluated using tower-based SIF observations, tower-based GPP observations, the MODIS monthly GPP product, and other similar SIF products. The long-term, high-resolution SIF dataset generated here provides a promising tool for monitoring vegetation dynamics in the China region, which will improve our understanding of vegetation photosynthetic activity in the East Asian region.

2 Materials and methods

2.1 Datasets for the generation of CNSIF

The remote sensing data we collected in this study are shown in Table 1.

GOSIF is a 0.05 ° SIF product obtained by Li and Xiao (2019) using a machine learning approach to fill in the discrete OCO-

5 2 SIF dataset based on MODIS EVI data and meteorological reanalysis data from MERRA-2. They used the Cubist regression

tree model to create a rule-based predictive model from OCO-2 SIF training samples, with the final prediction being the

average of multiple output values. This method produced comparable results with other machine learning approaches and has

been widely applied to estimate carbon fluxes and biophysical variables. This dataset exhibits similar seasonality to OCO-2

SIF and has a remarkable ability to highlight the crop regions with the highest global daily productivity. Therefore, we used

10 the monthly GOSIF dataset as the reference SIF for the CNN-driven downscaling method.

The surface reflectance and thermal infrared data were sourced from the Enhanced Thematic Mapper Plus (ETM+) sensor on

the Landsat 7 satellite, the Thermal Infrared Sensor (TIRS) on the Landsat 8 satellite, and the Multispectral Instrument (MSI)

on the Sentinel-2 satellite. The limitations of data coverage, the temporal span of the CNSIF dataset, and the advantages of

new sensors were considered. We collected visible and near-infrared reflectance data from ETM+ sensors, along with thermal

15 infrared data, covering the period from January 2003 to March 2017. Subsequently, we gathered visible and near-infrared

reflectance data from MSI sensors and thermal infrared data from TIRS sensors for the period from April 2017 to December

2022.

The MODIS Version 6.1 land cover product (MCD12Q1) provides annual global land cover maps at the 500-m spatial

resolution, with classification accuracy remaining stable over a long period. Therefore, we used the International Geosphere-

20 Biosphere Programme (IGBP) classification layer from MCD12Q1 to extract land cover types.

2.2 Tower-based measurements

2.2.1 Tower-based SIF for the verification of CNSIF

Ground validation is a necessary step in verifying the accuracy of almost all remote sensing products. Therefore, we used

tower-based SIF observations to validate the CNSIF dataset. The tower-based SIF validation data were collected from nine

25 flux sites in the ChinaSpec network (AR, DM, GC, HL, JR, PYH, SQ, XTS, and YX) (Du et al., 2023b; Du et al., 2019; Zhang

et al., 2020b; Li et al., 2020), which include some publicly available SIF observations from 2017 to 2021. The locations of these sites are shown in Figure 1, and site information is detailed in Table 2.

The nine tower-based SIF observation sites are distributed across various regions of China, with the main observed land cover types including four drylands, one paddy field, two grasslands, and two wetlands. Among them, the DM site grows dryland maize only in the summer, while the SQ site practices irrigated maize cultivation. The GC and XTS sites implement wheat-maize crop rotation, while the JR site rotates rice and wheat, growing rice in the summer and wheat in the winter. The underlying surface of the AR site consisted of alpine grassland, while the HL site featured a mixture of grassland and a few begonia trees. The underlying surface observed at the PYH site was tussock grass, which represents the characteristics of the ecological wetland of a freshwater lake. The underlying surface observed at the YX site was mangrove forest, which means the characteristics of the environmental type of subtropical mangrove forests and the transition zone of land-sea intersection.

2.2.2 Tower-based GPP for the verification of CNSIF

Tower-based GPP validation data were collected from 10 flux sites within China (Cha, Cng, Dan, Din, Du2, Du3, Ha2, HaM, Qia, and Sw2) as part of the FLUXNET 2015 network, which includes some publicly available GPP measurements from 2003 to 2012 (Pastorello et al., 2020). The locations of these sites are displayed in Figure 1, with detailed site information provided in Table 3.

The ten flux sites are distributed across various ecological types, including temperate mixed needle and broadleaf forests, subtropical evergreen broadleaf forests, artificial evergreen coniferous forests, typical grasslands, alpine meadows, desert steppes, and degraded grassland ecosystems.

2.3 Data processing and SIF retrieval method

2.3.1 Data processing for satellite observation

To obtain monthly data, we processed the collected high-resolution surface reflectance and thermal infrared data. First, the thermal infrared band data were processed into land surface temperature (LST) data based on the Google Earth Engine (GEE) platform (Ermida et al., 2020). Second, we performed a cloud masking process using a 65% cloud probability threshold to generate clean pixels without cloud shadows for each image (Pazur et al., 2022; Xu et al., 2022). Next, we calculated the

median surface reflectance (SR) and median LST on a monthly basis. Finally, we applied a Savitzky-Golay (SG) filter with a

window size of 7 and a polynomial order of 2 to further eliminate null values and outliers (You and Dong, 2020; You et al.,

2021). The processed SR and LST data were then resampled to 50-m and 500-m spatial resolutions, respectively.

Recent studies have shown that near-infrared reflectance of vegetation (NIRv) and the enhanced vegetation index (EVI) can

5   serve as proxies for SIF (Du et al., 2023a; Zeng et al., 2022; Bandopadhyay et al., 2021). Additionally, the near-infrared band

plays a crucial role in driving SIF variations in SIF product modeling (Zhang et al., 2018). Therefore, we used reflectance data

from three visible light bands (blue, green, and red) and one near-infrared band to calculate two vegetation indices. Specifically,

we calculated NIRv and EVI using Eq. (1) and (2), respectively, and incorporated these indices along with near-infrared

reflectance (R_{NIR}) and LST as input data for the CNN model in this study.

10

$$\text{NIRv} = R_{NIR} \times \text{NDVI} = R_{NIR} \times [(R_{NIR} - R_{Red})/(R_{NIR} + R_{Red})] \tag{1}$$

$$\text{EVI} = 2.5 \times [(R_{NIR} - R_{Red})/(R_{NIR} + 6 \times R_{Red} - 7.5 \times R_{Blue} + 1)] \tag{2}$$

2.3.2 Data processing for tower-based measurements

The tower-based SIF data collected from the ChinaSpec network includes SIF estimates calculated using various retrieval

methods, such as the 3FID method, spectral fitting method (SFM), and singular value decomposition (SVD) method. Liu and

15   Liu (2015) evaluated the performance of several SIF retrieval methods. They found that the 3FLD method is the most robust

for ground-based observations with a spectral resolution of 0.3 nm. Therefore, we used tower-based SIF values retrieved using

the 3FLD method for validation purposes. The hourly tower-based SIF data were aggregated to a daily timescale, and median

statistics were calculated monthly, resulting in the final monthly tower-based SIF dataset.

The tower-based GPP data collected from FLUXNET 2015 includes GPP estimates for both daytime (GPP_DT_VUT_REF)

20   and nighttime (GPP_NT_VUT_REF) periods, available on daily and monthly scales. The monthly tower-based GPP data were

obtained by calculating the average of the daytime and nighttime GPP values.

Considering the spatial scale differences between the tower-based observations and the spatial resolution (500-m) of the

developed SIF product, we applied the method proposed by Du et al. (2023b) (Eq. 3), using NIRv to bridge the gap between

the tower-based observations and CNSIF samples, thereby enabling more effective validation efforts.

25

$$SIF_{in-situ\_NIRv} = SIF_{in-situ} \times NIRv_{CNSIF}/NIRv_{tower} \tag{3}$$

7

where $SIF_{in\text{-}situ\_NIRv}$ represents the tower-based SIF value scaled by NIRv; $SIF_{in\text{-}situ}$ is the original tower-based SIF observation; $NIRv_{CNSIF}$ refers to the NIRv value of the CNSIF pixel matched to each tower site; and $NIRv_{tower}$ represents the NIRv value of the Sentinel-2 pixel centered on the tower location.

2.3.3 Data-driven-based SIF retrieval method

5   As mentioned earlier, most existing SIF products are developed using machine learning models, which typically train SIF values by selecting multiple features at the same pixel location and resolution (Fig. 2 a-c-d) (Zhang et al., 2018; Li and Xiao, 2019). These models do not account for the internal spatial structure information within the SIF pixel. In contrast, deep learning techniques, due to their deep, nonlinear structures, have recently been recognized as a promising downscaling technology (Sun and Lan, 2021; Jeong et al., 2022; Xu et al., 2021). In the field of remote sensing, several studies have demonstrated that
10   convolutional neural networks (CNNs) can automatically learn and extract features from complex data, thereby enhancing the spatial accuracy of remote sensing data (Sha et al., 2020; Bano-Medina et al., 2021).

We believe that the downscaling design is key to generating high-resolution SIF products. Therefore, we adopted a 2D CNN-based approach, taking into account the relationship between SIF pixels and their internal spatial structure to develop the SIF model (Fig. 2 b-c-d). We trained the model using multi-year (2003-2022) NIRv, R_{NIR}, EVI, LST, and GOSIF data respectively,
15   used the 2019 monthly GOSIF data for validation, and selected the best-performing CNN-driven SIF model as the target model. Finally, through transfer learning, we transferred the target model from a coarse resolution (0.05°) to a finer resolution (500-m) (Fig. 2e). The constructed CNN consists of four convolutional layers (kernel size: 3×3) and one fully connected layer (Fig. 2c). All CNNs were trained using an initial learning rate of 0.001, with stochastic gradient descent with momentum (SGDM) as the optimization algorithm.

20   ## 3 Results

3.1 Model performance

Figure. 3 illustrates the performance of the CNN-driven SIF model trained on 2018 data, with 2019 GOSIF data used for validation. We find that the SIF estimated by the model considering the information of the internal spatial structure of the pixel shows strong agreement with the referenced SIF (Fig. 3a), with a coefficient of determination (R^2) of 0.856 and a root mean

square error (RMSE) of 0.089 mW m^{-2} nm^{-1} sr^{-1}. Compared to other SIF models trained on multi-year data, the deviation between the referenced SIF and estimated SIF is smaller across the entire range. Additionally, the fit slope between the referenced SIF and estimated SIF is close to 1, indicating the absence of systematic bias (Fig. 3b).

3.2 Spatial distribution of the CNSIF dataset

5 Figure. 4 illustrates the spatial distribution of monthly CNSIF for the summer and winter of 2016 and 2020. We observed that the monthly CNSIF effectively captures the spatial patterns of vegetation photosynthetic activity. In July, high SIF values were mainly observed in the temperate coniferous forests and cropping areas of the Northeast Plain in northeastern China, the temperate deciduous broadleaf forests and cropping areas of the Huang-Huai-Hai Plain in central China, and the subtropical evergreen broadleaf forests and tropical rainforests in the southern part of the country (Fig. 4c, d). In December, high SIF

10 values were observed in the evergreen forests and tropical rainforests along the southern coastal regions (Fig. 4e). These two vegetation types maintain photosynthesis throughout the year, resulting in naturally higher SIF values during winter compared to the deciduous forests in the north. In fact, the Huang-Huai-Hai Plain in December also shows a marked SIF difference compared to its surroundings (Fig. 4f). This is because the higher resolution of the CNSIF data is capable of capturing the photosynthetic activity of winter crops, specifically winter wheat, in the region. In addition, the standard errors (σ_{SIF}) of the

15 weighted averages of monthly CNSIF for 2016 and 2020 range from 0 to 0.36 mW m^{-2} sr^{-1} nm^{-1} (Fig. 4g, h). The mean values of the standard errors for the two years are 0.159 mW m^{-2} sr^{-1} nm^{-1} and 0.154 mW m^{-2} sr^{-1} nm^{-1}, respectively, indicating that the error in the monthly gridded averages of CNSIF does not exceed 16%.

The zoomed-in spatial distribution map of CNSIF in the Huang-Huai-Hai Plain region showed a spatial pattern similar to that of GOSIF. However, the 500-m resolution CNSIF offers more effective pixel samples and finer details compared to the 0.05 °

20 resolution GOSIF (Fig. 4a, b), thereby reducing the uncertainty in monitoring vegetation photosynthesis, especially in small-scale or fragmented vegetation areas.

3.3 Temporal pattern in the CNSIF dataset

We further investigated the time series variation of CNSIF and compared it with GOSIF. The results showed that the CNSIF dataset exhibited evident seasonal variation between 2003 and 2022 (Fig. 5c). Additionally, the monthly average SIF in the

China region displayed a positive trend over the past 20 years (CNSIF: 0.054), indicating a gradual increase in vegetation greenness across the region. GOSIF also observed this phenomenon (GOSIF: 0.073), but CNSIF exhibited more significant interannual variability. Additionally, CNSIF showed slightly higher SIF values than GOSIF during the autumn and winter seasons. This may be due to GOSIF's coarse spatial resolution, which is more susceptible to mixed pixels, diluting the strength

5 of the SIF signal. Moreover, in autumn and winter, crops such as winter wheat may still be growing in some regions of China, and CNSIF is able to more accurately detect the photosynthetic activity of vegetation in these small-scale areas.

Figure. 5a shows the time trend differences between CNSIF and GOSIF. We observed that in most regions of China, CNSIF exhibits slower growth compared to GOSIF (purple areas), indicating that the SIF trend in these areas has been overestimated. The reconstructed CNSIF reduces this overestimation, particularly between 42 °N and 55 °N and between 18 °N and 25 °N (Fig.

10 5b). Additionally, compared to GOSIF, CNSIF shows less trend variability between 28 °N and 44 °N.

3.4 Temporal pattern in the CNSIF dataset

Recently, some publicly available tower-based SIF observations have provided valuable opportunities to validate the vegetation dynamics observed in CNSIF. For comparison, we used SIF results derived by the 3FLD method from the ChinaSpec network and aggregated the daily observations into monthly data (Zhang et al., 2021b). The tower-based

15 observations were then scaled using the NIRv method proposed by Du et al. (2023b).

The results show that CNSIF exhibits good agreement with the scaled tower-based SIF observations (Fig. 6h), with an R^2 of 0.489 and an RMSE of 0.149 mW m^{-2} sr^{-1} nm^{-1}. We observed that scatter points are more densely concentrated in the low SIF value range (0 to 0.2 mW m^{-2} sr^{-1} nm^{-1}) and the mid-value range (0.5 to 0.7 mW m^{-2} sr^{-1} nm^{-1}). This is because the dominant vegetation types at the nine observation sites are grasslands, farmlands, and wetlands, with observations restricted to periods

20 of pronounced vegetation phenology (grasslands: May-September, farmlands: April-October, wetlands: year-round). This also validates the accuracy of CNSIF.

Additionally, we validated CNSIF's ability to monitor vegetation phenology changes. We compared tower-based SIF data from sites with two or more observation periods against CNSIF (Fig. 6 a-g). The underlying surfaces of the seven sites included one alpine meadow, one grass-tree mixture, two wetlands, and three farmlands. We found that CNSIF was slightly underestimated

25 compared to the tower-based SIF. However, CNSIF successfully captured the dynamic trends of different vegetation types at

six sites and accurately reproduced the seasonal patterns of enhanced photosynthesis in spring and summer and reduced photosynthesis in autumn and winter (Fig. 6a, c-g). However, the comparison results for the HL site were significantly worse than those for the other sites (Fig. 6b). Ignoring differences in the number of tower-based observation samples, we attribute this to the vegetation heterogeneity of the HL site's underlying surface. The proportion of the primary land cover type at the

5   HL site is only 0.25 (Table 2), with the observed surface being a mix of grass and a few crabapple trees. In contrast, the proportion of the primary land cover type at the other six sites is much higher than at HL.

3.5 Comparison of CNSIF with tower-based GPP

Considering the spatial scale differences between flux tower observations and the CNSIF dataset, we utilized the 30m resolution GLC_FCS30 land cover product (Zhang et al., 2021a) to select flux tower sites where the land cover type within

10   the CNSIF grid was homogeneous (with the proportion of the primary land cover type being greater than or equal to 0.5). The observation periods at these flux sites span from 2003 to 2012.

The strong linear relationship between CNSIF and tower-based GPP (Fig. 7k, R^2=0.530) demonstrates the reliability of the dataset we developed. They are closely correlated on a monthly timescale. Additionally, we compared the relationship between CNSIF and tower-based GPP across different biomes (underlying surfaces of various observation sites) (Fig. 7a-j). Except for

15   the degraded grassland, CNSIF was significantly correlated with monthly GPP across nearly all biomes, demonstrating its applicability across different ecosystem types. The degraded grassland, characterized by its unique vegetation-soil degradation synergy, experiences declines in productivity, height, and cover as the degree of degradation intensifies. As a result, the GPP values observed at the Du3 and Sw2 flux sites (-1 to 2) are the lowest among the 10 sites. Despite this, CNSIF and GPP still exhibited significant correlations in both the degraded grassland (Fig. 7f) and the desert grassland zonal vegetation biome (Fig.

20   7j). Additionally, we observed larger slopes in the mixed needle-broadleaf forest (Fig. 7a, slope=0.053), evergreen broadleaf forest (Fig. 7d, slope=0.078), and artificial coniferous forest (Fig. 7i, slope=0.065). This is due to the canopy structure of forests being different from that of grasslands, and the stronger reabsorption of SIF caused the larger slopes.

3.6 Comparison of CNSIF with MODIS GPP products

We applied the mean compositing method to synthesize the 8-day MODIS GPP product (MOD17A2H) into monthly data to evaluate CNSIF's ability to monitor ecosystem function across different periods.

We conducted a correlation analysis between the CNSIF and GPP values for each pixel location across different months. The results indicated that the downscaled CNSIF has a strong correlation with monthly GPP. The average R^2 between CNSIF and GPP was 0.768 in 2016 (Fig. 8a) and 0.743 in 2020 (Fig. 8b). Additionally, over 62.9% and 57.4% of pixels across the study area had an R^2 greater than 0.80 in these two independent years, respectively. However, the R^2 between CNSIF and GPP is relatively low in southern China. The growth cycles and biomass production capacities of various vegetation types, from subtropical evergreen forests to seasonal rainforests, differ significantly. In particular, during drought seasons, severe meteorological stress events caused by prolonged heat and low rainfall may not always be consistently reflected in GPP data.

4 Discussion

4.1 Reducing the uncertainty of the downscaling method

In this study, we used the spatiotemporally continuous GOSIF product, which has been successfully applied to research in various ecological fields, as the data source for downscaling SIF. During model training, we found that the size of the input data significantly impacts the model's performance.

In fact, we experimented with four different input data formats: 7×7 and 9×9 pixel sizes that considered only internal spatial structure, and 11×11 and 13×13 pixel sizes that took both internal and neighboring information into account. Obviously, the results trained with different input data sizes revealed significant differences in the spatial distribution of SIF. The closer the input data size is to the pixel size of the data source, the higher the fitting accuracy of the downscaling model, and the closer the fitting slope is to 1 (Fig. 9b, Slope = 1.002; Fig. 9c, Slope = 0.971). Further comparison between models trained with 9×9 and 11×11 pixel sizes revealed that the SIF model considering only internal information performed significantly better than the model that also incorporated neighboring information (9×9 pixels: $R^2 = 0.856$, RMSE = 0.089; 11×11 pixels: $R^2 = 0.812$, RMSE = 0.126). Therefore, this study ultimately selected the 9×9 pixel, which maximizes the inclusion of internal information,

as the input data size for the model. This choice enables the model to converge more easily and achieve higher estimation performance.

4.2 Mitigating the spatial scale differences between CNSIF and tower-based SIF

There is a significant difference in spatial coverage between satellite-based and tower-based observations. This spatial mismatch increases the difficulty of ground validation for nearly all remote sensing products. Previous studies have demonstrated a strong linear correlation between SIF and NIRv (Du et al., 2023a; Zeng et al., 2019; Badgley et al., 2019; Dechant et al., 2022). Therefore, we attempted to use NIRv as a bridge to reduce the uncertainties caused by spatial scale differences.

As shown in Figure. 4, CNSIF exhibits good agreement with the scaled tower-based SIF observations and successfully captures the dynamic trends of photosynthetic activity across different vegetation types. The only exception is the HL site, where vegetation heterogeneity in the underlying surface led to lower validation accuracy. Therefore, using the NIRv-based method to scale tower-based SIF to match the CNSIF grid size is a reasonable approach.

5 Data availability

The monthly SIF dataset for the China region at 500-m spatial resolution from 2003 to 2022 (CNSIF) is publicly available at https://doi.org/10.6084/m9.figshare.27075145 (Du et al., 2024). The CNSIF dataset is stored in TIFF format and organized chronologically. The files are named CNSIF_<YYYY>_<MM>.tif, where "YYYY" and "MM" represent the year and month, respectively.

6 Conclusions

We conducted a comprehensive evaluation of the reconstructed CNSIF dataset and validated its performance. Overall, CNSIF exhibits a positive trend (Slope = 0.054), shows clear seasonality over time, and demonstrates more significant interannual variability. Spatially, it effectively captures the patterns of vegetation photosynthetic activity, with the mean standard error of the monthly CNSIF grid controlled within 16%. By comparing with tower-based observations, CNSIF successfully captured the dynamic trends of various vegetation types and accurately reproduced seasonal variations in photosynthetic intensity. The

strong correlations between tower-based GPP, MODIS GPP, and CNSIF further demonstrate CNSIF's applicability across different biomes. Additionally, the reconstructed CNSIF mitigated the overestimation of SIF trend changes in the China region. Overall, the CNSIF dataset, with its higher-resolution estimation of photosynthetic activity, offers a promising tool for monitoring historical vegetation dynamics in the China region. It allows us to incorporate the effects of ecosystem

5 fragmentation into global carbon cycle estimates and more refined biosphere models.

Author contributions

KD and XJ designed the experiments, and KD carried them out. KD, XK and GX developed the model code and generated the products. KD, XL, YZ, QN, HG and JS prepared the manuscript with contributions.

Competing interests

10 One of the (co-)authors is a member of the editorial board of ESSD (Xuecao Li).

Acknowledgments

Financial support

References

Badgley, G., Anderegg, L. D. L., Berry, J. A., and Field, C. B.: Terrestrial gross primary production: Using NIRv to scale from site to globe, Global Change Biol., 25, 3731-3740, https://doi.org/10.1111/gcb.14729, 2019.

Bandopadhyay, S., Rastogi, A., Cogliati, S., Rascher, U., Gabka, M., and Juszczak, R.: Can Vegetation Indices Serve as Proxies for Potential
20 Sun-Induced Fluorescence (SIF)? A Fuzzy Simulation Approach on Airborne Imaging Spectroscopy Data, Remote Sens., 13, https://doi.org/10.3390/rs13132545, 2021.

Bano-Medina, J., Manzanas, R., and Manuel Gutierrez, J.: On the suitability of deep convolutional neural networks for continental-wide downscaling of climate change projections, Clim Dynam., 57, 2941-2951, https://doi.org/10.1007/s00382-021-05847-0, 2021.

Brudvig, L. A., Damschen, E. I., Haddad, N. M., Levey, D. J., and Tewksbury, J. J.: The influence of habitat fragmentation on multiple
25 plant-animal interactions and plant reproduction, Ecology, 96, 2669-2678, https://doi.org/10.1890/14-2275.1.sm, 2015.

Chen, X., Huang, Y., Nie, C., Zhang, S., Wang, G., Chen, S., and Chen, Z.: A long-term reconstructed TROPOMI solar-induced fluorescence dataset using machine learning algorithms, Sci. Data, 9, https://doi.org/10.1038/s41597-022-01520-1, 2022.

Dechant, B., Ryu, Y., Badgley, G., Kohler, P., Rascher, U., Migliavacca, M., Zhang, Y., Tagliabue, G., Guan, K., Rossini, M., Goulas, Y., Zeng, Y., Frankenberg, C., and Berry, J. A.: NIR$_v$P: A robust structural proxy for sun-induced chlorophyll fluorescence and photosynthesis across scales, Remote Sens. Environ., 268, https://doi.org/10.1016/j.rse.2021.112763, 2022.

Doughty, R., Kurosu, T. P., Parazoo, N., Kohler, P., Wang, Y., Sun, Y., and Frankenberg, C.: Global GOSAT, OCO-2, and OCO-3 solar-induced chlorophyll fluorescence datasets, Earth Syst. Sci. Data, 14, 1513-1529, https://doi.org/10.5194/essd-14-1513-2022, 2022.

Du, K., Huang, J., Wang, W., Zeng, Y., Li, X., and Zhao, F.: Monitoring Low-Temperature Stress in Winter Wheat Using TROPOMI Solar-Induced Chlorophyll Fluorescence, Ieee Trans. Geosci., 62, https://doi.org/10.1109/tgrs.2024.3351141, 2024.

Du, K., Jing, X., Zeng, Y., Ye, Q., Li, B., and Huang, J.: An Improved Approach to Monitoring Wheat Stripe Rust with Sun-Induced Chlorophyll Fluorescence, Remote Sens., 15, https://doi.org/10.3390/rs15030693, 2023a.

Du, K., Xiao, G., Huang, J., Kang, X., Li, X., Zeng, Y., Niu, Q., Guan, H., and Song, J.: CNSIF: A reconstructed monthly 500-m spatial resolution solar-induced chlorophyll fluorescence dataset in China, Figshare, https://doi.org/10.6084/m9.figshare.27075145, 2024.

Du, S., Liu, X., Chen, J., Duan, W., and Liu, L.: Addressing validation challenges for TROPOMI solar-induced chlorophyll fluorescence products using tower-based measurements and an NIRv-scaled approach, Remote Sens. Environ., 290, https://doi.org/10.1016/j.rse.2023.113547, 2023b.

Du, S., Liu, L., Liu, X., Guo, J., Hu, J., Wang, S., and Zhang, Y.: SIFSpec: Measuring Solar-Induced Chlorophyll Fluorescence Observations for Remote Sensing of Photosynthesis, Sensors, 19, https://doi.org/10.3390/s19133009, 2019.

Du, S., Liu, L., Liu, X., Zhang, X., Zhang, X., Bi, Y., and Zhang, L.: Retrieval of global terrestrial solar-induced chlorophyll fluorescence from TanSat satellite, Sci. Bull., 63, 1502-1512, https://doi.org/10.1016/j.scib.2018.10.003, 2018.

Duveiller, G., Filipponi, F., Walther, S., Kohler, P., Frankenberg, C., Guanter, L., and Cescatti, A.: A spatially downscaled sun-induced fluorescence global product for enhanced monitoring of vegetation productivity, Earth Syst. Sci. Data, 12, 1101-1116, https://doi.org/10.5194/essd-12-1101-2020, 2020.

Eldering, A., Taylor, T. E., O'Dell, C. W., and Pavlick, R.: The OCO-3 mission: measurement objectives and expected performance based on 1 year of simulated data, Atmos. Meas. Tech., 12, 2341-2370, https://doi.org/10.5194/amt-12-2341-2019, 2019.

Ermida, S. L., Soares, P., Mantas, V., Goettsche, F.-M., and Trigo, I. E.: Google Earth Engine Open-Source Code for Land Surface Temperature Estimation from the Landsat Series, Remote Sens., 12, https://doi.org/10.3390/rs12091471, 2020.

Frankenberg, C., O'Dell, C., Berry, J., Guanter, L., Joiner, J., Koehler, P., Pollock, R., and Taylor, T. E.: Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2, Remote Sens. Environ., 147, 1-12, https://doi.org/10.1016/j.rse.2014.02.007, 2014.

Gensheimer, J., Turner, A. J., Kohler, P., Frankenberg, C., and Chen, J.: A convolutional neural network for spatial downscaling of satellite-based solar-induced chlorophyll fluorescence (SIFnet), Biogeosciences, 19, 1777-1793, https://doi.org/10.5194/bg-19-1777-2022, 2022.

Jeong, S., Ko, J., and Yeom, J.-M.: Predicting rice yield at pixel scale through synthetic use of crop and deep learning models with satellite data in South and North Korea, Sci. Total Environ., 802, https://doi.org/10.1016/j.scitotenv.2021.149726, 2022.

Joiner, J., Yoshida, Y., Vasilkov, A. P., Yoshida, Y., Corp, L. A., and Middleton, E. M.: First observations of global and seasonal terrestrial chlorophyll fluorescence from space, Biogeosciences, 8, 637-651, https://doi.org/10.5194/bg-8-637-2011, 2011.

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A. P., Middleton, E. M., Huemmrich, K. F., Yoshida, Y., and Frankenberg, C.: Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: methodology, simulations, and application to GOME-2, Atmos. Meas. Tech., 6, 2803-2823, https://doi.org/10.5194/amt-6-2803-2013, 2013.

Kang, X., Huang, C., Zhang, L., Wang, H., Zhang, Z., and Lv, X.: Regional-scale cotton yield forecast via data-driven spatio-temporal prediction (STP) of solar-induced chlorophyll fluorescence (SIF), Remote Sens. Environ., 299, https://doi.org/10.1016/j.rse.2023.113861, 2023.

Köhler, P., Guanter, L., and Joiner, J.: A linear method for the retrieval of sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data, Atmos. Meas. Tech., 8, 2589-2608, https://doi.org/10.5194/amt-8-2589-2015, 2015.

Köhler, P., Frankenberg, C., Magney, T. S., Guanter, L., Joiner, J., and Landgraf, J.: Global Retrievals of Solar-Induced Chlorophyll Fluorescence With TROPOMI: First Results and Intersensor Comparison to OCO-2, Geophys. Res. Lett., 45, 10456-10463, https://doi.org/10.1029/2018gl079031, 2018.

Li, X. and Xiao, J.: A Global, 0.05-Degree Product of Solar-Induced Chlorophyll Fluorescence Derived from OCO-2, MODIS, and Reanalysis Data, Remote Sens., 11, https://doi.org/10.3390/rs11050517, 2019.

Li, X., Ryu, Y., Xiao, J., Dechant, B., Liu, J., Li, B., Jeong, S., and Gentine, P.: New-generation geostationary satellite reveals widespread midday depression in dryland photosynthesis during 2020 western US heatwave, Sci. Adv., 9, https://doi.org/10.1126/sciadv.adi0775, 2023.

Li, Z., Zhang, Q., Li, J., Yang, X., Wu, Y., Zhang, Z., Wang, S., Wang, H., and Zhang, Y.: Solar-induced chlorophyll fluorescence and its link to canopy photosynthesis in maize from continuous ground measurements, Remote Sens. Environ., 236, https://doi.org/10.1016/j.rse.2019.111420, 2020.

Liu, X. and Liu, L.: Improving Chlorophyll Fluorescence Retrieval Using Reflectance Reconstruction Based on Principal Components Analysis, Ieee Geosci. Remote S., 12, 1645-1649, https://doi.org/10.1109/lgrs.2015.2417857, 2015.

Mohammed, G. H., Colombo, R., Middleton, E. M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Perez-Priego, O., Damm, A., Meroni, M., Joiner, J., Cogliati, S., Verhoef, W., Malenovsky, Z., Gastellu-Etchegorry, J.-P., Miller, J. R., Guanter, L., Moreno, J., Moya,

I., Berry, J. A., Frankenberg, C., and Zarco-Tejada, P. J.: Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress, Remote Sens. Environ., 231, https://doi.org/10.1016/j.rse.2019.04.030, 2019.

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., Poindexter, C., Chen, J., Elbashandy, A., Humphrey, M., Isaac, P., Polidori, D., Ribeca, A., van Ingen, C., Zhang, L., Amiro, B., Ammann, C., Arain, M. A., Ardo, J., Arkebauer, T., Arndt, S. K.,
5   Arriga, N., Aubinet, M., Aurela, M., Baldocchi, D., Barr, A., Beamesderfer, E., Marchesini, L. B., Bergeron, O., Beringer, J., Bernhofer, C., Berveiller, D., Billesbach, D., Black, T. A., Blanken, P. D., Bohrer, G., Boike, J., Bolstad, P. V., Bonal, D., Bonnefond, J.-M., Bowling, D. R., Bracho, R., Brodeur, J., Bruemmer, C., Buchmann, N., Burban, B., Burns, S. P., Buysse, P., Cale, P., Cavagna, M., Cellier, P., Chen, S., Chini, I., Christensen, T. R., Cleverly, J., Collalti, A., Consalvo, C., Cook, B. D., Cook, D., Coursolle, C., Cremonese, E., Curtis, P. S., D'Andrea, E., da Rocha, H., Dai, X., Davis, K. J., De Cinti, B., de Grandcourt, A., De Ligne, A., De Oliveira, R. C., Delpierre, N., Desai, A.
10 R., Di Bella, C. M., di Tommasi, P., Dolman, H., Domingo, F., Dong, G., Dore, S., Duce, P., Dufrene, E., Dunn, A., Dusek, J., Eamus, D., Eichelmann, U., ElKhidir, H. A. M., Eugster, W., Ewenz, C. M., Ewers, B., Famulari, D., Fares, S., Feigenwinter, I., Feitz, A., Fensholt, R., Filippa, G., Fischer, M., Frank, J., Galvagno, M., Gharun, M., Gianelle, D., Gielen, B., Gioli, B., Gitelson, A., Goded, I., Goeckede, M., Goldstein, A. H., Gough, C. M., Goulden, M. L., Graf, A., Griebel, A., Gruening, C., Gruenwald, T., Hammerle, A., Han, S., Han, X., Hansen, B. U., Hanson, C., Hatakka, J., He, Y., Hehn, M., Heinesch, B., Hinko-Najera, N., Hoertnagl, L., Hutley, L., Ibrom, A., Ikawa, H.,
15 Jackowicz-Korczynski, M., Janous, D., Jans, W., Jassal, R., Jiang, S., Kato, T., Khomik, M., Klatt, J., Knohl, A., Knox, S., Kobayashi, H., Koerber, G., Kolle, O., Kosugi, Y., Kotani, A., Kowalski, A., Kruijt, B., Kurbatova, J., Kutsch, W. L., Kwon, H., Launiainen, S., Laurila, T., Law, B., Leuning, R., Li, Y., Liddell, M., Limousin, J.-M., Lion, M., Liska, A. J., Lohila, A., Lopez-Ballesteros, A., Lopez-Blanco, E., Loubet, B., Loustau, D., Lucas-Moffat, A., Lueers, J., Ma, S., Macfarlane, C., Magliulo, V., Maier, R., Mammarella, I., Manca, G., Marcolla, B., Margolis, H. A., Marras, S., Massman, W., Mastepanov, M., Matamala, R., Matthes, J. H., Mazzenga, F., McCaughey, H., McHugh, I.,
20 McMillan, A. M. S., Merbold, L., Meyer, W., Meyers, T., Miller, S. D., Minerbi, S., Moderow, U., Monson, R. K., Montagnani, L., Moore, C. E., Moors, E., Moreaux, V., Moureaux, C., Munger, J. W., Nakai, T., Neirynck, J., Nesic, Z., Nicolini, G., Noormets, A., Northwood, M., Nosetto, M., Nouvellon, Y., Novick, K., Oechel, W., Olesen, J. E., Ourcival, J.-M., Papuga, S. A., Parmentier, F.-J., Paul-Limoges, E., Pavelka, M., Peichl, M., Pendall, E., Phillips, R. P., Pilegaard, K., Pirk, N., Posse, G., Powell, T., Prasse, H., Prober, S. M., Rambal, S., Rannik, U., Raz-Yaseef, N., Reed, D., de Dios, V. R., Restrepo-Coupe, N., Reverter, B. R., Roland, M., Sabbatini, S., Sachs, T., Saleska, S.
25 R., Sanchez-Canete, E. P., Sanchez-Mejia, Z. M., Schmid, H. P., Schmidt, M., Schneider, K., Schrader, F., Schroder, I., Scott, R. L., Sedlak, P., Serrano-Ortiz, P., Shao, C., Shi, P., Shironya, I., Siebicke, L., Sigut, L., Silberstein, R., Sirca, C., Spano, D., Steinbrecher, R., Stevens, R. M., Sturtevant, C., Suyker, A., Tagesson, T., Takanashi, S., Tang, Y., Tapper, N., Thom, J., Tiedemann, F., Tomassucci, M., Tuovinen, J.-P., Urbanski, S., Valentini, R., van der Molen, M., van Gorsel, E., van Huissteden, K., Varlagin, A., Verfaillie, J., Vesala, T., Vincke, C., Vitale, D., Vygodskaya, N., Walker, J. P., Walter-Shea, E., Wang, H., Weber, R., Westermann, S., Wille, C., Wofsy, S., Wohlfahrt, G., Wolf,
30 S., Woodgate, W., Li, Y., Zampedri, R., Zhang, J., Zhou, G., Zona, D., Agarwal, D., Torn, M., and Papale, D.: The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data, Sci. Data, 7, https://doi.org/10.1038/s41597-020-0534-3, 2020.

Pazur, R., Huber, N., Weber, D., Ginzler, C., and Price, B.: A national extent map of cropland and grassland for Switzerland based on Sentinel-2 data, Earth Syst. Sci. Data, 14, 295-305, https://doi.org/10.5194/essd-14-295-2022, 2022.

Porcar-Castell, A., Malenovsky, Z., Magney, T., Van Wittenberghe, S., Fernandez-Marin, B., Maignan, F., Zhang, Y., Maseyk, K., Atherton,
35 J., Albert, L. P., Robson, T. M., Zhao, F., Garcia-Plazaola, J.-I., Ensminger, I., Rajewicz, P. A., Grebe, S., Tikkanen, M., Kellner, J. R., Ihalainen, J. A., Rascher, U., and Logan, B.: Chlorophyll *a* fluorescence illuminates a path connecting plant molecular biology to Earth-system science, Nat. Plants, 7, 998-1009, https://doi.org/10.1038/s41477-021-00980-4, 2021.

Sha, Y., Gagne, D. J., II, West, G., and Stull, R.: Deep-Learning-Based Gridded Downscaling of Surface Meteorological Variables in Complex Terrain. Part I: Daily Maximum and Minimum 2-m Temperature, J. Appl. Meteorol. Clim., 59, 2057-2073,
40 https://doi.org/10.1175/jamc-d-20-0057.1, 2020.

Sun, L. and Lan, Y.: Statistical downscaling of daily temperature and precipitation over China using deep learning neural models: Localization and comparison with other methods, Int. J. Climatol., 41, 1128-1147, https://doi.org/10.1002/joc.6769, 2021.

Turner, A. J., Koehler, P., Magney, T. S., Frankenberg, C., Fung, I., and Cohen, R. C.: A double peak in the seasonality of California's photosynthesis as observed from space, Biogeosciences, 17, 405-422, https://doi.org/10.5194/bg-17-405-2020, 2020.

45 Turner, A. J., Kohler, P., Magney, T. S., Frankenberg, C., Fung, I., and Cohen, R. C.: Extreme events driving year-to-year differences in gross primary productivity across the US, Biogeosciences, 18, 6579-6588, https://doi.org/10.5194/bg-18-6579-2021, 2021.

Xiao, J., Fisher, J. B., Hashimoto, H., Ichii, K., and Parazoo, N. C.: Emerging satellite observations for diurnal cycling of ecosystem processes, Nat. Plants, 7, 877-887, https://doi.org/10.1038/s41477-021-00952-8, 2021.

Xiao, J., Chevallier, F., Gomez, C., Guanter, L., Hicke, J. A., Huete, A. R., Ichii, K., Ni, W., Pang, Y., Rahman, A. F., Sun, G., Yuan, W.,
50 Zhang, L., and Zhang, X.: Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years, Remote Sens. Environ., 233, https://doi.org/10.1016/j.rse.2019.111383, 2019.

Xu, L., Herold, M., Tsendbazar, N.-E., Masiliunas, D., Li, L., Lesiv, M., Fritz, S., and Verbesselt, J.: Time series analysis for global land cover change monitoring: A comparison across sensors, Remote Sens. Environ., 271, https://doi.org/10.1016/j.rse.2022.112905, 2022.

Xu, Y., Liu, X., Cao, X., Huang, C., Liu, E., Qian, S., Liu, X., Wu, Y., Dong, F., Qiu, C.-W., Qiu, J., Hua, K., Su, W., Wu, J., Xu, H., Han,
55 Y., Fu, C., Yin, Z., Liu, M., Roepman, R., Dietmann, S., Virta, M., Kengara, F., Zhang, Z., Zhang, L., Zhao, T., Dai, J., Yang, J., Lan, L., Luo, M., Liu, Z., An, T., Zhang, B., He, X., Cong, S., Liu, X., Zhang, W., Lewis, J. P., Tiedje, J. M., Wang, Q., An, Z., Wang, F., Zhang,

L., Huang, T., Lu, C., Cai, Z., Wang, F., and Zhang, J.: Artificial intelligence: A powerful paradigm for scientific research, Innovation (Cambridge (Mass.)), 2, 100179-100179, https://doi.org/10.1016/j.xinn.2021.100179, 2021.

Yao, L., Yang, D., Liu, Y., Wang, J., Liu, L., Du, S., Cai, Z., Lu, N., Lyu, D., Wang, M., Yin, Z., and Zheng, Y.: A New Global Solar-induced Chlorophyll Fluorescence (SIF) Data Product from TanSat Measurements, Adv. Atmos. Sci., 38, 341-345,

5 https://doi.org/10.1007/s00376-020-0204-6, 2021.

You, N. and Dong, J.: Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine, Isprs J. Photogramm., 161, 109-123, https://doi.org/10.1016/j.isprsjprs.2020.01.001, 2020.

You, N., Dong, J., Huang, J., Du, G., Zhang, G., He, Y., Yang, T., Di, Y., and Xiao, X.: The 10-m crop type maps in Northeast China during 2017-2019, Sci. Data, 8, https://doi.org/10.1038/s41597-021-00827-9, 2021.

10 Zeng, Y., Badgley, G., Dechant, B., Ryu, Y., Chen, M., and Berry, J. A.: A practical approach for estimating the escape ratio of near-infrared solar-induced chlorophyll fluorescence, Remote Sens. Environ., 232, https://doi.org/10.1016/j.rse.2019.05.028, 2019.

Zeng, Y., Chen, M., Hao, D., Damm, A., Badgley, G., Rascher, U., Johnson, J. E., Dechant, B., Siegmann, B., Ryu, Y., Qiu, H., Krieger, V., Panigada, C., Celesti, M., Miglietta, F., Yang, X., and Berry, J. A.: Combining near-infrared radiance of vegetation and fluorescence spectroscopy to detect effects of abiotic changes and stresses, Remote Sens. Environ., 270, https://doi.org/10.1016/j.rse.2021.112856, 2022.

15 Zhang, L., Zhang, Z., Luo, Y., Cao, J., and Tao, F.: Combining Optical, Fluorescence, Thermal Satellite, and Environmental Data to Predict County-Level Maize Yield in China Using Machine Learning Approaches, Remote Sens., 12, https://doi.org/10.3390/rs12010021, 2020a.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., and Mi, J.: GLC_FCS30: global land-cover product with fine classification system at 30m using time-series Landsat imagery, Earth Syst. Sci. Data, 13, 2753-2776, https://doi.org/10.5194/essd-13-2753-2021, 2021a.

Zhang, Y., Joiner, J., Alemohammad, S. H., Zhou, S., and Gentine, P.: A global spatially contiguous solar-induced fluorescence (CSIF)

20 dataset using neural networks, Biogeosciences, 15, 5779-5800, https://doi.org/10.5194/bg-15-5779-2018, 2018.

Zhang, Y., Fang, J., Smith, W. K., Wang, X., Gentine, P., Scott, R. L., Migliavacca, M., Jeong, S., Litvak, M., and Zhou, S.: Satellite solar-induced chlorophyll fluorescence tracks physiological drought stress development during 2020 southwest US drought, Global Change Biol., 29, 3395-3408, https://doi.org/10.1111/gcb.16683, 2023.

Zhang, Y., Zhang, Q., Liu, L., Zhang, Y., Wang, S., Ju, W., Zhou, G., Zhou, L., Tang, J., Zhu, X., Wang, F., Huang, Y., Zhang, Z., Qiu, B.,

25 Zhang, X., Wang, S., Huang, C., Tang, X., and Zhang, J.: ChinaSpec: A Network for Long-Term Ground-Based Measurements of Solar-Induced Fluorescence in China, J. Geophys. Res-Biogeo., 126, https://doi.org/10.1029/2020jg006042, 2021b.

Zhang, Z., Xu, W., Qin, Q., and Long, Z.: Downscaling Solar-Induced Chlorophyll Fluorescence Based on Convolutional Neural Network Method to Monitor Agricultural Drought, Ieee Trans. Geosci., 59, 1012-1028, https://doi.org/10.1109/tgrs.2020.2999371, 2021c.

Zhang, Z., Zhang, Y., Zhang, Q., Chen, J. M., Porcar-Castell, A., Guanter, L., Wu, Y., Zhang, X., Wang, H., Ding, D., and Li, Z.: Assessing

30 bi-directional effects on the diurnal cycle of measured solar-induced chlorophyll fluorescence in crop canopies, Agr. Forest. Meteorol., 295, https://doi.org/10.1016/j.agrformet.2020.108147, 2020b.

Zou, C., Du, S., Liu, X., and Liu, L.: TCSIF: a temporally consistent global Global Ozone Monitoring Experiment-2A (GOME-2A) solar-induced chlorophyll fluorescence dataset with the correction of sensor degradation, Earth Syst. Sci. Data, 16, 2789-2809, https://doi.org/10.5194/essd-16-2789-2024, 2024.
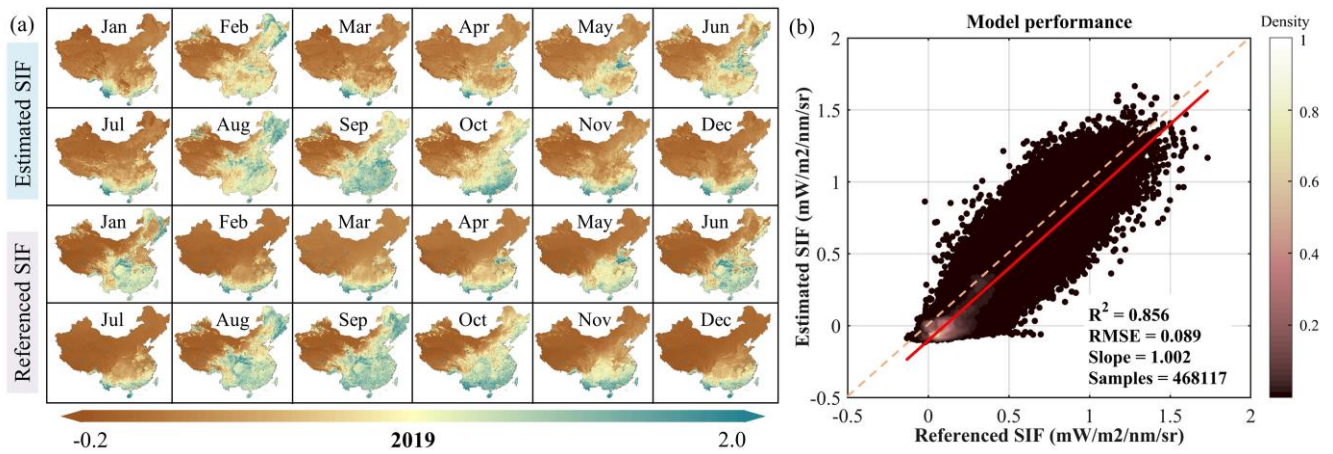
35
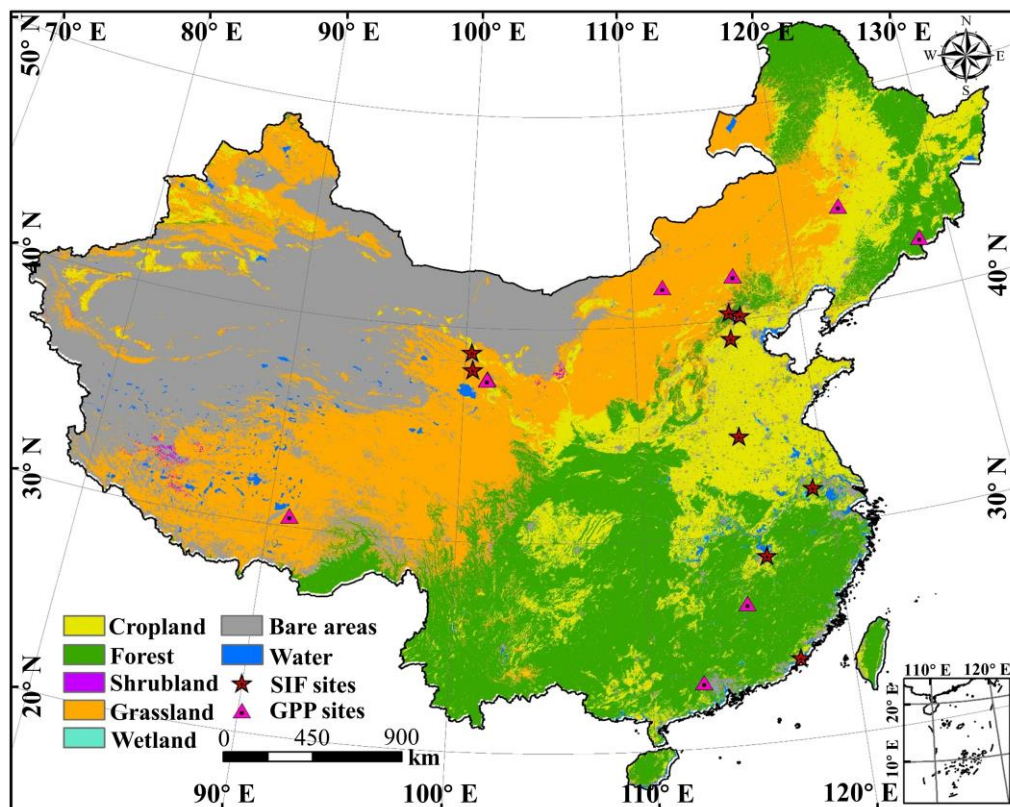
Fig. 1: Land cover types of the study area and the locations of all tower-based observation sites used in this research.
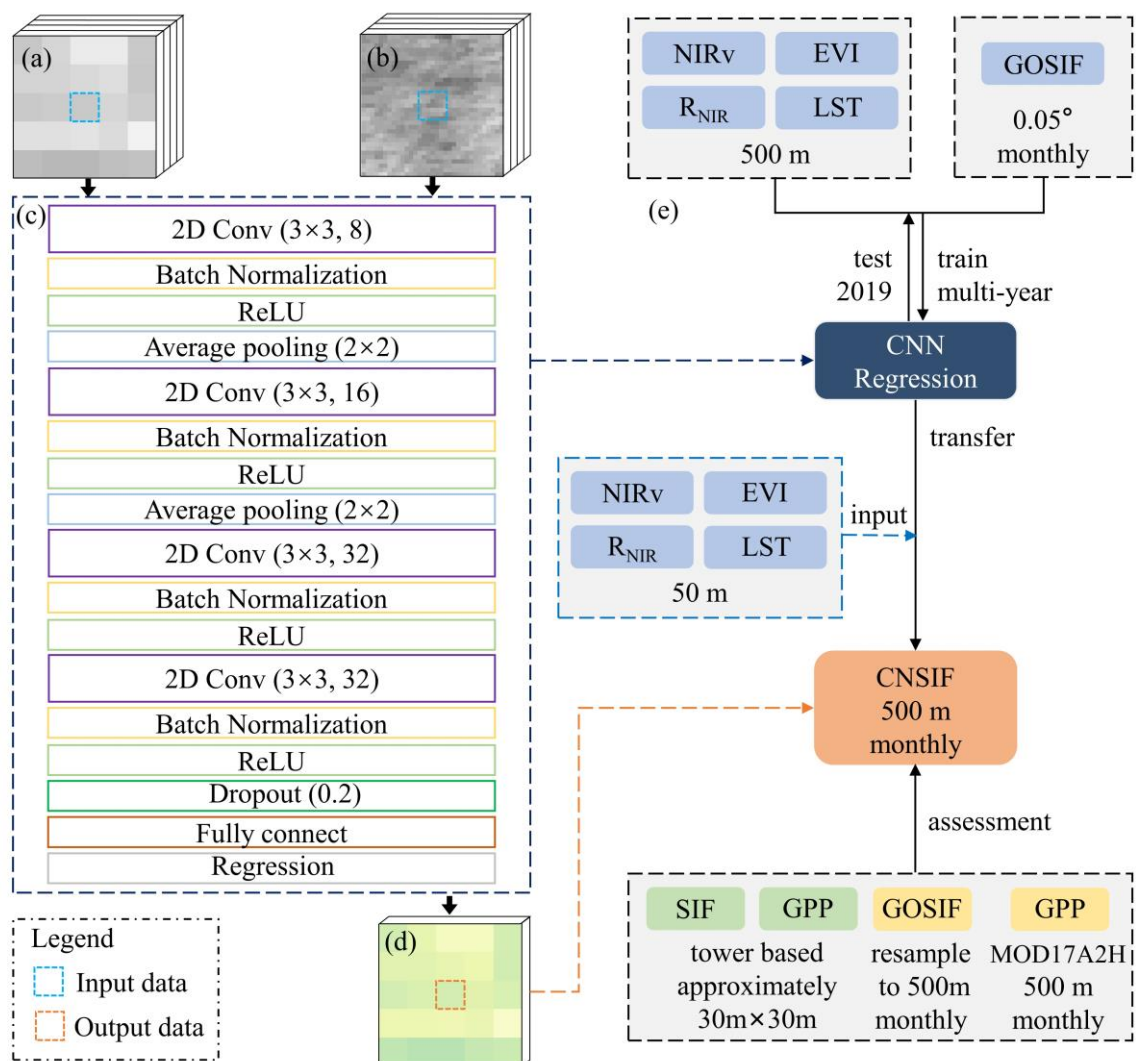
Fig. 2: CNN-based SIF retrieval framework design. (a) 1×1 pixel at a coarse scale, (b) 9×9 pixels at a fine scale, (c) schematic diagram of the CNN network structure, (e) flowchart of the retrieval framework.

Fig. 3: (a) Comparison between the reference results and the reconstructed predicted results, (b) model performance.
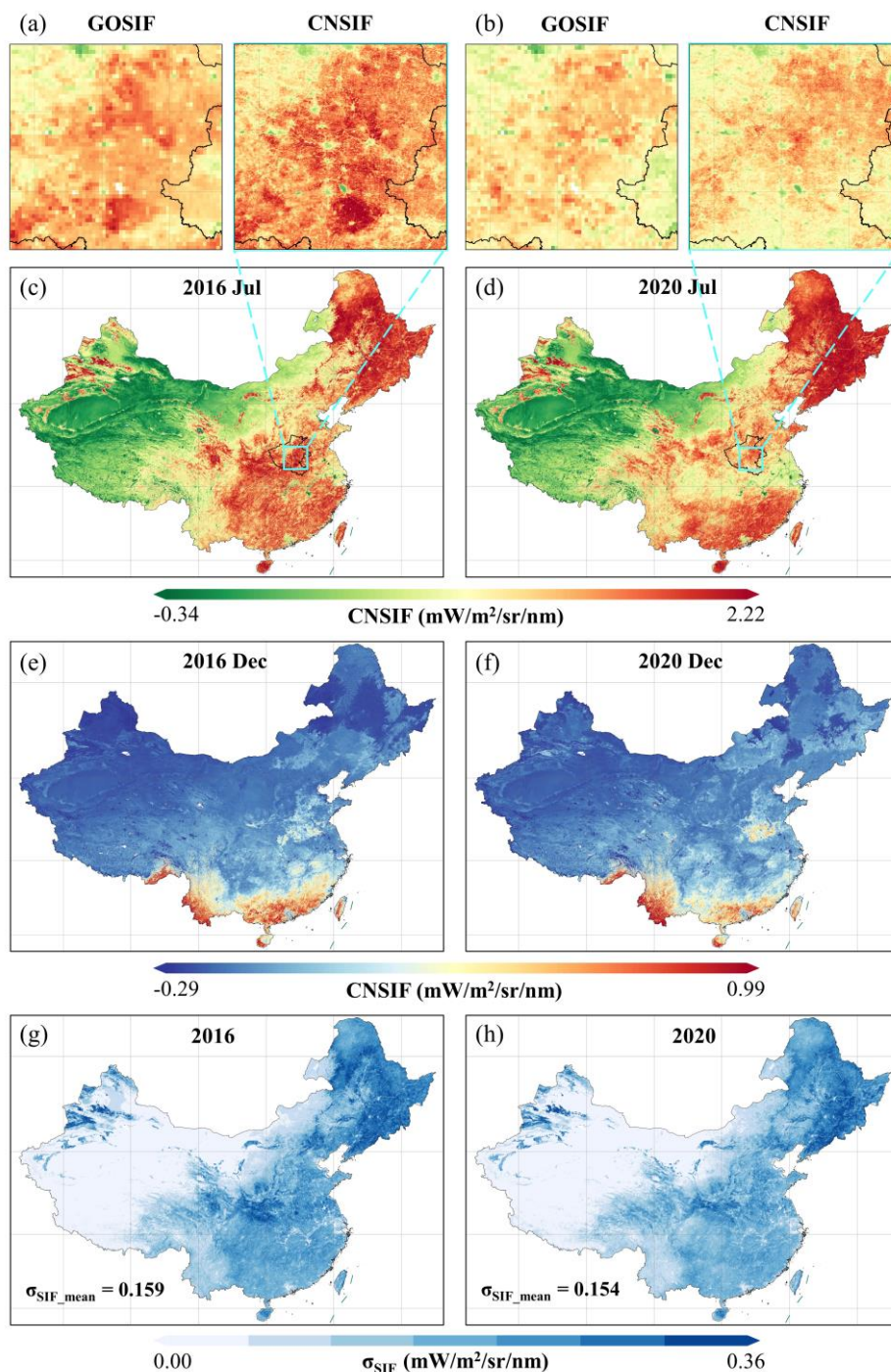
Fig. 4: Spatial distribution of CNSIF. (a), (b) local zoomed-in views of dense crop areas; (c), (d) summer spatial distribution characteristics of CNSIF; (e), (f) winter spatial distribution characteristics of CNSIF; (g), (h) standard errors of the weighted means of CNSIF.
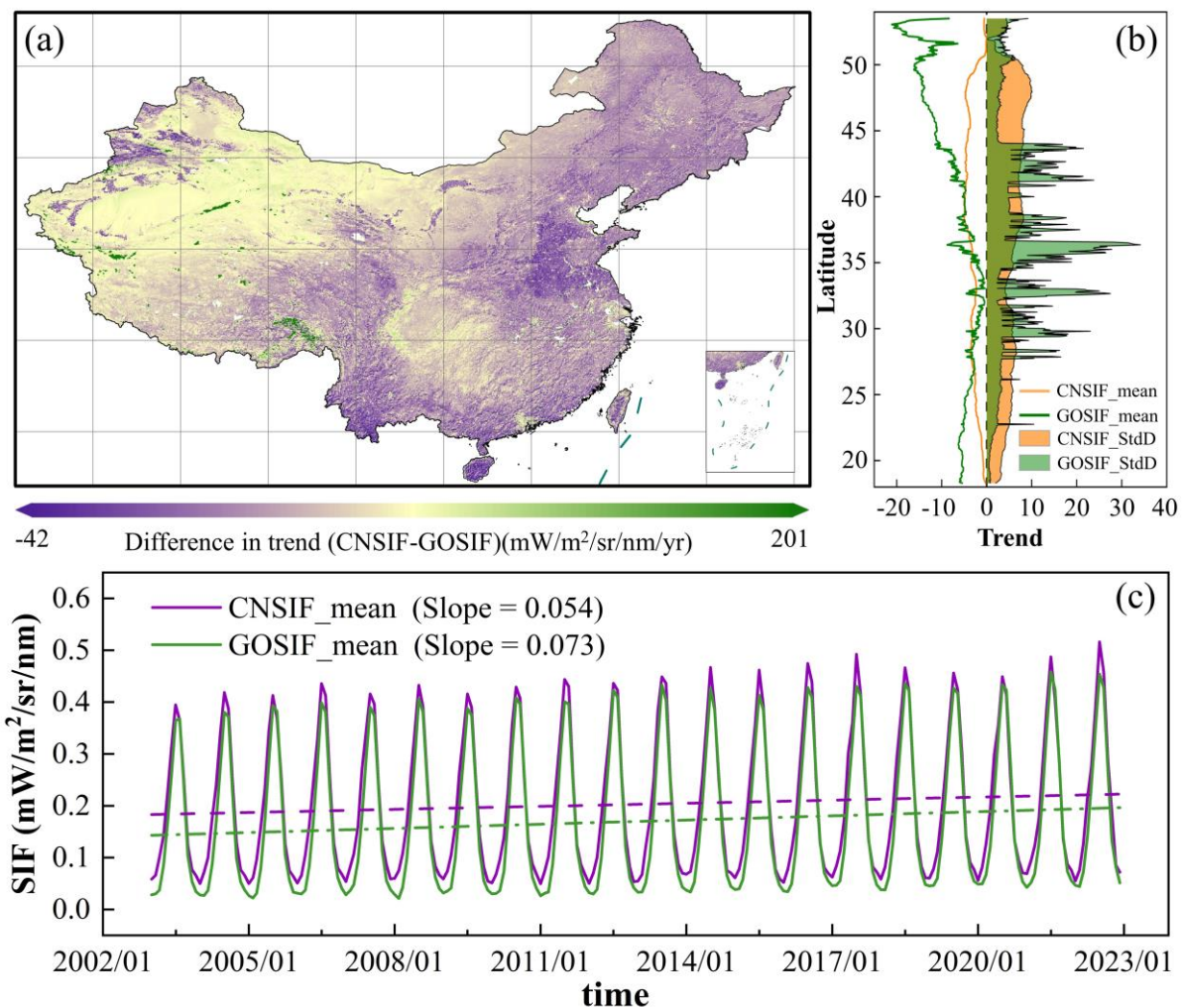
Fig. 5: (a) Time trend differences between the CNSIF and GOSIF datasets, (b) latitudinal distribution from 2003 to 2022, and (c) time series variation of the mean CNSIF from 2003 to 2022. The yellow and green shaded areas in (b) represent the standard errors of the CNSIF and GOSIF trends, respectively.

5
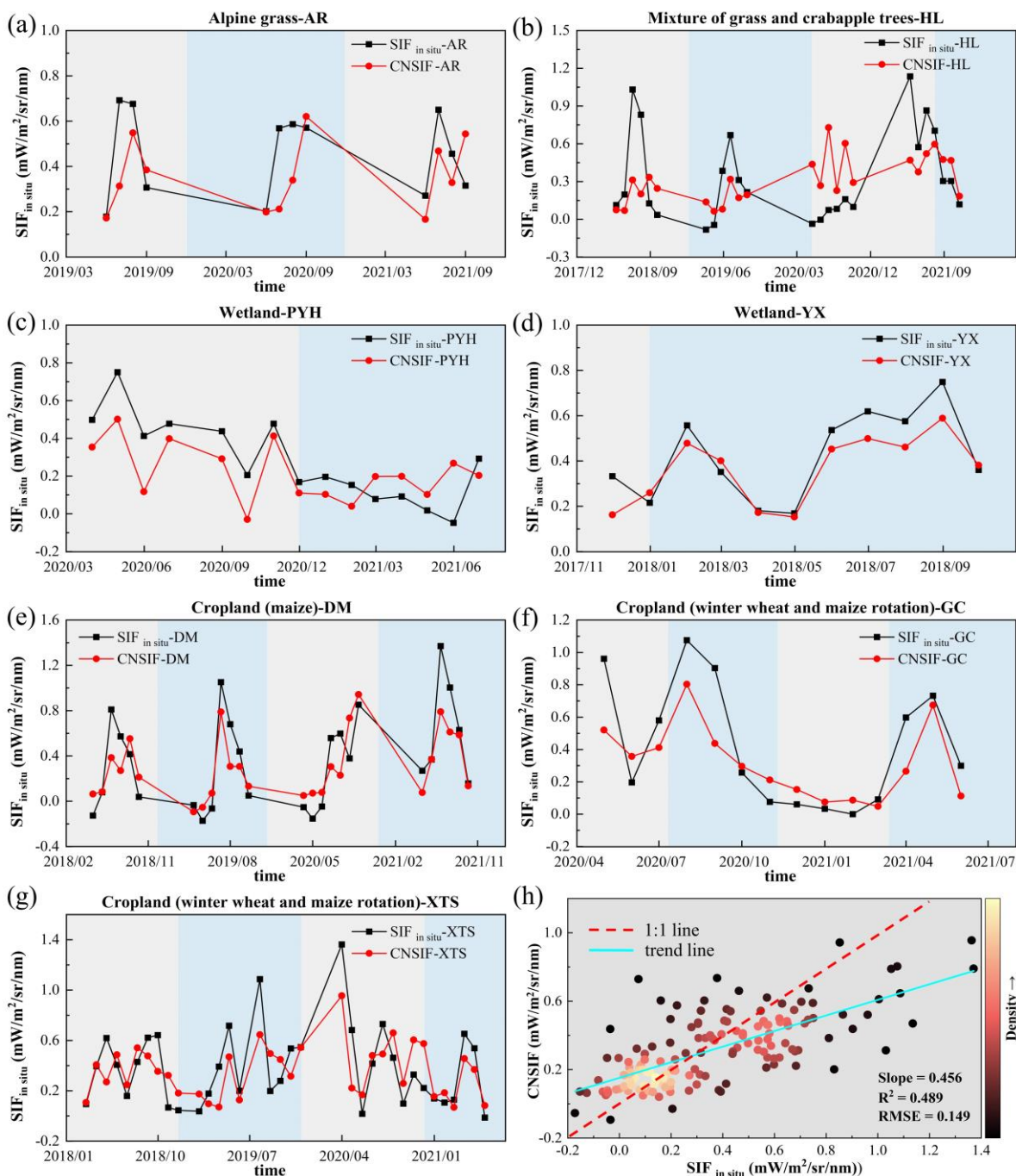
Fig. 6: (a)-(g) Comparison between tower-based SIF observations and CNSIF for sites with two or more observation periods, (h) comparison between CNSIF and the scaled tower-based SIF observations across all sites.
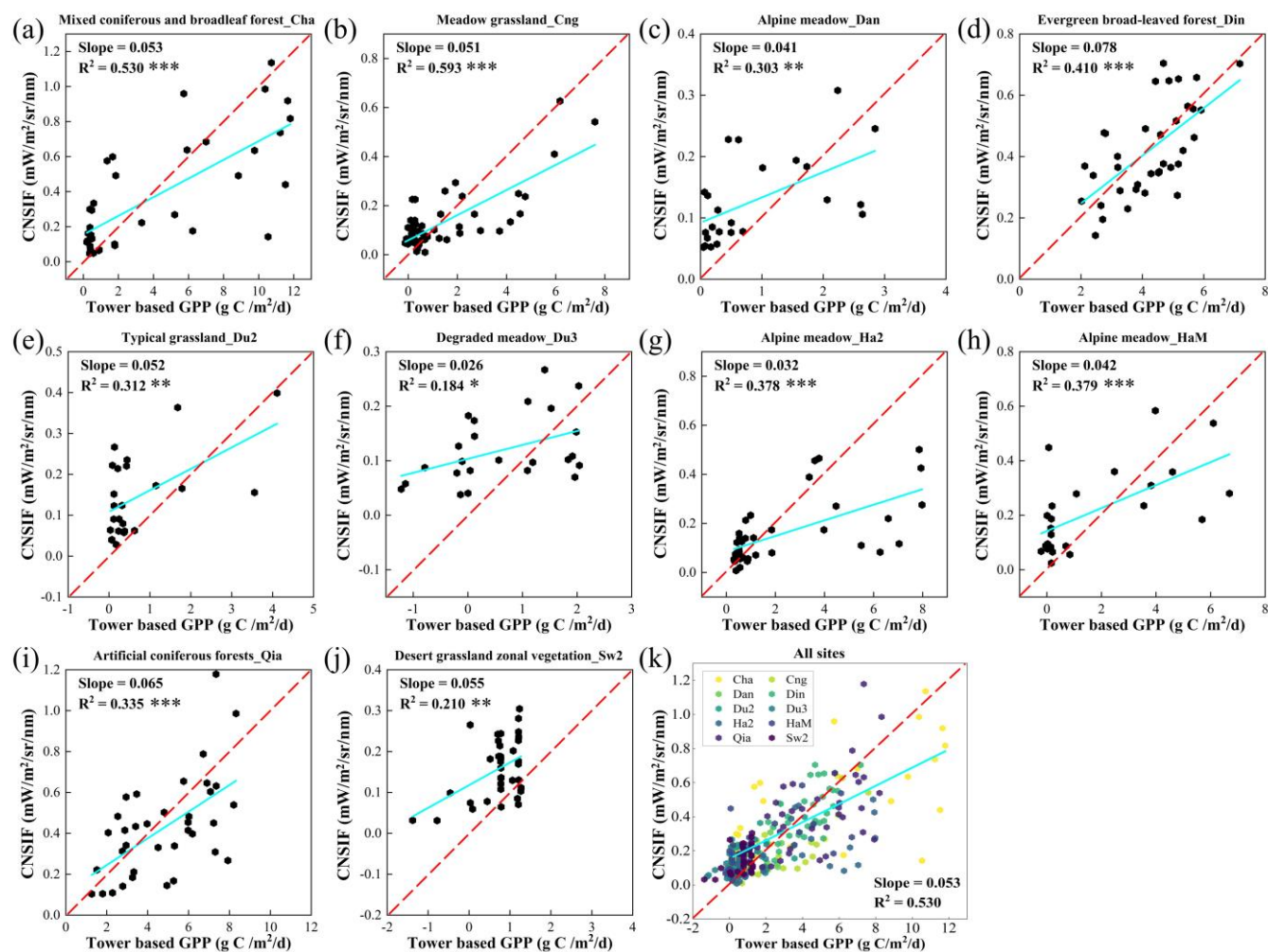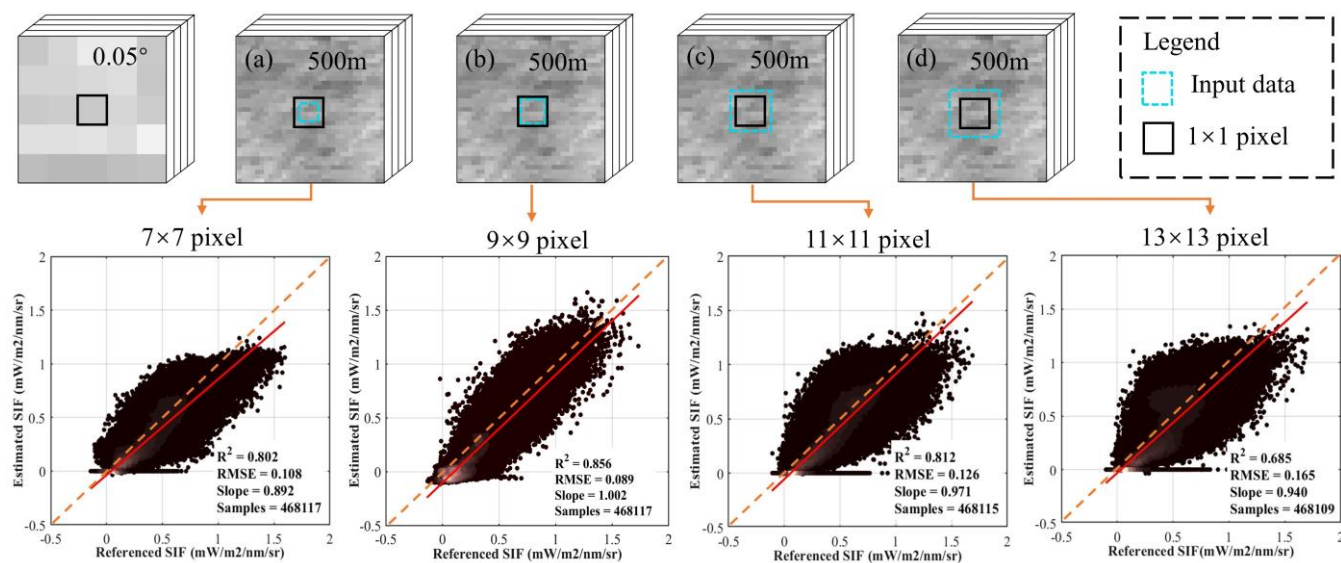
Fig. 7: (a)–(j) Comparison between CNSIF and tower-based GPP observations across different biomes, (k) comparison between CNSIF and tower-based GPP observations across all sites.
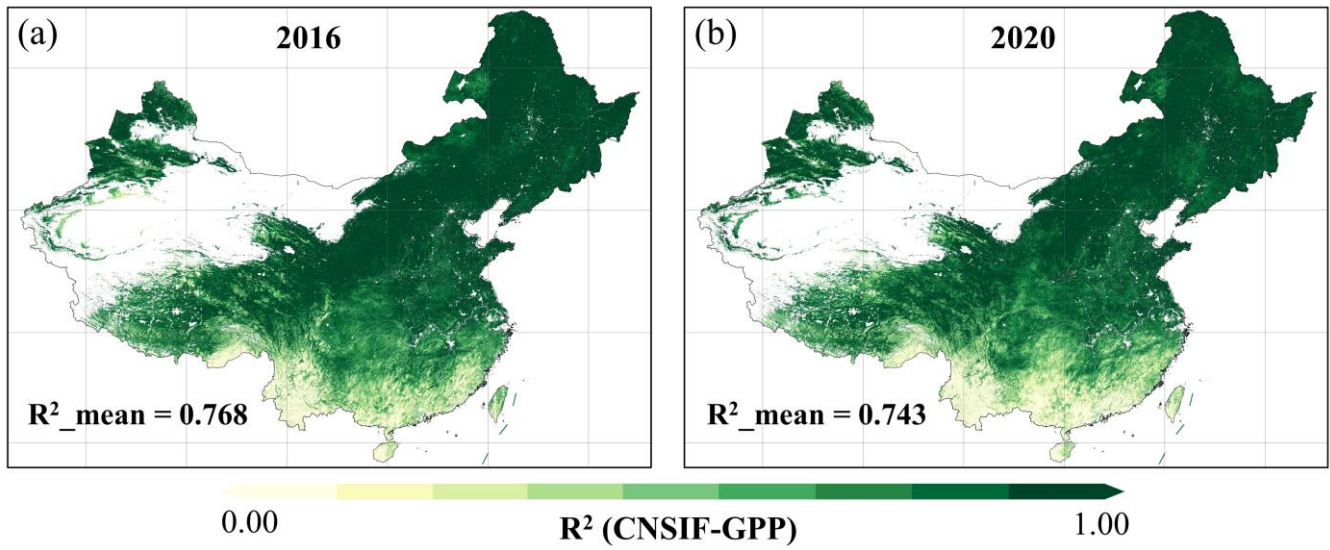
Fig. 8: Comparison between CNSIF and the MODIS GPP product.

Fig. 9: Model performance based on different input data sizes. (a) 7×7 pixels considering only internal information, (b) 9×9 pixels considering only internal information, (c) 11×11 pixels considering both internal and neighboring information, (d) 13×13 pixels considering both internal and neighboring information.

5

Table 1

An overview of the collected datasets in this study.

Dataset	Data source	Variables/Band	Spatial resolution	Temporal resolution	Temporal range
GOSIF	OCO-2 SIF	SIF	0.05°× 0.05°	Monthly	2001.01-present
Surface reflectance	Landsat 7 ETM+	SR_B1, SR_B2, SR_B3, SR_B4	30-m	16-day	1999.05-2022.04
	Sentinel 2 MSI	B2, B3, B4, B8	50-m	5-day	2017.03-present
Thermal infrared data	Landsat 7 ETM+	ST_B6	60-m	16-day	1999.05-2022.04
	Landsat 8 TIRS	ST_B10	30-m	16-day	2013.04-present
Land cover	MCD12Q1	Land cover	500-m	Yearly	2001.01-2022.01

Table 2

Detailed information of SIF sites used in this study.

Site Name	ID	Latitude Longitude	Ecosystem type	Fraction of primary land cover type	Temporal range
ARou	AR	38.0444 N 100.4647 E	Alpine grass	0.81	2019 - 2021
DaMan	DM	38.8555 N 100.3722 E	Cropland (maize)	0.92	2018 - 2021
GuCheng	GC	39.1486 N 115.7351 E	Cropland (winter wheat and maize rotation)	0.93	2020 - 2021
HuaiLai	HL	40.3489 N 115.7882 E	Mixture of grass and crabapple trees	0.25	2018 - 2021
JuRong	JR	31.8068 N 119.2173 E	Cropland (rice and wheat rotation)	0.93	2017
PoYangHu	PYH	28.8949 N 116.3371 E	Wetland	0.75	2020 - 2021
ShangQiu	SQ	34.5870 N 115.5753 E	Cropland (maize on irrigated land)	0.92	2017 - 2018
XiaoTangshan	XTS	40.1786 N 116.4432 E	Cropland (winter wheat and maize rotation)	0.37	2018 - 2021
YunXiao	YX	23.9064 N 117.4028 E	Wetland	0.83	2017 - 2018

Table 3

Detailed information on GPP sites was used in this study.

Site Name	ID	Latitude Longitude	Ecosystem type	Fraction of primary land cover type	Temporal range
ChangBaiShan	Cha	42.4025 N 128.0958 E	Mixed coniferous and broadleaf forest	1	2003 - 2005
ChangLing	Cng	44.5934 N 123.5092 E	Meadow grassland	0.8	2007 - 2010
DangXiong	Dan	30.4978 N 91.0664 E	Alpine meadow	0.5-0.8	2004 - 2005
DingHuShan	Din	23.1733 N 112.5361 E	Evergreen broad-leaved forest	1	2003 - 2005
DuoLun_Grassland	Du2	42.0467 N 116.2836 E	Typical grassland	1	2007 - 2008
DuoLun_Degraded meadow	Du3	42.0551 N 116.2809 E	Degraded meadow	1	2009 - 2010
HaiBei_Shrubland	Ha2	37.6652 N 101.3312 E	Alpine meadow	1	2003 - 2005
HaiBei_Alpine Tibet site	HaM	37.6128 N 101.3128 E	Alpine meadow	1	2003 - 2004
QianYanZhou	Qia	26.7414 N 115.0581 E	Artificial coniferous forests	1	2003 - 2005
SiZiWang_Grazed	Sw2	41.7902 N 111.8971 E	Desert grassland zonal vegetation	1	2010 - 2012