

## Simulations of solar-induced chlorophyll fluorescence over crop canopies using the integrated APSIM model

Yakai Wang<sup>a</sup>, Zhunqiao Liu<sup>b</sup>, Qiang Yu<sup>b</sup>, Liangyun Liu<sup>c</sup>, Xinjie Liu<sup>c</sup>, Linchao Li<sup>a</sup>, Qianlan Jia<sup>a</sup>, Chenhui Guo<sup>a</sup>, Xiaoliang Lu<sup>b,\*</sup>

<sup>a</sup> College of Natural Resources and Environment, Northwest A&F University, Yangling, Shaanxi 712100, China

<sup>b</sup> State Key Laboratory of Soil Erosion and Dry-land Farming on the Loess Plateau, Northwest A&F University, Yangling, Shaanxi 712100, China

<sup>c</sup> Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China

### ARTICLE INFO

#### Keywords:

Solar-induced chlorophyll fluorescence  
Crop model  
Photosynthesis  
Remote sensing

### ABSTRACT

Agricultural production models predict crop yield by accounting for a variety of species, cultivar, farming management, and environmental impacts on crop photosynthesis. Without suitable constraints, however, large uncertainties may exist in simulations of crop photosynthesis. Recent advances in retrieving solar-induced chlorophyll fluorescence (SIF) at the top-of-canopy (TOC) have provided a promising measurement for crop photosynthesis. Within the framework of the APSIM (Agricultural Production Systems sIMulator) model, a SIF module was developed to connect crop photosynthesis to TOC SIF emission ( $SIF_{toc}$ ) which can be measured by remote sensing platforms. The new model (APSIM-SIF) first estimates the leaf-level chlorophyll fluorescence emitted over the full SIF spectrum ( $SIF_{tot\_full}$ ) according to  $CO_2$  assimilation in crops. The model then mechanically decomposes the conversion from  $SIF_{tot\_full}$  to  $SIF_{toc}$  into two factors: the SIF band conversion factor ( $\epsilon$ ) and the fluorescence escape ratio ( $f_{esc}$ ) that represent the impact of leaf physiological status and plant structure properties, respectively.  $\epsilon$  can be estimated using leaf structural and biochemical parameters as inputs;  $f_{esc}$  for near-infrared SIF can be expressed as a function of directional reflectance in the near-infrared region ( $R_{NIR}$ ), Normalized Difference Vegetation Index (NDVI), and the fraction of PAR absorbed by crops ( $f_{APAR}$ ). The APSIM-SIF model determined more than 90% of the variation in gross primary productivity (GPP), aboveground biomass and leaf area index (LAI) measurements for maize (*Zea mays* L.) at two AmeriFlux sites in the U.S. Midwest and it also captured the seasonality of SIF ( $R^2 = 0.84$ ) and GPP ( $R^2 = 0.81$ ) well at an irrigated maize site in China. The APSIM-SIF model was also applied to the simulation of TOC SIF emission of maize and soybean (*Glycine max* L.) in the U.S. Midwest during the 2018 growing season. The simulated  $SIF_{toc}$  accounted for more than 75% of the variability of daily satellite SIF observations for grid squares with more than 70% crop area. The main contribution of this study lies in two aspects: (1) a physically-based framework is proposed to incorporate the SIF module to the APSIM-DCaPST model, and (2) the two important factors used in this framework ( $\epsilon$  and  $f_{esc}$ ) remains largely constant during the peak growing season. These findings provide a theoretically robust and operational basis for linking SIF observations with crop growth.

### 1. Introduction

Agricultural production models have been successfully applied to diverse research topics such as predicting crop yield, evaluating various management strategies, and assessing the impact of climate change on crops. One such model, the APSIM (Agricultural Production Systems sIMulator) model, has evolved into a platform containing modules that simulate various aspects of agricultural production systems and allow

flexible specification of management scenarios (Holzworth et al., 2014). One of the key drivers of crop growth and yield is canopy photosynthesis. In the standard APSIM crop model, daily potential photosynthesis rate (biomass accumulation) is estimated based on radiation interception and a radiation use efficiency (RUE) (Sinclair and Muchow, 1999) scheme. RUE is species specific and varies between different phenological stages. The impact of air temperature, water supply, vapor pressure deficit,  $CO_2$  concentration, and plant nutrition are determined

\* Corresponding author.

E-mail address: [luxiaoliang@nwfau.edu.cn](mailto:luxiaoliang@nwfau.edu.cn) (X. Lu).

<https://doi.org/10.1016/j.compag.2022.107494>

Received 2 June 2022; Received in revised form 4 November 2022; Accepted 9 November 2022

Available online 19 November 2022

0168-1699/© 2022 Elsevier B.V. All rights reserved.

by using a series of stress functions that reduce the potential photosynthesis rate. Although the RUE type of canopy photosynthesis modelling is simple and straightforward, it is not adequate to explicitly quantify the specific contribution of these limiting factors in photosynthesis.

Solar-induced chlorophyll fluorescence (SIF) is energy emitted by one dissipative pathway of absorbed photosynthetically active radiation (APAR). About 1% of APAR is re-emitted by chloroplasts in the longer wavelength spectral range of 640 to 850 nm (Porcar-Castell et al., 2014). Recent advances in spectroscopy and retrieval algorithms make it possible to measure SIF by either remote sensing or near-surface systems (Frankenberg et al., 2011; Joiner et al., 2011; Yang et al., 2015), and many studies have shown that SIF emission has a close link with plant photosynthetic rate. For example, Guanter et al. (2014) show that measurements of SIF from satellites can be used to estimate gross primary productivity (GPP) of croplands, while Guan et al. (2017) show that SIF has an advantage over traditional greenness indices estimated from reflectance-based indices for detecting the physiological status of crops. Several satellite SIF products have been retrieved from space-borne spectrometers, including GOME-2 (Joiner et al., 2011), TANSO-FTS (Frankenberg et al., 2011), and SCIAMACHY (Sun et al., 2018). Among such products, the recent SIF data from the Tropospheric Monitoring Instrument (TROPOMI, Köhler et al., 2018) has a notable advantage in spatial and temporal resolution (up to 7 km × 3.5 km with daily revisits).

Many researchers are now using data assimilation or machine learning approaches to better constrain crop models. For example, Huang et al. (2016) used a Kalman filter algorithm to assimilate remotely sensed leaf area index (LAI) time series data with the WOFOST crop model, thereby enhancing regional wheat yield predictability under climate fluctuations. Despite these advances, remote sensing indices, such as Normalized Difference Vegetation Index can only detect the “greenness”, rather than the actual photosynthetic activity of vegetation, and thus their use results in a certain lag in detecting a response to environmental stresses (Lee et al., 2013; Lu et al., 2018). Studies have already proven that remote sensing of SIF is a promising indicator of photosynthetic status and related stress effects in terrestrial vegetation (see the review provided by Mohammed et al., 2019). Combining SIF observations with crop models provides a new opportunity to better track crop status and growth at much shorter time scales, which may allow more accurate predictions of crop yield than previous approaches which use directional reflectance in the near-infrared region (NIR) to detect plant status.

SIF emission can be simulated with the canopy radiative transfer models (RTMs) or the terrestrial biosphere models (TBMs). To track the propagation of SIF photons within the vegetation canopy, a variety of SIF RTMs with different degrees of complexity have been developed. For example, three-dimensional SIF RTMs including Discrete Anisotropic Radiative Transfer (DART) model (Gastellu-Etchegorry et al., 2015), the FluorWPS model (Zhao et al., 2016), the FluorFLIES model (Gao et al., 2022) and the FluorFLIGHT model (Hernández-Clemente et al., 2017) provide a strong basis to quantitatively interpret the interaction of SIF photons with the different components of the forest scene (i.e., tree crowns, understory vegetation, soils). By contrast, the soil canopy observation photosynthesis energy (SCOPE) model (Van der Tol et al., 2009), a one-dimensional SIF RTM, shows a strong potential to simulate SIF emission in ecosystems with homogeneous canopies (e.g. crops). Although producing accurate results, almost all these SIF RTMs have high requirements on inputs and are computationally too expensive, making them less suitable for global applications. Attempts to add SIF modules to TBMs have been made in recent years. Lee et al. (2015) developed the equations to simulate SIF emission in the Community Land Model version 4 (CLM4) following van der Tol et al. (2014). Qiu et al. (2019) implemented a SIF module in the Boreal Ecosystem Productivity Simulator (BEPS) to address the impacts of canopy scattering on observed SIF. A SIF module has two main jobs: (1) to estimate

fullband SIF emission (640–850 nm,  $SIF_{tot,full}$ ) from photosynthetic activity, and (2) to calculate the top-of-canopy (TOC) narrowband SIF emission ( $SIF_{toc}$ ) in the observation direction. In previous efforts (Lee et al., 2015; Qiu et al., 2019), a single correction factor was used to convert  $SIF_{tot,full}$  into  $SIF_{toc}$ . However, this highly aggregated factor did not explicitly account for the conversion from broadband to narrowband SIF or the strong dependence of  $SIF_{toc}$  on observation direction (Liu et al., 2016). Thus, it may have limited application because both the broadband to narrowband SIF conversion and the angle effect of  $SIF_{toc}$  vary across time and space.

A cross-scale model for photosynthesis, namely the diurnal canopy photosynthesis–stomatal conductance module (DCaPST, Wu et al., 2019), was recently added to APSIM (APSIM v. 7.8). Compared with the empirical RUE approach, the DCaPST module mechanistically simulates crop photosynthesis at both leaf and canopy scales (see Section 2.1 for details). It enables an evaluation of changes in photosynthetic properties and their impact on canopy-level carbon assimilation and crop yield. By explicitly representing the key processes of photosynthesis, the new APSIM model equipped with the DCaPST module (referred to as APSIM-DCaPST) provides an opportunity to connect crop growth with the chlorophyll fluorescence emitted by crops. Here, we describe the development of a SIF module and its integration into the framework of APSIM-DCaPST. All of the steps in the conversion from  $SIF_{tot,full}$  to  $SIF_{toc}$  are explicitly formulated, and the impact of the controlling factors is quantitatively analyzed. Hereafter, we refer to the APSIM-DCaPST model with the addition of the SIF module as APSIM-SIF.

## 2. Methods and data

### 2.1. Brief description of DCaPST

APSIM is a process-based crop model that includes plant, soil, climate, and management modules (Holzworth et al., 2014). Wu et al. (2019) developed a new photosynthesis module (APSIM-DCaPST) under the APSIM crop modelling framework (v. 7.8). APSIM-DCaPST is driven by canopy leaf area index, specific leaf nitrogen, and crop water supply as well as daily meteorological data (i.e., solar radiation, maximum and minimum air temperature, and precipitation), and generates outputs of daily growth of shoot biomass ( $BIO_{shoot,day}$ , g biomass  $m^{-2} day^{-1}$ ) that are fed back into the crop model. Photosynthesis of both the sunlit and shaded leaf fractions of the crop canopy are calculated separately at an hourly time step, and summed for the canopy and over the diurnal period to obtain daily canopy  $CO_2$  assimilation ( $A$ ,  $\mu mol CO_2 m^{-2} day^{-1}$ ).  $A$  is then used to calculate  $BIO_{shoot,day}$ :

$$BIO_{shoot,day} = A \times \frac{44g/mol}{1,000,000} \times \frac{B}{1 + root : shoot} \quad (1)$$

where  $B$  is a conversion factor for biochemical conversion of  $CO_2$  to biomass and maintenance respiration which is set to 0.41 and 0.34 (g biomass (g  $CO_2$ )<sup>-1</sup>) for maize and soybean, respectively (Sinclair and Horie, 1989); root:shoot is the ratio of biomass allocation between roots and above-ground shoots and varies with crop growth stage. Typically, this ratio increases in the early stages of crop development and decreases towards flowering (Wu et al., 2019). The DCaPST module also uses specific leaf nitrogen (SLN, g N  $m^{-2}$ ) to represent the crop nitrogen status which constrains a variety of key photosynthesis-related parameters such as the maximum rate of rubisco carboxylation ( $V_{cmax}$ ) and the maximum rate of electron transport ( $J_{max}$ ). The APSIM-DCaPST model has been validated against field-observed data, and has demonstrated a strong predictive power by explaining nearly 90% of the variance in the crop yields of C3 wheat and C4 sorghum (Wu et al., 2019).

### 2.2. Incorporation of chlorophyll fluorescence into APSIM-DCaPST

The total SIF emitted from leaves can be thought of as the product of

the absorbed photosynthetically active radiation (APAR,  $\text{mW m}^{-2}$ ) multiplied by the plant fluorescence yield (Guanter et al., 2014):

$$\text{SIF}_{\text{tot\_full}} = \text{PAR} \times f_{\text{APAR}} \times \Phi_{\text{F}} = \text{APAR} \times \Phi_{\text{F}} \quad (2)$$

where  $\text{SIF}_{\text{tot\_full}}$  ( $\text{mW m}^{-2}$ ) is the total (all leaves in the canopy) SIF flux density within the full SIF spectrum (640–850 nm); PAR ( $\text{mW m}^{-2}$ ) is the flux of photosynthetically active radiation at the top of the canopy, and is converted from solar radiation by assuming that the fraction of PAR to solar radiation is 50%;  $f_{\text{APAR}}$  is the fraction of PAR absorbed by the plant, and  $\Phi_{\text{F}}$  is the fluorescence yield (number of photons that fluoresce per absorbed photon). In the DCaPST module, APAR (and thus  $f_{\text{APAR}}$ ) is calculated using the procedures set out by de Pury and Farquhar (1997).

In addition to SIF emission, APAR has two pathways: photosynthesis (photochemical quenching) and non-photochemical quenching (NPQ) or heat. In analogy to  $\Phi_{\text{F}}$ , their yields can be expressed as  $\Phi_{\text{P}}$  and  $\Phi_{\text{NPQ}}$ . Moreover, NPQ is the sum of heat loss in light-adapted conditions ( $\Phi_{\text{N}}$ ) and in dark-adapted conditions ( $\Phi_{\text{D}}$ ). So, according to the conservation of energy, we have:

$$\Phi_{\text{F}} + \Phi_{\text{P}} + \Phi_{\text{N}} + \Phi_{\text{D}} = 1 \quad (3)$$

The fraction of each item can be expressed by their rate constants ( $k$ ) as:

$$\begin{aligned} \Phi_{\text{F}} &= \frac{k_{\text{F}}}{\sum K} \\ \Phi_{\text{P}} &= \frac{k_{\text{P}}}{\sum K} \\ \Phi_{\text{N}} &= \frac{k_{\text{N}}}{\sum K} \\ \Phi_{\text{D}} &= \frac{k_{\text{D}}}{\sum K} \end{aligned} \quad (4)$$

$$\sum K = k_{\text{F}} + k_{\text{P}} + k_{\text{N}} + k_{\text{D}}$$

and  $\Phi_{\text{F}}$  can be written as (Lee et al., 2015):

$$\Phi_{\text{F}} = \frac{k_{\text{F}}}{k_{\text{F}} + k_{\text{N}} + k_{\text{D}}} (1 - \Phi_{\text{P}}) \quad (5)$$

Van der Tol et al (2014) suggested that  $k_{\text{F}} = 0.05$  and  $k_{\text{D}} = \max(0.03T_{\text{air}} + 0.0773, 0.87)$  where  $T_{\text{air}}$  is air temperature ( $^{\circ}\text{C}$ );  $k_{\text{N}}$  can be represented as a function of the degree of light saturation (Van der Tol et al., 2014):

$$k_{\text{N}} = \frac{(1 + \beta)x^{\alpha}}{\beta + x^{\alpha}} \times k_{\text{N}}^{\circ} \quad (6)$$

where  $k_{\text{N}}^{\circ}$ ,  $\alpha$ , and  $\beta$  are fitted parameters. Based on measurements of cotton growing under unstressed conditions, these parameters have been given values of 2.48, 2.83, and 0.114, respectively (Van der Tol et al., 2014).  $x$  ranges from 0 (photochemistry operating at full efficiency) to 1 (photochemistry totally blocked by feedback), and  $x$  can be defined as (Lee et al., 2013):

$$x = 1 - \frac{\Phi_{\text{P}}}{\Phi_{\text{P0}}} = 1 - \frac{J_e}{J_0} \rightarrow \Phi_{\text{P}} = \frac{J_e}{J_0} \Phi_{\text{P0}} \quad (7)$$

where  $\Phi_{\text{P0}}$  is the maximum photochemical yield as observed under dark adapted, low light conditions;  $J_e$  is the actual electron transport rate ( $\mu\text{mol m}^{-2} \text{s}^{-1}$ ); and  $J_0$  is the maximum potential electron transport rate at  $25^{\circ}\text{C}$  ( $\mu\text{mol m}^{-2} \text{s}^{-1}$ ) calculated from APAR and the dark-adapted rate constants (Lee et al., 2015).  $J_e$  can be estimated by (Lee et al., 2015):

$$J_e = \begin{cases} A \frac{C_i + 2\Gamma_*}{C_i - \Gamma_*} & \text{C3} \\ A & \text{C4} \end{cases} \quad (8)$$

where  $C_i$  ( $\mu\text{mol mol}^{-1}$ ) is the  $\text{CO}_2$  concentration within the intercellular air space;  $\Gamma_*$  ( $\mu\text{mol mol}^{-1}$ ) is the  $\text{CO}_2$  concentration at the compensation point. By combining Eq. (2), (5), (6), (7), and (8), one can obtain  $\text{SIF}_{\text{tot\_full}}$  for C3 and C4 crops, as follows:

$$\text{SIF}_{\text{tot\_full}} = \frac{k_{\text{F}}}{k_{\text{F}} + \frac{(1+\beta)x^{\alpha}}{\beta+x^{\alpha}} \times k_{\text{N}}^{\circ} + k_{\text{D}}} \left( 1 - \Phi_{\text{P0}} \times \frac{A \frac{C_i + 2\Gamma_*}{C_i - \Gamma_*}}{J_0} \right) \times \text{APAR C3} \quad (9)$$

$$\text{SIF}_{\text{tot\_full}} = \frac{k_{\text{F}}}{k_{\text{F}} + \frac{(1+\beta)x^{\alpha}}{\beta+x^{\alpha}} \times k_{\text{N}}^{\circ} + k_{\text{D}}} \left( 1 - \Phi_{\text{P0}} \times \frac{A}{J_0} \right) \times \text{APAR C4} \quad (10)$$

where  $\Gamma_*$  can be estimated by using the Farquhar-von Caemmerer-Berry (FvCB) model (Farquhar et al., 1980) of the leaf photosynthesis- $\text{CO}_2$  response curve;  $C_i$ ,  $J_0$ ,  $A$ , and APAR can be obtained from the APSIM-DCaPST model.

### 2.3. Conversion of $\text{SIF}_{\text{tot\_full}}$ into $\text{SIF}_{\text{toc}}(\lambda)$

$\text{SIF}_{\text{tot\_full}}$ , introduced in Section 2.2, is broadband SIF emission at the leaf scale. In contrast, top-of-canopy SIF radiance at a wavelength of  $\lambda$  nm ( $\text{SIF}_{\text{toc}}(\lambda)$ ,  $\text{mW m}^{-2} \text{nm}^{-1} \text{sr}^{-1}$ ) represents SIF signals propagated throughout the canopy and measured by remote sensing or near-surface platforms. Due to the scattering/(re)absorption processes inside the leaves, as well as in the canopy,  $\text{SIF}_{\text{toc}}(\lambda)$  measured at the canopy level is only a small proportion of  $\text{SIF}_{\text{tot\_full}}$ . Two more steps are needed to obtain  $\text{SIF}_{\text{toc}}(\lambda)$  from  $\text{SIF}_{\text{tot\_full}}$ : (1) converting  $\text{SIF}_{\text{tot\_full}}$  into  $\text{SIF}_{\text{tot}}(\lambda)$  ( $\text{mW m}^{-2} \text{nm}^{-1}$ ), i.e., the total fluorescence emitted by all leaves at a wavelength of  $\lambda$  nm, and (2) converting  $\text{SIF}_{\text{tot}}(\lambda)$  into  $\text{SIF}_{\text{toc}}(\lambda)$  by accounting for scattering/(re)absorption of SIF signals in the observation direction. Note that TOC SIF emission is a function of the canopy structure, sun-canopy-sensor geometry, and leaf optical properties. In this study, for simplicity, we expressed it as  $\text{SIF}_{\text{toc}}(\lambda)$ .

#### 2.3.1. Conversion of $\text{SIF}_{\text{tot\_full}}$ into $\text{SIF}_{\text{tot}}(\lambda)$

First, we used the leaf excitation-fluorescence matrices (EF-matrices, Van der Tol et al., 2019) simulated by the Fluspect model (Vilfan et al., 2016) to convert  $\text{SIF}_{\text{tot\_full}}$  into  $\text{SIF}_{\text{tot}}(\lambda)$ . The EF-matrices ( $351 \times 211$ ) are responsible for converting excitation spectra (400–750 nm) into fluorescence emission spectra (640–850 nm) with a 1-nm resolution. Fluspect is an extension of the leaf optical model of PROSPECT (Jacquemoud and Baret, 1990) which incorporates the simulation of scattering and absorption of chlorophyll fluorescence within the leaf. Fluspect produces four EF-matrices with one backward scattering matrix and one forward scattering matrix for both photosystem I (PSI) and photosystem II (PSII) as a function of fluorescence quantum efficiency for PSI (Fqe1) and PSII (Fqe2), leaf structural, and biochemical parameters. Hence, a total of four EF-matrices at 1 nm spectral resolution were calculated: one for each photosystem for each side of the leaf.

In Fluspect,  $\text{SIF}_{\text{tot}}(\lambda)$  is the product of the summation of these four EF-matrices ( $EF_{\text{sum}}$ ) and the irradiance (400–750 nm). As the APSIM model does not simulate the solar irradiance spectra,  $\text{SIF}_{\text{tot}}(\lambda)$  cannot be directly obtained from the product of the EF-matrices and the irradiance spectra. Alternatively, a SIF band conversion factor ( $\epsilon(\lambda)$ ,  $\text{nm}^{-1}$ ) linking  $\text{SIF}_{\text{tot\_full}}$  to  $\text{SIF}_{\text{tot}}(\lambda)$  can be defined:

$$\epsilon(\lambda) = \frac{\sum_{i=400}^{750} EF_{\text{sum}}(i, \lambda)}{\sum_{\lambda=640}^{850} \sum_{i=400}^{750} EF_{\text{sum}}(i, \lambda)} \quad \lambda \in [640\text{nm}..850\text{nm}] \quad (11)$$

Eq. (11) shows that  $\epsilon$  was actually calculated as SIF radiance at a given wavelength divided by the integral of SIF radiance over the fluorescence spectrum.  $\epsilon$  ( $1 \times 211$ ) describes the spectral distribution of emitted fluorescence at the leaf scale. The applicability of Eq. (11) was supported by the strong relationship ( $R^2 = 1.00$ ) between  $\epsilon$  and the ratio of single-band fluorescence to broadband fluorescence simulated by the

Fluspect model (see Fig. S1). Fig. 1 shows an example of  $\epsilon$  in the spectral range of 640–850 nm. In this example,  $\epsilon$  at 740 nm was approximately equal to 0.012, suggesting that SIF emission at 740 nm (the thin grey band in Fig. 1) accounted for only about 1.2% of SIF emitted from the full SIF spectrum (the shaded red area in Fig. 1).

The shape of  $\epsilon$  depends on leaf structural and biochemical parameters, including leaf thickness parameter ( $N$ ), leaf chlorophyll content ( $C_{ab}$ ,  $\mu\text{g cm}^{-2}$ ), carotenoid content ( $C_{car}$ ,  $\mu\text{g cm}^{-2}$ ), dry matter content ( $C_{dm}$ ,  $\mu\text{g cm}^{-2}$ ), anthocyanins ( $C_{ant}$ ,  $\mu\text{g cm}^{-2}$ ), senescence material fraction ( $C_s$ , fraction), leaf water equivalent layer ( $C_w$ , cm), and fluorescence quantum efficiency for photosystem I (Fqe1) and photosystem II (Fqe2). The sensitivity of  $\epsilon$  to these inputs is detailed in Text S1. By using  $\epsilon$ , one can convert  $SIF_{tot\_full}$  into  $SIF_{tot}(\lambda)$ :

$$SIF_{tot}(\lambda) = SIF_{tot\_full} \times \epsilon(\lambda) \quad (12)$$

### 2.3.2. Conversion of $SIF_{tot}(\lambda)$ into $SIF_{toc}(\lambda)$

Due to the scattering/(re)absorption processes occurring in the plant canopy, only part of  $SIF_{tot}(\lambda)$  can escape and thus be observed by a sensor above the canopy. That part is  $SIF_{toc}(\lambda)$ . The fluorescence escape probability ( $f_{esc}$ , Mohammed et al., 2019), the ratio between  $SIF_{toc}$  multiplied by  $\pi$  and  $SIF_{tot}$ , is typically used to determine the fraction of total emitted SIF that escapes the vegetation canopy and is detected remotely (Lu et al., 2020):

$$SIF_{toc}(\lambda) = \frac{f_{esc}(\lambda) \times SIF_{tot}(\lambda)}{\pi} \quad (13)$$

$f_{esc}$  is determined by the sun-canopy-sensor geometry, leaf optical properties, and canopy structure. Recent studies (Liu et al., 2016; Yang and van der Tol, 2018) have shown that the scattering/(re)absorption processes of NIR SIF can be well expressed by directional reflectance. Also, based on spectral invariant theory (Knyazikhin et al., 2011), Zeng et al (2019) showed that  $f_{esc}$  in the NIR region can be estimated as:

$$f_{esc} = \frac{R_{NIR} \times NDVI}{f_{APAR}} \quad (14)$$

where  $R_{NIR}$  represents directional reflectance in the NIR region and NDVI is the normalized difference vegetation index. In this study,  $f_{APAR}$  was taken from the DCaPST module.

The schematic representation of the construction and calibration of

the APSIM-SIF model is illustrated in Fig. 2. The leaf physiological information for the fluorescence simulation is provided by the DCaPST module, while the fluorescence upscaling from leaf level to canopy level depends on  $\epsilon$  and  $f_{esc}$ . The required inputs include the soil, weather, cultivar, crop management information and measurements of directional reflectance, and the main outputs of APSIM-SIF are LAI, above-ground biomass,  $A$  and SIF emissions (e.g.  $SIF_{tot\_full}$ ,  $SIF_{tot}$ , and  $SIF_{toc}$ ). Observations of LAI ( $LAI_{obs}$ ) and above-ground biomass ( $Biomass_{obs}$ ) can be used to calibrate the cultivar parameters in the APSIM-SIF model. Further, its DCaPST module may be also calibrated using observations of  $A$  ( $A_{obs}$ ) and remotely sensed TOC SIF ( $SIF_{toc\_obs}$ ).

## 2.4. Data

### 2.4.1. Study region

The study region (Fig. S2) consisted of 12 states in the U.S. Midwest: North Dakota, South Dakota, Nebraska, Kansas, Minnesota, Iowa, Missouri, Wisconsin, Illinois, Michigan, Indiana, and Ohio. The APSIM-SIF model was used to simulate two major crops (maize and soybean) in this region. The area of maize in the Midwest in 2018 was 29,522,461 ha, while the soybean area was 25,439,093 ha. The study region is one of the most intense areas of agricultural production in the world, accounting for about 33% and 34% of global maize and soybean production, respectively.

### 2.4.2. Data for model calibration/evaluation

We used observations from six crop flux sites located in the United States and China for model calibration and evaluation (Table S1). The eddy flux data and ground-based crop growth observations were measured at two AmeriFlux agricultural sites (US-Ne2 and US-Ne3) located at the University of Nebraska Eastern Nebraska Research and Extension Center near Mead, Nebraska (Ne), USA. US-Ne2 (41.1649 N, 96.4701 W) is an irrigated site equipped with a center pivot system; US-Ne3 (41.1797 N, 96.4397 W) is a rainfed site. Both sites use rotated maize-soybean cropping systems. Because of the completeness of their data collection, which includes eddy-covariance flux and meteorological measurements, along with detailed ground-based crop growth observations (e.g., aboveground biomass and LAI) (Peng et al., 2018), US-Ne2 and US-Ne3 are the main sites providing data to calibrate the APSIM-SIF model for irrigated and rainfed maize. The other ground-based

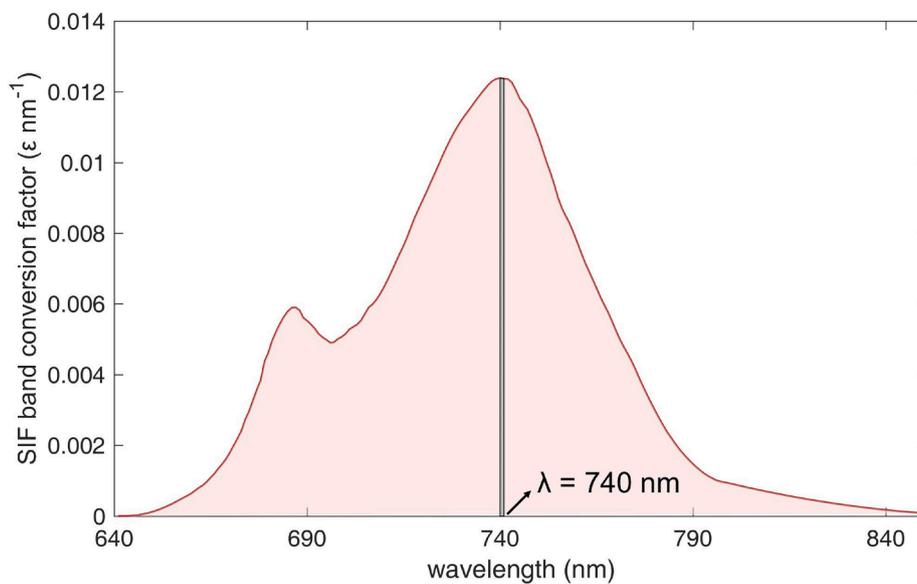
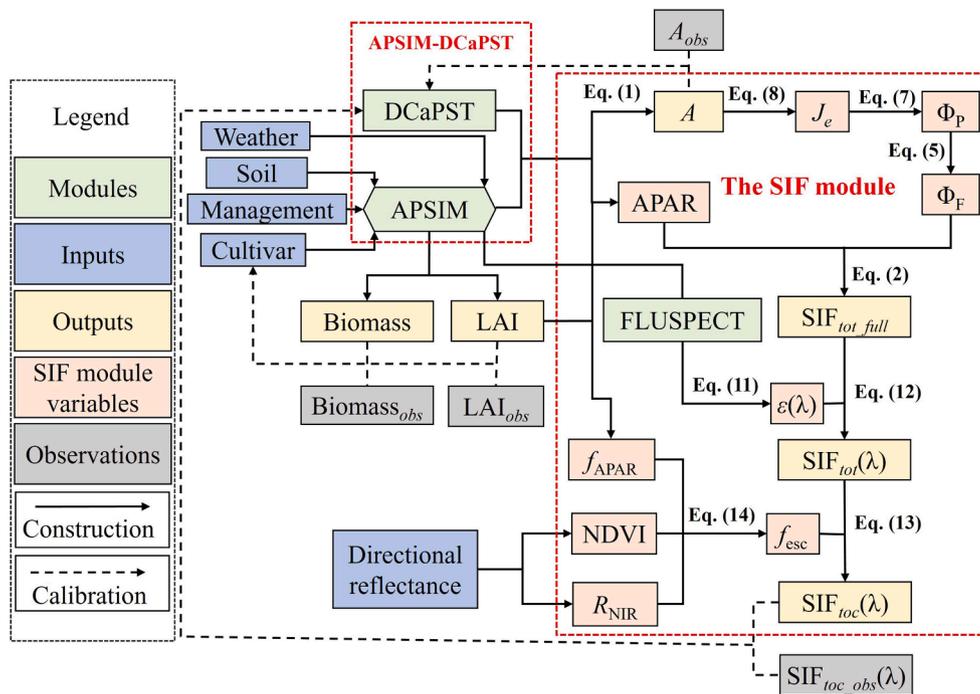


Fig. 1. An example of the SIF band conversion factor ( $\epsilon$ ,  $\text{nm}^{-1}$ , Eq. (11)) simulated by the Fluspect model. The marker indicates the 740 nm wavelength and the gray line has a width of 1 nm. The input parameters were set as follows:  $C_{ab} = 80 \mu\text{g cm}^{-2}$ ,  $C_{car} = 20 \mu\text{g cm}^{-2}$ ,  $C_{dm} = 0.012 \mu\text{g cm}^{-2}$ ,  $C_{ant} = 0 \mu\text{g cm}^{-2}$ ,  $C_s = 0$ ,  $C_w = 0.009$  cm,  $N = 1.4$ ,  $Fqe1 = 0.002$  and  $Fqe2 = 0.01$ . The definitions of these inputs are provided in the main text.



**Fig. 2.** Flowchart of the construction and calibration process of the APSIM-SIF model. The two red dashed rectangles represent the two major components of the APSIM-SIF model: the APSIM-DCaPST model and the SIF module. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

measurements of GPP and SIF at 760 nm were from the Daman site, located south of Zhangye, Gansu Province, China (CN-Dm, 100.3722°E, 38.8555°N). The major crop type at this site is irrigated maize, which was usually sown in early May and harvested in mid-September. The SIF observation system at Daman was installed 25 m above the ground on an eddy covariance tower. The main component of the SIF system was a spectrometer (QE65PRO, Ocean Optics, Inc.) which covers the spectral range of 645 – 805 nm with a spectral resolution of 0.34 nm, spectral sampling interval of 0.17 nm and signal-to-noise ratio of 1000. Details of the SIF measurements can be found in Liu et al., 2021. SIF measurements and the eddy flux data during the growing season were collected in 2018 from day 150 to day 230. Daily maximum temperature, minimum temperature, precipitation, and solar radiation collected at these three flux tower sites were used as APSIM-SIF forcing data. The soil parameters for the study sites were estimated from the GSDE dataset (see below).

For model calibration, the cultivar parameters for maize (Table S2) were determined by minimizing the root mean square error (RMSE) between simulated aboveground biomass and LAI and corresponding observations at US-Ne2 and US-Ne3. Due to the lack of measurements of aboveground biomass and LAI for soybean, the soybean cultivar parameters were set to default values (Table S3). Three important parameters ( $\chi_{V_{cmax25}}$ ,  $\chi_{J_{max25}}$ ,  $\chi_{gm}$ , Table S4) in the DCaPST module were also optimized for maize and soybean by using daily GPP time series at these two sites. The planting date and density were optimized according to the daily GPP data, and the setting of other management parameters referred to the typical agricultural management measures in the area (Kukul and Irmak, 2020). These parameters were assumed to remain constant across the study region (Table S5). The 30-min GPP values were extracted from the difference between ecosystem respiration and the net ecosystem exchange of CO<sub>2</sub> (NEE) using air temperature, following the method proposed by Reichstein et al (2005). The gap-filling and flux partitioning processes were performed using the REddyProc tool (<http://www.bgc-jena.mpg.de/bgi/index.php/Services/REddyProcWeb>), provided by the Max Planck Institute for Biogeochemistry, Germany (Wutzler et al., 2018). The daily average of GPP was calculated using 30-

minute data from 6:00 to 18:00 to match the daily GPP simulations of APSIM-DCaPST (Wu et al., 2019).

The performance of APSIM-DCaPST for maize was evaluated by comparing simulated GPP, aboveground biomass, and LAI with corresponding measurements at the US-Ne2 and US-Ne3 sites during the 2011 and 2013 growing seasons (Table S1). Because US-Ne2 and US-Ne3 were also the calibration sites, daily mean GPP from three other sites: Bondville, Illinois (Bo1) near Champaign, Illinois, Brooks Field Site 10 (Br1) near Ames Iowa, and Rosemount-G21 (Ro1) near Minneapolis, Minnesota were also used to evaluate the photosynthesis processes of APSIM-DCaPST more independently. These three sites were rainfed rotational maize-soybean cropping systems (Table S1). Ground-based measurements of TOC SIF at 760 nm and *in situ* measured canopy GPP at the CN-Dm site were also used to assess the SIF module.

#### 2.4.3. Data for the regional simulation

The spatial distribution of maize and soybean in 2018 (Fig. S2) was extracted from the USDA NASS cropland data layer (CDL) dataset. The CDL contains annually mapped crop types at a 30-m spatial resolution for the contiguous United States since 2008. Daily meteorological inputs, consisting of maximum and minimum temperature, precipitation, and solar radiation, were obtained from the Daymet V3 dataset (Thornton et al., 2016). The required soil data were estimated from the gridded Global Soil Dataset for use in Earth System Models (GSDE, Shangguan et al. 2014). GSDE provides global soil information at a horizontal spatial resolution of 30 arc-seconds. The vertical variation of soil properties is represented by eight layers to a maximum depth of 230 cm (Table S6). Bulk density (BD, g cm<sup>-3</sup>), drained lower limit (LL15, mm mm<sup>-1</sup>), drained upper limit (DUL, mm mm<sup>-1</sup>), and saturated water content (SAT, mm mm<sup>-1</sup>) were acquired from GSDE for each soil layer (Fig. S3). BD was directly obtained from GSDE; LL15, DUL, and SAT were estimated from the percentages of sand and clay, as described by Saxton et al (1986).

We used the SIF data product provided by TROPOMI on board the Copernicus Sentinel-5 Precursor satellite to evaluate TOC SIF simulated by the APSIM-SIF model at the regional scale. Köhler et al. (2018)

developed a data-driven approach to retrieve SIF emission at 740 nm from TROPOMI radiance measurements, based on the singular value decomposition technique. The TROPOMI SIF product provides daily global observations at a high spatial resolution of  $3.5 \times 7$  km at nadir. We used data from May to September in 2018 with cloud cover  $< 0.15$ . The correction factor (Köhler et al., 2018) provided in the TROPOMI SIF data product was used to convert instantaneous SIF into daily mean SIF.  $R_{NIR}$  (Eq. (14)) was calculated as the ratio of the canopy continuous-level radiance to the solar irradiance obtained from the TROPOMI SIF product (Zhang et al., 2019) to ensure that the observation direction remained the same as for the satellite SIF observations. NDVI (Eq. (14)) was calculated from the MODIS Nadir BRDF-Adjusted Reflectance (NBAR) product (MCD43A4.006), which has a spatial resolution of 500 m, in order to minimize potential seasonal effects under different observation and illumination conditions (Strahler et al., 1999). The MODIS NDVI product was resampled to the spatial resolution of TROPOMI SIF ( $7 \text{ km} \times 7 \text{ km}$ ) using the Google Earth Engine (GEE, <https://earthengine.google.com>) platform.

#### 2.4.4. The APSIM-SIF simulation

To match with the TROPOMI SIF product, we conducted the APSIM-

SIF simulation for maize and soybean in the U.S. Midwest at a spatial resolution of  $7 \times 7$  km and at a daily time step. The crop grid squares for maize and soybean were resampled from the CDL dataset based on the nearest neighbor method. Pixels with a low crop fraction ( $< 10\%$ ) were not considered. In total, we obtained 6114 grid squares for maize and 5955 grid squares for soybean (Fig. S2). A three-year period was simulated, consisting of one spin-up year (2017) and two years (2018 and 2019) for the regional simulation. All the required inputs for the regional simulation were converted to the APSIM format and aggregated into a spatial resolution of  $7 \times 7$  km.

### 3. Results

#### 3.1. APSIM-SIF evaluation

##### 3.1.1. Validation of the APSIM-DCaPST model

The APSIM-SIF model accounted for most of the variability in daily GPP for maize at US-Ne2 and US-Ne3, with average  $R^2$  values of 0.93 (RMSE =  $2.61 \text{ g C m}^{-2} \text{ day}^{-1}$ ) and 0.94 (RMSE =  $3.13 \text{ g C m}^{-2} \text{ day}^{-1}$ ), respectively (Fig. 3a and b). Although the simulated aboveground biomass had a strong correlation with the observations ( $R^2 > 0.99$ ),

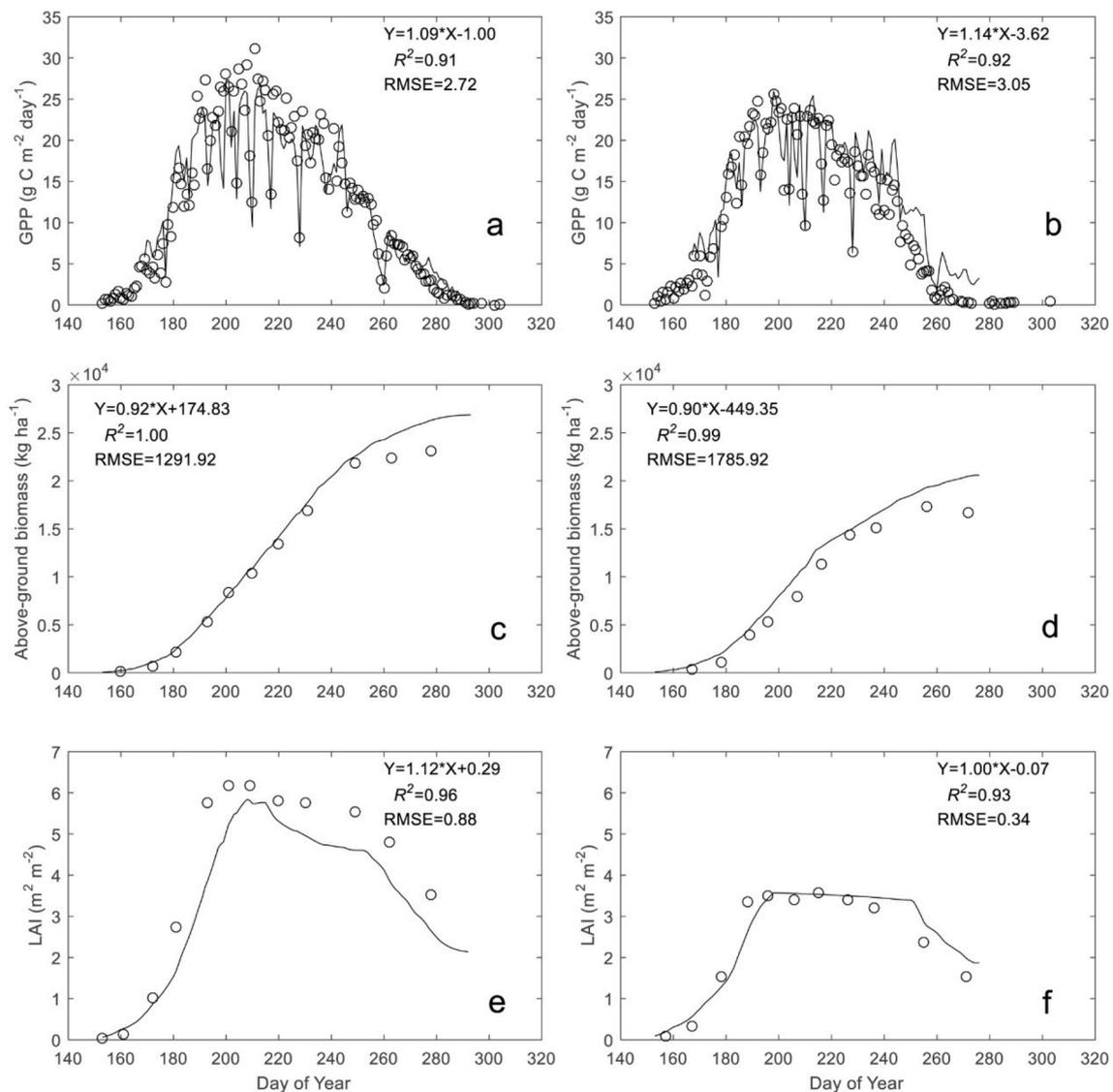


Fig. 3. Observations (circles) of daily mean gross primary productivity (GPP,  $\text{g C m}^{-2} \text{ day}^{-1}$ , a, b), above-ground biomass ( $\text{kg ha}^{-1}$ ) (c, d), and leaf area index (LAI,  $\text{m}^2 \text{ m}^{-2}$ , e, f) for maize at US-Ne2 (left column) and US-Ne3 (right column) in 2011 compared with values simulated by APSIM-SIF (lines). Similar results for 2013 are provided in Fig. S4.

APSIM-SIF tended to overestimate above-ground biomass in the later growing season by a magnitude of around  $1500 \text{ kg ha}^{-1}$  with a relative error  $< 7.5\%$  (Fig. 3c and d). The APSIM-SIF model also accounted for more than 90% of the variation in LAI measurements at these two sites (Fig. 3e and f), but it appears to have underestimated LAI at high observed LAI values.

The performance of APSIM-SIF in simulating daily GPP for maize and soybean was further evaluated at Bo1, Br1, and Ro1 (Table S1, Fig. S5). At these three sites, simulated daily maize GPP was well correlated with observed GPP, with the model accounting for more than 65% of the variability in GPP: 66% (RMSE =  $5.35 \text{ g C m}^{-2} \text{ day}^{-1}$ ), 85% (RMSE =  $3.46 \text{ g C m}^{-2} \text{ day}^{-1}$ ), and 83% (RMSE =  $3.59 \text{ g C m}^{-2} \text{ day}^{-1}$ ) at Bo1, Br1, and Ro1, respectively (Fig. S5a, b, c). The model also captured the dynamics of photosynthetic activity of soybean well, explaining 64% (RMSE =  $3.04 \text{ g C m}^{-2} \text{ day}^{-1}$ ), 80% (RMSE =  $2.65 \text{ g C m}^{-2} \text{ day}^{-1}$ ), and 71% (RMSE =  $3.43 \text{ g C m}^{-2} \text{ day}^{-1}$ ) of the variability in daily GPP at the respective three sites (Fig. S5d, e, f).

### 3.1.2. Validation of the SIF module

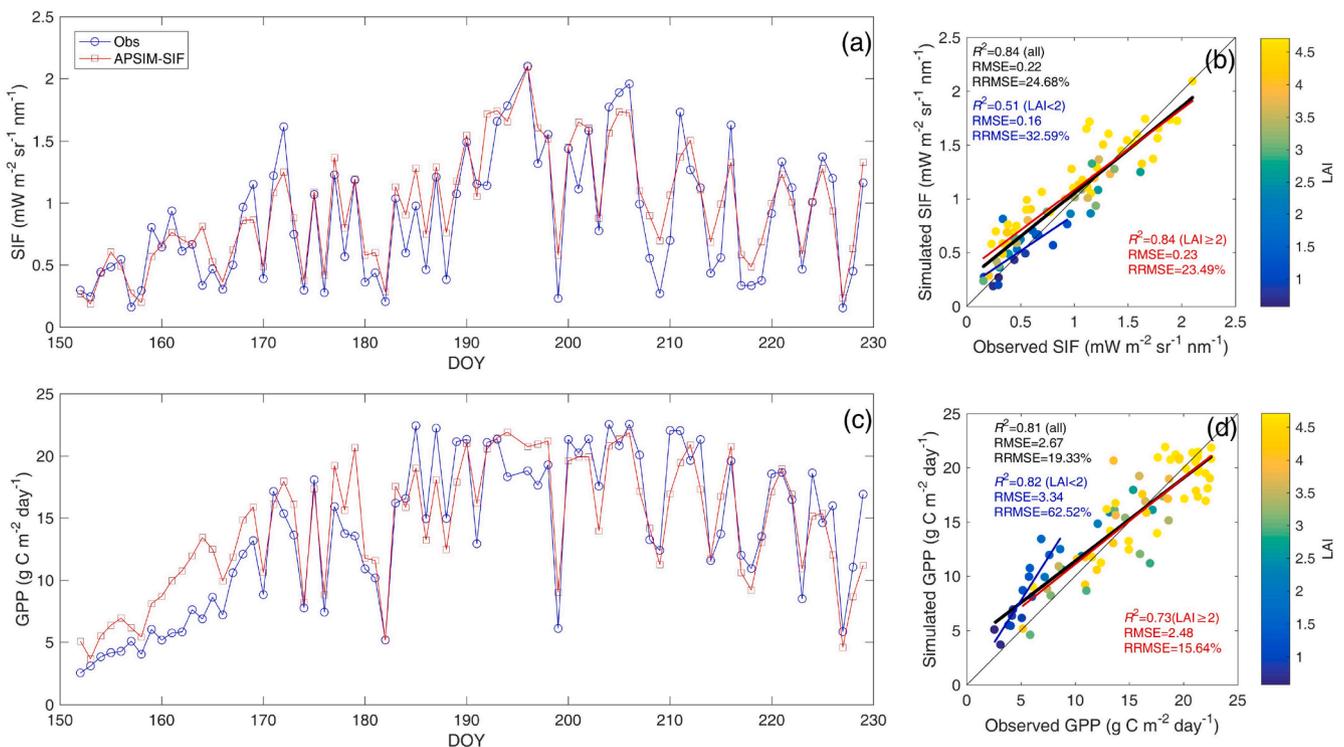
The APSIM-SIF model captures the seasonality of TOC SIF (Fig. 4a) and GPP (Fig. 4c) at the Daman site well with  $R^2$  values of 0.84 (RMSE =  $0.22 \text{ mW m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$ , RRMSE = 24.68%) and 0.81 (RMSE =  $2.67 \text{ g C m}^{-2} \text{ day}^{-1}$ , RRMSE = 19.33%), respectively. The  $R^2$  between simulated and observed TOC SIF ( $\text{SIF}_{\text{toc,m}}$  vs.  $\text{SIF}_{\text{toc,o}}$ ) showed an increasing trend as LAI increased (Fig. 4b).  $\text{SIF}_{\text{toc,m}}$  determined 51% of  $\text{SIF}_{\text{toc,o}}$  variability (RMSE =  $0.16 \text{ mW m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$ , RRMSE = 32.59%) in the low LAI values (LAI  $< 2$ ), indicating the negative effect of soil reflectance caused by the sparse canopy. In contrast,  $\text{SIF}_{\text{toc,m}}$  accounted for 84% of the variance (RMSE =  $0.22 \text{ mW m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$ , RRMSE = 23.49%) in  $\text{SIF}_{\text{toc,o}}$  under high LAI levels. We also found that GPP was overestimated by the model when LAI values were low: the RRMSE was larger than 60% for LAI  $< 2$  (Fig. 4c and d).

## 3.2. Regional SIF simulation

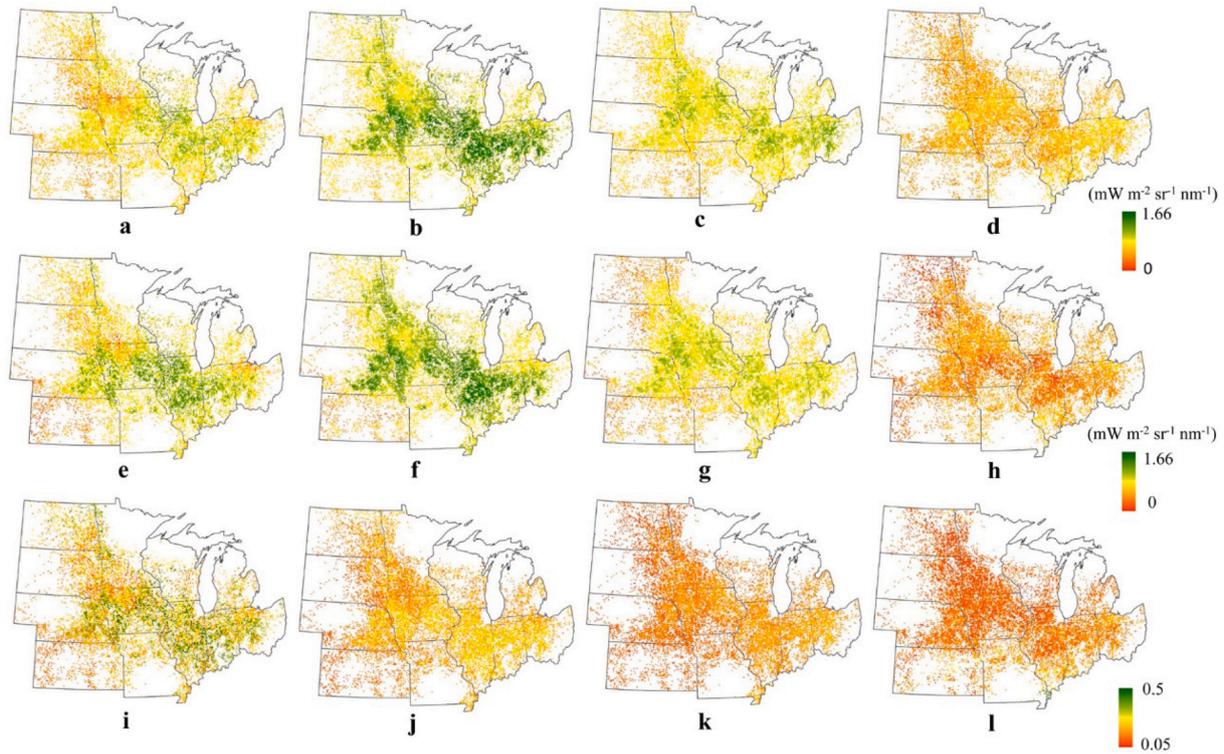
### 3.2.1. Variability of $\text{SIF}_{\text{toc}}$ and $f_{\text{esc}}$

We produced monthly mean top-of-canopy SIF values ( $\text{SIF}_{\text{toc,o}}$ ,  $\text{mW m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$ ) from daily TROPOMI SIF observations, for the period from June to September (Fig. 5 a-d). We were also able to simulate top-of-canopy SIF ( $\text{SIF}_{\text{toc,m}}$ , Fig. 5e-h) from  $\text{SIF}_{\text{tot}}$  and  $f_{\text{esc}}$ . Overall,  $\text{SIF}_{\text{toc,m}}$  values were of comparable magnitude to  $\text{SIF}_{\text{toc,o}}$  for soybean and maize: the largest difference between  $\text{SIF}_{\text{toc,m}}$  and  $\text{SIF}_{\text{toc,o}}$  being  $< 0.1 \text{ mW m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$  (11%) in the summer (July and August, Fig. 5f and g), confirming that the estimation of  $\text{SIF}_{\text{tot}}$  and  $f_{\text{esc}}$  was reasonably good. However, the difference between  $\text{SIF}_{\text{toc,m}}$  and  $\text{SIF}_{\text{toc,o}}$  increased in September (Fig. 5h):  $\text{SIF}_{\text{toc,m}}$  was underestimated by 0.16 and 0.13 for maize and soybean (29% and 25%), respectively (Table 1). To check the spatiotemporal correspondence between simulated and observed TOC SIF, an empirical orthogonal function (EOF) analysis (Sun et al., 2017) was performed on the monthly averaged  $\text{SIF}_{\text{toc,m}}$  and  $\text{SIF}_{\text{toc,o}}$  in the study region during the growing season (June–September), and their temporal and spatial coherences for the orthogonal components were investigated. The first two EOFs (EOF1 and EOF2) explained 94% (Fig. S6a and e) and 98% (Fig. S6b and f) of the variability in  $\text{SIF}_{\text{toc,m}}$  and  $\text{SIF}_{\text{toc,o}}$ , respectively. For the first leading mode,  $\text{SIF}_{\text{toc,m}}$  and  $\text{SIF}_{\text{toc,o}}$  showed similar spatial patterns, especially in the Corn Belt (Fig. S6a and b), and the Pearson correlation coefficient ( $r$ ) between them was consistently high across maize (0.73, Fig. S6d) and soybean (0.63, Fig. S6d). In the second mode,  $\text{SIF}_{\text{toc,m}}$  had a higher correlation with  $\text{SIF}_{\text{toc,o}}$  ( $r = 0.78$ , Fig. S6e and f), with  $r = 0.59$  and 0.85 for maize and soybean (Fig. S6h), respectively. The time series of the first two components also showed a high similarity between  $\text{SIF}_{\text{toc,m}}$  and  $\text{SIF}_{\text{toc,o}}$  (Fig. S6c and g). Taken together, the EOF analysis showed that the spatiotemporal variations of  $\text{SIF}_{\text{toc,m}}$  and  $\text{SIF}_{\text{toc,o}}$  were highly consistent, indicating that the APSIM-SIF model had a satisfactory performance in simulating TOC SIF in maize and soybean.

Daily  $f_{\text{esc}}$  values were also aggregated into monthly values (Fig. 5i-l). Poor quality reflectance data may cause large uncertainties in  $f_{\text{esc}}$ , and



**Fig. 4.** Comparisons of daily SIF (a, b) and GPP (c, d) between ground measurements and simulations from APSIM-SIF at the Daman site in 2018. The performance of APSIM-SIF with the coefficient of determination ( $R^2$ ), root mean square error (RMSE,  $\text{mW m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$  for SIF and  $\text{g C m}^{-2} \text{ day}^{-1}$  for GPP), relative root mean square error (RRMSE, %) under different LAI levels is shown in the subplots on the right.



**Fig. 5.** Monthly mean top-of-canopy SIF emission provided by TROPOMI observations ( $SIF_{toc,o}$ ,  $mW m^{-2} sr^{-1} nm^{-1}$ ), top-of-canopy SIF emission simulated by the APSIM-SIF model ( $SIF_{toc,m}$ ,  $mW m^{-2} sr^{-1} nm^{-1}$ ), and fluorescence escape ratio ( $f_{esc}$ ) in June (a, e, i), July (b, f, j), August (c, g, k), and September (d, h, l), 2018.  $SIF_{toc,o}$ ,  $f_{esc}$ , and  $SIF_{toc,m}$  are for 740 nm.

**Table 1**

Daily mean and standard deviations (in parentheses) of fluorescence escape ratio ( $f_{esc}$ ) and top-of-canopy SIF emission ( $SIF_{toc,m}$ ,  $mW m^{-2} sr^{-1} nm^{-1}$ ), photosynthetically active radiation (PAR,  $MJ m^{-2} day^{-1}$ ), fraction of absorbed photosynthetically active radiation ( $f_{APAR}$ ), and leaf area index (LAI,  $m^2 m^{-2}$ ) of maize (M) and soybean (S) simulated by the APSIM-SIF model for U.S. Midwest in June, July, August, and September 2018.  $SIF_{toc,o}$  represents daily mean top-of-canopy SIF emission provided by TROPOMI.  $SIF_{toc,m}$ ,  $SIF_{toc,o}$ , and  $f_{esc}$  are for 740 nm.

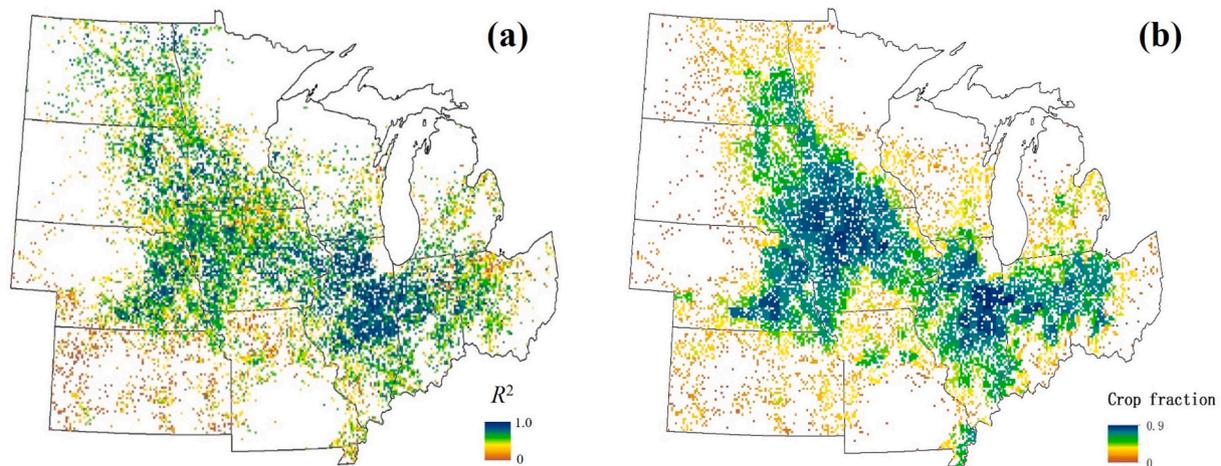
|               | June            |                 | July            |                 | August          |                 | September       |                 |
|---------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|               | M               | S               | M               | S               | M               | S               | M               | S               |
| $SIF_{toc,o}$ | 0.75<br>(0.19)  | 0.80<br>(0.25)  | 1.08<br>(0.24)  | 1.01<br>(0.22)  | 0.80<br>(0.16)  | 0.87<br>(0.19)  | 0.55<br>(0.13)  | 0.51<br>(0.12)  |
| $f_{esc}$     | 0.18<br>(0.04)  | 0.32<br>(0.09)  | 0.15<br>(0.03)  | 0.21<br>(0.05)  | 0.14<br>(0.03)  | 0.13<br>(0.03)  | 0.12<br>(0.03)  | 0.13<br>(0.05)  |
| $SIF_{toc,m}$ | 0.78<br>(0.24)  | 0.95<br>(0.29)  | 1.01<br>(0.27)  | 1.00<br>(0.24)  | 0.78<br>(0.21)  | 0.77<br>(0.21)  | 0.39<br>(0.13)  | 0.38<br>(0.14)  |
| PAR           | 19.98<br>(1.82) | 20.07<br>(2.58) | 21.71<br>(0.89) | 21.84<br>(0.83) | 18.43<br>(1.10) | 18.33<br>(1.15) | 15.81<br>(1.66) | 15.87<br>(1.71) |
| $f_{APAR}$    | 0.63<br>(0.06)  | 0.34<br>(0.03)  | 0.92<br>(0.02)  | 0.66<br>(0.03)  | 0.92<br>(0.03)  | 0.94<br>(0.03)  | 0.81<br>(0.05)  | 0.78<br>(0.18)  |
| LAI           | 1.97<br>(0.42)  | 0.68<br>(0.07)  | 4.67<br>(0.28)  | 2.04<br>(0.19)  | 4.11<br>(0.45)  | 4.99<br>(0.82)  | 2.27<br>(0.42)  | 2.84<br>(1.67)  |

~ 2% of the available data was discarded by adopting  $f_{esc} < 0.05$  and  $f_{esc}$  greater than 0.5 as thresholds. Maize had a relatively stable  $f_{esc}$  during the growing season, varying in the range of 0.12–0.18 (Table 1). The leaf angle distribution in maize was more planophile during early stages of crop development, although a shift to erectophile canopies may occur with increased LAI to avoid overlap between leaves (Dechant et al., 2020). Because erectophile canopies have smaller  $f_{esc}$  than planophile canopies (Zeng et al., 2019), this adjustment in leaf angle distribution tended to offset the increase in  $f_{esc}$  due to increase in LAI, leading to a relatively stable pattern of  $f_{esc}$  in maize (Table 1). In contrast, the  $f_{esc}$  of soybean showed a strong increase during the early phase, with a moderate decrease in the peak growth period of the growing season: soybean  $f_{esc}$  reached a maximum value of 0.32 in June (Fig. 5i, Table 1), decreased to 0.21 in July (Fig. 5j, Table 1), and remained almost

constant at 0.13 in the final two months (Fig. 5k and l, Table 1). The leaf angle distribution in soybean largely remained erectophile throughout the growing season, which may explain the observed decreasing pattern of  $f_{esc}$  when the canopy closed (Table 1). In addition, soybean had a consistently higher  $f_{esc}$  than maize in June and July: soybean  $f_{esc}$  was almost twice as large as maize  $f_{esc}$  in June (Table 1). During senescence, the cropland became more heterogeneous as LAI decreased, such that  $f_{esc}$  derived under the 1-D assumption (see Limitations) may have caused more uncertainties (Wang et al., 2020; Zhang et al., 2019).

### 3.2.2. Effects of cropland area on the performance of APSIM-SIF

The spatial distribution of the  $R^2$  between daily  $SIF_{toc,m}$  and  $SIF_{toc,o}$  (Fig. 6) showed that the performance of APSIM-SIF in simulating TOC SIF emission may deteriorate for grid squares with small cropland area.



**Fig. 6.** The spatial distribution of the coefficient of determination ( $R^2$ ) between the daily simulated and observed TOC SIF ( $\text{mW m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$ ) (a) and the crop fraction (b) within  $7 \times 7$  km grid squares across the U.S. Midwest in 2018.

For example, among the states where cropland covers most of the area (e.g., Illinois and Iowa), a much stronger correlation between  $\text{SIF}_{\text{TOC},m}$  and  $\text{SIF}_{\text{TOC},o}$  was observed than for those with small, fragmented cropland areas (e.g., Kansas and Missouri) (Fig. 6a). The  $R^2$  between  $\text{SIF}_{\text{TOC},m}$  and  $\text{SIF}_{\text{TOC},o}$  increased from 40% to more than 85% when crop fraction increased from 30% to 80% (Fig. 6a and b). In particular,  $R^2$  reached more than 0.75 for grid squares with more than 70% crop area, which accounted for around 28% of the total number of grid squares (Fig. 6b). In contrast, most  $R^2$  values were less than 0.5 for grid squares with cropland area of less than 30% (Fig. 6a and b).

## 4. Discussion

### 4.1. From photosynthesis to TOC SIF

In related previously published studies, a single correction factor ( $k$ ), which was either a constant or an empirical function of LAI, was used to convert leaf-level full broadband SIF emission into a TOC narrowband SIF signal. However, as we have shown above, this conversion involves two major steps that are very different in their physical basis and controlling factors. Although  $k$  is computationally efficient, its use makes it difficult to identify the key variables, and may limit development of a deeper understanding of the underlying mechanisms operating.

We mechanistically decomposed this conversion into two factors: the SIF band conversion factor  $\epsilon$ , and the fluorescence escape ratio,  $f_{\text{esc}}$ . As the conversion is explicitly formulated, any further advances in estimating  $\epsilon$  and  $f_{\text{esc}}$  can be easily employed to improve the performance of APSIM-SIF. Further, we found that  $\epsilon$  in the NIR region was controlled primarily by the quantum efficiency of photosystem II (Fqe2) and leaf chlorophyll concentration ( $C_{\text{ab}}$ ), reflecting the impact of plant status and growth stage (Fig. S7). Considering that the variation in  $C_{\text{ab}}$  is relatively low in the peak growth season, it is plausible that the estimation of  $\epsilon$  in the NIR wavelengths can be greatly simplified.

$f_{\text{esc}}$  represents the impact of canopy structure on radiation transfer of the SIF signal within plant canopies. It is a complex function of the sun-canopy-sensor geometry and canopy structural parameters (e.g., leaf area index and leaf inclination distribution) such that use of a simple factor, such as  $k$ , without observing constraints may not estimate  $f_{\text{esc}}$  with reasonable accuracy. Recent studies (Yang and van der Tol, 2018; Liu et al., 2016; Zeng et al., 2019; Lu et al 2020) have shown that directional reflectance in the NIR domain has good potential to represent the scattering/(re)absorption processes in the propagation of NIR SIF through vegetation canopies. In other words, the impacts of these processes on  $f_{\text{esc}}$  are implicitly taken into account by incorporating directional reflectance. As directional reflectance is concurrently

measured with TOC SIF and the atmospheric effect is limited in the NIR domain (Zhang et al., 2019), the reflectance-based method has a practical advantage. We also found that  $f_{\text{esc}}$  remained relatively constant across maize (0.14–0.15) and soybean (0.17–0.21) dominated cropland during the summer crop growth period (July–August, Table 1). Thus it is feasible to determine  $f_{\text{esc}}$  by using a constant value derived from the peak growing season, which may considerably simplify the application of the APSIM-SIF model for crop canopies.

### 4.2. Limitations

First, there is room for improvement in the calculation of  $\epsilon$  in future work. The excitation-fluorescence matrices used in this study were developed for maize (PSI) and barley (PSII), respectively (Franck et al., 2002), and their level of suitability for maize and soybean is not clear. Also, there is a lack of understanding regarding how these matrices deal with the response to stressful conditions such as high air temperature, strong light intensity, high VPD, etc. Additional experiments will be needed to evaluate the performance of the matrices under stressful conditions.

Second, the derivation of  $f_{\text{esc}}$  (Eq. (14)) is based on the spectral invariant theory (Knyazikhin et al., 2011) that assumes that the contribution of soil to TOC directional reflectance is negligible (i.e., ‘black soil’). However, for row crops, such as maize and soybean, the strips of bare soil between rows can make a substantial contribution to TOC radiance prior to the soil surface being completely covered by the crop canopy (Zhao et al., 2010). Early or late in the growing season, when LAI values are low,  $f_{\text{esc}}$  may also have large uncertainties due to strong soil contamination.

## 5. Conclusions

In this study, the APSIM-SIF model was developed within the framework of the new APSIM model and its DCaPST module. Leaf-level broadband SIF emission was first estimated according to photosynthetic carbon assimilation in crops. In contrast to previous studies, we then mechanistically decomposed the conversion from broadband SIF emitted from leaves to narrowband TOC SIF into two factors, namely  $\epsilon$  and  $f_{\text{esc}}$ . We found that  $\epsilon$  is responsible for estimating narrowband SIF from the broadband SIF at the leaf scale, and can reveal the impact of physiological status. The second factor,  $f_{\text{esc}}$ , on the other hand, represents the impact of canopy structure on the radiative transfer of emitted SIF. TOC directional reflectance in the NIR region demonstrated good potential for estimating  $f_{\text{esc}}$ . We also show that  $f_{\text{esc}}$  remained relatively constant during the peak growth period of the crop growing season,

which promotes the APSIM-SIF model's efficiency. The validation results show that APSIM-SIF can accurately track dynamic changes in GPP and SIF of crops at the site scale. Our findings highlight the importance in considering of cropland area on the performance of APSIM-SIF. This work contributes to the understanding of the observed signals from satellites and crop photosynthesis at the regional scale. Further research is needed to combine SIF observations with the APSIM-SIF model to improve crop yield prediction. This study not only establishes the connection between crop photosynthesis and SIF emission, but also provides a general scheme to integrate SIF emission into the photosynthesis process in agricultural system models.

#### CRedit authorship contribution statement

**Yakai Wang:** Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft. **Zhunqiao Liu:** Data curation, Formal analysis, Methodology, Software. **Qiang Yu:** Data curation, Methodology, Writing – review & editing. **Liangyun Liu:** Data curation. **Xinjie Liu:** Data curation. **Linchao Li:** Writing – review & editing. **Qianlan Jia:** Writing – review & editing. **Chenhui Guo:** Writing – review & editing. **Xiaoliang Lu:** Conceptualization, Formal analysis, Methodology, Project administration, Software, Writing – review & editing, Funding acquisition.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Acknowledgments

The authors acknowledge DCaPST code support from Alex Wu at the University of Queensland, Brisbane, Queensland, Australia. Acknowledgment is made to the APSIM Initiative which takes responsibility for quality assurance and a structured innovation programme for APSIM's modelling software, which is provided free for research and development use (see [www.apsim.info](http://www.apsim.info) for details). The authors thank the following AmeriFlux sites for their data records: US-Ne2, US-Ne3, US-Bo1, US-Br1, and US-Ro1. The TROPOMI SIF datasets are available at <ftp://fluo.gps.caltech.edu/data/tropomi/>. This research was funded by the National Natural Science of China (42071328) and the Chinese Universities Scientific Fund (24520212452021125).

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.compag.2022.107494>.

#### References

- De Pury, D., Farquhar, G., 1997. Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. *Plant, Cell Environ.* 20, 537–557. <https://doi.org/10.1111/j.1365-3040.1997.00094.x>.
- Dechant, B., Ryu, Y., Badgley, G., Zeng, Y., Berry, J.A., Zhang, Y., Goulas, Y., Li, Z., Zhang, Q., Kang, M., 2020. Canopy structure explains the relationship between photosynthesis and sun-induced chlorophyll fluorescence in crops. *Remote Sens. Environ.* 241, 111733 <https://doi.org/10.1016/j.rse.2020.111733>.
- Farquhar, G.D., von Caemmerer, S.V., Berry, J.A., 1980. A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C<sub>3</sub> species. *Planta* 149, 78–90. <https://doi.org/10.1007/BF00386231>.
- Franck, F., Juneau, P., Popovic, R., 2002. Resolution of the photosystem I and photosystem II contributions to chlorophyll fluorescence of intact leaves at room temperature. *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1556, 239–246. [https://doi.org/10.1016/S0005-2728\(02\)00366-3](https://doi.org/10.1016/S0005-2728(02)00366-3).
- Frankenberg, C., Fisher, J.B., Worden, J., Badgley, G., Saatchi, S.S., Lee, J.E., Toon, G.C., Butz, A., Jung, M., Kuze, A., 2011. New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. *Geophys. Res. Lett.* 38 <https://doi.org/10.1029/2011GL048738>.
- Gao, S., Huete, A., Kobayashi, H., Doody, T.M., Liu, W., Wang, Y., Zhang, Y., Lu, X., 2022. Simulation of solar-induced chlorophyll fluorescence in a heterogeneous forest using 3-D radiative transfer modelling and airborne LiDAR. *ISPRS J. Photogramm. Remote Sens.* 191, 1–17. <https://doi.org/10.1016/j.isprsjprs.2022.07.004>.
- Gastellu-Etchegorry, J.-P., Yin, T., Lauret, N., Cajfinger, T., Gregoire, T., Grau, E., Feret, J.-B., Lopes, M., Guilleux, J., Dedieu, G., 2015. Discrete anisotropic radiative transfer (DART 5) for modeling airborne and satellite spectroradiometer and LIDAR acquisitions of natural and urban landscapes. *Remote Sensing* 7, 1667–1701. <https://doi.org/10.3390/rs70201667>.
- Guan, K., Wu, J., Kimball, J.S., Anderson, M.C., Frolking, S., Li, B., Hain, C.R., Lobell, D. B., 2017. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. *Remote Sens. Environ.* 199, 333–349. <https://doi.org/10.1016/j.rse.2017.06.043>.
- Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J.A., Frankenberg, C., Huete, A.R., Zarco-Tejada, P. and Lee, J.-E., 2014. Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. *Proceedings of the National Academy of Sciences*, 111, E1327–E1333. <https://doi.org/10.1073/pnas.1320008111>.
- Hernández-Clemente, R., North, P.R., Hornero, A., Zarco-Tejada, P.J., 2017. Assessing the effects of forest health on sun-induced chlorophyll fluorescence using the FluorFLIGHT 3-D radiative transfer model to account for forest structure. *Remote Sens. Environ.* 193, 165–179. <https://doi.org/10.1016/j.rse.2017.02.012>.
- Holzworth, D.P., Huth, N.I., deVoil, P.G., Zurcher, E.J., Herrmann, N.I., McLean, G., Chenu, K., van Oosterom, E.J., Snow, V., Murphy, C., 2014. APSIM-evolution towards a new generation of agricultural systems simulation. *Environ. Model. Softw.* 62, 327–350. <https://doi.org/10.1016/j.envsoft.2014.07.009>.
- Huang, J., Sedano, F., Huang, Y., Ma, H., Li, X., Liang, S., Tian, L., Zhang, X., Fan, J., Wu, W., 2016. Assimilating a synthetic Kalman filter leaf area index series into the WOFOST model to improve regional winter wheat yield estimation. *Agric. For. Meteorol.* 216, 188–202. <https://doi.org/10.1016/j.agrformet.2015.10.013>.
- Jacquemoud, S., Baret, F., 1990. PROSPECT: A model of leaf optical properties spectra. *Remote Sens. Environ.* 34 (2), 75–91. [https://doi.org/10.1016/0034-4257\(90\)90100-Z](https://doi.org/10.1016/0034-4257(90)90100-Z).
- Joiner, J., Yoshida, Y., Vasilkov, A., Middleton, E., 2011. First observations of global and seasonal terrestrial chlorophyll fluorescence from space. *Biogeosciences* 8, 637. <https://doi.org/10.5194/bg-8-637-2011>.
- Knyazikhin, Y., Schull, M.A., Xu, L., Myneni, R.B., Samanta, A., 2011. Canopy spectral invariants. Part 1: A new concept in remote sensing of vegetation. *J. Quant. Spectrosc. Radiat. Transf.* 112, 727–735. <https://doi.org/10.1016/j.jqsrt.2010.06.014>.
- 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. *Geophys. Res. Lett.* 45, 10456–10463. <https://doi.org/10.1029/2018GL079031>.
- Kukal, M., Irmak, S., 2020. Light interactions, use and efficiency in row crop canopies under optimal growth conditions. *Agric. For. Meteorol.* 284, 107887 <https://doi.org/10.1016/j.agrformet.2019.107887>.
- 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 Biol.* 21, 3469–3477. <https://doi.org/10.1111/gcb.12948>.
- Lee, J.E., Frankenberg, C., van der Tol, C., Berry, J.A., Guanter, L., Boyce, C.K., Fisher, J. B., Morrow, E., Worden, J.R., Asefi, S., 2013. Forest productivity and water stress in Amazonia: Observations from GOSAT chlorophyll fluorescence. *Proc. R. Soc. B.* 280, 20130171. <https://doi.org/10.1098/rspb.2013.0171>.
- Liu, L., Liu, X. and Guan, L., 2016. Uncertainties in linking solar-induced chlorophyll fluorescence to plant photosynthetic activities, 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 4414–4417. <https://doi.org/10.1109/IGARSS.2016.7730150>.
- 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. *Agric. For. Meteorol.* 307, 108522 <https://doi.org/10.1016/j.agrformet.2021.108522>.
- Lu, X., Liu, Z., Zhao, F., Tang, J., 2020. Comparison of total emitted solar-induced chlorophyll fluorescence (SIF) and top-of-canopy (TOC) SIF in estimating photosynthesis. *Remote Sens. Environ.* 251, 112083 <https://doi.org/10.1016/j.rse.2020.112083>.
- Lu, X., Liu, Z., Zhou, Y., Liu, Y., An, S., Tang, J., 2018. Comparison of phenology estimated from reflectance-based indices and solar-induced chlorophyll fluorescence (SIF) observations in a temperate forest using GPP-based phenology as the standard. *Remote Sensing* 10, 932. <https://doi.org/10.3390/rs10060932>.
- Mohammed, G.H., Colombo, R., Middleton, E.M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., 2019. Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. *Remote Sens. Environ.* 231, 111177 <https://doi.org/10.1016/j.rse.2019.04.030>.
- Peng, B., Guan, K., Chen, M., Lawrence, D.M., Pokhrel, Y., Suyker, A., Arkebauer, T., Lu, Y., 2018. Improving maize growth processes in the community land model: Implementation and evaluation. *Agric. For. Meteorol.* 250, 64–89. <https://doi.org/10.1016/j.agrformet.2017.11.012>.
- Porcar-Castell, A., Tyystjärvi, E., Atherton, J., Van der Tol, C., Flexas, J., Pfundel, E.E., Moreno, J., Frankenberg, C., Berry, J.A., 2014. Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: mechanisms and challenges. *J. Exp. Bot.* 65, 4065–4095. <https://doi.org/10.1093/jxb/eru191>.

- Qiu, B., Chen, J.M., Ju, W., Zhang, Q., Zhang, Y., 2019. Simulating emission and scattering of solar-induced chlorophyll fluorescence at far-red band in global vegetation with different canopy structures. *Remote Sens. Environ.* 233, 111373 <https://doi.org/10.1016/j.rse.2019.111373>.
- Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., 2005. On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm. *Global Change Biol.* 11, 1424–1439. <https://doi.org/10.1111/j.1365-2486.2005.001002.x>.
- Saxton, K., Rawls, W.J., Romberger, J., Papendick, R., 1986. Estimating generalized soil-water characteristics from texture. *Soil Sci. Soc. Am. J.* 50, 1031–1036. <https://doi.org/10.2136/sssaj1986.03615995005000040039x>.
- Shangguan, W., Dai, Y., Duan, Q., Liu, B., Yuan, H., 2014. A global soil data set for earth system modeling. *J. Adv. Model. Earth Syst.* 6, 249–263. <https://doi.org/10.1002/2013MS000293>.
- Sinclair, T., Horie, T., 1989. Leaf nitrogen, photosynthesis, and crop radiation use efficiency: a review. *Crop Sci.* 29, 90–98. <https://doi.org/10.2135/cropsci1989.0011183X002900010023x>.
- Sinclair, T.R., Muchow, R.C., 1999. Radiation use efficiency. *Adv. Agron.* 65, 215–265. [https://doi.org/10.1016/S0065-2113\(08\)60914-1](https://doi.org/10.1016/S0065-2113(08)60914-1).
- Strahler, A.H., Muller, J., Lucht, W., Schaaf, C., Tsang, T., Gao, F., Li, X., Lewis, P., Barnsley, M.J., 1999. MODIS BRDF/albedo product: algorithm theoretical basis document version 5.0. MODIS documentation 23, 42–47.
- 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 Sens. Environ.* 209, 808–823. <https://doi.org/10.1016/j.rse.2018.02.016>.
- Sun, Y., Frankenberg, C., Wood, J.D., Schimel, D., Jung, M., Guanter, L., Drewry, D., Verma, M., Porcar-Castell, A., Griffis, T.J., 2017. OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. *Science* 358, eaam5747. <https://doi.org/10.1126/science.aam5747>.
- Thornton, P., Thornton, M., Vose, R., 2016. Daymet: Annual Tile Summary Cross-Validation Statistics for North America, Version 3. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAAC/1348>.
- Van der Tol, C., Berry, J., Campbell, P., Rascher, U., 2014. Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence. *J. Geophys. Res. Biogeosci.* 119, 2312–2327. <https://doi.org/10.1002/2014JG002713>.
- Van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., Su, Z., 2009. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. *Biogeosciences* 6, 3109–3129. <https://doi.org/10.5194/bg-6-3109-2009>.
- Van der Tol, C., Vilfan, N., Dauwe, D., Cendrero-Mateo, M.P., Yang, P., 2019. The scattering and re-absorption of red and near-infrared chlorophyll fluorescence in the models Fluspect and SCOPE. *Remote Sens. Environ.* 232, 111292 <https://doi.org/10.1016/j.rse.2019.111292>.
- Vilfan, N., Van der Tol, C., Muller, O., Rascher, U., Verhoef, W., 2016. Fluspect-B: A model for leaf fluorescence, reflectance and transmittance spectra. *Remote Sens. Environ.* 186, 596–615. <https://doi.org/10.1016/j.rse.2016.09.017>.
- Wang, C., Guan, K., Peng, B., Chen, M., Jiang, C., Zeng, Y., Wu, G., Wang, S., Wu, J., Yang, X., 2020. Satellite footprint data from OCO-2 and TROPOMI reveal significant spatio-temporal and inter-vegetation type variabilities of solar-induced fluorescence yield in the US Midwest. *Remote Sens. Environ.* 241, 111728 <https://doi.org/10.1016/j.rse.2020.111728>.
- Wu, A., Hammer, G.L., Doherty, A., von Caemmerer, S., Farquhar, G.D., 2019. Quantifying impacts of enhancing photosynthesis on crop yield. *Nat. Plants* 5, 380–388. <https://doi.org/10.1038/s41477-019-0398-8>.
- Wutzler, T., Lucas-Moffat, A., Migliavacca, M., Knauer, J., Sickel, K., Šigut, L., Menzer, O., Reichstein, M., 2018. Basic and extensible post-processing of eddy covariance flux data with REdDyProc. *Biogeosciences* 15, 5015–5030. <https://doi.org/10.5194/bg-15-5015-2018>.
- Yang, P., van der Tol, C., 2018. Linking canopy scattering of far-red sun-induced chlorophyll fluorescence with reflectance. *Remote Sens. Environ.* 209, 456–467. <https://doi.org/10.1016/j.rse.2018.02.029>.
- 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. *Geophys. Res. Lett.* 42, 2977–2987. <https://doi.org/10.1002/2015GL063201>.
- Zeng, Y., Badgley, G., Dechant, B., Ryu, Y., Chen, M., Berry, J.A., 2019. A practical approach for estimating the escape ratio of near-infrared solar-induced chlorophyll fluorescence. *Remote Sens. Environ.* 232, 111209 <https://doi.org/10.1016/j.rse.2019.05.028>.
- Zhang, Z., Chen, J.M., Guanter, L., He, L., Zhang, Y., 2019. From canopy-leaving to total canopy far-red fluorescence emission for remote sensing of photosynthesis: First results from TROPOMI. *Geophys. Res. Lett.* 46, 12030–12040. <https://doi.org/10.1029/2019GL084832>.
- Zhao, F., Dai, X., Verhoef, W., Guo, Y., van der Tol, C., Li, Y., Huang, Y., 2016. FluorWPS: A Monte Carlo ray-tracing model to compute sun-induced chlorophyll fluorescence of three-dimensional canopy. *Remote Sens. Environ.* 187, 385–399. <https://doi.org/10.1016/j.rse.2016.10.036>.
- Zhao, F., Gu, X., Verhoef, W., Wang, Q., Yu, T., Liu, Q., Huang, H., Qin, W., Chen, L., Zhao, H., 2010. A spectral directional reflectance model of row crops. *Remote Sens. Environ.* 114, 265–285. <https://doi.org/10.1016/j.rse.2009.09.018>.