Data assimilation of solar-induced chlorophyll fluorescence improves gross primary production estimation by a process-based VISIT-SIF model in a rice paddy

水田におけるプロセスベースモデル VISIT-SIF を用いた太陽光誘起クロロフィル蛍光の統計的データ同化による総一次生産（GPP）推定精度向上

北海道大学 大学院国際食資源学院

国際食資源学専攻 博士後期課程

Hokkaido University Graduate School of Global Food Resources

Division of Global Food Resources Doctoral Course

Fan Liangxian

Contents

Abstract

Understanding gross primary production (GPP) is crucial for the global carbon cycle, and many studies have shown that solar-induced chlorophyll fluorescence (SIF) is a reliable proxy for GPP. This study integrates SIF data into the process-based Vegetation Integrative Simulator for Trace gases (VISIT-SIF) model to enhance GPP simulations in a rice paddy field in Tsukuba, Japan. Using data assimilation techniques (Bayesian Optimization) with ground-based SIF data and satellite-derived SIF (CSIF) from June 2019 to September 2020, I optimized key model parameters to improve GPP simulation accuracy. Sensitivity analysis via SHapley Additive exPlanations (SHAP) revealed that absorbed photosynthetically active radiation (APAR)-related parameters are most critical for SIF simulations, while nitrogen-related parameters associated with the maximum rate of carboxylation (Vcmax) strongly influence GPP.

The optimization significantly improved the VISIT-SIF model's performance across multiple timescales. At the half-hourly scale, SIF R^2 increased from 0.37 to 0.60, while GPP R^2 improved from 0.47 to 0.68, with substantial reductions in relative errors. Similar improvements were observed at daily and weekly scales, demonstrating the model's enhanced ability to capture diurnal and seasonal variations. Comparisons between sunny and cloudy conditions highlighted the model's stronger performance under direct sunlight, with reduced accuracy under diffuse light. Notably, the optimized model showed improved correlations between SIF and GPP at all temporal scales, achieving R^2 values as high as 0.99 at the weekly scale. However, Stage II of the

growing season presented challenges, with relatively weaker SIF-GPP relationships even after optimization.

I further investigated the relationship between SIF and GPP and the impacts of environmental factors. Results showed that the SIF-GPP correlation strengthened with increasing temporal aggregation, from half-hourly to weekly scales, reflecting the integration of short-term variations over longer periods. Environmental conditions, particularly light variability, played a significant role in shaping this relationship. The model performed better under stable, direct sunlight, while cloud cover introduced complexities that reduced simulation accuracy. These findings underscore the importance of considering environmental factors, such as light quality and cloud effects, in improving SIF-GPP modeling.

Overall, this study demonstrates the value of using SIF as a proxy for GPP and highlights the effectiveness of integrating ground-based and satellite SIF data into process-based models. While data assimilation significantly improved model performance, challenges such as SIF overestimation and limitations under cloudy conditions remain. These findings contribute to advancing the understanding of SIF-GPP dynamics and emphasize the potential of SIF for enhancing GPP estimations across varying environmental conditions.

Results and conclusions of this study are summarized as follows:

Chapter 1: Sensitivity Analysis of Parameters to SIF and GPP Outputs

The sensitivity analysis using SHAP identified distinct influences of model parameters on SIF and GPP simulations. For SIF, parameters related to absorbed photosynthetically active radiation (APAR), such as f, ρ_{cd} and k_d, were found to have the largest contributions. These parameters govern radiative transfer processes, with lower values of f positively impacting SIF simulation accuracy. Conversely, parameters like H showed minimal effects, reflecting their lower importance in SIF dynamics.

For GPP, nitrogen-related parameters such as N_o, x_n and N_b, linked to the maximum rate of carboxylation (Vcmax), were identified as the most critical. Higher canopy-top leaf nitrogen content (N_o) significantly improved GPP outputs, emphasizing the importance of physiological and biochemical processes in photosynthesis. The SHAP analysis ultimately identified 14 key parameters, contributing over 95% to the model outputs, which were selected for further optimization.

Chapter 2: Improvement of SIF and GPP Simulations

At half-hourly Scale, model optimization significantly enhanced the accuracy of SIF and GPP simulations. At the half-hourly scale, R^2 for SIF increased from 0.37 to 0.60, with rRMSE decreasing from 124.21% to 63.39%. GPP simulations showed a similar improvement, with R^2 increasing from 0.47 to 0.68 and rRMSE decreasing from

150.00% to 47.85%. These improvements were particularly pronounced in Stage III of the growing season.

At the daily scale, R^2 for SIF improved to 0.69, with a notable reduction in rRMSE from 66.83% to 50.12%. GPP simulations exhibited similar enhancements, with R^2 values for 2018 reaching 0.87 and a substantial reduction in relative errors. These findings demonstrate the optimized model's ability to better represent both diurnal and seasonal dynamics across multiple timescales.

The model performed better under sunny conditions compared to cloudy conditions. For SIF, R^2 under sunny conditions reached 0.76, while cloudy conditions had a lower R^2 of 0.53. For GPP, R^2 improved from 0.55 under cloudy conditions to 0.79 under sunny conditions. Error metrics, including rRMSE and PBIAS, also showed significant reductions under sunny conditions. These results suggest the model is better suited for simulating conditions with stable direct sunlight, while its performance under diffuse light remains a challenge.

Both SIF and GPP displayed characteristic hump-shaped diurnal patterns. After optimization, the model achieved R^2 values exceeding 0.98 for both SIF and GPP across the growing season and individual stages. Despite these improvements, the model underestimated variability (CV) in the observed data, particularly during Stages I and II, indicating areas for further refinement in representing finer-scale dynamics.

Chapter 3: Data Assimilation with Sparse Satellite SIF

Data assimilation using satellite-derived SIF (CSIF) improved the model's performance, albeit less effectively than ground-based data. For GPP, the CSIF-optimized model achieved an R^2 of 0.65 and a near-neutral PBIAS (-2.72%), indicating effective bias reduction. However, for SIF, the R^2 was 0.51, with notable overestimation reflected in a PBIAS of 30.87%. While CSIF data can be a valuable substitute for ground-based observations in regions without site-based measurements, its coarser resolution and inherent uncertainties limit its effectiveness.

Chapter 4: Relationships Between SIF and GPP Across Scales

The relationship between SIF and GPP strengthened with increasing temporal aggregation. At the half-hourly scale, the optimized model achieved R^2 values up to 0.88, while at daily and weekly scales, R^2 reached 0.92 and 0.99, respectively. These improvements highlight the effectiveness of data assimilation in enhancing model performance across time scales. However, the model still struggled with Stage II dynamics, where the SIF-GPP relationship remained weaker compared to other stages, indicating the need for further optimization.

Chapter 5: Environmental Factors and SIF-GPP Dynamics

Environmental conditions significantly influenced the SIF-GPP relationship. The model performed better under sunny conditions, with higher R^2 values and lower errors,

compared to cloudy conditions, where light variability posed challenges. The results underscore the importance of direct sunlight in facilitating accurate simulations. Furthermore, the model faced difficulties in capturing seasonal SIF-GPP dynamics and large SIF values under complex conditions. These findings suggest potential directions for improving model performance, including better representation of diffuse light and environmental interactions.

Keywords: Gross primary production; Solar-induced chlorophyll fluorescence; Data assimilation; SHapley Additive exPlanations; SIF-GPP relationship.

I. General Introduction

I-1 Introduction

This study addresses these challenges by enhancing the VISIT model through the integration of a radiative transfer module to simulate SIF and incorporating SHapley Additive exPlanations (SHAP) for parameter sensitivity analysis. Using both ground-based SIF and satellite-derived CSIF data, I performed Bayesian optimization to refine 14 key model parameters. Our study aimed to improve GPP simulation accuracy at half-hourly, daily, and seasonal scales in a rice paddy field in Japan. Additionally, I explored the relationship between SIF and GPP across different temporal scales and environmental conditions, providing insights into how factors such as light variability influence SIF-GPP dynamics. These findings contribute to advancing process-based model accuracy and understanding the potential of SIF for GPP estimation under diverse environmental scenarios.

I-2 Backgrounds

I-2-1 Importance of Gross Primary Production (GPP) Simulation

Gross primary production (GPP), the total amount of carbon fixed by vegetation through photosynthesis, is a fundamental component of the global carbon cycle. Understanding GPP is essential for quantifying carbon fluxes between ecosystems and the atmosphere, which is critical for addressing climate change and managing ecosystems sustainably (Aubinet et al., 1999; Baldocchi, 2003; Chen et al., 2009; Damm et al., 2010; Z. Li et al., 2020). At small spatial scales, GPP is often estimated

using eddy covariance flux measurements combined with ecosystem respiration models. However, scaling these estimates to regional or global levels presents significant challenges, necessitating proxies derived from remote sensing techniques.

I-2-2 Limitations of Vegetation Indices for GPP Estimation

Traditional remote sensing approaches, such as the normalized difference vegetation index (NDVI) and the enhanced vegetation index (EVI), have been widely used to estimate GPP (Huete et al., 2006; Running et al., 2004). These indices primarily reflect vegetation greenness and canopy structure, which are indirectly linked to photosynthetic activity. However, they are less sensitive to short-term physiological changes in photosynthesis (Sun et al., 2018), leading to significant uncertainties in GPP estimation, especially under conditions where vegetation structure remains stable despite dynamic changes in photosynthetic activity (Guanter et al., 2014a; Garbulsky et al., 2014). As a result, these indices often fail to capture rapid or subtle variations in plant physiological responses to environmental changes.

I-2-3 Solar-Induced Chlorophyll Fluorescence (SIF) as a Proxy for GPP

Solar-induced chlorophyll fluorescence (SIF) has emerged as a robust proxy for GPP, offering advantages over traditional vegetation indices. SIF is emitted as a byproduct of absorbed photosynthetically active radiation (PAR) during photosynthesis, making it a direct indicator of photosynthetic activity (Frankenberg and Berry, 2018; Meroni et al., 2009). Unlike reflectance-based indices, SIF captures real-time variations

in physiological processes, providing a more accurate measure of plant photosynthesis (Wood et al., 2017; Zhang et al., 2018; Liu et al., 2021). Studies have shown strong correlations between SIF and GPP across temporal and spatial scales, highlighting its potential for improving GPP estimation (Li et al., 2018; Buareal et al., 2023; Morozumi et al., 2023). Recent advancements in remote sensing have enabled SIF retrieval from both satellite and ground-based platforms. Satellites such as GOSAT, OCO-2, and TROPOMI provide global SIF measurements with increasingly fine spatial and temporal resolution (Guanter et al., 2021; Köhler et al., 2018). However, satellite data are often limited by atmospheric interference, coarse resolution, and infrequent temporal coverage, particularly in heterogeneous landscapes (Bacour et al., 2019; R. Li et al., 2022). Ground-based SIF observations, on the other hand, offer higher accuracy and temporal resolution, capturing half-hourly dynamics that are invaluable for detailed studies (Cogliati et al., 2015; Drolet et al., 2014; Z. Li et al., 2020). Despite their smaller spatial coverage, ground-based data provide essential validation for satellite-derived products and insights into localized SIF-GPP dynamics.

I-2-4 Challenges in Integrating SIF with Process-Based Models

Process-based models like the Vegetation Integrative Simulator for Trace gases (VISIT) simulate carbon fluxes by integrating ecological, physiological, and biogeochemical processes (Ito and Oikawa, 2002; Ito et al., 2017). Although these models have been validated against eddy covariance data, incorporating SIF into the

modeling framework poses significant challenges. Radiative transfer processes, the distribution of absorbed PAR, and chlorophyll fluorescence dynamics require precise representation, which is often lacking in existing models (Koffi et al., 2015; Bacour et al., 2019). Recent studies have demonstrated the utility of integrating SIF into terrestrial biosphere models to improve GPP simulations (Lee et al., 2015; Norton et al., 2019; Thum et al., 2017), yet limitations in parameter sensitivity analysis and optimization hinder further progress.

II. Materials and Methods

II-1 Site and Data

II-1-1 Study site

Our study was conducted in a rice paddy field of approximately 200 ha named the AsiaFlux Mase site at 36.05°N, 140.03°E (site code "MSE" in the AsiaFlux network; http://asiaflux.net/index.php?page_id=83) in Tsukuba, Ibaraki, Japan. Eddy covariance, a high-resolution spectrometer and micrometeorological instruments were installed approximately 3 m above the ground. It has a maritime subtropical monsoon climate, with an annual mean air temperature of 13.5 °C and an annual mean precipitation of 1235 mm (Ono et al., 2015). Rice seedlings were transplanted around day of year (DOY) 121, and grain was harvested around DOY 255. Following the classification method of Ono (Ono et al., 2013), the rice growing season was divided into three stages: Stage I is the vegetative stage (DOY 120-180), Stage II is the reproductive stage (DOY 181-210), and Stage III is the ripening stage (DOY 211-harvest).

II-1-2 Measurements of eddy flux, meteorological data and SIF data

The data used in this study included model input data, data for data assimilation, and data for model validation. The VISIT input data included downward solar radiation, downward longwave radiation, air temperature, relative humidity, barometric pressure, wind speed and precipitation from 2018 to2020. The ground-based SIF data and satellite SIF data (both from June 29th, 2019 to September 10th, 2020) were used for

data assimilation. The ground-based SIF data (from June 29th, 2019 to September 10th, 2020) and GPP data (from 2018 to 2020) were used to verify the accuracy of the model.

Eddy flux instruments were installed at our study site. The measurement of turbulent fluxes of CO_2 and water vapor, the wind speed and air temperature are by an eddy covariance system consisting of an open-path infrared gas analyzer (LI-7500; LI-COR, Lincoln, Nebraska, USA), a closed-path infrared gas analyzer (LI-7000; LI-COR), and a sonic anemometer-thermometer (DA600–3TV; Kaijo, Tokyo, Japan). Data were collected at a frequency of 10 Hz with a data logger (CR3000; Campbell Scientific, Logan, Utah, USA) (Ono et al., 2008). The net ecosystem CO_2 exchange, derived from the CO_2 flux measured over the rice field by the EC system, was mathematically partitioned into GPP and ecosystem respiration (Ono et al., 2008; Ono & Maruyama, 2015). Finally, half-hourly non-gap-filled GPP data were used for further analysis. The barometric pressure measured by the barometer, and the downward shortwave and downward longwave radiation reaching the ground are measured by a net radiometer. Our precipitation data were from the Nagamine Tsukuba Areological Observatory in the Automated Meteorological Data Acquisition System (AMeDAS; https://www.jma.go.jp/jma/indexe.html), located at 36.06°N, 140.13°E. The precipitation data were sampled every hour, and I resampled the precipitation data to a half-hourly resolution, averaging the one-hourly data into two subsets. For example, if the precipitation during the hour from 10:00 to 11:00 is 10 mm, I attribute a value of 5

mm to the two half-hourly precipitation amounts from 10:00-10:30 and 10:30-11:00.

For ground-based SIF data, I installed a high-resolution spectrometer (QEpro; Ocean Insight, Orlando, Florida, United States of America [USA]) with spectral coverage between 649 and 807 nm (sampling interval of 0.15 nm/pixel) and a full width at half maximum of 0.30 nm besides eddy flux tower. The QEpro spectrometer was connected via a fiber switch (FSM1 × 8, Piezosystem Jena GmbH, Jena, Germany; MOL-1 × 8–600-H, Leoni Fiber Optics GmbH, Fo¨ritztal, Germany) to three optical fiber channels. The near-infrared SIF was retrieved from paired measurements of downward solar irradiance and upward reflected radiance by the spectral fitting method. Refer to Buareal et al. (2023) for SIF calculation details.

For satellite SIF data, the global spatially contiguous solar-induced fluorescence (CSIF) dataset is a gap-filled dataset generated from MODIS surface reflectance and the OCO-2 SIF dataset (Zhang et al., 2018). The CSIF data are at 757 nm and have a 4-day temporal resolution and spatial resolution of 0.05° × 0.05°. In this study, the latest version (Version 2) of the CSIF under all-sky conditions from June 29th, 2019 to September 10th, 2020 (corresponding to ground-based SIF data) was used. The satellite SIF that I utilized was the value of the pixel where our study site was located, which was largely covered by the representative rice paddy. The CSIF not only shows high accuracy when validated against the satellite-retrieved OCO-2 SIF but also exhibits reasonably high consistency with both the reconstructed and satellite-retrieved GOME-

2 SIF (Zhang et al., 2018).

II-1-3 Sky conditions

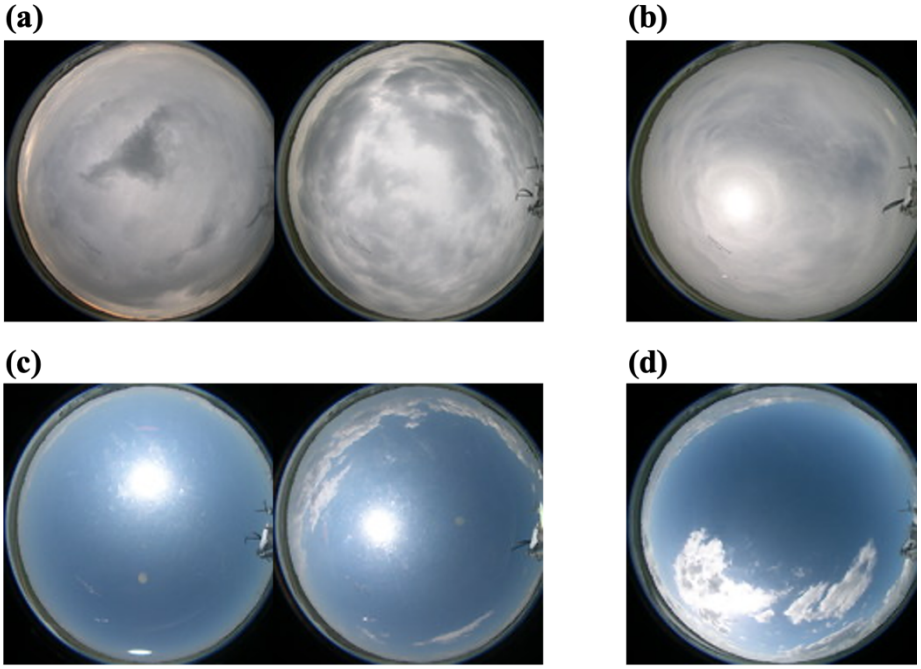

Figure 1. An example of (a) a cloudy condition, (b) not a cloudy condition (c) a sunny condition, (d) not a sunny condition. The criteria for cloudy conditions: 1. the sky observed by the weather camera must be completely covered by clouds; 2. the clouds are thick and have no obvious translucency; and 3. all six photos within half an hour need to satisfy the above requirements. The criterion for sunny conditions: 1. The sky observed by the weather camera has a maximum of about 30% cloud cover; 2. the sun cannot be blocked by clouds when the sun is within the observation range; 3. all six photos within half an hour need to fulfill the above requirements.

Automatic digital fisheye cameras (ADFCs) and hemispherical spectroradiometers (HSSRs) from the Phenological Eyes Network (PEN; http://pen.envr.tsukuba.ac.jp/DataPolicy.html; Nagai et al., 2018; Nasahara and Nagai, 2015) were used to detect the radiance and variability of the weather Conditions during the rice growing season. All datasets were acquired at half-hourly resolution. The camera monitored the sky every 5 minutes from 6:15 am to 7:55 pm. Unfortunately, the camera was not working throughout 2019 and between the 200th and 300th day of 2020 due to a malfunction. Therefore, no weather information was recorded during these periods. As for how sunny and cloudy conditions are determined, I used Okta scale method which commonly used in meteorology. The sky captured in the photo is divided into ten equal parts, and the number of parts covered by clouds is estimated. For example, if three out of ten parts are covered, the cloud cover is considered to be 30%. To minimize misinterpretations between sunny and cloudy weather, we set strict criteria for determining sunny and cloudy weather. Figure 1 (a) shows an example of a cloudy condition. The criterion for a sunny condition is as follows: 1. The sky observed by the weather camera has a maximum of 30% cloud cover; 2. the sun cannot be blocked by clouds when the sun is within the observation range; and 3. all six photos within half an hour need to fulfill the above requirements. Figure 1 (c) shows an example of sunny conditions. Initially, I intended to set stricter criteria for clear days, requiring no cloud cover in the field of view of the weather camera. However, practical checks revealed this condition to be rare, resulting in insufficient sunny condition samples.

Consequently, I adjusted the criteria to the current standards. The criterion for sunny conditions: 1. The sky observed by the weather camera has a maximum of about 30% cloud cover; 2. the sun cannot be blocked by clouds when the sun is within the observation range; 3. all six photos within half an hour need to fulfill the above requirements. The method used in this research is a basic method of calculating cloud cover in meteorology. The photo of the sky is divided into ten equal parts, and then the number of parts of the sky covered by clouds is estimated. If about three parts are covered, the cloud cover is 30%.

II-2 Model

VISIT is a process-based model based on the dry-matter production theory for simulating trace gas fluxes between the atmosphere and an ecosystem Ito and Oikawa (2002), applies Farquhar's sun/shade model (De Pury & Farquhar, 1997; Farquhar et al., 1980) to simulate photosynthetic processes. More details and the performance of VISIT can be found in previous studies (Ito, 2019; Ito et al., 2017; Ito & Inatomi, 2012; Ito & Oikawa, 2002). To associate the photosynthesis process of VISIT with SIF, I coupled the SCOPE geometric module with VISIT. This new version of VISIT was named VISIT-SIF(refer to Miyauchi et al. (2024) for details). I simulated SIF in the 760 nm range, as this wavelength exhibits a strong SIF signal. Additionally, this range aligns with ground-observed SIF data, which were collected at the same wavelength band (Buareal et al., 2023).

In the VISIT-SIF (Miyauchi et al., 2024), SIF is calculated as follows:

$$\mathrm{SIF}_{760} = F\, r \quad (1)$$

where F (mW m^{-2} sr^{-1} nm^{-1}) is the TOC (top of canopy) chlorophyll fluorescence in the observation direction calculated by VISIT. r is the fraction of SIF at 760 nm to total emitted SIF. r is calculated by using an approximation obtained from the relationships between the chlorophyll a+b content (C_{ab}) and r via SCOPE presimulation. I used a regression formula, $r = 1.2 \cdot 10^{-3}\, ln\,(C_{ab}) + 4.7 \cdot 10^{-3}$, to estimate r for entire biomes regardless of changes in the LAI, with biome-specific values of C_{ab} according to Norton et al.(2019).

F is the chlorophyll fluorescence emitted from the sunlit and shaded leaves. Therefore, F can be calculated as follows:

$$F = F_{\mathrm{sun}} + F_{\mathrm{shade}} \quad (2)$$

$$F_{\text{shade}} = F_{\text{sun}}\, r_{\text{shade/sun}} \qquad (3)$$

where F_{sun} and F_{shade} are the chlorophyll fluorescence from sunlit and shaded leaves calculated by VISIT, respectively. $r_{\text{shade/sun}}$ (see equation (14)) is the ratio of F_{shade} to F_{sun} calculated by SCOPE. F_{sun} is calculated as follows:

$$F_{sun} = \frac{APAR_{sun}\, \Phi_{F,sun}\, r_{oz/sz}\, f_u}{\pi} \quad (4)$$

where $APAR_{sun}$ is the absorbed photosynthetically active radiation, and $\Phi_{F,sun}$ is the quantum yield of chlorophyll fluorescence in sunlit leaves. $r_{oz/sz}$ (see equation (13)) is a correction factor used to convert the sunlit leaves chlorophyll fluorescence to remote-sensed chlorophyll fluorescence in an arbitrary observation direction. f_u is the fraction of upward-emitted SIF to total energy emitted as SIF in the canopy, which is estimated as the mean of the upward fraction SIF in each canopy layer calculated by SCOPE for applying VISIT's single layer canopy. $\Phi_{F,sun}$ is calculated as follows:

$$\Phi_{F,sun} = \Phi_{Fm',sun}\left(1 - \Phi_{P,sun}\right) \quad (5)$$

where $\Phi_{Fm',sun}$ and $\Phi_{P,sun}$ are the quantum yield of fluorescence at light-acclimated leaves exposed to saturated irradiance and by actual photochemistry, respectively. $\Phi_{Fm',sun}$ is defined as follows:

$$\Phi_{Fm',sun} = \frac{k_F}{k_F + k_D + k_{N,sun}} \quad (6)$$

where k_F and k_D (0.05 and 0.95, respectively; van der Tol et al., 2014) represent relative rate constants for chlorophyll fluorescence (F), NPQ when dark-acclimate (D), and NPQ (N). $k_{N,sun}$ is calculated by an empirical equation (Flexas et al., 2002) as follows:

$$k_{N,sun} = (6.2473x - 0.5944)\, x \quad (7)$$

$$x = 1 - \frac{\Phi_{P,sun}}{\Phi_{P0}} \quad (8)$$

where x is the degree of light saturation (Flexas et al., 2002) and Φ_{P0} is the quantum yield for photochemistry for leaves under dark-acclimated conditions. Φ_{P0} and $\Phi_{P,sun}$ are defined as follows:

$$\Phi_{P0} = \frac{k_P}{k_F+k_D+k_P} \quad (9)$$

$$\Phi_{P,sun} = \Phi_{P0} \frac{J_{a,sun}}{J_{e,sun}} \quad (10)$$

In Equation (9), k_P (= 4.0) is the rate coefficient of irradiance emanated during photochemistry to irradiance total (van der Tol et al., 2014). In Equation (10), $J_{a,sun}$ (μmol m^{-2} s^{-1}) and $J_{e,sun}$ (μmol m^{-2} s^{-1}) are the actual and potential electron transport rates, respectively. $J_{a,sun}$ is calculated by VISIT based on van der Tol et al. (2014) :

$$J_{a,sun} = \frac{4A_{sun}(C_i+2\Gamma^*)}{C_i-\Gamma^*} (11)$$

where C_i (Pa) is the intercellular CO_2 partial pressure, and Γ^* (Pa) is the CO_2

compensation point. A_{sun} (μmol m^{-2} s^{-1}) is the CO_2 assimilation in sunlit leaves, which is calculated by applying the Farquhar model while considering environmental conditions. $J_{e,sun}$ is calculated as follow:

$$J_{e,sun} = I_{sun} \quad \Phi_{P0} \qquad (12)$$

where I_{sun} (μmol m^{-2} s^{-1}) is the absorbed photosynthetic photon flux density in sunlit leaves.

The geometric coefficients $r_{oz/sz}$ and $r_{\text{shade}/sun}$ vary with changing solar zenith angle (SZ), observation zenith angle (OZ), relative azimuth angle of solar and observation direction (AZ), and LAI. The VISIT-SIF applied a simplified method which uses the look-up table to compute the geometric coefficients for arbitrary observation direction (nadir in this study). The look-up tables of $r_{oz/sz}$ and $r_{\text{shade}/sun}$ were generated by simulations with varying angular parameters (SZ, OZ, and AZ), LAI, and solar radiation (SRAD; W m^{-2}) using the SCOPE model, as follows:

$$r_{oz/sz} = \frac{F_{SCOPE,sun}(LAI,SZ,OZ,AZ,SRAD)}{F_{SCOPE,sun,sz=oz}(LAI,SZ,OZ',AZ',SRAD)} \quad (13)$$

$$r_{\text{shade}/sun} = \frac{F_{SCOPE,shade}(LAI,SZ,OZ,AZ,SRAD)}{F_{SCOPE,sun}(LAI,SZ,OZ,AZ,SRAD)} \quad (14)$$

where $F_{SCOPE,sun}$, $F_{SCOPE,sun,sz=oz}$, and $F_{SCOPE,shade}$ are the chlorophyll fluorescence from sunlit leaves in the situation when the observation direction and the incident direction of the sun are located along one axis (OZ' = SZ, AZ' = 0) and from shaded leaves. In the simulation of SIF variability in the target site, the LAI and SRAD given by VISIT and the angular parameters obtained from the SIF observation camera are used for $r_{oz/sz}$ and $r_{\text{shade}/sun}$ in Eqs. (3) and (4), and in our study area, the observation camera was positioned perpendicular to the rice field.

II-3 Sensitivity test (SHAP analysis)

In this study, I applied the SHapley Additive exPlanation (SHAP) method for sensitivity analysis. SHAP is based on the concept of Shapley values from game theory, which assigns a contribution to each feature for a model's output (Lundberg & Lee, 2017). Unlike traditional feature importance methods, SHAP not only quantifies the importance of each feature but also shows its positive or negative effect on different data points. Lundberg et al. (2020) extended the SHAP method with the TreeExplainer model, which enables direct measurement of interaction effects between features,

allowing for a more comprehensive local explanation. The SHAP values are computed by calculating the marginal contribution of a specific feature to the model's output. In the game-theoretic framework, each feature is treated as a "contributor" and its contribution is summed across all possible feature combinations. Specifically, SHAP decomposes a model's prediction using the following equation:

$$f(x) = g(x') = \emptyset_0 + \sum_{i=1}^{M} \emptyset_i x_i' \qquad (15)$$

where f and g are the original model and explanation model, respectively, and g should match the output of the original model for a simplified input x' of x. $\emptyset_0$ represents the model's baseline prediction which is the mean value of the prediction values (Carlsson et al., 2020), $\emptyset_i$ is the SHAP value of feature i, and x_i' denotes the simplified input feature.

This study employed the SHAP method to detect the model parameters effect on SIF and GPP simulation. Here, I randomly selected 6000 parameter sets as SHAP input samples.

II-4 Bayesian Optimization

In this study, I employed Bayesian Optimization (BO) to optimize 14 key parameters in the VISIT-SIF model using half-hourly SIF observations and CSIF dataset. BO is a probabilistic model-based strategy that is particularly suitable for optimizing expensive black-box functions where evaluations are costly, and the objective function is noisy or unknown. This makes it especially appropriate for our complex VISIT-SIF model, where each function evaluation is computationally intensive.

Bayesian Optimization constructs a surrogate model of the objective function, typically using a Gaussian Process (GP). A GP provides probabilistic predictions of the function's behavior by modeling the relationship between previously evaluated points and their corresponding objective values. The key advantage of this approach is that it not only predicts the mean function value but also provides a measure of uncertainty, allowing the algorithm to balance exploration of unknown regions and exploitation of known promising areas in the parameter space.Process Overview:

1. Constructing the Surrogate Model with a Gaussian Process:

Objective Function f(x)

I aim to optimize the objective function f(x), where $x \in Rd$ is the input vector in a d-dimensional space. The function f(x) is unknown and expensive to evaluate, potentially requiring significant time or resources.

Gaussian Process GP($\mu(x)$,$k(x,x')$)):

In Bayesian Optimization, I model this unknown function f(x) using a Gaussian Process (GP). A GP is a non-parametric statistical model that treats the function as a random process, where the function values at any set of input points follow a multivariate normal distribution. The GP is characterized by two key components: the mean function $\mu(x)$ and the covariance (kernel) function $k(x,x')$.

Mean Function $\mu(x)$:

The mean function $\mu(x)$ represents the average predicted value of the function at a given input x. Typically, in the absence of prior knowledge, I initialize the mean function to zero, assuming the function is unbiased and has no specific trend.

Covariance Function $k(x,x')$:

The covariance function $k(x,x')$ measures the similarity between two input points x and x'. Through the covariance function, the GP captures the relationships between input points, modeling how the function varies.

2. Acquisition Function:

Bayesian Optimization iteratively selects new points by maximizing an acquisition function, balancing exploration and exploitation. In this study, I used the Expected Improvement (EI) acquisition function (Jones et al., 1998). The EI function selects points that are likely to provide an improvement over the current best-known value while considering the uncertainty in unexplored regions.

3. Cost Function:

I minimized a cost function designed to measure the discrepancy between the simulated and observed SIF values:

$$J_{(\mathrm{x})} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(A_i - B_i)^2} \Big/ \left(1 - \frac{\sum_{i=1}^{n}(A_i-\bar{A})^2}{\sum_{i=1}^{n}(B_i-\bar{B})^2}\right) \qquad (16)$$

Where x represents the parameter vector to be optimized in the model. A_i and B_i are the simulated and observed SIF values at the i-th time point, respectively. $\bar{A}$ and $\bar{B}$ are the mean values of A_i and B_i. n is the total number of data points.

This cost function integrates the Root Mean Square Error (RMSE) and a variance-based ratio, aiming to minimize the discrepancy between the model output and

observations and ensuring that the model not only performs well in terms of error, but also has a good overall fit. The Gaussian Process uses this cost function to guide the optimization process.

In this study, I used the GPyOpt module (Gonzalez et al., 2016; González et al., 2015) to perform Bayesian Optimization on the model parameters. The initial set of points was randomly chosen within predefined bounds based on previous studies (Koffi et al., 2015; Rayner et al., 2005; Scholze et al., 2007; Wu et al., 2020) and I see the bounds as large as possible to avoid errors from subjective views. As new points were evaluated, the surrogate model (GP) was iteratively updated, allowing the algorithm to intelligently navigate the search space and efficiently find the optimal parameter set.

II-5 Metrics for evaluating models

II-5-1 Relative Root Mean Square Error (rRMSE)

The relative root mean square error (rRMSE) is a standard metric used to evaluate the accuracy of model predictions relative to observations. It quantifies the discrepancy between predicted and observed values, normalizing the error by the mean of the observed data. rRMSE is expressed as a percentage, with lower values indicating better model performance. The rRMSE is calculated using the following equation:

$$rRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}}{\bar{O}} \times 100 \qquad (17)$$

Where P_i is the predicted value, O_i is the observed value, $\bar{O}$ is the mean of the observed values, and n is the number of observations.

II-5-2 Percent Bias (PBIAS)

Percent bias (PBIAS) is a metric that measures the average tendency of the predicted values to be larger or smaller than the observed values. A PBIAS value close to zero indicates a model with little systematic bias, while larger positive or negative values suggest significant overestimation or underestimation, respectively. PBIAS is often used to assess systematic biases in model predictions.

The equation for PBIAS is:

$$PBIAS = \frac{\sum_{i=1}^{n}(P_i - O_i)^2}{\sum_{i=1}^{n} O_i} \times 100 \qquad (18)$$

Where P_i is the predicted value, O_i is the observed value, and n is the number of observations.

II-5-3 Coefficient of Variation (CV)

The coefficient of variation (CV) is a measure of relative variability. It expresses the standard deviation of the data as a percentage of the mean, providing an indication of the extent of variability in relation to the average value. CV is especially useful for comparing the degree of variability between different datasets. A higher CV indicates

greater relative variability, while a lower CV suggests more consistent data. The formula for CV is:

$$CV = \frac{\sigma}{\mu} \times 100 \qquad (19)$$

Where σ is the standard deviation of the observed values, and μ is the mean of the observed values.

III. Results

III-1 Sensitivity analysis of parameters to SIF and GPP outputs

The SHAP analysis of the model parameters (see Table 1 for the description of parameters) provides a detailed view of the individual parameter contributions and their effects on the accuracy of GPP and SIF simulations. From Figure 2 (a) and (b), f is the most influential parameter, followed by ρ_{cd} and k_d in SIF's simulations. This suggests that parameters related to the absorbed photosynthetically active radiation (APAR) are critical to the simulation of SIF, and that variations in the value of f can significantly modulate the simulation output of the model. In contrast, parameters like H exhibit relatively lower SHAP values, suggesting a lesser impact on SIF simulations in the studied rice paddy ecosystem. The average contribution of each parameter can be seen in Figure 2 (a). For example, the contribution of f to SIF is 0.07 mW m^{-2} nm^{-1} sr-^{1} which is greater compared with other parameters. Figure 2 (b) also shows how high and low values of each parameter affect the SIF simulation. For example, a low value of f contributes more positively to the SIF simulation, while a low value of k_d reduces the SIF simulation. In addition, the SHAP values of f and ρ_{cd} showed a relatively balanced pattern under high and low feature values. From the long left tails of SHAP values of N_o indicate that a low N_o could induce larger variations in SIF dynamics. Figure 2 (c) shows a clustering of samples by their local explanation embedding. The shade of the color reflects the magnitude of the SHAP values. I can find that SIF values are usually driven by f, ρ_{cd} and k_d, especially f. From Figure 2 (d) and (e), the parameter N_o has the

greatest influence on the GPP simulation results, followed by x_n, f and N_b. The average contribution of N_o to GPP is 208.07 gC m^{-2} $year^{-1}$ followed by x_n (163.95 gC m^{-2} $year^{-1}$) which is greater compared with other parameters (Figure 2 (d)). The parameter N_o, x_n and N_b are all related the maximum rate of carboxylation (V_{cmax}). Most of the parameters ranked in front are related to key physiological and biochemical factors in the photosynthesis process, such as nitrogen content, carboxylation efficiency (Kc), and so on. Figure 2 (e) shows how the value (high/low) of each parameter affects GPP output. For example, a high value of N_0 significantly improves the simulated value of GPP, suggesting that higher canopy top leaf nitrogen content contributes to higher GPP output. Figure 2 (f) reveals the effect of each parameter on GPP over 6000 instances. When GPP output is high, it is always accompanied by a higher value of N_o and/or x_n.

In summary, the SHAP analysis reveals that the GPP and SIF models are sensitive to different sets of parameters. Nitrogen-related parameters dominate in the GPP model, while radiative transfer parameters, such as f, ρ_{cd}, and k_d, are more influential in the SIF model. According to the SHAP analysis, I selected these 14 parameters in front to do the model optimization, these 14 parameters contributed more than 95% to the model's simulation results.

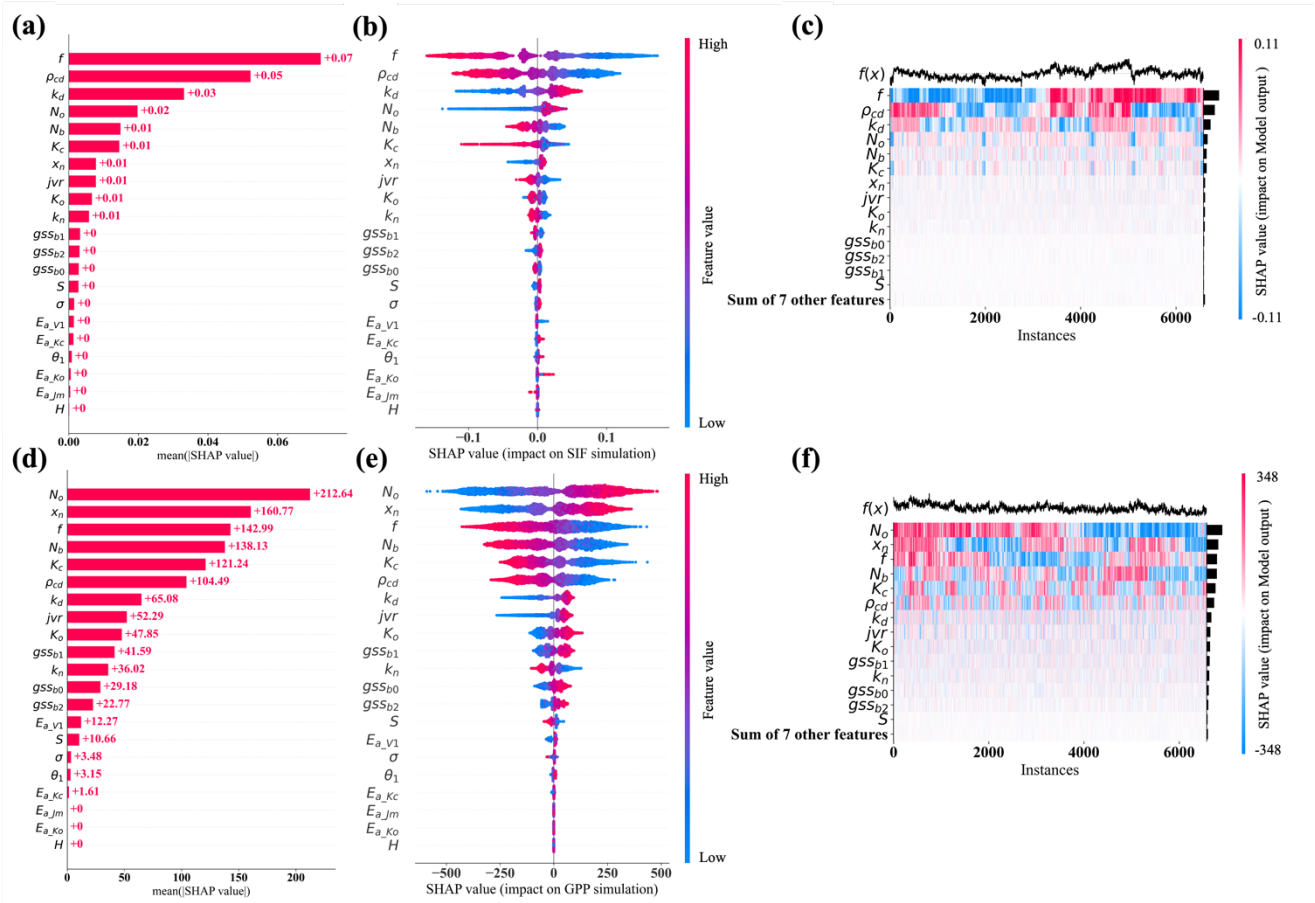

Figure 2. SHAP values of parameters for simulating SIF and GPP. (a) Bar plot of the mean absolute SHAP values of parameters for simulating SIF. (b) A set of bee swarm plots; in which each dot corresponds to a sample (A sample consists of 21 parameter values and the GPP obtained by these parameters (the annual average of the sum of the GPPs at the half-hour scale from 2018 to 2020)) in this study. The dot position on x axis displays the impact that parameter has on the SIF for that sample. When multiple dots are located in the same x position, they pile up to mean density. (c) A clustering of samples by their local explanation embedding. Columns are samples and rows are SHAP values. (d)-(f) are for SIF.

Table 1

Description of 21 input parameters of VISIT-SIF.

Related process	Parameter	Description	Unit	Prior	Post (site)	Post (CSIF)	Relative change(site)	Relative change(CSIF)	Range
V_{cmax}	k_n*	nitrogen content attenuation	-	0.713	0.832	0.682	16.69%	-4.35%	(0.1, 0.9)
	x_n*	nitrogen-photosynthetic capacity	mmol mol^{-1} s^{-1}	0.46	0.48	0.84	4.35%	82.61%	(0.2, 0.8)
	N_o*	canopy-top leaf nitrogen	mmol m^{-2}	120	199	168	65.83%	40.00%	(70, 250)
	N_b*	nonphotosynthetic leaf nitrogen	mmol m^{-2}	35	64	49	82.86%	40.00%	(15, 120)
	E_{a_V1}	activation energy for carboxylation	J mol^{-1}	64800					
APAR	k_d*	diffuse PAR attenuation	-	0.719	0.559	0.877	-22.25%	21.97%	(0.1, 0.9)
	f*	spectral adjustment	-	0.15	0.27	0.12	80.00%	-20.00%	(0.1, 0.9)
	ρ_{cd}*	diffuse reflectance	-	0.036	0.163	0.043	352.78%	19.44%	(0.01, 0.9)
	σ	scattering	-	0.15					
Stomatal conductance	gss_{b0}*	stomatal conductance parameter	-	5	5.4	6.5	8.00%	30.00%	(2, 10)
	gss_{b1}*	stomatal conductance parameter	-	4500	4412	4316	-1.96%	-4.09%	(1500, 8000)
	gss_{b2}*	stomatal conductance parameter	-	6.9	4	7.3	-42.03%	5.80%	(2, 10)

Michaelis–Menten constant and activation energy	Kc^*	Michaelis coefficient for carboxylation	- Pa	40	35	32	-12.50%	-20.00%	(20, 60)
	Ko^*	Michaelis coefficient for oxygenation	Pa	24800	22092	25176	-10.92%	1.52%	(20000, 40000)
	E_{a_Kc}	activation energy for kc	J mol^{-1}	59400					
	E_{a_Ko}	activation energy for ko	J mol^{-1}	36000					
Jmax	S^*	temperature response	J mol^{-1} K^{-1}	710	701	700	-1.27%	-1.41%	(700, 800)
	jvr^*	J_{max}/V_{cmax} ratio	-	2.1	3.8	3.3	80.95%	57.14%	(1, 6)
	E_{a_Jm}	activation energy for electron transport	J mol^{-1}	37000					
	H	curvature parameter	J mol^{-1}	220000					
	θ_1	convexity for electron transport	-	0.7					

These are the 21 input parameters of VISIT-SIF, which are related to V_{cmax} (the maximum rate of carboxylation), APAR (absorbed photosynthetically active radiation), stomatal conductance, the Michaelis–Menten constant, and Jmax (maximum rate of electron transport).

The parameters marked with an asterisk (*) in the table are those identified through the sensitivity test as having a significant impact on SIF and GPP. These parameters were selected for optimization, while the remaining parameters were excluded from the optimization process.

"Prior" refers to the parameter values before optimization. "Post (Site)" indicates the parameter values after optimization using ground-based SIF data, while "Post (CSIF)" represents the values after optimization using satellite-derived SIF data (CSIF). The "Range" specifies the parameter bounds set for the optimization process.

Relative change(site) = (Post(Site)- Prior) / Prior; Relative change(CSIF) = (Post(CSIF- Prior) / Prior

III-2 Improvement of SIF and GPP simulation by optimization at the half-hourly and daily scales

III-2-1 Improvement of SIF and GPP simulations at half-hourly scale

To assess the improvement of half-hourly simulations after model optimization, I compared the model performance for SIF and GPP over the whole growing season and across different growth stages (Stage I, Stage II, and Stage III). This analysis provides a comprehensive evaluation of how optimization improved the model's ability to capture temporal variations in photosynthesis and fluorescence dynamics under varying physiological and environmental conditions.

For SIF before optimization (Figure 3 (a)-(d)), the R^2 for the entire growing season was relatively low (0.37) (Figure 3 (a)), indicating a weak correlation between observed and simulated values. Among the individual growth stages, Stage III exhibited the highest R^2 (0.52) (Figure 3 (d)), reflecting relatively better model performance during the ripening stage when canopy structure and chlorophyll content are more stable. However, the model struggled most during Stage II, with R^2 dropping to 0.14 (Figure 3 (c)), likely due to the complex physiological changes and high variability in environmental conditions during the reproductive phase. The high values of relative root mean square error (rRMSE) across all stages, ranging from 120.18% to 132.43%, further highlight the significant deviations between simulated and observed SIF values. Additionally, percent bias (PBIAS) analysis revealed systematic underestimation of SIF, particularly in Stage III, where PBIAS was -29.05%. This underestimation suggests that the pre-optimization model had difficulty capturing the high fluorescence emission during peak photosynthetic activity.

After optimization (Figure 3 (e)-(h)), there was a noticeable improvement in values for SIF simulations. The R^2 for the entire growing season increased significantly to 0.60 (Figure 3 (e)), reflecting a stronger overall agreement between observed and simulated values. Stage III showed the largest improvement, with R^2 increasing from 0.52 to 0.69 (Figure 3 (h)), demonstrating the model's enhanced ability to capture fluorescence dynamics during this stage. Stage I and Stage II also experienced

improvements, with R^2 increasing to 0.61 and 0.43, respectively (Figure 3 (f) and (g)). The substantial decrease in rRMSE values after optimization, particularly for Stage III (from 120.18% to 55.25%), indicates a marked reduction in model error. However, the PBIAS values post-optimization suggest some overestimation of SIF, particularly in Stage II (43.23%) and Stage I (37.77%), which may point to overcompensation in the optimization process.

For GPP before optimization (Figure 3 (i)-(l)), the initial model performance was also limited. The R^2 for the entire growing season was 0.47 (Figure 3 (i)), indicating moderate correlation between observed and simulated GPP. Stage I performed slightly better, with an R^2 of 0.50 (Figure 3 (j)), likely due to the simpler physiological processes in the early vegetative stage. However, Stage II had the lowest R^2 (0.12) (Figure 3 (k)), reflecting the model's poor fit during this critical reproductive stage. The high rRMSE values across all stages, particularly in Stage III where rRMSE reached 173.12% (Figure 3 (l)), highlight the model's significant underestimation of GPP. Similarly, PBIAS analysis revealed large negative biases, especially in Stage III (-49.75%), indicating systematic underestimation of peak photosynthetic activity during this stage.

After optimization (Figure 3 (m)-(p)), the GPP simulations showed substantial improvements. The R^2 for the entire growing season increased to 0.68 (Figure 3 (m)), demonstrating a stronger overall correlation between observed and simulated GPP values. Stage I and Stage III also saw notable enhancements, with R^2 values increasing

to 0.68 and 0.66, respectively (Figure 3 (n) and (p)). While Stage II remained relatively weaker compared to other stages, its R^2 still improved from 0.12 to 0.49 (Figure 3 (o)), indicating better model performance during the reproductive phase. The rRMSE values decreased significantly across all stages, particularly in Stage III, where rRMSE dropped from 173.12% to 36.30%, reflecting a substantial reduction in model error. Furthermore, the PBIAS values improved dramatically, with near-neutral biases across all stages. For example, the PBIAS for the entire growing season was reduced to 0.26%, and in Stage III, it was minimized to 1.42%. These results highlight the effectiveness of the optimization process in reducing both systematic and random errors in GPP simulations.

Overall, the optimization process significantly improved the model's ability to simulate half-hourly SIF and GPP across all growth stages, with the most notable improvements observed in Stage III for both SIF and GPP. While rRMSE values decreased substantially after optimization, indicating a reduction in model error, the PBIAS for SIF suggests some remaining issues with overestimation, particularly in Stage I and Stage II. For GPP, however, the optimization process successfully reduced both rRMSE and PBIAS, leading to more accurate and reliable simulations.

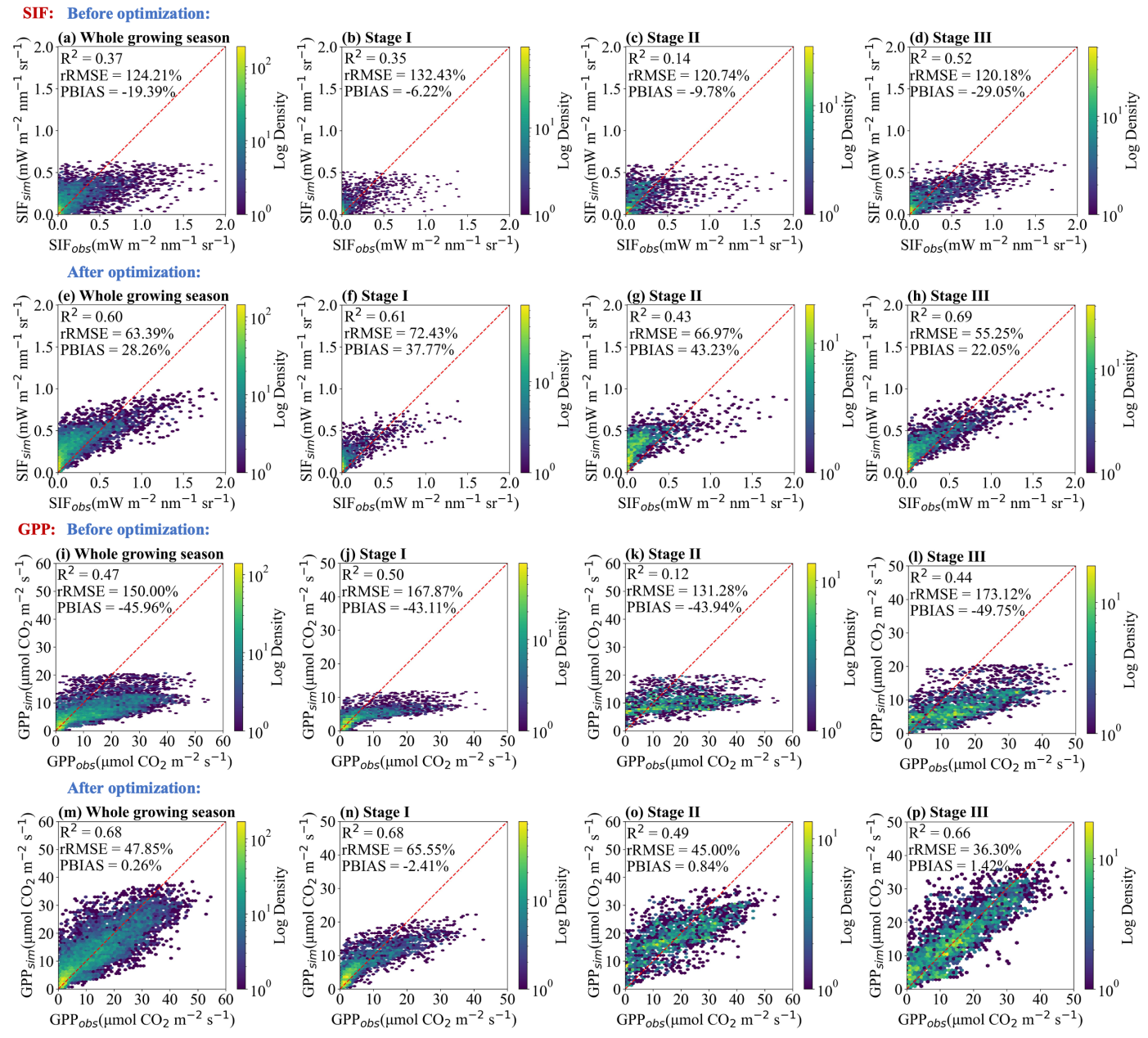

Figure 3. Comparisons between the simulated and observed half-hourly SIF for the entire growing season and at the I-III growth stages (a-d) before and (e-h) after optimization from 7:00-17:00 and between the GPP for the entire growing season and at the I-III growth stages (i-l) before and (m-p) after optimization at the half-hourly scale from 7:00-17:00. The data points are colored according to density. Because of the

limitation of SIF observation data, SIF data showed here is from June 29th, 2019 to September 10th, 2020 including one entire growing season. GPP is from 2018-2020 including three entire growing seasons.

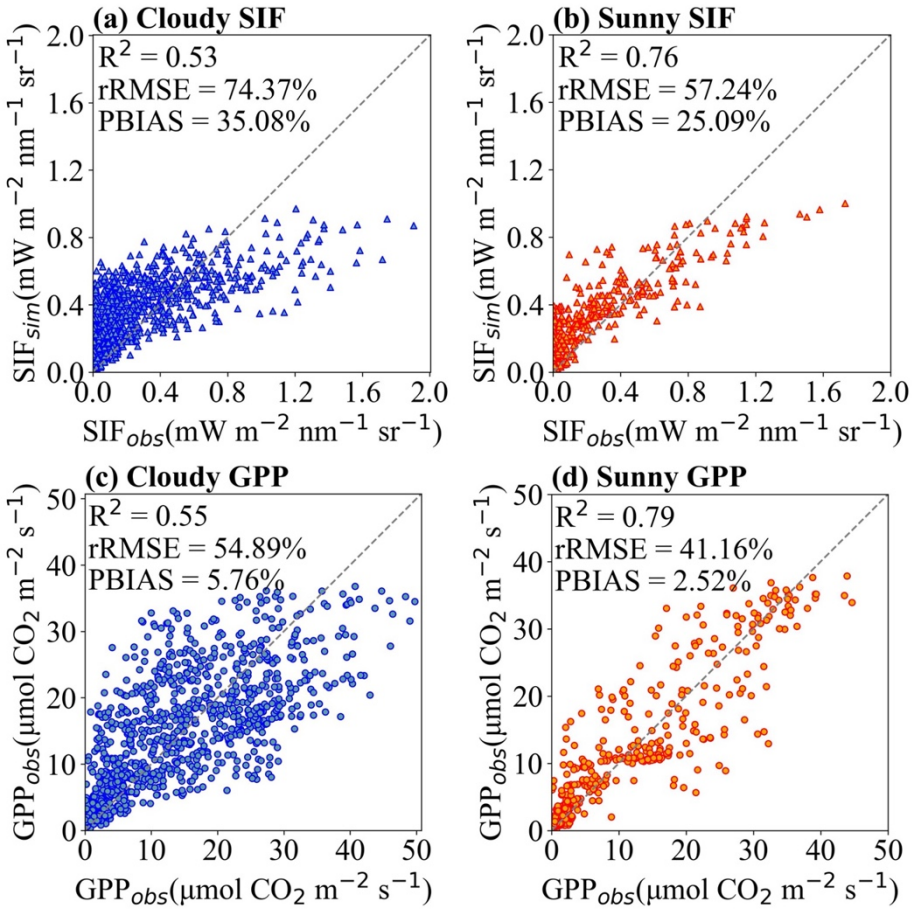

Figure 4. Comparison of optimized simulated and observed SIF data under (a) cloudy and (b) sunny conditions at the half-hourly scale from 7:00-17:00, between June 29th 2019 and September 10th 2020. Comparison of optimized simulated and observed GPP data under (c) cloudy and (d) sunny conditions at the half-hourly scale for 2018-2020. The blue points represent cloudy conditions, the orange points represent sunny conditions. Weather conditions were defined in section 2.1.3.

III-2-2 SIF and GPP under sunny and cloudy conditions

To further evaluate the performance of the optimized model, I examined the relationship between half-hourly simulated data (after optimization) and observed data under different weather conditions, specifically comparing cloudy and sunny conditions for SIF and GPP. This analysis allows for a detailed understanding of the model's ability to simulate photosynthetic dynamics and fluorescence under varying light environments, highlighting its strengths and limitations.

For SIF (Figure 4 (a), (b)), the model performed better under sunny conditions compared to cloudy conditions. The R^2 for sunny conditions reached 0.76 (Figure 4 (a)), indicating a strong correlation between simulated and observed SIF values. This high R^2 reflects the model's capability to accurately capture diurnal variations in SIF under conditions of stable, direct sunlight. In contrast, the R^2 for cloudy conditions was 0.53 (Figure 4 (b)), suggesting reduced model performance under diffuse light conditions. This discrepancy likely arises from the complexities of radiative transfer under cloud cover, where light scattering modifies the absorption and emission processes that influence SIF. The rRMSE and PBIAS values further support this trend, with rRMSE decreasing significantly from 74.37% on cloudy days to 57.24% on sunny days, and PBIAS improving from 35.08% to 25.09%. These results indicate that the optimized model not only captures overall patterns in SIF more effectively under sunny conditions but also reduces systematic errors associated with overestimation.

For GPP (Figure 4 (c), (d)), the model also exhibited better performance under sunny conditions. The R^2 under sunny conditions was 0.79 (Figure 4 (c)), compared to 0.55 under cloudy conditions (Figure 4 (d)). The lower R^2 under cloudy conditions suggests that the model struggles to simulate GPP accurately when diffuse light dominates, likely due to changes in photosynthetic efficiency and canopy light distribution. Similar to SIF, the rRMSE and PBIAS for GPP showed clear improvements under sunny conditions. The rRMSE decreased from 54.89% on cloudy days to 41.16% on sunny days, reflecting a substantial reduction in model error. Meanwhile, the PBIAS decreased from 5.76% to 2.52%, indicating that the model's bias in GPP estimation is much lower when direct sunlight is present. These findings suggest that the model can better simulate photosynthetic activity when light conditions are stable and predictable.

Overall, the model performed better under sunny conditions for both SIF and GPP, as demonstrated by higher R^2, lower rRMSE, and reduced PBIAS values. The stronger performance on sunny days underscores the model's ability to simulate photosynthetic dynamics under direct sunlight, where radiative transfer and light use efficiency processes are more straightforward. However, the lower performance under cloudy conditions highlights the model's limitations in accounting for the complexities introduced by diffuse light. Cloud cover affects the partitioning of absorbed photosynthetically active radiation (APAR) between photochemical and non-

photochemical pathways, altering the relationships between SIF, GPP, and light use efficiency. These results indicate potential areas for further improvement in the model, such as enhancing the representation of diffuse light effects and radiative transfer processes under variable atmospheric conditions.

III-2-3 Diurnal cycle of SIF and GPP

The averaged diurnal cycle data for each time point shown in Figure 5 were obtained by averaging the data for that time point over the entire growing season and over each growth stage of 2020. The simulated and observed SIF and GPP data all exhibited a hump shape at the diurnal step, with SIF and GPP values lower in the morning and evening than at midday. Therefore, the R^2 values between simulations and observations were high (all $R^2 > 0.84$). For SIF (Figure 5 (a)-(d)) before optimization, the overall R^2 for SIF during the entire growing season was 0.93, and across different growth stages, it ranged from 0.90 to 0.94. The highest R^2 was observed in Stage III (0.94) (Figure 5 (d)), indicating a good fit between the modeled and observed SIF. However, the model showed substantial errors in terms of rRMSE and PBIAS. In particular, the rRMSE was highest in Stage I (57.70%) (Figure 5 (b)), and PBIAS showed a significant overestimation during Stage I (43.23%) and underestimation in Stage III (-30.42%). Additionally, the model significantly underestimated the variability in SIF, particularly in Stages II and III, where the CV values were only 17.56% and 18.42%, respectively, compared to the observed 37.43% and 44.37%. After

optimization, the R^2 values improved significantly, reaching 0.99 for the entire season and consistently high values (0.98) across all growth stages. This indicates that the optimization effectively improved the model's ability to capture the SIF dynamics throughout the day. However, while R^2 improved, the rRMSE for Stage II increased from 20.14% to 51.50% (Figure 5 (c)), and PBIAS remained relatively high across all stages, particularly in Stage II (48.15%). In terms of variability, the CV values for the whole season decreased to 25.40%, showing that the model better captured the variability across stages, although the simulated variability remained lower than observed in Stage I (28.13% vs. 47.35%).

For GPP (Figure 6 (e)-(h)), the initial model performance showed an R^2 of 0.93 for the whole growing season (Figure 6 (e)), with slight variations across stages. The model performed best during Stage II ($R^2 = 0.95$) (Figure 6 (g)) and exhibited lower accuracy in Stage III ($R^2 = 0.84$) (Figure 6 (h)). The errors in rRMSE were fairly consistent across stages, with values ranging between 51.05% and 58.22%. PBIAS indicated a significant underestimation of GPP across all stages, with the largest underestimation in Stage III (-52.3%). The CV values before model optimization were much lower than observed, with the whole season CV at 6.10% compared to the observed 26.41%, indicating that the model underestimated the variability in GPP. After optimization, the GPP model performance improved significantly, achieving an R^2 of 0.97 for the entire growing season and even higher values during individual

growth stages, especially in Stage II (R^2 = 0.98). The optimization also reduced the rRMSE across all stages, particularly in Stage I (11.12%) and Stage II (14.65%), indicating a substantial reduction in model errors. Furthermore, the PBIAS improved dramatically, with the model closely aligning with observed GPP values, especially in Stage III, where PBIAS reached near 0 (0.12%). The optimization process also improved the model's ability to capture GPP variability, with the CV for the whole season increasing to 14.27%. While this is still below the observed value, it represents a notable improvement from the pre-optimization results.

However, long-term averaged data may not fully represent the model's overall performance due to the loss of details during averaging. This is highlighted in Figure 6, where the observed and simulated SIF (Figure 6 (a)) and GPP (Figure 6 (b)) at half-hour resolutions from 7:00–17:00 PM during DOY196–203 in 2020 are compared. I selected this period randomly. Notably, both the SIF and GPP observation showed intense fluctuations (CV(SIF) = 63.49%, CV(GPP) = 48.98%) throughout the day, different from the smooth hump-shaped patterns shown in Figure 5. In Figure 6 (a), the optimized SIF exhibits a more fluctuating hump-shaped pattern, closely aligning with the observed SIF (R^2 values improved from 0.19 to 0.48, and rRMSEs decreased from 79.89% to 70.95%) but with slightly weaker fluctuations than the observed SIF. However, the overestimation of optimized SIF is more significant than the initial SIF with PBIAS increased from 34.62% to 64.17%. The CV further emphasizes the model's

underestimation of variability. The observed SIF has a CV of 63.49%, whereas the model's CV before optimization was only 35.94%. After optimization, the CV decreased to 28.96%, indicating that while the model captured the general trend better, it still underestimates the true variability of SIF.

In Figure 6 (b), the prior GPP values are notably lower than the observed values, showing mild fluctuations that generally correspond to the observed data, but still with a relatively high R^2 value of 0.75, and the rRMSE was 54.14%. After optimization, the simulated GPP was more consistent with the observed data, with an R^2 value of 0.82, an rRMSE of 29.66%, and underestimation have been significantly improved with PBIAS changed from -31.60% to -2.36%. The CV analysis supports this improvement, with the observed GPP having a CV of 48.98%. Before optimization, the simulated GPP had a much lower CV of 8.66%, indicating a severe underestimation of variability. After optimization, the CV increased to 23.74%, showing that the model better captures the day-to-day fluctuations in GPP but still underestimates the full extent of variability present in the observed data.

In summary, the optimization process improved the model's performance for both SIF and GPP simulations, particularly in terms of R^2, which increased across all stages. However, despite the improvement in the correlation between modeled and observed data, some challenges remain in reducing the rRMSE and PBIAS, particularly for SIF in Stage II and in improving the model ability to capture the variability of SIF and GPP.

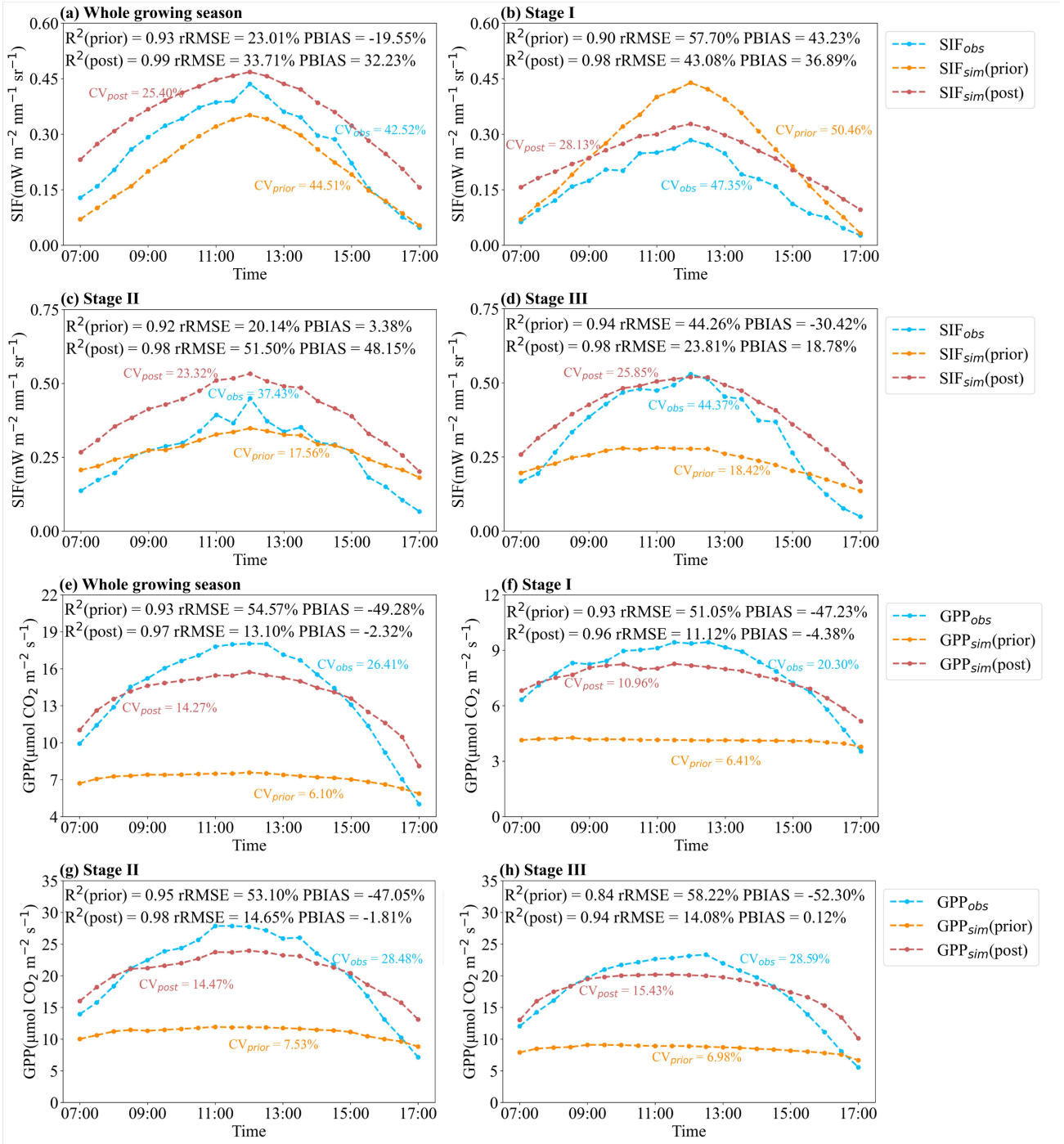

Figure 5. Diurnal variations (7:00–17:00) averaged over the entire growing season and three growth stages of 2020 for (a-d) SIF and (e-h) GPP. The blue points represent the observed data, the orange points represent the results of the model simulation before

optimization, and the red points represent the results of the model simulation after optimization. In the figure, "prior" means before optimization, and "post" means after optimization.

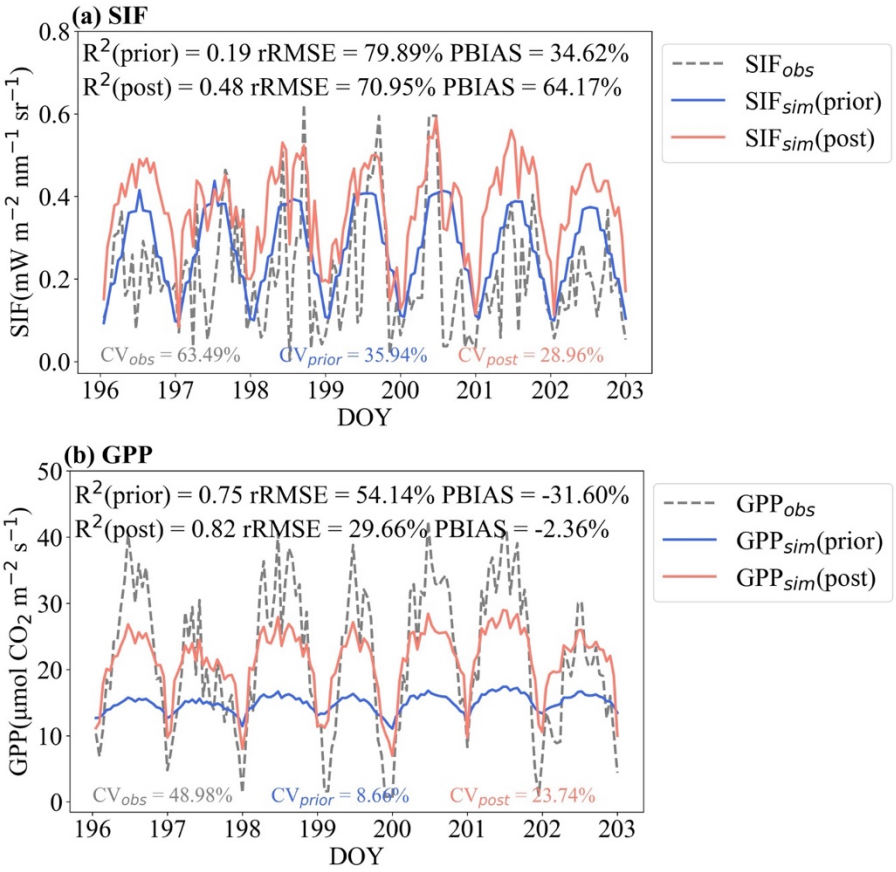

Figure 6. Observed and simulated (a) SIF and (b) GPP at half-hour resolution from 7:00 to 17:00 during DOY196-203 in 2020. The gray dotted line represents the observed data, the blue line represents the model simulation before optimization, and the red line represents the model simulation after optimization.

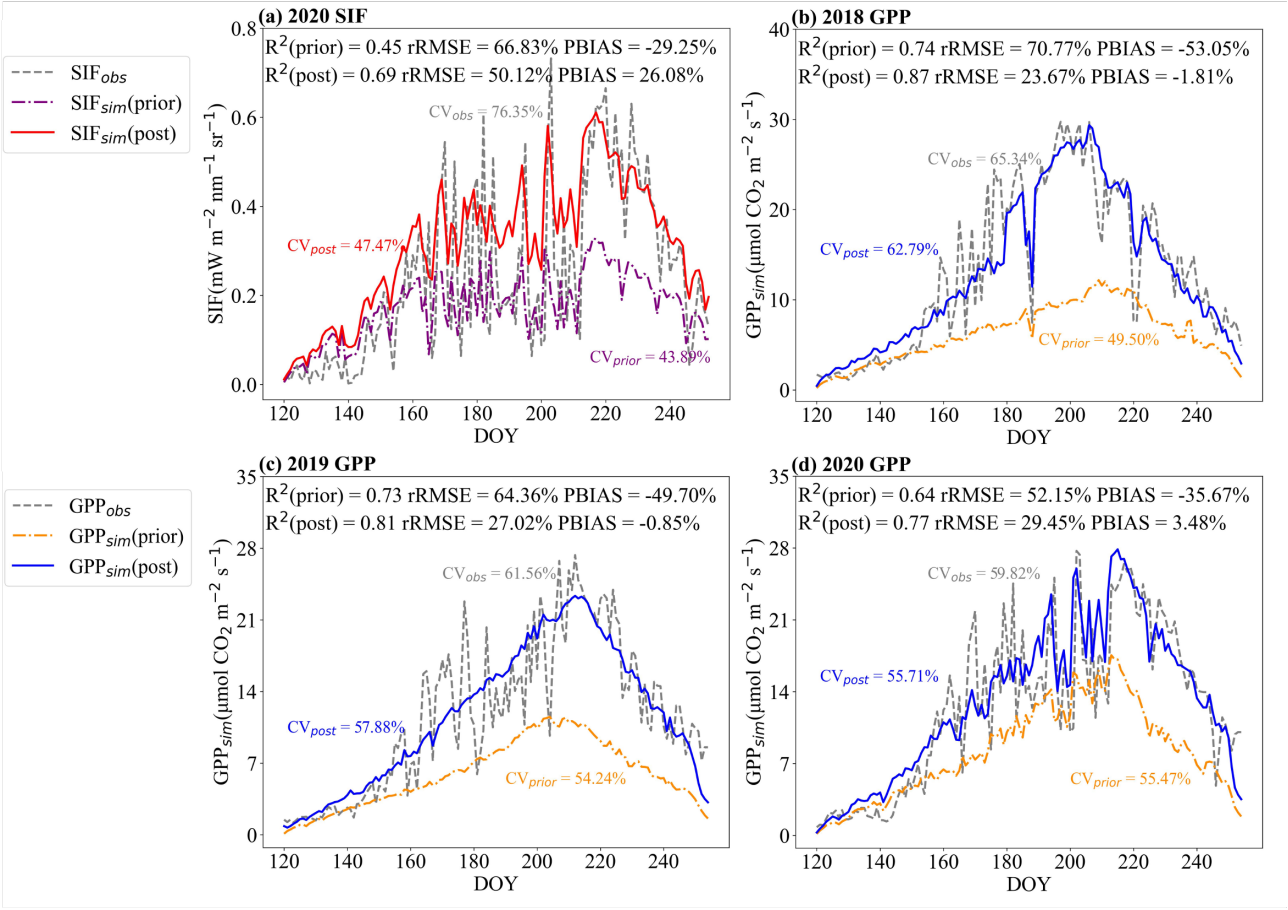

Figure 7. Comparison between (a) simulated SIF and observed SIF data for 2020 and (b-d) simulated GPP and observed GPP data before and after optimization for 2018-2020 at the daily scale. The gray dotted line shows the observed SIF in (a) and the observed GPP in (b-d). The purple line shows the simulated SIF before optimization, and the red dotted line shows the simulated SIF after optimization. The orange line shows the simulated GPP before optimization, and the blue dotted line shows the simulated GPP after optimization. The daily SIF and GPP data were obtained by averaging half-hourly data from 7:00 to 17:00 each day.

III-2-4 Improvement of SIF and GPP simulations at daily scale

The daily SIF and GPP values shown in Figure 7 were calculated by averaging half-hourly data from 7:00 to 17:00 for each day. Since complete SIF data were only available for 2020, the analysis of SIF performance focuses exclusively on this year. By averaging over daily timescales, this analysis highlights seasonal and stage-specific patterns while reducing the influence of short-term variability. Across the entire growing season, the optimization process resulted in significant improvements for both SIF and GPP simulations (Figure 7).

For SIF (Figure 7 (a)), the overall R^2 increased substantially from 0.45 to 0.69 after optimization, reflecting a much stronger correlation between simulated and observed daily SIF values. This improvement demonstrates that the optimization effectively reduced systematic discrepancies between the model and observations. Furthermore, the rRMSE decreased from 66.83% to 50.12%, indicating a significant reduction in overall error. However, the PBIAS showed a shift from underestimation (-29.25%) to overestimation (26.08%), highlighting a trade-off in bias correction post-optimization. This overestimation suggests that while the optimization improved the fit between simulated and observed SIF, it introduced a tendency to overpredict SIF values, particularly during periods of peak fluorescence.

For GPP (Figure 7 (b)-(d)), the improvements spanned three years (2018–2020), with the R^2 increasing for all years. The highest R^2 was observed in 2018 (0.87),

followed by 2019 (0.81) and 2020 (0.77), indicating consistent enhancement in model performance across years. The rRMSE showed particularly notable reductions, especially in 2018, where it decreased from 70.77% to 23.67%, indicating a marked reduction in error. Similarly, PBIAS improved from significant underestimation (-53.05%) to a near-neutral bias (-1.81%), demonstrating the optimization's effectiveness in aligning simulated GPP with observations. These results verify that the optimization process was robust, even for years without direct SIF data assimilation, as observed in 2018.

The CV analysis further supports these findings. For GPP, the pre-optimization model consistently underestimated variability across years and growth stages. After optimization, the model's ability to capture variability improved significantly. For instance, in 2018, the CV for GPP increased from 49.50% to 62.79%, approaching the observed CV of 65.34% (Figure 7 (b)). For SIF, however, the optimized model continued to underestimate variability compared to observations, as the simulated CV remained lower than the observed values. This underestimation indicates that while the model captured overall trends more accurately after optimization, it still struggled to reflect the full extent of natural fluctuations in SIF.

For the I-III growth stages, the evaluation indicators for daily simulations are presented in Table 2. Before optimization, the daily SIF simulations exhibited relatively weak correlations with observations, especially in Stages I and II, where R^2 values were

0.38 and 0.23, respectively. These stages are characterized by high variability in physiological and environmental conditions, which likely contributed to the poor model performance. In Stage III, the R^2 improved to 0.69, indicating better performance during the ripening stage, where canopy conditions stabilize. However, the model underestimated SIF significantly in Stage III, with a negative PBIAS of -41.43%. The rRMSE values were also high across stages, particularly in Stage I (79.91%) and Stage II (73.31%), highlighting substantial discrepancies between simulated and observed daily SIF values.

After optimization, the SIF simulation performance improved across all growth stages. The R^2 values increased markedly, particularly in Stage I (from 0.38 to 0.67), demonstrating enhanced model accuracy during the early vegetative phase. The rRMSE values also decreased substantially, especially in Stage III (from 52.36% to 23.74%), reflecting significant error reductions. However, the PBIAS showed an overestimation trend post-optimization, with values increasing in Stage I (from -6.26% to 47.44%) and Stage II (from -29.08% to 39.82%). Despite these biases, the model captured overall variability better, as indicated by improvements in CV values. For example, the CV in Stage I increased from 53.36% to 59.78%, moving closer to observed variability.

Table 2

The metrics of SIF and GPP at each growth stages at daily scale

	SIF Stage I	SIF Stage II	SIF Stage III	2018 GPP Stage I	2018 GPP Stage II	2018 GPP Stage III	2019 GPP Stage I	2019 GPP Stage II	2019 GPP Stage III	2020 GPP Stage I	2020 GPP Stage II	2020 GPP Stage III
R^2 (prior)	0.38	0.23	0.69	0.72	0.28	0.62	0.73	0.35	0.84	0.70	0.21	0.60
R^2 (post)	0.67	0.41	0.77	0.80	0.71	0.78	0.82	0.49	0.85	0.80	0.46	0.76
rRMSE (prior)(%)	79.91	73.31	52.36	81.34	64.06	55.58	79.79	47.34	61.69	67.39	38.87	48.84
rRMSE (post)(%)	75.17	68.11	23.74	46.66	11.39	20.87	41.83	28.21	16.72	40.89	33.09	18.37
PBIAS (prior)(%)	-6.26	-29.08	-41.43	-46.3	-61.35	-48.42	-48.86	-38.69	-57.23	-34.44	-23.79	-43.9
PBIAS (post)(%)	47.44	39.82	8.02	-6.02	0.84	-1.82	-7.75	16.04	-7.91	1.13	12.05	-0.68
CV (obs)(%)	100.37	62.90	44.95	87.95	20.91	40.52	84.72	31.78	37.24	85.42	28.82	33.60
CV (prior)(%)	53.36	29.01	29.50	49.58	15.64	40.27	53.39	15.55	38.63	56.59	14.73	43.04
CV (post)(%)	59.78	24.02	30.46	58.38	17.22	44.89	62.14	15.41	39.76	62.16	17.64	37.78

"prior" means before optimization, and "post" means after optimization.

R^2, rRMSE, PBIAS between simulations (both before and after optimization) and observations for different stages at daily scale, and CV for observations (CV(obs)), CV for simulation before optimization (CV(prior)) and CV for simulation after optimization (CV(post)).

For GPP (Table 2), the pre-optimization model performed better than for SIF, but it still exhibited significant underestimation across all stages. Before optimization, the R^2 values were relatively high in Stage I (0.72) and Stage III (0.62) but were notably

lower in Stage II ($R^2 = 0.28$), reflecting challenges in simulating photosynthetic activity during this transitional phase. Severe underestimation of GPP was evident from the negative PBIAS, particularly in Stage II (-61.35%) and Stage III (-48.42%). The rRMSE values were high across all stages, with Stage I showing the highest error (81.34%).

After optimization, the GPP simulation improved significantly across all growth stages. The R^2 values increased in all stages, with the most substantial improvement in Stage II (R^2 increasing from 0.28 to 0.71), reflecting better model accuracy during this critical phase. The rRMSE values also decreased sharply, particularly in Stage II (from 64.06% to 11.39%) and Stage III (from 55.58% to 20.87%). Furthermore, the PBIAS improved dramatically, especially in Stage II, where it shifted from -61.35% to 0.84%, indicating a substantial reduction in bias. The CV for GPP increased closer to observed values, particularly in Stage I (from 49.58% to 58.38%), further demonstrating the model's enhanced ability to capture variability.

In summary, the optimization process significantly improved the model's performance for daily SIF and GPP simulations, reducing errors and increasing R^2 across the growing season. Notably, the optimization also enhanced the model's ability to simulate GPP in 2018, a year without SIF observations used in the assimilation process, thereby validating the robustness of the optimized model. However, despite these improvements, the model continues to overestimate SIF in certain stages and

underestimate variability, particularly for SIF, indicating the need for further refinements.

III-3 Data assimilation of the sparsely satellite SIF dataset

Satellite-derived SIF datasets, such as CSIF, have emerged as valuable tools for estimating gross primary production (GPP) at regional and global scales. These datasets offer unique opportunities to expand the spatial and temporal scope of GPP studies, providing a bridge to regions where ground-based measurements are sparse or unavailable. However, their application to process-based models like VISIT-SIF is not without challenges. The sparse temporal resolution, coarse spatial coverage, and inherent uncertainties caused by atmospheric interference in satellite-derived SIF data can hinder their accuracy and limit their effectiveness for direct model optimization. Despite these limitations, the assimilation of CSIF data into process-based models represents a promising step towards improving simulations of both SIF and GPP at larger scales.

As shown in Figure 8 (a), the comparison between CSIF and ground-based SIF observations yielded an R^2 of 0.53, an rRMSE of 47.71%, and a PBIAS of -22.06%. These metrics indicate that while CSIF data successfully capture general trends in SIF dynamics, discrepancies remain, particularly in terms of underestimation. This underestimation is likely due to the coarser resolution of CSIF, which averages out fine-scale variations, as well as uncertainties in the algorithms used to derive SIF from

satellite measurements. Additionally, atmospheric effects and sensor-specific errors can introduce noise, further contributing to the observed discrepancies. Despite these limitations, the moderate R^2 suggests that CSIF provides a reasonable basis for model optimization, particularly in regions lacking ground-based observations.

Following data assimilation, the VISIT-SIF model was optimized using the CSIF data, with parameter adjustments detailed in Table 1. The optimization process aimed to improve the model's ability to simulate both SIF and GPP, and the results were validated against site-observed data. For SIF, the optimized model achieved an R^2 of 0.51, an rRMSE of 68.79%, and a PBIAS of 30.87% when compared to site observations (Figure 8 (b)). While the optimization improved the model's correlation with observed SIF, as indicated by the R^2, the increase in PBIAS reflects a tendency to overestimate SIF post-optimization. This overestimation suggests that the optimization process, while effective in reducing underestimation biases, may have introduced a new bias, particularly in capturing peak SIF values.

For GPP, the CSIF-optimized model showed better performance, with an R^2 of 0.65, an rRMSE of 52.07%, and a near-neutral PBIAS of -2.72% when compared to site-observed GPP (Figure 8 (c)). These results indicate that the model effectively reduced bias in GPP simulations, as evidenced by the low PBIAS, which represents a significant improvement compared to pre-optimization values. The R^2 value demonstrates that the model captures a substantial portion of the variability in observed

GPP data, although it remains slightly lower than the R^2 achieved through optimization with site-observed SIF. The rRMSE values, while indicating room for further refinement, highlight the model's capability to reproduce general patterns and variability in GPP at daily scales.

Although the performance of the CSIF-optimized model was slightly inferior to that achieved with site-observed SIF data, the results are still promising. For instance, the lower R^2 for GPP simulations compared to site-optimized results underscores the limitations of satellite-derived data in capturing fine-scale variability. However, the near-neutral PBIAS and improved rRMSE suggest that CSIF data can still serve as a valuable substitute for ground-based data in regions where such measurements are unavailable. The capability of CSIF-optimized models to reproduce GPP dynamics with reasonable accuracy demonstrates the potential of satellite-derived SIF data to support large-scale carbon cycle studies.

For the parameter values(Figure 9), our results suggest significant differences in the parameters optimized using ground-based SIF data compared to those optimized using satellite-derived CSIF data. Notably, f (spectral adjustment parameter) and ρ_{cd} (diffuse reflectance) showed substantial changes when ground-based SIF data were used for optimization, increasing by 80% and 353%, respectively. These parameters were ranked first and second in the SHAP analysis regarding their contribution to SIF simulations. Their large changes suggest that ground-based data, with its higher

precision and finer spatial resolution, allows for more detailed adjustments of local optical and physiological processes, leading to more substantial improvements in the model. However, when we used satellite-derived CSIF data for optimization, the changes in these parameters were more modest. For instance, f decreased by 20%, and ρ_{cd} experienced much smaller adjustments.

Overall, the results highlight the utility of satellite-derived data for data assimilation in the VISIT-SIF model. While the performance of the CSIF-optimized model does not match that of site-observed SIF optimization, it remains an effective alternative for regions where in-situ observations are limited or unavailable. The study underscores the importance of improving satellite SIF retrieval algorithms and reducing uncertainties associated with atmospheric and sensor effects to enhance the reliability of CSIF data. Future efforts should focus on integrating high-resolution remote sensing datasets or combining multiple data sources to further refine the performance of process-based models like VISIT-SIF.

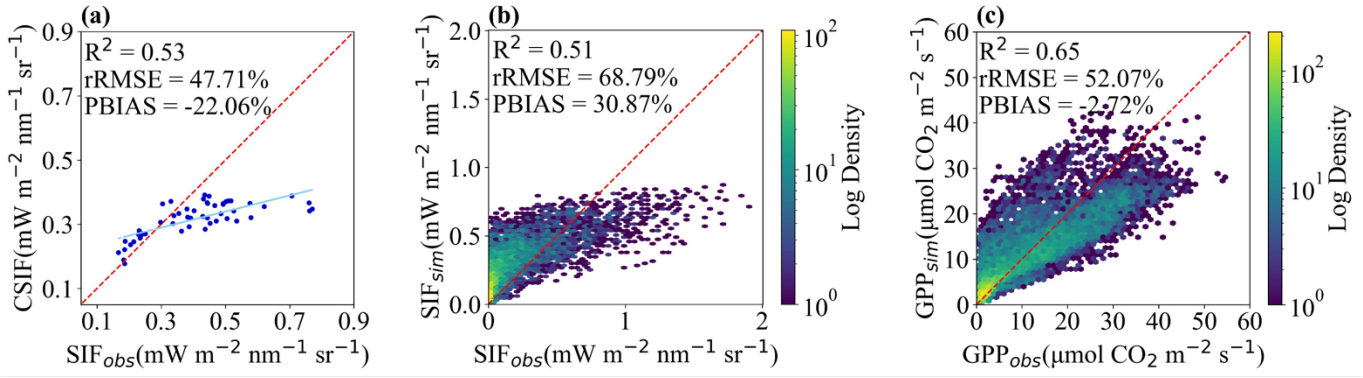

Figure 8. (a) Comparison between the CSIF and SIF_{obs}, (b) comparison between the SIF simulated by the model optimized by the CSIF and observed SIF data at the half-hourly scale, and (c) comparison between the GPP simulated by the model optimized by the CSIF and observed GPP data at the half-hourly scale.

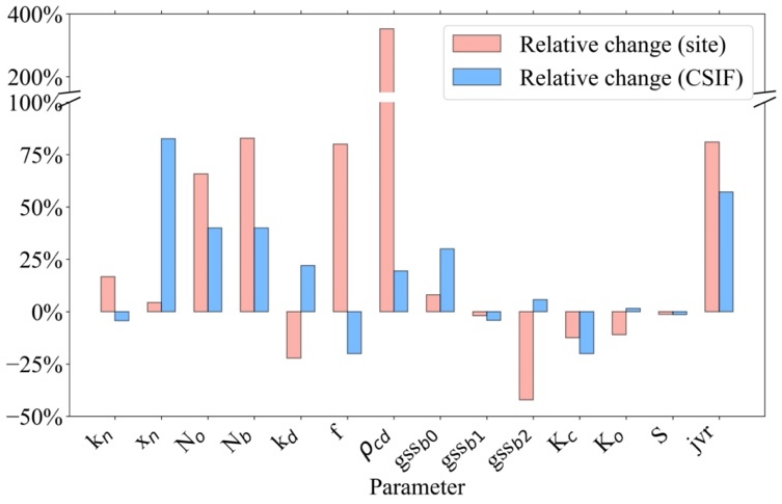

Figure 9. Comparison of parameter values optimized by site SIF data and CSIF data.

III-4 The relationships between SIF and GPP across different time scales

For the ground-based SIF-GPP relationship (Figure 10 (a-d)), the data shows consistently strong linear correlations between SIF and GPP across different timescales (half-hourly, daily, and weekly) and growth stages. These correlations highlight the robustness of SIF as a proxy for GPP under varying temporal resolutions and phenological conditions. At the half-hourly scale, the R^2 values are relatively high across the whole growing season and individual stages, ranging from 0.70 in Stage I to 0.78 in Stage III. These results reflect the model's capability to capture real-time photosynthetic activity, especially during the ripening stage, when the canopy structure and chlorophyll content are stable. At the daily scale, the correlations improve significantly with temporal aggregation, reaching 0.92 in Stage III. This improvement aligns with the smoothing effect of daily averaging, which reduces noise and short-term fluctuations, thereby enhancing the linear relationship between SIF and GPP. At the weekly scale, the relationship is strongest, with R^2 values as high as 0.96 in Stage I and 0.95 in Stage III. This indicates that longer temporal aggregation provides a more integrated view of the relationship between SIF and GPP, capturing overall trends while minimizing the impact of transient environmental variability.

Before optimization (Figure 10 (e-h)), the model's SIF-GPP relationship displays much weaker linear correlations, especially at finer timescales. At the half-hourly scale, the R^2 values are low, ranging from 0.32 to 0.42 across different stages, suggesting that

the pre-optimization model struggles to capture rapid fluctuations in photosynthetic activity. These weak correlations likely reflect inaccuracies in the model's representation of diurnal SIF and GPP dynamics, which are influenced by factors such as light availability, canopy structure, and physiological processes. At the daily scale, the R^2 values improve slightly with temporal aggregation but still indicate poor performance, particularly in Stage II, where the R^2 drops to 0.17. This stage, characterized by high variability in physiological and environmental conditions, poses significant challenges for the pre-optimization model. At the weekly scale, the model shows some improvement, achieving an R^2 of 0.86 in Stage I. However, discrepancies remain significant in Stage II ($R^2 = 0.15$), highlighting persistent model deficiencies in capturing the relationship between SIF and GPP during this critical growth phase.

After optimization (Figure 10 (i-l)), the model's performance improves substantially across all timescales. At the half-hourly scale, the R^2 values increase significantly, reaching as high as 0.88 in Stage II and 0.87 in Stage III. These improvements reflect the effectiveness of the optimization process in refining parameter representations and reducing systematic errors in real-time simulations. At the daily scale, the optimization results in further enhancements, with R^2 values increasing to 0.92 in Stages I and III. However, Stage II still shows relatively weaker performance ($R^2 = 0.73$), indicating lingering challenges in accurately simulating SIF-GPP dynamics during this transitional stage. The weekly scale shows the most dramatic

improvements, with the optimized model achieving very strong correlations, such as an R^2 of 0.99 in Stage I. This near-perfect correlation underscores the model's ability to capture aggregated trends and reduce variability-induced errors at longer temporal scales.

In conclusion, data assimilation significantly improves the model's ability to simulate the SIF-GPP relationship, particularly at finer timescales like half-hourly and daily. The enhanced correlations across growth stages and temporal resolutions demonstrate the value of optimization in refining process-based models like VISIT-SIF. However, the model's performance in Stage II remains weaker compared to other stages, even after optimization. This suggests the need for further refinement in the model's representation of physiological and environmental processes during the reproductive phase, which is marked by high variability in canopy structure and light use efficiency. Future work could focus on improving parameterizations specific to this stage to further enhance the model's accuracy.

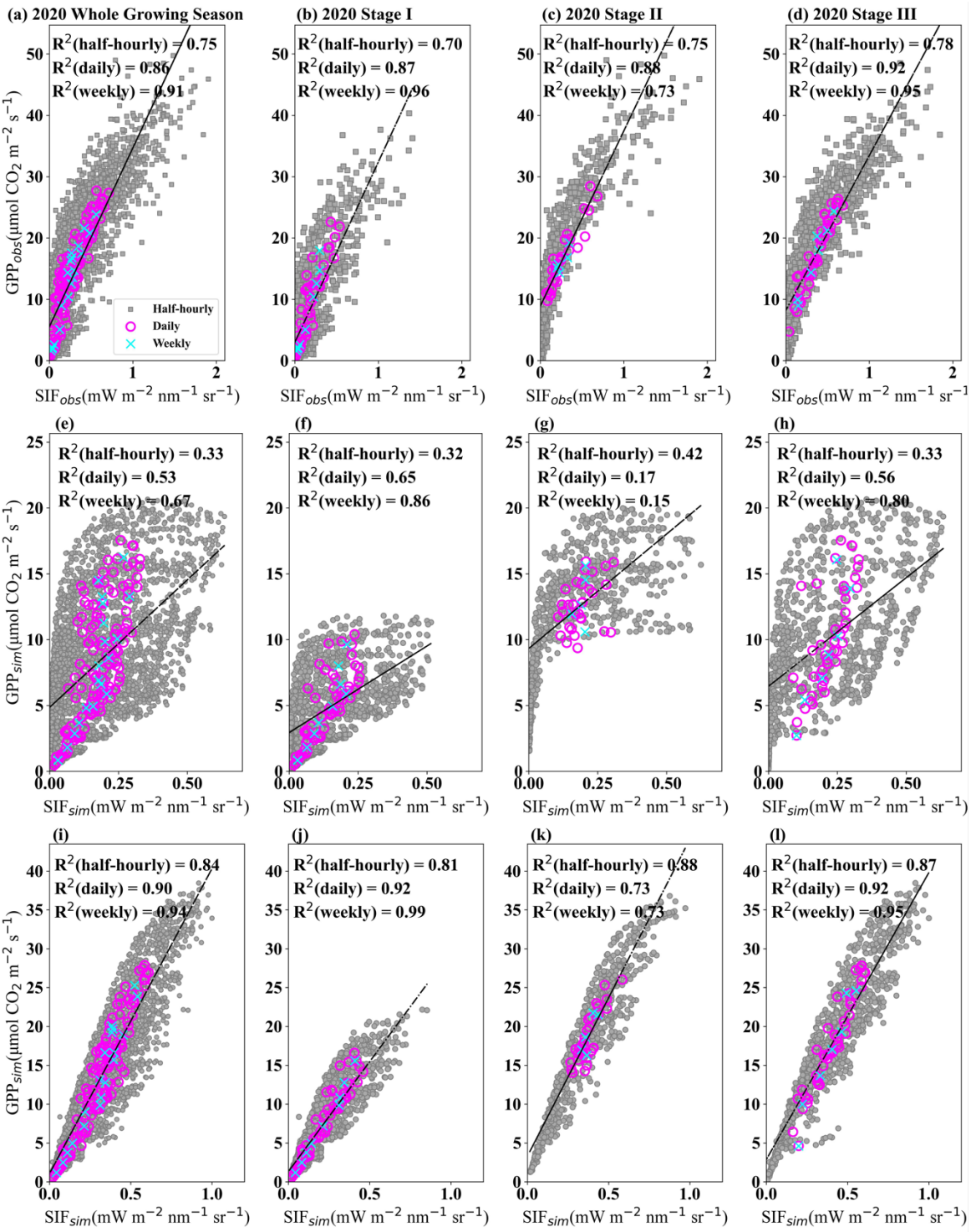

Figure 10. Relationships between ground-based SIF and GPP for (a) whole growing season and (b-d) each growth stage from half-hourly to daily to weekly steps for 2020 for daytime (7:00-17:00); (e-h) for simulated SIF and GPP by initial VISIT (before

optimization); (i-l) for simulated SIF and GPP by optimized VISIT. The grey, pink, and cyan in the figures represent half-hourly, daily (1 d) and weekly scales, respectively. The black line is a linear regression for the half-hourly data, and linear fit for other time scales is not shown in the figures.

III-5 The relationship between SIF and GPP and possible impacts of environmental factors

The relationship between observed SIF and GPP is complex (Guanter et al., 2012; Paul-Limoges et al., 2018), and the sensitivity of GPP to SIF has not yet received sufficient attention (Wang et al., 2020). To improve our understanding of the relationship between GPP and SIF, it is important to study the dynamic relationship between SIF and GPP at the ecosystem scale. In this section, I investigated the effects of environmental factors (temperature, precipitation, and sunlight) on the SIF–GPP relationship.

To test the influence of temperature (Figure 11 (a-c)), I divided the rice growing season into three periods with different temperatures: May (average daytime (7:00-17:00) temperature: 20.00 °C), June and July (24.08 °C), and August to mid-September (29.45 °C). As the temperature increased, the linear correlation of SIF–GPP strengthened, with R^2 values changing from 0.44 in May to 0.80 in August-September, and rRMSEs decreased from 0.53 to 0.27. With increasing temperature, the slopes of

SIF–GPP initially increased and then decreased slightly (13.75 in May, 27.57 in June and July, and 25.05 in August and September).

To investigate the impact of precipitation, I categorized the growing season into 2 periods with and without rainfall (Figure 11 (d-e)). The observed SIF–GPP exhibited a weaker linear relationship during rainfall (R^2 = 0.65, rRMSE = 0.54) than during nonrainfall periods (R^2 = 0.76, rRMSE = 0.41), but the slope of SIF–GPP under rainy conditions (36.39) was greater than that under nonrainfall conditions (28.92).

To examine the impact of sunlight on SIF–GPP (Figure 11 (f-g)), I compared SIF-GPP under cloudy and sunny conditions (identified in Section 2.6). A stronger linear relationship of SIF–GPP was found under sunny conditions (R^2 = 0.80, rRMSE = 0.41) than under cloudy conditions (R^2 = 0.72, rRMSE = 0.45). However, there were no significant differences between the two slopes, as the slope under sunny conditions was 30.99, which was slightly greater than 29.94 for cloudy conditions.

In summary, the SIF–GPP correlation was affected by environmental factors (Verma et al., 2017; Paul-Limoges et al., 2018). I found that the greater the temperature was, the stronger the correlation between the rice SIF and GPP (Figure 11 (a-c)). I also noted that the slope of the linear model was greatest in June and July, which is consistent with the conclusion reached by Chen et al.(2021) in which the most pronounced seasonal amplitude of GPP/SIF occurred in intermediate temperature ranges. The

correlation between SIF and GPP was better under nonrainy conditions than under rainy conditions (Figure 11 (d) and (e)). The slopes of the SIF and GPP equations were significantly greater under rainy conditions than under nonrainy conditions. However, I found that the slopes under sunny and cloudy (cloudy but not rainy) conditions were not significantly different. Therefore, it is concluded that temperature and precipitation have the most significant effects on SIF–GPP. Thus, to improve the accuracy of GPP inversion using SIF, it is important to consider the spatial and temporal variations in the SIF–GPP relationship, as well as environmental factor constraints and PFTs, especially when conducting long-term large-scale studies (Chen et al., 2021b; Guanter et al., 2012; Jeong and Park, 2021). However, I should consider that precipitation not only affects vegetation but also affects the instruments that detect SIF and GPP, especially the instrument for detecting SIF, which is obtained by inverting vegetation spectra. The effect of precipitation on the reflectance spectra of vegetation should also be significant, and the GPP, which is calculated by calculating gas exchange, is affected by precipitation less than that of SIF. In other words, the results I obtained were not only accompanied by changes in the physiological state of the vegetation but also by instrumental errors, and it is challenging to estimate which factor had a more significant impact on the results.

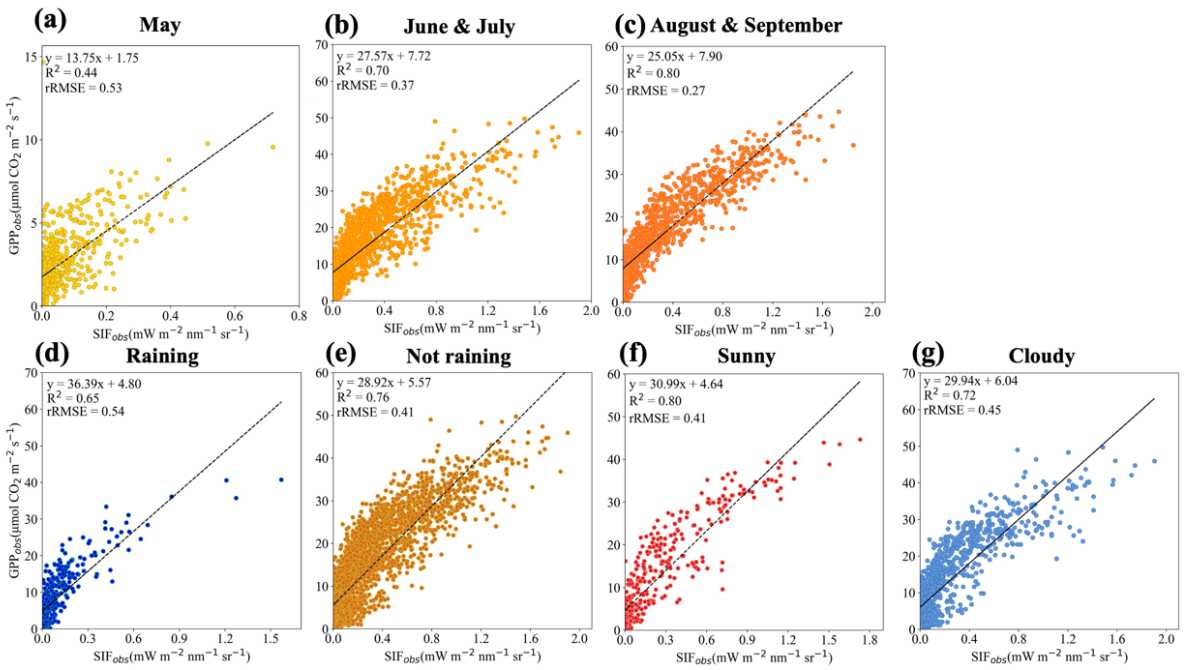

Figure 11. The relationship between ground-based SIF and GPP (a) in May (average daytime temperature of 7:00-17:00, 20.00 °C), (b) in June and July (24.08 °C), and (c) in August to mid-September (29.45 °C) under different weather conditions, including (a) rain, (b) no rain (sunny or cloudy but not rainy conditions), (c) sunny, and (d) cloudy (cloudy but not rainy conditions), at a half-hourly scale from 7:00-17:00. Linear regression lines are shown in black, and regression statistics are given. R^2 values and rRMSEs represent the coefficient of determination and the relative RMSE of the fitted linear regressions, respectively.

IV. Discussion

IV-1 Key parameters of SIF and GPP simulation

The SHAP analysis of model parameters provides a comprehensive view of the contributions and influence of individual parameters on the simulation accuracy of both GPP and SIF. This analysis reveals several key insights into the underlying processes that drive model performance and provides direction for future model improvements.

Our SHAP analysis for GPP simulations demonstrates that parameters related to nitrogen content which effect V_{cmax} a lot, particularly canopy-top leaf nitrogen (N_o), play a critical role in determining the accuracy of GPP estimates. This aligns with previous studies, which have emphasized that terrestrial carbon cycle models exhibit high sensitivity to V_{cmax} (Bonan et al., 2011; Farquhar et al., 1980; Kattge & Knorr, 2007; Rogers, 2014; Sargsyan et al., 2014; Walker et al., 2017), and the importance of N_o aligns with existing studies (Ackerly, 1992; Kattge et al., 2009; Terashima & Hikosaka, 1995; Wright et al., 2004) that emphasize the role of nitrogen in enhancing photosynthetic capacity, as nitrogen is directly involved in the synthesis of key photosynthetic proteins, including Rubisco.

In addition to nitrogen-related parameters, the spectral adjustment parameter (f) and carboxylation efficiency (Kc) also exhibit high SHAP values. These parameters directly influence the model's capacity to capture variations in light use efficiency (LUE) and overall photosynthetic rates. This partially explains why LUE-based models have

been shown to successfully simulate GPP with relatively high accuracy (Bao et al., 2022; Huang et al., 2021; Monteith, 1972; Zhou et al., 2016). The prominent impact of these parameters underlines the model's sensitivity to both biochemical and radiative transfer processes.

The SHAP analysis for SIF simulations highlights a different set of dominant parameters, with the parameter f and diffuse reflectance (ρ_{cd}) emerging as the most influential. These parameters are directly linked to the energy partitioning processes within the canopy, affecting how APAR is distributed between fluorescence emission, heat dissipation, and photosynthesis. This is consistent with previous studies showing that SIF is more directly related to APAR than GPP under certain conditions (Frankenberg et al., 2012; Jenkins et al., 2007; Knyazikhin et al., 1998; Yang et al., 2018; Yang et al., 2015). For example, Yang et al. (2018) found SIF is a better proxy of APAR than GPP at half-hourly resolution in a rice paddy. Yang et al., (2015) also found that SIF has a stronger correlation with APAR than with GPP at diurnal and seasonal scales in forests. The dominance of these parameters suggests that the SIF model is particularly sensitive to radiative transfer processes, more so than to the biochemical processes that drive GPP. The strong impact of the diffuse PAR attenuation (k_d) on SIF simulation further emphasizes the importance of accurately modeling light interactions within the canopy. The influence of k_d, along with ρ_{cd}, indicates that variations in light scattering and absorption significantly affect the amount of energy

re-emitted as fluorescence, making these factors critical for SIF predictions. This divergence in parameter sensitivity highlights the complex nature of SIF-GPP relationships, as different processes dominate the simulation of each variable.

IV-2 Impact of data assimilation on model performance

In this study, I observed that the data assimilation process brought notable improvements to the model's performance in simulating both SIF and GPP. However, one interesting result is that GPP simulations were generally more accurate than SIF simulations across the entire growing season and during each growth stage, even though the data assimilation process specifically used SIF data to optimize the model parameters. Bacour et al. (2019) also observed similar findings, where GPP simulations outperformed SIF simulations across various plant functional types on a monthly timescale. Similarly, Parazoo et al. (2020) observed that CLM5.0 demonstrated stronger capability in simulating GPP compared to SIF evaluated by ground-based data. Several factors may contribute to this outcome, which warrants a more detailed discussion.

The assimilation of ground-based SIF data helped to optimize key parameters related to APAR like f and ρ_{cd} (see Table 1), but the SHAP analysis revealed that GPP is more directly influenced by parameters related to V_{cmax} than SIF. Verma et al. (2017b) found that SIF provided very weak constrains on V_{cmax} estimates. Yang et al. (2018) found a stronger relationship between SIF and APAR ($R^2 = 0.82$) than SIF and GPP ($R^2 = 0.76$). Therefore, the assimilation of SIF data, while improving both GPP and SIF

simulations, had a more pronounced effect on GPP because it relies more heavily on physiological processes that were well-represented and calibrated in our model by previous studies (Ichii et al., 2013; Ito, 2008, 2019; Ito et al., 2005, 2005; Ito & Inatomi, 2012; Ito & Oikawa, 2002). Most ecosystem models, including the one used in this study, are designed with the primary goal of accurately simulating GPP, so that the model's structure and parameterization are better optimized for GPP than for SIF. This suggests that equating SIF and photosynthesis is an oversimplification that could diminish the effectiveness of SIF as a biophysical parameter in GPP models. A more accurate estimation of GPP may be achieved by integrating SIF or APAR with additional environmental variables.

Interestingly, the parameters f and ρ_{cd} showed significant changes when ground-based SIF data were used for data assimilation, increasing by 80% and 353%, respectively. These two parameters ranked first and second in SHAP analysis in terms of their contribution to SIF simulations. However, when satellite-derived CSIF data were used, the changes in these parameters were more modest, with f decreasing by 20% and ρ_{cd} experiencing much smaller changes. This suggests that ground-based data may provide higher precision and better inform the model about local optical and physiological processes, leading to more significant parameter adjustments. The higher resolution and precision of ground-based data enable the model to capture complex canopy interactions more effectively, further supporting the advantages of ground-

based SIF data for optimizing simulations at the site level. In contrast, satellite data, while valuable for large-scale applications, may not drive parameter adjustments as effectively due to its coarser resolution.

Both ground and satellite-based SIF observations are influenced by several factors, including atmospheric conditions, sensor errors, and viewing angles. These sources of uncertainty can propagate through the data assimilation process, potentially reducing the accuracy of the SIF simulations. For example, the uncertainties in satellite-derived CSIF data may result from the lower spatial resolution and more noises captured by satellite sensors. This could partly explain the relatively smaller changes in key parameters when CSIF data were used for optimization. In contrast, ground-based data, being more localized and precise, could lead to more substantial parameter adjustments, as seen in the optimization process. These uncertainties in SIF data may ultimately limit the accuracy of SIF simulations compared to GPP, which is more reliant on internal model representations of ecological processes and is less affected by external measurement errors. However, the representativeness of in situ SIF data is limited due to the small number of instrumented sites and their restricted coverage. The choice between in situ and satellite data should depend on the context and objectives of the research.

IV-3 Temporal scale of SIF-GPP relationships

In this study, the aggregation of half-hourly data into daily and weekly time scales revealed a clear trend of stronger linear correlations between SIF and GPP. This pattern aligns with previous research (Damm et al., 2015; Yang et al., 2018; Zhang et al., 2016), which suggests that at larger temporal scales, environmental noise and physiological variability are smoothed out, allowing for a more direct comparison between these two variables. Zhang et al. (2016) observed that the SIF-GPP relationship often exhibits non-linear behavior at hourly time scales, particularly during the early part of the day when light saturation and variations in photochemical quenching (PQ) and non-photochemical quenching (NPQ) are most influential (Parazoo et al., 2020). Similarly, Gu et al. (2019), through the derivation of fundamental equations, explored the complex dynamic relationship between SIF and photosynthesis, demonstrating that the nonlinearity in the instantaneous SIF-GPP relationship arises from fundamental differences in the light response behaviors of SIF and GPP. Additionally, Damm et al. (2015) highlighted that the applicability of the SIF-GPP relationship varies significantly across ecosystems, with the R^2 values ranging from 0.48 in temperate mixed forests to 0.59 in grasslands and 0.88 in croplands. These findings help explain the relatively strong linear relationship observed between SIF and GPP, even at the half-hourly scale, in the rice paddy analyzed in this study. Furthermore, Li et al. (2020) found the relationship between SIF and V_{cmax} in rice is dynamic throughout the growing season.

Before flowering, the correlation between SIF and V_{cmax} is relatively weak, but after flowering, this correlation strengthens significantly, particularly between far-red SIF (760 nm) and V_{cmax}. This dynamic is also evident at the half-hourly scale in this study, as indicated by the higher R^2 values of SIF-GPP during Stages II and III compared to Stage I.

These results suggest that while SIF can serve as a proxy for GPP at shorter time scales, its reliability improves considerably when data are aggregated over longer temporal periods. However, despite its potential as a robust tool for estimating GPP, SIF's effectiveness varies significantly across ecosystems. Further research is needed to better understand how canopy structure, photosynthetic efficiency, and temporal scale influence the relationship between SIF and GPP.

IV-4 Limitations and future perspectives

While this study demonstrates the effectiveness of data assimilation techniques in improving the simulation accuracy of both SIF and GPP, several limitations need to be addressed. These limitations arise from the model's sensitivity to specific parameters, some simplified calculation processes, and the inherent challenges in large-scale ecological modeling.

As discussed in Section 4.1, the parameters most influential for SIF simulation are not necessarily those most critical for GPP simulation. This discrepancy highlights the

complex relationship between SIF and GPP resulting from the different physiological processes of photosynthesis and chlorophyll fluorescence, and indicates the possible non-linear SIF-GPP relationship. Future research could focus on more detailed modeling of these biophysical processes, improving the model's sensitivity and parameterization for both SIF and GPP.

As noted in Section 3.2.2, the model's performance is less robust under cloudy conditions. This suggests that further refinement of the radiative transfer module is necessary, particularly to improve the model's handling of diffuse radiation under cloudy skies. It is also important to introduce parameterizations specific to different growth stages and environmental conditions, which could better capture the variability in SIF and GPP across seasons. Integrating machine learning techniques alongside process-based modeling could further enhance parameter tuning and model performance.

Additionally, as discussed in Section 3.1 and 3.2, the model's performance in simulating SIF during the stage II was concerning, and I think it may be that VISIT has systematic errors in simulations with large SIF fluctuations and large SIF values, a similar challenge encountered in other SIF simulation studies (Norton et al., 2019). The underestimation of strong SIF values by the model may be due to the lack of precision in the estimation of key physiological parameters such as photosynthetic efficiency and

chlorophyll content. These parameters need to be further optimized in future studies, especially when modeling ecosystems with significant seasonal variation.

In the VISIT model, I introduced r (the fraction of SIF at 760 nm) and two look-up tables $r_{oz/sz}$ (equation(13)) and $r_{\mathrm{shade/sun}}$ (equation(14)) from SCOPE to simplify the calculation. Although this approach is effective in simplifying the calculation, this approximation may introduce some errors, especially in scenarios with complex vegetation structure and optical properties. For r, future improvements can try a more refined parametric model that directly simulates the nonlinear relationship between C_{ab} and r to further improve the accuracy of the model. In addition, I used the two look-up tables to simplify the simulation of the SIF emission of shaded leaves. This simplified treatment may not be sufficiently accurate under some complex observational geometries, especially when solar radiation intensity or vegetation structure is highly heterogeneous. Future studies could consider using more complex geometric models or integrating real-time optical measurements to improve the accuracy of geometric coefficient calculations.

The study was conducted in a specific rice paddy ecosystem, which may limit the generalizability of the findings to other vegetation types. Future research should test the optimized VISIT-SIF model in different ecosystems with varied canopy structures and environmental conditions to assess its broader applicability.

V. Conclusions

In this study, SHAP analysis revealed the key parameters driving the model's performance. Nitrogen-related parameters were identified as crucial for GPP simulations, while radiative transfer parameters played a significant role in SIF modeling. This analysis highlighted the model's sensitivity to specific physiological and environmental processes, providing deeper insights into the factors influencing GPP and SIF dynamics. Then, I applied data assimilation of both ground-measured SIF and satellite SIF data (CSIF) to improve the GPP simulation accuracy in a process-based VISIT model for a rice paddy. By incorporating ground-measured, the optimized model demonstrated significant improvements in simulating GPP, across whole growing season and different growth stages and at both half-hourly and daily scales. The model's ability to capture diurnal and seasonal variations was markedly enhanced, with improved performance during critical growth phases such as the early, mid, and late growth periods. Although the model optimized by CSIF was not as good as the results from the site-observed data optimization, the results are still promising, suggesting that satellite-derived data can serve as a valuable substitute for ground-based data in regions where in-situ measurements are unavailable. Further, I performed an analysis of the SIF-GPP relationship on half-hourly scale, daily scale, and weekly scale and found that the larger the time scale, the stronger the linear relationship of SIF-GPP.

Despite the improvements achieved through data assimilation, challenges still persist in fully capturing the variability of SIF and GPP in simulations, particularly

under changing environmental conditions, such as cloudy weather. Additionally, while the model performed well in capturing general trends, some biases were introduced, notably in the overestimation of SIF in certain growth stages. Future studies should aim to refine the parameterization of the model, incorporate higher-resolution satellite data, and further explore the integration of local and global datasets to improve model robustness and applicability on larger spatial scal

VI. References

Ackerly, D. D. (1992). Light, leaf age, and leaf nitrogen concentration in a tropical vine. *Oecologia*, *89*(4), 596–600. https://doi.org/10.1007/BF00317169

Aubinet, M., Grelle, A., Ibrom, A., Rannik, Ü., Moncrieff, J., Foken, T., Kowalski, A. S., Martin, P. H., Berbigier, P., Bernhofer, Ch., Clement, R., Elbers, J., Granier, A., Grünwald, T., Morgenstern, K., Pilegaard, K., Rebmann, C., Snijders, W., Valentini, R., & Vesala, T. (1999). Estimates of the Annual Net Carbon and Water Exchange of Forests: The EUROFLUX Methodology. In A. H. Fitter & D. G. Raffaelli (Eds.), *Advances in Ecological Research* (Vol. 30, pp. 113–175). Academic Press. https://doi.org/10.1016/S0065-2504(08)60018-5

Bacour, C., Maignan, F., MacBean, N., Porcar-Castell, A., Flexas, J., Frankenberg, C., Peylin, P., Chevallier, F., Vuichard, N., & Bastrikov, V. (2019). Improving Estimates of Gross Primary Productivity by Assimilating Solar-Induced Fluorescence Satellite Retrievals in a Terrestrial Biosphere Model Using a Process-Based SIF Model. *Journal of Geophysical Research: Biogeosciences*, *124*(11), 3281–3306. https://doi.org/10.1029/2019JG005040

Baker, N. R. (2008). Chlorophyll Fluorescence: A Probe of Photosynthesis In Vivo. *Annual Review of Plant Biology*, *59*(1), 89–113. https://doi.org/10.1146/annurev.arplant.59.032607.092759

Baldocchi, D. D. (2003). Assessing the eddy covariance technique for evaluating carbon

dioxide exchange rates of ecosystems: Past, present and future. *Global Change Biology*, *9*(4), 479–492. https://doi.org/10.1046/j.1365-2486.2003.00629.x

Bao, S., Ibrom, A., Wohlfahrt, G., Koirala, S., Migliavacca, M., Zhang, Q., & Carvalhais, N. (2022). Narrow but robust advantages in two-big-leaf light use efficiency models over big-leaf light use efficiency models at ecosystem level. *Agricultural and Forest Meteorology*, *326*, 109185. https://doi.org/10.1016/j.agrformet.2022.109185

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., Lawrence, D. M., & Swenson, S. C. (2011). Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data. *Journal of Geophysical Research: Biogeosciences*, *116*(G2). https://doi.org/10.1029/2010JG001593

Buareal, K., Kato, T., Morozumi, T., Ono, K., & Nakashima, N. (2023). Red solar-induced chlorophyll fluorescence as a robust proxy for ecosystem-level photosynthesis in a rice field. *Agricultural and Forest Meteorology*, *336*, 109473. https://doi.org/10.1016/j.agrformet.2023.109473

Carlsson, L. S., Samuelsson, P. B., & Jönsson, P. G. (2020). Interpretable Machine Learning—Tools to Interpret the Predictions of a Machine Learning Model Predicting the Electrical Energy Consumption of an Electric Arc Furnace. *Steel*

Research International, *91*(11), 2000053. https://doi.org/10.1002/srin.202000053

Chen, B., Black, T. A., Coops, N. C., Hilker, T., (Tony) Trofymow, J. A., & Morgenstern, K. (2009). Assessing Tower Flux Footprint Climatology and Scaling Between Remotely Sensed and Eddy Covariance Measurements. *Boundary-Layer Meteorology*, *130*(2), 137–167. https://doi.org/10.1007/s10546-008-9339-1

Cogliati, S., Verhoef, W., Kraft, S., Sabater, N., Alonso, L., Vicent, J., Moreno, J., Drusch, M., & Colombo, R. (2015). Retrieval of sun-induced fluorescence using advanced spectral fitting methods. *Remote Sensing of Environment*, *169*, 344–357. https://doi.org/10.1016/j.rse.2015.08.022

Damm, A., Elbers, J., Erler, A., Gioli, B., Hamdi, K., Hutjes, R., Kosvancova, M., Meroni, M., Miglietta, F., Moersch, A., Moreno, J., Schickling, A., Sonnenschein, R., Udelhoven, T., Van Der LINDEN, S., Hostert, P., & Rascher, U. (2010). Remote sensing of sun-induced fluorescence to improve modeling of diurnal courses of gross primary production (GPP). *Global Change Biology*, *16*(1), 171–186. https://doi.org/10.1111/j.1365-2486.2009.01908.x

Damm, A., Guanter, L., Laurent, V. C. E., Schaepman, M. E., Schickling, A., & Rascher, U. (2014). FLD-based retrieval of sun-induced chlorophyll fluorescence from medium spectral resolution airborne spectroscopy data. *Remote Sensing of Environment*, *147*, 256–266. https://doi.org/10.1016/j.rse.2014.03.009

Damm, A., Guanter, L., Paul-Limoges, E., van der Tol, C., Hueni, A., Buchmann, N., Eugster, W., Ammann, C., & Schaepman, M. E. (2015). Far-red sun-induced chlorophyll fluorescence shows ecosystem-specific relationships to gross primary production: An assessment based on observational and modeling approaches. *Remote Sensing of Environment*, *166*, 91–105. https://doi.org/10.1016/j.rse.2015.06.004

Daumard, F., Champagne, S., Fournier, A., Goulas, Y., Ounis, A., Hanocq, J.-F., & Moya, I. (2010). A Field Platform for Continuous Measurement of Canopy Fluorescence. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(9), 3358–3368. IEEE Transactions on Geoscience and Remote Sensing. https://doi.org/10.1109/TGRS.2010.2046420

De Pury, D. G. G., & Farquhar, G. D. (1997). Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. *Plant, Cell & Environment*, *20*(5), 537–557. https://doi.org/10.1111/j.1365-3040.1997.00094.x

Drolet, G., Wade, T., Nichol, C. J., MacLellan, C., Levula, J., Porcar-Castell, A., Nikinmaa, E., & Vesala, T. (2014). A temperature-controlled spectrometer system for continuous and unattended measurements of canopy spectral radiance and reflectance. *International Journal of Remote Sensing*, *35*(5), 1769–1785. https://doi.org/10.1080/01431161.2014.882035

Du, S., Liu, L., Liu, X., & Hu, J. (2017). Response of Canopy Solar-Induced Chlorophyll Fluorescence to the Absorbed Photosynthetically Active Radiation Absorbed by Chlorophyll. *Remote Sensing*, *9*(9), Article 9. https://doi.org/10.3390/rs9090911

Duveiller, G., & Cescatti, A. (2016). Spatially downscaling sun-induced chlorophyll fluorescence leads to an improved temporal correlation with gross primary productivity. *Remote Sensing of Environment*, *182*, 72–89. https://doi.org/10.1016/j.rse.2016.04.027

Farquhar, G. D., von Caemmerer, S., & Berry, J. A. (1980). A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. *Planta*, *149*(1), 78–90. https://doi.org/10.1007/BF00386231

Flexas, J., Escalona, J. M., Evain, S., Gulías, J., Moya, I., Osmond, C. B., & Medrano, H. (2002). Steady-state chlorophyll fluorescence (Fs) measurements as a tool to follow variations of net CO2 assimilation and stomatal conductance during water-stress in C3 plants. *Physiologia Plantarum*, *114*(2), 231–240. https://doi.org/10.1034/j.1399-3054.2002.1140209.x

Frankenberg, C., & Berry, J. (2018). Solar Induced Chlorophyll Fluorescence: Origins, Relation to Photosynthesis and Retrieval. In S. Liang (Ed.), *Comprehensive Remote Sensing* (pp. 143–162). Elsevier. https://doi.org/10.1016/B978-0-12-

409548-9.10632-3

Frankenberg, C., Fisher, J. B., Worden, J., Badgley, G., Saatchi, S. S., Lee, J.-E., Toon, G. C., Butz, A., Jung, M., Kuze, A., & Yokota, T. (2011). New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophysical Research Letters*, *38*(17). https://doi.org/10.1029/2011GL048738

Frankenberg, C., O'Dell, C., Berry, J., Guanter, L., Joiner, J., Köhler, P., Pollock, R., & Taylor, T. E. (2014). Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2. *Remote Sensing of Environment*, *147*, 1–12. https://doi.org/10.1016/j.rse.2014.02.007

Frankenberg, C., O'Dell, C., Guanter, L., & McDuffie, J. (2012). Remote sensing of near-infrared chlorophyll fluorescence from space in scattering atmospheres: Implications for its retrieval and interferences with atmospheric CO_2 retrievals. *Atmospheric Measurement Techniques*, *5*(8), 2081–2094. https://doi.org/10.5194/amt-5-2081-2012

Garbulsky, M. F., Filella, I., Verger, A., & Peñuelas, J. (2014). Photosynthetic light use efficiency from satellite sensors: From global to Mediterranean vegetation. *Environmental and Experimental Botany*, *103*, 3–11. https://doi.org/10.1016/j.envexpbot.2013.10.009

Gonzalez, J., Dai, Z., Hennig, P., & Lawrence, N. (2016). Batch Bayesian Optimization via Local Penalization. *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, 648–657. https://proceedings.mlr.press/v51/gonzalez16a.html

González, J., Longworth, J., James, D. C., & Lawrence, N. D. (2015). *Bayesian Optimization for Synthetic Gene Design* (arXiv:1505.01627). arXiv. http://arxiv.org/abs/1505.01627

Gu, L., Han, J., Wood, J. D., Chang, C. Y.-Y., & Sun, Y. (2019). Sun-induced Chl fluorescence and its importance for biophysical modeling of photosynthesis based on light reactions. *New Phytologist*, *223*(3), 1179–1191. https://doi.org/10.1111/nph.15796

Guanter, L., Bacour, C., Schneider, A., Aben, I., van Kempen, T. A., Maignan, F., Retscher, C., Köhler, P., Frankenberg, C., Joiner, J., & Zhang, Y. (2021). The TROPOSIF global sun-induced fluorescence dataset from the Sentinel-5P TROPOMI mission. *Earth System Science Data*, *13*(11), 5423–5440. https://doi.org/10.5194/essd-13-5423-2021

Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J. A., Frankenberg, C., Huete, A. R., Zarco-Tejada, P., Lee, J.-E., Moran, M. S., Ponce-Campos, G., Beer, C., Camps-Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker,

J. M., & Griffis, T. J. (2014a). Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. *Proceedings of the National Academy of Sciences*, *111*(14), E1327–E1333. https://doi.org/10.1073/pnas.1320008111

Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J. A., Frankenberg, C., Huete, A. R., Zarco-Tejada, P., Lee, J.-E., Moran, M. S., Ponce-Campos, G., Beer, C., Camps-Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J. M., & Griffis, T. J. (2014b). Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. *Proceedings of the National Academy of Sciences*, *111*(14), E1327–E1333. https://doi.org/10.1073/pnas.1320008111

Huang, X., Xiao, J., Wang, X., & Ma, M. (2021). Improving the global MODIS GPP model by optimizing parameters with FLUXNET data. *Agricultural and Forest Meteorology*, *300*, 108314. https://doi.org/10.1016/j.agrformet.2020.108314

Huete, A. R., Didan, K., Shimabukuro, Y. E., Ratana, P., Saleska, S. R., Hutyra, L. R., Yang, W., Nemani, R. R., & Myneni, R. (2006). Amazon rainforests green-up with sunlight in dry season. *Geophysical Research Letters*, *33*(6). https://doi.org/10.1029/2005GL025583

Ichii, K., Kondo, M., Lee, Y.-H., Wang, S.-Q., Kim, J., Ueyama, M., Lim, H.-J., Shi,

H., Suzuki, T., Ito, A., Kwon, H., Ju, W., Huang, M., Sasai, T., Asanuma, J., Han, S., Hirano, T., Hirata, R., Kato, T., … Zhao, F.-H. (2013). Site-level model–data synthesis of terrestrial carbon fluxes in the CarboEastAsia eddy-covariance observation network: Toward future modeling efforts. *Journal of Forest Research*, *18*(1), 13–20. https://doi.org/10.1007/s10310-012-0367-9

Ito, A. (2008). The regional carbon budget of East Asia simulated with a terrestrial ecosystem model and validated using AsiaFlux data. *Agricultural and Forest Meteorology*, *148*(5), 738–747. https://doi.org/10.1016/j.agrformet.2007.12.007

Ito, A. (2010). Changing ecophysiological processes and carbon budget in East Asian ecosystems under near-future changes in climate: Implications for long-term monitoring from a process-based model. *Journal of Plant Research*, *123*(4), 577–588. https://doi.org/10.1007/s10265-009-0305-x

Ito, A. (2019). Disequilibrium of terrestrial ecosystem CO_2 budget caused by disturbance-induced emissions and non-CO_2 carbon export flows: A global model assessment. *Earth System Dynamics*, *10*(4), 685–709. https://doi.org/10.5194/esd-10-685-2019

Ito, A., & Inatomi, M. (2012). Use of a process-based model for assessing the methane budgets of global terrestrial ecosystems and evaluation of uncertainty. *Biogeosciences*, *9*(2), 759–773. https://doi.org/10.5194/bg-9-759-2012

Ito, A., Inatomi, M., Mo, W., Lee, M., Koizumi, H., Saigusa, N., Murayama, S., & Yamamoto, S. (2007). Examination of model-estimated ecosystem respiration using flux measurements from a cool-temperate deciduous broad-leaved forest in central Japan. *Tellus B*, *59*(3), 616–624. https://doi.org/10.1111/j.1600-0889.2007.00258.x

Ito, A., Muraoka, H., Koizumi, H., Saigusa, N., Murayama, S., & Yamamoto, S. (2006). Seasonal variation in leaf properties and ecosystem carbon budget in a cool-temperate deciduous broad-leaved forest: Simulation analysis at Takayama site, Japan. *Ecological Research*, *21*(1), 137–149. https://doi.org/10.1007/s11284-005-0100-7

Ito, A., Nishina, K., Reyer, C. P. O., François, L., Henrot, A.-J., Munhoven, G., Jacquemin, I., Tian, H., Yang, J., Pan, S., Morfopoulos, C., Betts, R., Hickler, T., Steinkamp, J., Ostberg, S., Schaphoff, S., Ciais, P., Chang, J., Rafique, R., … Zhao, F. (2017). Photosynthetic productivity and its efficiencies in ISIMIP2a biome models: Benchmarking for impact assessment studies. *Environmental Research Letters*, *12*(8), 085001. https://doi.org/10.1088/1748-9326/aa7a19

Ito, A., & Oikawa, T. (2002). A simulation model of the carbon cycle in land ecosystems (Sim-CYCLE): A description based on dry-matter production theory and plot-scale validation. *Ecological Modelling*, *151*(2), 143–176.

https://doi.org/10.1016/S0304-3800(01)00473-2

Ito, A., Saigusa, N., Murayama, S., & Yamamoto, S. (2005). Modeling of gross and net carbon dioxide exchange over a cool-temperate deciduous broad-leaved forest in Japan: Analysis of seasonal and interannual change. *Agricultural and Forest Meteorology*, *134*(1), 122–134. https://doi.org/10.1016/j.agrformet.2005.11.002

Jenkins, J. P., Richardson, A. D., Braswell, B. H., Ollinger, S. V., Hollinger, D. Y., & Smith, M.-L. (2007). Refining light-use efficiency calculations for a deciduous forest canopy using simultaneous tower-based carbon flux and radiometric measurements. *Agricultural and Forest Meteorology*, *143*(1), 64–79. https://doi.org/10.1016/j.agrformet.2006.11.008

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A. P., Middleton, E. M., Huemmrich, K. F., Yoshida, Y., & Frankenberg, C. (2013). Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: Methodology, simulations, and application to GOME-2. *Atmospheric Measurement Techniques*, *6*(10), 2803–2823. https://doi.org/10.5194/amt-6-2803-2013

Joiner, J., Yoshida, Y., Guanter, L., Lamsal, L., Li, C., Fasnacht, Z., Köhler, P., Frankenberg, C., Sun, Y., & Parazoo, N. (2024). Noise Reduction for Solar-Induced Fluorescence Retrievals Using Machine Learning and Principal

Component Analysis: Simulations and Applications to GOME-2 Satellite Retrievals. *Artificial Intelligence for the Earth Systems*, *3*(3). https://doi.org/10.1175/AIES-D-23-0085.1

Joiner, J., Yoshida, Y., Vasilkov, A. P., Yoshida, Y., Corp, L. A., & Middleton, E. M. (2011). First observations of global and seasonal terrestrial chlorophyll fluorescence from space. *Biogeosciences*, *8*(3), 637–651. https://doi.org/10.5194/bg-8-637-2011

Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, *13*(4), 455–492. https://doi.org/10.1023/A:1008306431147

Kattge, J., & Knorr, W. (2007). Temperature acclimation in a biochemical model of photosynthesis: A reanalysis of data from 36 species. *Plant, Cell & Environment*, *30*(9), 1176–1190. https://doi.org/10.1111/j.1365-3040.2007.01690.x

Kattge, J., Knorr, W., Raddatz, T., & Wirth, C. (2009). Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models. *Global Change Biology*, *15*(4), 976–991. https://doi.org/10.1111/j.1365-2486.2008.01744.x

Knyazikhin, Y., Martonchik, J. V., Myneni, R. B., Diner, D. J., & Running, S. W. (1998).

Synergistic algorithm for estimating vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from MODIS and MISR data. *Journal of Geophysical Research: Atmospheres*, *103*(D24), 32257–32275. https://doi.org/10.1029/98JD02462

Koffi, E. N., Rayner, P. J., Norton, A. J., Frankenberg, C., & Scholze, M. (2015). Investigating the usefulness of satellite-derived fluorescence data in inferring gross primary productivity within the carbon cycle data assimilation system. *Biogeosciences*, *12*(13), 4067–4084. https://doi.org/10.5194/bg-12-4067-2015

Köhler, P., Frankenberg, C., Magney, T. S., Guanter, L., Joiner, J., & Landgraf, J. (2018). Global Retrievals of Solar-Induced Chlorophyll Fluorescence With TROPOMI: First Results and Intersensor Comparison to OCO-2. *Geophysical Research Letters*, *45*(19), 10,456-10,463. https://doi.org/10.1029/2018GL079031

Lee, J.-E., Berry, J. A., van der Tol, C., Yang, X., Guanter, L., Damm, A., Baker, I., & Frankenberg, C. (2015). Simulations of chlorophyll fluorescence incorporated into the Community Land Model version 4. *Global Change Biology*, *21*(9), 3469–3477. https://doi.org/10.1111/gcb.12948

Li, J., Zhang, Y., Gu, L., Li, Z., Li, J., Zhang, Q., Zhang, Z., & Song, L. (2020). Seasonal variations in the relationship between sun-induced chlorophyll fluorescence and photosynthetic capacity from the leaf to canopy level in a rice crop. *Journal of*

Experimental Botany, *71*(22), 7179–7197. https://doi.org/10.1093/jxb/eraa408

Li, R., Lombardozzi, D., Shi, M., Frankenberg, C., Parazoo, N. C., Köhler, P., Yi, K., Guan, K., & Yang, X. (2022). Representation of Leaf-to-Canopy Radiative Transfer Processes Improves Simulation of Far-Red Solar-Induced Chlorophyll Fluorescence in the Community Land Model Version 5. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002747. https://doi.org/10.1029/2021MS002747

Li, X., Xiao, J., He, B., Altaf Arain, M., Beringer, J., Desai, A. R., Emmel, C., Hollinger, D. Y., Krasnova, A., Mammarella, I., Noe, S. M., Ortiz, P. S., Rey-Sanchez, A. C., Rocha, A. V., & Varlagin, A. (2018). Solar-induced chlorophyll fluorescence is strongly correlated with terrestrial photosynthesis for a wide variety of biomes: First global analysis based on OCO-2 and flux tower observations. *Global Change Biology*, *24*(9), 3990–4008. https://doi.org/10.1111/gcb.14297

Li, Z., Zhang, Q., Li, J., Yang, X., Wu, Y., Zhang, Z., Wang, S., Wang, H., & Zhang, Y. (2020). Solar-induced chlorophyll fluorescence and its link to canopy photosynthesis in maize from continuous ground measurements. *Remote Sensing of Environment*, *236*, 111420. https://doi.org/10.1016/j.rse.2019.111420

Liu, X., Guanter, L., Liu, L., Damm, A., Malenovský, Z., Rascher, U., Peng, D., Du, S., & Gastellu-Etchegorry, J.-P. (2019). Downscaling of solar-induced chlorophyll

fluorescence from canopy level to photosystem level using a random forest model. *Remote Sensing of Environment*, *231*, 110772. https://doi.org/10.1016/j.rse.2018.05.035

Liu, X., Liu, Z., Liu, L., Lu, X., Chen, J., Du, S., & Zou, C. (2021). Modelling the influence of incident radiation on the SIF-based GPP estimation for maize. *Agricultural and Forest Meteorology*, *307*, 108522. https://doi.org/10.1016/j.agrformet.2021.108522

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, *30*. https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html

Meroni, M., Rossini, M., Guanter, L., Alonso, L., Rascher, U., Colombo, R., & Moreno, J. (2009). Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications. *Remote Sensing of Environment*, *113*(10), 2037–2051. https://doi.org/10.1016/j.rse.2009.05.003

Miao, G., Guan, K., Yang, X., Bernacchi, C. J., Berry, J. A., DeLucia, E. H., Wu, J., Moore, C. E., Meacham, K., Cai, Y., Peng, B., Kimm, H., & Masters, M. D. (2018). Sun-Induced Chlorophyll Fluorescence, Photosynthesis, and Light Use Efficiency of a Soybean Field from Seasonally Continuous Measurements. *Journal of Geophysical Research: Biogeosciences*, *123*(2), 610–623. https://doi.org/10.1002/2017JG004180

Migliavacca, M., Perez-Priego, O., Rossini, M., El-Madany, T. S., Moreno, G., van der Tol, C., Rascher, U., Berninger, A., Bessenbacher, V., Burkart, A., Carrara, A., Fava, F., Guan, J.-H., Hammer, T. W., Henkel, K., Juarez-Alcalde, E., Julitta, T., Kolle, O., Martín, M. P., … Reichstein, M. (2017). Plant functional traits and canopy structure control the relationship between photosynthetic CO2 uptake and far-red sun-induced fluorescence in a Mediterranean grassland under different nutrient availability. *New Phytologist*, *214*(3), 1078–1091. https://doi.org/10.1111/nph.14437

Miyauchi, T., Saito, M., Noda, H. M., Ito, A., Kato, T., & Matsunaga, T. (2024). Process-based Modeling of Solar-induced Chlorophyll Fluorescence with VISIT-SIF version 1.0. *EGUsphere*, 1–27. https://doi.org/10.5194/egusphere-2024-1542

Monteith, J. L. (1972). Solar Radiation and Productivity in Tropical Ecosystems. *Journal of Applied Ecology*, *9*(3), 747–766. https://doi.org/10.2307/2401901

Morozumi, T., Kato, T., Kobayashi, H., Sakai, Y., Nakashima, N., Buareal, K., Nasahara, K. N., Akitsu, T. K., Murayama, S., Noda, H. M., & Muraoka, H. (2023). Contributions of the understory and midstory to total canopy solar-induced chlorophyll fluorescence in a ground-based study in conjunction with seasonal gross primary productivity in a cool-temperate deciduous broadleaf forest. *Remote Sensing of Environment*, *284*, 113340. https://doi.org/10.1016/j.rse.2022.113340

Nagai, S., Akitsu, T., Saitoh, T. M., Busey, R. C., Fukuzawa, K., Honda, Y., Ichie, T., Ide, R., Ikawa, H., Iwasaki, A., Iwao, K., Kajiwara, K., Kang, S., Kim, Y., Khoon, K. L., Kononov, A. V., Kosugi, Y., Maeda, T., Mamiya, W., … Nasahara, K. N. (2018). 8 million phenological and sky images from 29 ecosystems from the Arctic to the tropics: The Phenological Eyes Network. *Ecological Research*, *33*(6), 1091–1092. https://doi.org/10.1007/s11284-018-1633-x

Nasahara, K. N., & Nagai, S. (2015). Review: Development of an in situ observation network for terrestrial ecological remote sensing: the Phenological Eyes Network (PEN). *Ecological Research*, *30*(2), 211–223. https://doi.org/10.1007/s11284-014-1239-x

Norton, A. J., Rayner, P. J., Koffi, E. N., Scholze, M., Silver, J. D., & Wang, Y.-P. (2019). Estimating global gross primary productivity using chlorophyll fluorescence and a data assimilation system with the BETHY-SCOPE model. *Biogeosciences*,

16(15), 3069–3093. https://doi.org/10.5194/bg-16-3069-2019

Ono, K., Mano, M., Han, G. H., Nagai, H., Yamada, T., Kobayashi, Y., Miyata, A., Inoue, Y., & Lal, R. (2015). Environmental controls on fallow carbon dioxide flux in a single-crop rice paddy, Japan. *Land Degradation and Development*, *26*(4), 331–339. https://doi.org/10.1002/ldr.2211

Ono, K., & Maruyama, A. (2015). Development of an onsite computation scheme of eddy-covariance fluxes. 農業気象, *71*(4), 318–329. https://doi.org/10.2480/agrmet.D-15-00003

Ono, K., Maruyama, A., Kuwagata, T., Mano, M., Takimoto, T., Hayashi, K., Hasegawa, T., & Miyata, A. (2013). Canopy-scale relationships between stomatal conductance and photosynthesis in irrigated rice. *Global Change Biology*, *19*(7), 2209–2220. https://doi.org/10.1111/gcb.12188

Ono, K., Miyata, A., & Yamada, T. (2008). Apparent downward CO2 flux observed with open-path eddy covariance over a non-vegetated surface. *Theoretical and Applied Climatology*, *92*(3), 195–208. https://doi.org/10.1007/s00704-007-0323-3

Parazoo, N. C., Bowman, K., Fisher, J. B., Frankenberg, C., Jones, D. B. A., Cescatti, A., Pérez-Priego, Ó., Wohlfahrt, G., & Montagnani, L. (2014). Terrestrial gross primary production inferred from satellite fluorescence and vegetation models.

Global Change Biology, *20*(10), 3103–3121. https://doi.org/10.1111/gcb.12652

Parazoo, N. C., Magney, T., Norton, A., Raczka, B., Bacour, C., Maignan, F., Baker, I., Zhang, Y., Qiu, B., Shi, M., MacBean, N., Bowling, D. R., Burns, S. P., Blanken, P. D., Stutz, J., Grossmann, K., & Frankenberg, C. (2020). Wide discrepancies in the magnitude and direction of modeled solar-induced chlorophyll fluorescence in response to light conditions. *Biogeosciences*, *17*(13), 3733–3755. https://doi.org/10.5194/bg-17-3733-2020

Porcar-Castell, A., Tyystjärvi, E., Atherton, J., van der Tol, C., Flexas, J., Pfündel, E. E., Moreno, J., Frankenberg, C., & Berry, J. A. (2014). Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: Mechanisms and challenges. *Journal of Experimental Botany*, *65*(15), 4065–4095. https://doi.org/10.1093/jxb/eru191

Qiu, B., Xue, Y., Fisher, J. B., Guo, W., Berry, J. A., & Zhang, Y. (2018). Satellite Chlorophyll Fluorescence and Soil Moisture Observations Lead to Advances in the Predictive Understanding of Global Terrestrial Coupled Carbon-Water Cycles. *Global Biogeochemical Cycles*, *32*(3), 360–375. https://doi.org/10.1002/2017GB005744

Rascher, U., Agati, G., Alonso, L., Cecchi, G., Champagne, S., Colombo, R., Damm, A., Daumard, F., de Miguel, E., Fernandez, G., Franch, B., Franke, J., Gerbig, C.,

Gioli, B., Gómez, J. A., Goulas, Y., Guanter, L., Gutiérrez-de-la-Cámara, Ó., Hamdi, K., … Zaldei, A. (2009). CEFLES2: The remote sensing component to quantify photosynthetic efficiency from the leaf to the region by measuring sun-induced fluorescence in the oxygen absorption bands. *Biogeosciences*, *6*(7), 1181–1198. https://doi.org/10.5194/bg-6-1181-2009

Rascher, U., Alonso, L., Burkart, A., Cilia, C., Cogliati, S., Colombo, R., Damm, A., Drusch, M., Guanter, L., Hanus, J., Hyvärinen, T., Julitta, T., Jussila, J., Kataja, K., Kokkalis, P., Kraft, S., Kraska, T., Matveeva, M., Moreno, J., … Zemek, F. (2015). Sun-induced fluorescence – a new probe of photosynthesis: First maps from the imaging spectrometer HyPlant. *Global Change Biology*, *21*(12), 4673–4684. https://doi.org/10.1111/gcb.13017

Rayner, P. J., Scholze, M., Knorr, W., Kaminski, T., Giering, R., & Widmann, H. (2005). Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS). *Global Biogeochemical Cycles*, *19*(2). https://doi.org/10.1029/2004GB002254

Rogers, A. (2014). The use and misuse of Vc,max in Earth System Models. *Photosynthesis Research*, *119*(1), 15–29. https://doi.org/10.1007/s11120-013-9818-1

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., & Hashimoto, H.

(2004). A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. *BioScience*, *54*(6), 547–560. https://doi.org/10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2

Sargsyan, K., Safta, C., Najm, H. N., Debusschere, B. J., Ricciuto, D., & Thornton, P. (2014). DIMENSIONALITY REDUCTION FOR COMPLEX MODELS VIA BAYESIAN COMPRESSIVE SENSING. *International Journal for Uncertainty Quantification*, *4*(1). https://doi.org/10.1615/Int.J.UncertaintyQuantification.2013006821

Scholze, M., Kaminski, T., Rayner, P., Knorr, W., & Giering, R. (2007). Propagating uncertainty through prognostic carbon cycle data assimilation system simulations. *Journal of Geophysical Research: Atmospheres*, *112*(D17). https://doi.org/10.1029/2007JD008642

Sun, Y., Frankenberg, C., Jung, M., Joiner, J., Guanter, L., Köhler, P., & Magney, T. (2018). Overview of Solar-Induced chlorophyll Fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, cross-mission comparison, and global monitoring for GPP. *Remote Sensing of Environment*, *209*, 808–823. https://doi.org/10.1016/j.rse.2018.02.016

Terashima, I., & Hikosaka, K. (1995). Comparative ecophysiology of leaf and canopy photosynthesis. *Plant, Cell & Environment*, *18*(10), 1111–1128.

https://doi.org/10.1111/j.1365-3040.1995.tb00623.x

Thum, T., Zaehle, S., Köhler, P., Aalto, T., Aurela, M., Guanter, L., Kolari, P., Laurila, T., Lohila, A., Magnani, F., Van Der Tol, C., & Markkanen, T. (2017). Modelling sun-induced fluorescence and photosynthesis with a land surface model at local and regional scales in northern Europe. *Biogeosciences*, *14*(7), 1969–1987. https://doi.org/10.5194/bg-14-1969-2017

van der Tol, C., Berry, J. A., Campbell, P. K. E., & Rascher, U. (2014). Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence. *Journal of Geophysical Research: Biogeosciences*, *119*(12), 2312–2327. https://doi.org/10.1002/2014JG002713

Verma, M., Schimel, D., Evans, B., Frankenberg, C., Beringer, J., Drewry, D. T., Magney, T., Marang, I., Hutley, L., Moore, C., & Eldering, A. (2017a). Effect of environmental conditions on the relationship between solar-induced fluorescence and gross primary productivity at an OzFlux grassland site. *Journal of Geophysical Research: Biogeosciences*, *122*(3), 716–733. https://doi.org/10.1002/2016JG003580

Verma, M., Schimel, D., Evans, B., Frankenberg, C., Beringer, J., Drewry, D. T., Magney, T., Marang, I., Hutley, L., Moore, C., & Eldering, A. (2017b). Effect of environmental conditions on the relationship between solar-induced fluorescence

and gross primary productivity at an OzFlux grassland site. *Journal of Geophysical Research: Biogeosciences*, *122*(3), 716–733. https://doi.org/10.1002/2016JG003580

Walker, A. P., Quaife, T., van Bodegom, P. M., De Kauwe, M. G., Keenan, T. F., Joiner, J., Lomas, M. R., MacBean, N., Xu, C., Yang, X., & Woodward, F. I. (2017). The impact of alternative trait-scaling hypotheses for the maximum photosynthetic carboxylation rate (Vcmax) on global gross primary production. *New Phytologist*, *215*(4), 1370–1386. https://doi.org/10.1111/nph.14623

Wang, H., Yan, S., Ciais, P., Wigneron, J.-P., Liu, L., Li, Y., Fu, Z., Ma, H., Liang, Z., Wei, F., Wang, Y., & Li, S. (2022). Exploring complex water stress–gross primary production relationships: Impact of climatic drivers, main effects, and interactive effects. *Global Change Biology*, *28*(13), 4110–4123. https://doi.org/10.1111/gcb.16201

Wood, J. D., Griffis, T. J., Baker, J. M., Frankenberg, C., Verma, M., & Yuen, K. (2017). Multiscale analyses of solar-induced florescence and gross primary production. *Geophysical Research Letters*, *44*(1), 533–541. https://doi.org/10.1002/2016GL070775

Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J.,

Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., … Villar, R. (2004). The worldwide leaf economics spectrum. *Nature*, *428*(6985), 821–827. https://doi.org/10.1038/nature02403

Wu, M., Scholze, M., Kaminski, T., Voßbeck, M., & Tagesson, T. (2020). Using SMOS soil moisture data combining CO2 flask samples to constrain carbon fluxes during 2010–2015 within a Carbon Cycle Data Assimilation System (CCDAS). *Remote Sensing of Environment*, *240*, 111719. https://doi.org/10.1016/j.rse.2020.111719

Yang, K., Ryu, Y., Dechant, B., Berry, J. A., Hwang, Y., Jiang, C., Kang, M., Kim, J., Kimm, H., Kornfeld, A., & Yang, X. (2018). Sun-induced chlorophyll fluorescence is more strongly related to absorbed light than to photosynthesis at half-hourly resolution in a rice paddy. *Remote Sensing of Environment*, *216*, 658–673. https://doi.org/10.1016/j.rse.2018.07.008

Yang, P., van der Tol, C., Verhoef, W., Damm, A., Schickling, A., Kraska, T., Muller, O., & Rascher, U. (2019). Using reflectance to explain vegetation biochemical and structural effects on sun-induced chlorophyll fluorescence. *Remote Sensing of Environment*, *231*, 110996. https://doi.org/10.1016/j.rse.2018.11.039

Yang, X., Tang, J., Mustard, J. F., Lee, J.-E., Rossini, M., Joiner, J., Munger, J. W., Kornfeld, A., & Richardson, A. D. (2015). Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a

temperate deciduous forest. *Geophysical Research Letters*, *42*(8), 2977–2987. https://doi.org/10.1002/2015GL063201

Zarco-Tejada, P. J., Berni, J. A. J., Suárez, L., Sepulcre-Cantó, G., Morales, F., & Miller, J. R. (2009). Imaging chlorophyll fluorescence with an airborne narrow-band multispectral camera for vegetation stress detection. *Remote Sensing of Environment*, *113*(6), 1262–1275. https://doi.org/10.1016/j.rse.2009.02.016

Zhang, Y., Guanter, L., Berry, J. A., van der Tol, C., Yang, X., Tang, J., & Zhang, F. (2016). Model-based analysis of the relationship between sun-induced chlorophyll fluorescence and gross primary production for remote sensing applications. *Remote Sensing of Environment*, *187*, 145–155. https://doi.org/10.1016/j.rse.2016.10.016

Zhang, Y., Joiner, J., Alemohammad, S. H., Zhou, S., & Gentine, P. (2018). A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks. *Biogeosciences*, *15*(19), 5779–5800. https://doi.org/10.5194/bg-15-5779-2018

Zhang, Y., Xiao, X., Zhang, Y., Wolf, S., Zhou, S., Joiner, J., Guanter, L., Verma, M., Sun, Y., Yang, X., Paul-Limoges, E., Gough, C. M., Wohlfahrt, G., Gioli, B., van der Tol, C., Yann, N., Lund, M., & de Grandcourt, A. (2018). On the relationship between sub-daily instantaneous and daily total gross primary production:

Implications for interpreting satellite-based SIF retrievals. *Remote Sensing of Environment*, *205*, 276–289. https://doi.org/10.1016/j.rse.2017.12.009

Zhang, Z., Zhang, Y., Joiner, J., & Migliavacca, M. (2018). Angle matters: Bidirectional effects impact the slope of relationship between gross primary productivity and sun-induced chlorophyll fluorescence from Orbiting Carbon Observatory-2 across biomes. *Global Change Biology*, *24*(11), 5017–5020. https://doi.org/10.1111/gcb.14427

Zhou, Y., Wu, X., Ju, W., Chen, J. M., Wang, S., Wang, H., Yuan, W., Andrew Black, T., Jassal, R., Ibrom, A., Han, S., Yan, J., Margolis, H., Roupsard, O., Li, Y., Zhao, F., Kiely, G., Starr, G., Pavelka, M., … Varlagin, A. (2016). Global parameterization and validation of a two-leaf light use efficiency model for predicting gross primary production across FLUXNET sites. *Journal of Geophysical Research: Biogeosciences*, *121*(4), 1045–1072. https://doi.org/10.1002/2014JG002876

VII. Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Kato Tomomichi. His patience, kindness, and unwavering support have been a constant source of inspiration throughout my doctoral journey. He has created an open and respectful environment where I could freely express my ideas, even disagreeing with him when necessary, without fear of judgment. His ability to listen, encourage, and guide without imposing has allowed me to grow both as a researcher and as an individual. For this, I am profoundly thankful.

I am also deeply grateful to the members of our laboratory for their collaboration, insightful discussions, and continuous encouragement. Your support and camaraderie have made my doctoral studies not only productive but also truly enjoyable.

Lastly, I would like to extend my heartfelt appreciation to my family and friends for their unconditional love and encouragement, which have been my pillar of strength throughout this journey.

Thank you all for making this experience meaningful and fulfilling.