

LETTER • OPEN ACCESS

Forests for forests: combining vegetation indices with solar-induced chlorophyll fluorescence in random forest models improves gross primary productivity prediction in the boreal forest

To cite this article: Zoe Amie Pierrat *et al* 2022 *Environ. Res. Lett.* **17** 125006

View the [article online](#) for updates and enhancements.

You may also like

- [Nitrogen deposition shows no consistent negative nor positive effect on the response of forest productivity to drought across European FLUXNET forest sites](#)
S C van der Graaf, T A J Janssen, J W Erisman *et al.*
- [Probabilistic impacts of compound dry and hot events on global gross primary production](#)
Xinying Wu and Dabang Jiang
- [Photosynthetic productivity and its efficiencies in ISIMIP2a biome models: benchmarking for impact assessment studies](#)
Akihiko Ito, Kazuya Nishina, Christopher P O Reyer *et al.*

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

OPEN ACCESS

RECEIVED
28 July 2022REVISED
18 November 2022ACCEPTED FOR PUBLICATION
24 November 2022PUBLISHED
12 December 2022

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.

Forests for forests: combining vegetation indices with
solar-induced chlorophyll fluorescence in random forest models
improves gross primary productivity prediction in the boreal forestZoe Amie Pierrat^{1,*} , Jacob Bortnik¹ , Bruce Johnson² , Alan Barr² , Troy Magney³ ,
David R Bowling⁴ , Nicholas Parazoo⁵ , Christian Frankenberg⁶ , Ulli Seibt¹ and Jochen Stutz¹ ¹ Department of Atmospheric and Oceanic Sciences, University of California Los Angeles, Los Angeles, CA, United States of America² Global Institute for Water Security, University of Saskatchewan, Saskatoon, SK, Canada³ Department of Plant Sciences, University of California Davis, Davis, CA, United States of America⁴ Department of Biology, University of Utah, Salt Lake City, UT, United States of America⁵ Jet Propulsion Laboratory, Pasadena, CA, United States of America⁶ Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, United States of America

* Author to whom any correspondence should be addressed.

E-mail: zpierrat@g.ucla.edu**Keywords:** remote sensing, random forest models, boreal forest, gross primary production, solar-induced fluorescence, light use efficiency, photosynthesisSupplementary material for this article is available [online](#)

Abstract

Remote sensing is a powerful tool for understanding and scaling measurements of plant carbon uptake via photosynthesis, gross primary productivity (GPP), across space and time. The success of remote sensing measurements can be attributed to their ability to capture valuable information on plant structure (physical) and function (physiological), both of which impact GPP. However, no single remote sensing measure provides a universal constraint on GPP and the relationships between remote sensing measurements and GPP are often site specific, thereby limiting broader usefulness and neglecting important nuances in these signals. Improvements must be made in how we connect remotely sensed measurements to GPP, particularly in boreal ecosystems which have been traditionally challenging to study with remote sensing. In this paper we improve GPP prediction by using random forest models as a quantitative framework that incorporates physical and physiological information provided by solar-induced fluorescence (SIF) and vegetation indices (VIs). We analyze 2.5 years of tower-based remote sensing data (SIF and VIs) across two field locations at the northern and southern ends of the North American boreal forest. We find (a) remotely sensed products contain information relevant for understanding GPP dynamics, (b) random forest models capture quantitative SIF, GPP, and light availability relationships, and (c) combining SIF and VIs in a random forest model outperforms traditional parameterizations of GPP based on SIF alone. Our new method for predicting GPP based on SIF and VIs improves our ability to quantify terrestrial carbon exchange in boreal ecosystems and has the potential for applications in other biomes.

1. Introduction

Uncertainty in future climate projections is largely driven by terrestrial ecosystem feedbacks on the carbon cycle (Friedlingstein *et al* 2014). A major contributor to terrestrial ecosystem feedbacks is plant carbon uptake via photosynthesis. Plant carbon

uptake can be estimated locally at the tower/site level, as gross primary productivity (GPP), but remote sensing is necessary to scale and understand carbon exchange across space and time (Anav *et al* 2015). This is especially relevant in arctic-boreal ecosystems which play a major, but highly uncertain, role in the global carbon cycle (Bonan 2008, Thurner

et al 2014). Arctic-boreal carbon cycle uncertainty can be attributed to unevenly distributed field observations and challenges in using remote sensing (e.g. high sun angles, snow cover, short growing seasons, persistent greenness) (Nelson *et al* 2022). Therefore, improvements must be made in how we connect remotely sensed measurements to GPP in arctic-boreal ecosystems.

GPP can be described using the light-use-efficiency model (Monteith 1972):

$$\text{GPP} = \text{PAR} \times f\text{PAR} \times \text{LUE}_p \quad (1)$$

where PAR is the photosynthetically active radiation, $f\text{PAR}$ is the fraction of photosynthetically active radiation absorbed by chlorophyll, and LUE_p is the light-use-efficiency of photosynthesis. Proposed remote sensing metrics for approximating GPP typically track either the physical ($f\text{PAR}$) or the physiological (LUE_p) components of this equation.

Solar-induced chlorophyll fluorescence is an especially powerful remote sensing metric for understanding GPP over a variety of ecosystems (Sun *et al* 2017, 2018) and across a variety of scales (Yang *et al* 2017, Mohammed *et al* 2019, Magney *et al* 2020, Porcar-Castell *et al* 2021, Pierrat *et al* 2022) due to its connection to both the physical ($\text{PAR} \times f\text{PAR}$) and the physiological (LUE_p) components of GPP. Solar-induced fluorescence (SIF) can be described using the light-use-efficiency model:

$$\text{SIF} = \text{PAR} \times f\text{PAR} \times \text{LUE}_F \times f_{\text{esc}} \quad (2)$$

where LUE_F is the light-use-efficiency of fluorescence, and f_{esc} is the fraction of emitted SIF photons which escape the canopy. SIF and GPP share the physical drivers $\text{PAR} \times f\text{PAR}$, which largely explains the relationship between SIF and GPP in cropping systems (Dechant *et al* 2020). Equations (1) and (2) can be combined to eliminate the absorbed photosynthetically active radiation ($\text{APAR} = \text{PAR} \times f\text{PAR}$):

$$\text{GPP} = \text{SIF} \times \frac{\text{LUE}_p}{\text{LUE}_F \times f_{\text{esc}}} \quad (3)$$

The physiological component of SIF, LUE_F covaries with LUE_p under moderate light conditions (Porcar-Castell *et al* 2014, van der Tol *et al* 2014) and averaged over broad spatio-temporal scales (Magney *et al* 2020, Pierrat *et al* 2022). This is the basis for SIF as a proxy for GPP when measured from space across a broad range of ecosystems (Sun *et al* 2017, 2018). However, these conditions are not necessarily always met (Marrs *et al* 2020), especially in the boreal forest where complex canopy structure and divergence between the light and carbon fixation reactions of photosynthesis can complicate these signals (Maguire *et al* 2020, Pierrat *et al* 2022). Therefore,

a more nuanced relationship between SIF and GPP that takes into consideration divergence between variations in LUE_F , LUE_p , and f_{esc} will further improve the utility of SIF as a proxy for GPP. Additional reflectance-based remotely sensed metrics, i.e. vegetation indices (VIs), can provide more information on both the structural and physiological processes impacting GPP and can therefore improve our ability to track and understand GPP with remote sensing.

Physically, SIF and GPP are impacted by illumination conditions and canopy structure (i.e. the combined effects of leaf-area index (LAI), vertical distribution of LAI, leaf orientation, clumping, etc) which mediate both $f\text{PAR}$ and f_{esc} . At a constant PAR and LUE_p , GPP amplifies under cloudy sky conditions because a higher diffuse fraction allows light to penetrate deeper into the canopy, thus increasing $f\text{PAR}$ (Gu *et al* 2002, Alton *et al* 2007, Durand *et al* 2021). Although the amplification of GPP under diffuse skies has been well documented, it is often not considered when approximating GPP with remote sensing proxies.

Canopy structure is often approximated using greenness based VIs which are sensitive to chlorophyll content in an instrument field of view. Thus, they are generally a good approximation for $f\text{PAR}$. The normalized difference vegetation index (NDVI) effectively tracks vegetation productivity in ecosystems where chlorophyll content and carbon uptake are closely correlated (i.e. larger variations in $f\text{PAR}$ than LUE_p) (Tucker 1979, Yang *et al* 2017, Wang and Friedl 2019). This is not the case in the boreal forest where changes in carbon uptake do not correlate with changes in chlorophyll content (Sims *et al* 2006, Garbulsky *et al* 2010, Gamon *et al* 2013). Additionally, NDVI is extremely sensitive to the presence of snow cover (Pierrat *et al* 2021a, 2022). The near-infrared reflectance from vegetation (NIRv) expands on NDVI by multiplying NDVI by the total scene NIR reflectance, thereby amplifying the vegetated signal. NIRv has shown stronger correlations with $f\text{PAR}$ and GPP than NDVI alone (Badgley *et al* 2017, 2019). NIRv can also be used as a proxy for f_{esc} (Zeng *et al* 2019). Both NDVI and NIRv provide useful information on the physical/structural influences on SIF and GPP, however, it is unclear how sensitive these indices are to changes in $f\text{PAR}$ and f_{esc} and if there exists a universal relationship across ecosystems.

Physiologically, the light-use-efficiencies of both fluorescence and photosynthesis are mediated by non-photochemical quenching—a heat dissipation mechanism that plants utilize to avoid damage from excess sunlight (Raczka *et al* 2019, Walter-McNeill *et al* 2021). The extent to which non-photochemical quenching is necessary to protect plant tissue depends on a host of environmental controls that determine a plant's ability to photosynthesize

(Demmig-Adams and Adams 2006). Of particular relevance in this respect is plant's sensitivity to temperature. Freezing over winter or heat waves over summer create stress conditions which subsequently impact LUE_P . Monitoring the extent of non-photochemical quenching thus provides insight into LUE_P (Adams *et al* 2004), and likely LUE_F , although those dependencies are not yet well quantified.

Non-photochemical quenching manifests in two forms of photoprotection in boreal ecosystems, both of which can be detected with remote sensing measures (Gamon *et al* 1997, 2016, Demmig-Adams *et al* 2008). The first is rapidly reversible and therefore important under short-term high light stress. It can be detected using the photochemical reflectance index (PRI). Therefore, PRI tracks changes in LUE_P over the diurnal cycle (Gamon *et al* 1997, Wong and Gamon 2015a, 2015b). The second form is sustained and thus important over longer time periods such as winter in regions where freezing temperatures limit the ability for plants to photosynthesize (Verhoeven 2014). This form can be tracked using the chlorophyll-carotenoid index (CCI) (Gamon *et al* 2016). CCI has successfully tracked variations in GPP over the seasonal cycle because it is sensitive to long-term changes in LUE_P .

Including information on heat dissipation dynamics through remotely sensed products (PRI and CCI) improves modeling of photosynthetic phenology (Wong *et al* 2022), especially when used in conjunction with SIF (Wang *et al* 2020, Hikosaka and Tsujimoto 2021). In addition, correcting the SIF signal for canopy and structural effects using structurally sensitive metrics also improves its relationship with GPP (Liu *et al* 2020, Lu *et al* 2020). It has been well documented and mechanistically explained why VIs are sensitive to changes in non-photochemical quenching or canopy structure but there are no universal quantitative relationships among them. To develop such a relationship, the impacts from canopy structure, view/sun angle effects, snowcover, and potential differences in instrumentation (field of view, instrument sensitivity) must all be accounted for. Until mechanistic models are able to effectively account for all the aforementioned effects, the capacity for these indices to inform SIF and GPP is limited to qualitative or site-specific empirical relationships. In order to make more effective use of remote sensing as a proxy for GPP, it is necessary to have a quantitative framework that can relate SIF, VIs, and GPP.

Advances in machine learning have provided exciting opportunities for data analysis and predictive modeling. In particular, random forest models are non-parametric in nature and are therefore well suited for approximating non-linear, multi-parameter relationships in complex systems

(Breiman 2001). Because random forest models do not assume functional dependencies between input variables and predicted output and rather 'learn' relationships based on input data, they are particularly useful for systems where statistical parameterized models either do not exist or are not well constrained. Random forest models are also highly interpretable compared with other machine learning techniques due to predictor importance estimates, which make them more useful for explaining and understanding observed relationships. Finally, random forest models have already been used to understand nuance in the SIF–GPP relationship (Jiao *et al* 2019, Bai *et al* 2022, Pierrat *et al* 2022). We propose that using random forest models as a tool to understand and predict GPP based on a combination of remote sensing metrics will present a significant improvement over traditional parameterized models (Monteith 1972, van der Tol *et al* 2014, Dechant *et al* 2020) or other machine learning approaches.

The central question of this study is therefore: **can we improve our ability to predict GPP from SIF by using random forest models as a quantitative framework that can incorporate additional physical and physiological information provided by VIs?** To answer this question, we present 2.5 years of tower-based remote sensing data across two boreal forest locations, qualitatively evaluate the seasonal and diurnal trends among them, and present an interpretation on the physical and physiological information contained in them (section 3.1). We justify the use of random forest models as a tool for understanding SIF, VI, and GPP dynamics by showing that they can accurately reproduce SIF–GPP–PAR relationships (section 3.2). Finally, we explore the utility of random forest models and remotely sensed products for improving GPP estimation by showing how random forest models driven by SIF, VIs, and temperature improve the prediction of GPP and shed light on important physical and physiological processes (sections 3.3 and 3.4).

2. Materials and methods

2.1. Site descriptions: Southern Old Black Spruce (SOBS) and National Ecological Observatory Network (NEON) Delta Junction (DEJU)

We collected data at the SOBS site (FLUXNET site code CA-Obs) and the National Ecological Observatory Network (NEON), DEJU which represent the northern and southern limits of the North American boreal forest and associated environmental conditions (figure 1, table 1). SOBS is located near the southern limit of the boreal forest in Saskatchewan, Canada (53.98°N, 105.12°W) (Jarvis *et al* 1997). It is a mixed forest stand with stem density predominantly (90%) black spruce (*Picea mariana*), and scattered

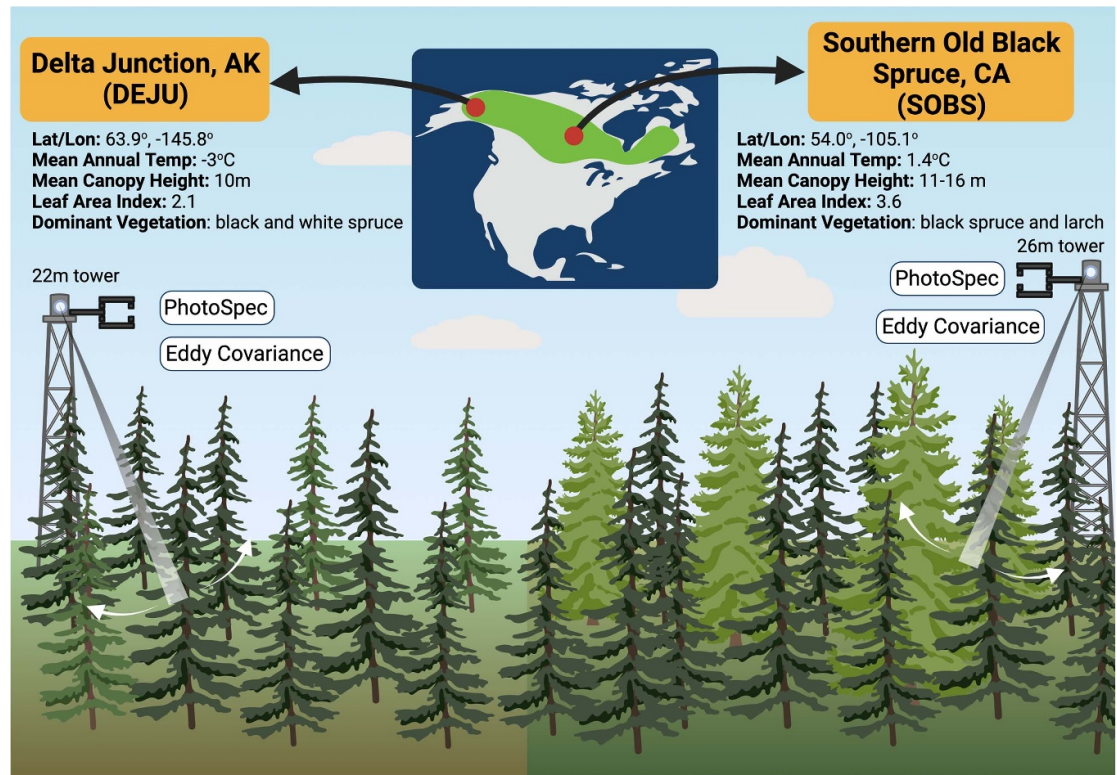


Figure 1. Experimental setup at the two boreal forest field locations. Site information for SOBS from (Chen *et al* 2006, Pappas *et al* 2020). Site information for DEJU from the National Ecological Observatory Network (NEON). Figure created with BioRender.com.

(10%) larch (*Larix laricina*). DEJU is located towards the northern end of the boreal forest outside DEJU, Alaska (63.88°N, 145.75°W). It is an evergreen forest dominated by black and white spruce (*P. mariana* and *Picea glauca*).

2.2. Data collection: tower-based remote sensing, GPP, and environmental variables

We collected tower-based remotely sensed measurements (far-red SIF, NDVI, NIRv, CCI, PRI) using PhotoSpec (see Grossmann *et al* 2018) for detailed instrument description) at both the SOBS and DEJU field sites. Measurements ran from August 2019 to December 2021 at both field locations (figures 1 and 2). At both sites, Photospec was installed atop the scaffolding tower facing due north. It has a narrow field of view (0.7°), 2D scanning capabilities, and simultaneously measures SIF and VIs at the same point in the canopy (Grossmann *et al* 2018). Individual measurements take approximately 20 s. We took canopy representative scans at both field locations within a 30 min window and averaged measurements together to compare with the temporal resolution of GPP and environmental variables. SIF was retrieved in the far-red (745–758 nm) wavelength range using a Fraunhofer-line based fitting algorithm (Grossmann *et al* 2018). The Fraunhofer-line based approach makes SIF

retrievals insensitive to atmospheric scattering and therefore robust even under cloudy sky conditions (Frankenberg *et al* 2011, Mohammed *et al* 2019, Chang *et al* 2020). We filtered data for low quality retrievals and retrievals with unstable sky conditions (Pierrat *et al* 2021a, 2022) for both sites. The VIs were calculated as follows, with $\rho_{nm:nm}$ = the average reflectance across a wavelength range in nm:

$$NDVI = \frac{\rho_{830:860} - \rho_{620:670}}{\rho_{830:860} + \rho_{620:670}} \quad (4)$$

(Tucker 1979)

$$NIRv = \frac{\rho_{830:860} - \rho_{620:670}}{\rho_{830:860} + \rho_{620:670}} \times \rho_{830:860} \quad (5)$$

(Badgley *et al* 2017)

$$PRI = \frac{\rho_{569:571} - \rho_{520:532}}{\rho_{569:571} + \rho_{520:532}} \quad (6)$$

(Gamon *et al* 1997)

$$CCI = \frac{\rho_{520:532} - \rho_{620:670}}{\rho_{520:532} + \rho_{620:670}} \quad (7)$$

(Gamon *et al* 2016).

In boreal ecosystems, the onset of photosynthesis often occurs prior to complete snowmelt (Starr and Oberbauer 2003, Parazoo *et al* 2018). Therefore, any

Table 1. Summary of methods for measurement and data processing at the two field locations.

Measurement	SOBS	DEJU
SIF, VIs, PAR, and Df	PhotoSpec mounted atop the tower and processed following (Grossmann <i>et al</i> 2018, Pierrat <i>et al</i> 2022).	PhotoSpec mounted atop the tower and processed following (Grossmann <i>et al</i> 2018, Pierrat <i>et al</i> 2022).
GPP		
Eddy-covariance	Taken using a 3D sonic anemometer (CSAT3, Campbell Scientific, Logan, UT) in combination with a closed-path infrared gas (CO ₂ /H ₂ O) analyzer (LI-7200 analyzer, Li-Cor, Lincoln, NE) operated in absolute mode. We performed quality assurance on the data using the standard Fluxnet-Canada method following (Barr <i>et al</i> 2006, Barr <i>et al</i> 2004).	Obtained from (National Ecological Observatory Network (NEON) 2022a) using a Campbell Scientific CSAT-3 3D Sonic Anemometer and LI-COR—LI7200 gas analyzer. We performed quality assurance on carbon fluxes based on turbulent and storage fluxes separately, using a bivariate statistical procedure for each, to overcome quality flag restrictions in the ‘expanded’ NEON eddy-covariance bundle. The 3% outliers (3% of rarest events from the tails of each distribution) were excluded from joint probability distributions for all available data for (a) turbulent flux and PAR, and separately for (b) storage flux and time of day. Net Ecosystem Exchange (NEE) data were considered valid if both the turbulent and storage fluxes passed this quality control step (and NEE is equal to their sum). Data for NEE and meteorological variables were filtered to remove low turbulence (low friction velocity) periods and then gap-filled via the R package REddyProc (Wutzler <i>et al</i> 2018). REddyProc was used to partition NEE into GPP and Reco using the method of Lasslop <i>et al</i> (2010), with air temperature used as the driving temperature for Reco.
Partitioning	Vaisala HMP45C probe at 6 m	Vaisala HUMICAP Humidity and Temperature Probe—HMP 155 at 22 m (National Ecological Observatory Network (NEON) 2022b)
Air temperature (T _{air}) and relative humidity (RH)	Type-T (copper-constantan) thermocouples at a 10 cm depth	Thermometrics—Climate RTD 100-ohm Probe at 6 cm depth (National Ecological Observatory Network (NEON) 2022c)
Soil temperature (T _{soil})	Cambell Scientific CS615 Water Content Reflectometers at a 7.5 cm depth	Sentek—EnviroSCAN TriSCAN at 6 cm depth (National Ecological Observatory Network (NEON) 2022d)
Soil water content (SWC)		

proposed approach for predicting GPP based on remote metrics must be effective even in the presence of snow. Because of this, we did not filter for snow but visually identified snow dates using PhenoCam imagery (Richardson *et al* 2018). Filtering for snow using an NDVI threshold >0.5 (Magney *et al* 2019, Cheng *et al* 2020) did not change the results of this study (figures S1–S4).

To include impacts of illumination conditions on our analysis, we determined a measure of direct vs diffuse radiation (Df). Df reflects the deviation of PAR at a given solar zenith angle from the expected PAR during a clear sky reference day so that Df = 1 is clear sky conditions (Pierrat *et al* 2021b). Df values were calculated for every PhotoSpec measurement (~20 s

resolution) and averaged together in 30 min windows to compare with GPP and environmental measurements (table 1). Values where Df > 0.8 are considered clear sky conditions.

Data collection and processing for eddy covariance and meteorological data for both SOBS and DEJU are summarized in table 1.

2.3. Data analysis: random forest and parameterized models

We trained and tested a variety of random forest models (table 2). All random forest models were produced using Matlab’s TreeBagger function (MATLAB 2019) which is based on the random forest algorithm from (Breiman 2001). We used the full data collection

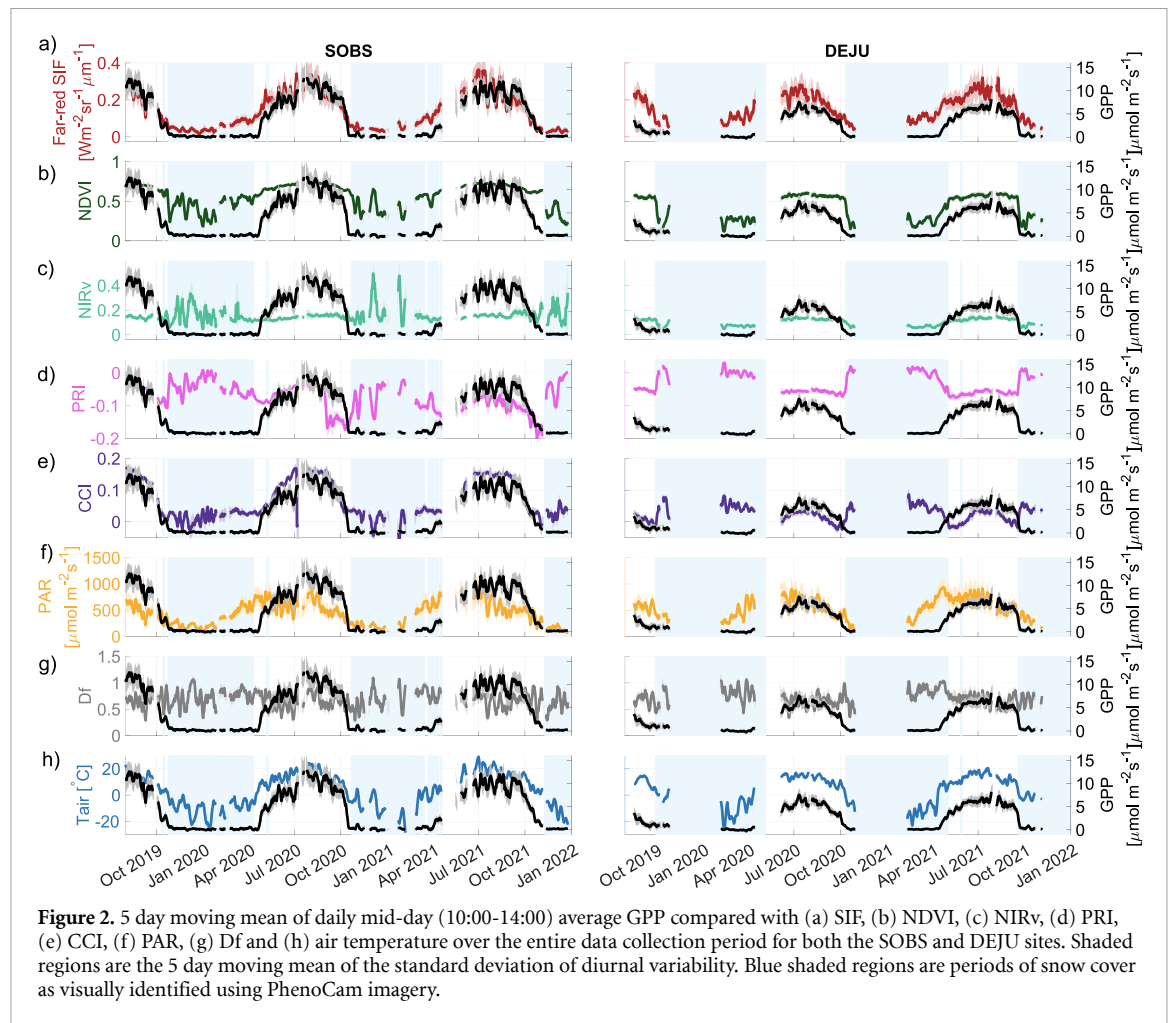
Table 2. Naming conventions and descriptions of the random forest models used in this study. ENV describes models driven with environmental inputs. RS describes models driven with remote sensing and meteorological inputs. Parentheses indicate the same model setup tested with a different variable or location.

Model name	Number of data points	Temporal resolution	Site	Inputs	Output	Purpose
ENV-SIF	12 482 (3969)	Half-hourly	SOBS (DEJU)	PAR, Tair, Tsoil, SWC, RH, Df	SIF	To test the ability of random forest models to reproduce quantitative SIF-GPP dynamics (section 3.2).
ENV-GPP	12 482 (3969)	Half-hourly	SOBS (DEJU)	PAR, Tair, Tsoil, SWC, RH, Df	GPP	To test the ability of random forest models to reproduce quantitative SIF-GPP dynamics (section 3.2).
RS-SOBS	12 482	Half-hourly	SOBS	SIF, CCI, PRI, GPP, NDVI, NIRv, Tair, Df	GPP	To test our ability to improve GPP prediction from SIF by including additional remote sensing metrics (section 3.3).
RS-DEJU	10 433	Half-hourly	DEJU	SIF, CCI, PRI, GPP, NDVI, NIRv, Tair, Df	GPP	To test our ability to improve GPP prediction from SIF by including additional remote sensing metrics (section 3.3).
RS-total	1157 (833 clear sky)	Daily midday (10:00–14:00) average	SOBS and DEJU	SIF, CCI, NDVI, NIRv, Tair, (site)	GPP	To test the applicability of our random forest approach across multiple sites (section 3.4).
RS-DSOBS	622	Daily midday (10:00–14:00) average	SOBS	SIF, CCI, NDVI, NIRv, Tair	GPP	To compare our random forest approach with existing models for GPP based on SIF (section 3.4).
RS-DDEJU	535	Daily midday (10:00–14:00) average	SOBS	SIF, CCI, NDVI, NIRv, Tair	GPP	To compare our random forest approach with existing models for GPP based on SIF (section 3.4).

window (August 2019–December 2021, figure 2) to train all models. All models were created using 100 regression trees and sampled with replacement on an in bag fraction of 0.7 (i.e. 70% of the data were randomly chosen to train the models saving 30% to test the models). Out-of-bag (OOB) predictor importance estimates were determined using the permuted predictor delta error following the standard Classification And Regression Tree (CART) algorithm (Breiman 2001) using the remaining 30% of data. Model performance was evaluated by calculating the Pearson's correlation coefficient (Dickinson Gibbons

et al 1985) between measurements and predictions on the full dataset (R^2) and the reserved test dataset (OOB R^2 score). Specific models are identified in section 3 with the naming conventions in table 2.

We compared random forest models with common parameterized light-use-efficiency models for the relationships among SIF, GPP, and PAR (equations (1)–(3)). All curve fitting and goodness of fit statistics (R^2 values) were done using Matlab's fit function (MATLAB 2019). Specifics of fitted equations are provided in section 3 and figure captions where relevant.



3. Results and discussion

3.1. Trends among SIF, VIs, and GPP

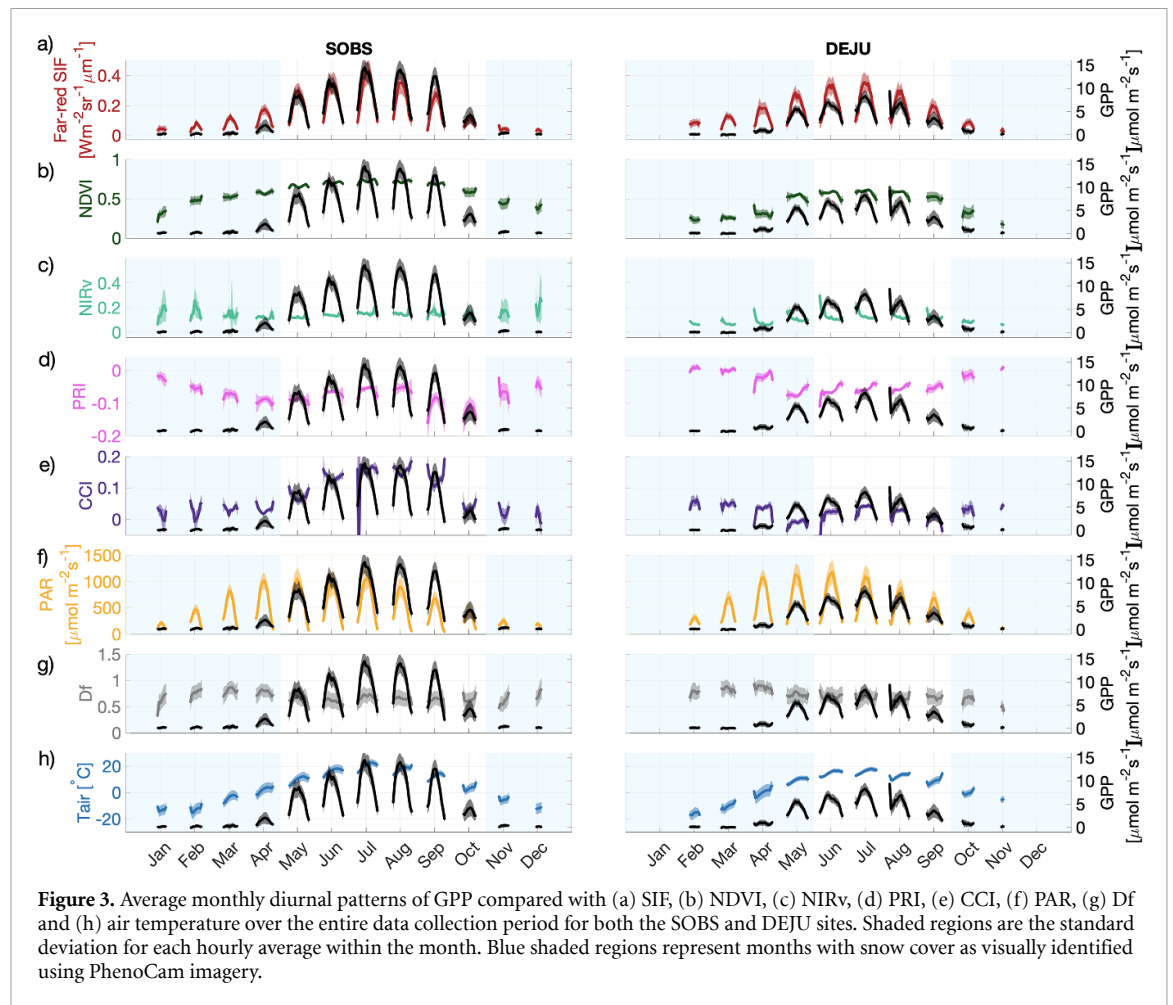
We present the relationships and trends among SIF, VIs, and environmental parameters across seasonal and diurnal scales for both field locations in figures 2 and 3. Compared with SOBS, DEJU generally exhibits a shorter growing season, a smaller summer maximum GPP, and lower values for all remotely sensed metrics (figures 2 and 3).

SIF tracks the seasonal cycle and daily-weekly variability in GPP (figure 2). Monthly diurnal profiles of SIF show good agreement with monthly diurnal profiles of GPP across both sites (figure 3 SIF increases in spring prior to changes in GPP (figure 2) and shows a small diurnal profile over winter at both sites, while GPP does not (figure 3). This early spring increase has been reported in other evergreen locations (Magney *et al* 2019, Pierrat *et al* 2022, Yang *et al* 2022) and is attributed to persistent photosystem II activity (Porcar-Castell 2011, Bowling *et al* 2018). This leads to a winter light response of SIF which means SIF increases coincident with PAR in winter/early spring in a way that does not reflect changes in GPP. This does not preclude the use of SIF as a

proxy for GPP, but it must be accounted for to prevent overestimation of GPP in winter (Pierrat *et al* 2021a, 2022).

Structurally sensitive indices (NDVI and NIRv) show little variability over summer and are sensitive to snow in winter across both sites (figure 2). Over summer at SOBS, NDVI and NIRv show greater changes than at DEJU, which reflects the fact that SOBS is a mixed-species forest with scattered deciduous larch trees. SOBS therefore experiences greater changes in canopy structure over the seasonal cycle than evergreen DEJU. NDVI peaks in mid-morning and mid-afternoon over the summer months at SOBS and has no clear pattern at DEJU (figure 3). NIRv is slightly higher in the morning at DEJU and has no clear pattern at SOBS (figure 3). These patterns suggest that NDVI and NIRv are useful for accounting for shifts in viewing geometry and illumination effects across the diurnal cycle.

Physiologically sensitive indices (CCI and PRI) generally track the seasonal cycle of GPP but show sensitivity to snow cover in winter. CCI is more sensitive to snow at DEJU due to a higher fraction of snow cover on the canopy. Both CCI and PRI increase in spring following GPP at both sites (figure 2). This is



consistent with a two-phased spring transition beginning with the onset of photosynthesis, followed by a reduction in sustained photoprotection (Pierrat *et al* 2021a). CCI at DEJU is lower than at SOBS which may indicate differences in chlorophyll:carotenoid ratios between the sites. Diurnally, PRI shows a small decrease around midday during summer months at both sites (figure 3). This highlights the sensitivity of PRI to rapidly reversible photoprotection which will occur under high-light, high-stress conditions. CCI shows a slight diurnal pattern at SOBS and does not show a consistent diurnal pattern at DEJU. The lack of a diurnal pattern at DEJU supports the idea that CCI is sensitive to sustained cold-season photoprotection which will not change over the course of the diurnal cycle. PRI and CCI thus provide information on plant heat dissipation processes at both seasonal and diurnal timescales.

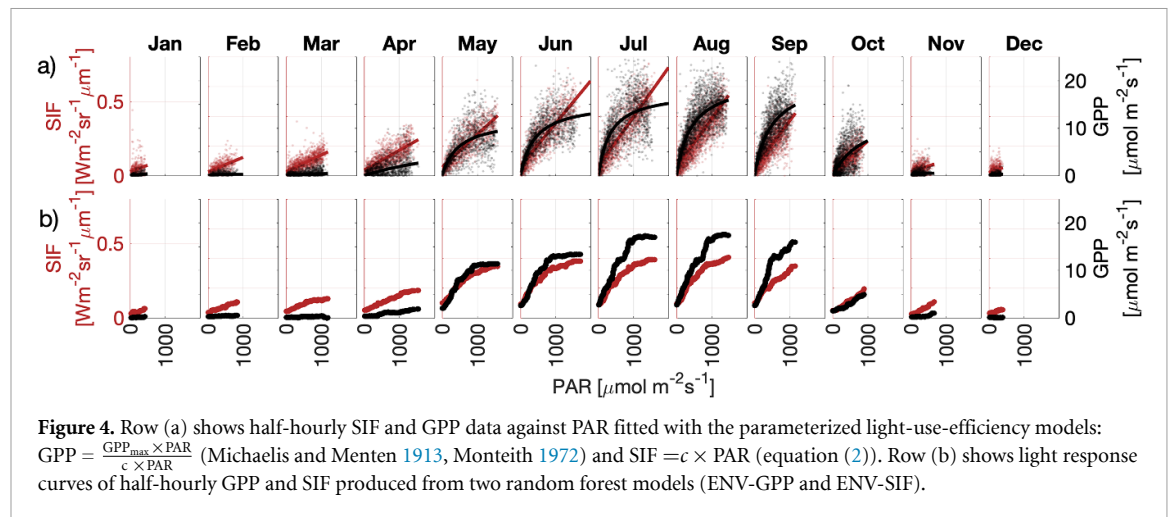
Environmental conditions at the two sites (PAR, Df, and air temperature) are consistent with expected environmental patterns. PAR increases prior to an increase in GPP and begins to decrease in fall prior to a decline in GPP at both sites. Df does not show a defined seasonal cycle, nor does it show a prominent diurnal pattern at either site. Thus, diffuse sky conditions are largely independent of season or time

of day. Air temperature tracks the seasonal cycle of GPP, and air temperatures above 0 °C are a good indicator of growing season length (Parazoo *et al* 2018, Pierrat *et al* 2021a, Stettz *et al* 2022). Air temperature peaks in the afternoon following the peak in SIF and GPP.

3.2. SIF and GPP light response curves: light-use-efficiency and random forest models

We tested the ability of random forest models to reproduce quantitative SIF-GPP dynamics by comparing light response curves following the parameterized light-use-efficiency model (figure 4 row (a)) with light response curves produced by environmentally driven random forest models (table 2, ENV-SIF and ENV-GPP, figure 4 row (b)), at a half-hourly resolution at the SOBS site. The same analysis tested at the DEJU site shows consistent results (figures S5 and S6).

The half-hourly data and parameterized light-use-efficiency model ($GPP = \frac{GPP_{max} \times PAR}{c \times PAR}$, (Michaelis and Menten 1913, Monteith 1972) and ($SIF = c \times PAR$, equation (2) show a curved light response for GPP, consistent with the light saturation of GPP, and a near linear light response for SIF (figure 4 row (a)). Both GPP and SIF show a seasonally dynamic light



response with minima over winter and maxima over summer. The light response of SIF approaches but does not go to zero over the winter months due to persistent winter photosystem II activity (Porcar-Castell 2011, Bowling *et al* 2018, Yang *et al* 2022).

Both ENV-SIF and ENV-GPP models had strong performance with high R^2 ($R^2 = 0.93$ and 0.94 respectively) and OOB R^2 scores (OOB $R^2 = 0.86$ and 0.89 respectively) (figure S7). This highlights the ability of random forest models to capture the environmental dependencies of both SIF and GPP. The two trained models were run for each month using the range of PAR values for that month and the monthly average for the rest of the environmental predictor variables to reproduce the light response curves of SIF and GPP (figure 4 row (b)). The monthly light response curves produced by the random forest models have the same patterns as both the data and the parameterized light-use-efficiency models. The light response of GPP is curved and the light response of SIF is largely linear. SIF shows a slight curvature at high PAR values which we attribute to changes in f_{esc} at high PAR (Pierrat *et al* 2022). Both GPP and SIF exhibit a monthly variable light response and the light response of SIF does not drop to zero over winter, consistent with the light-use-efficiency model.

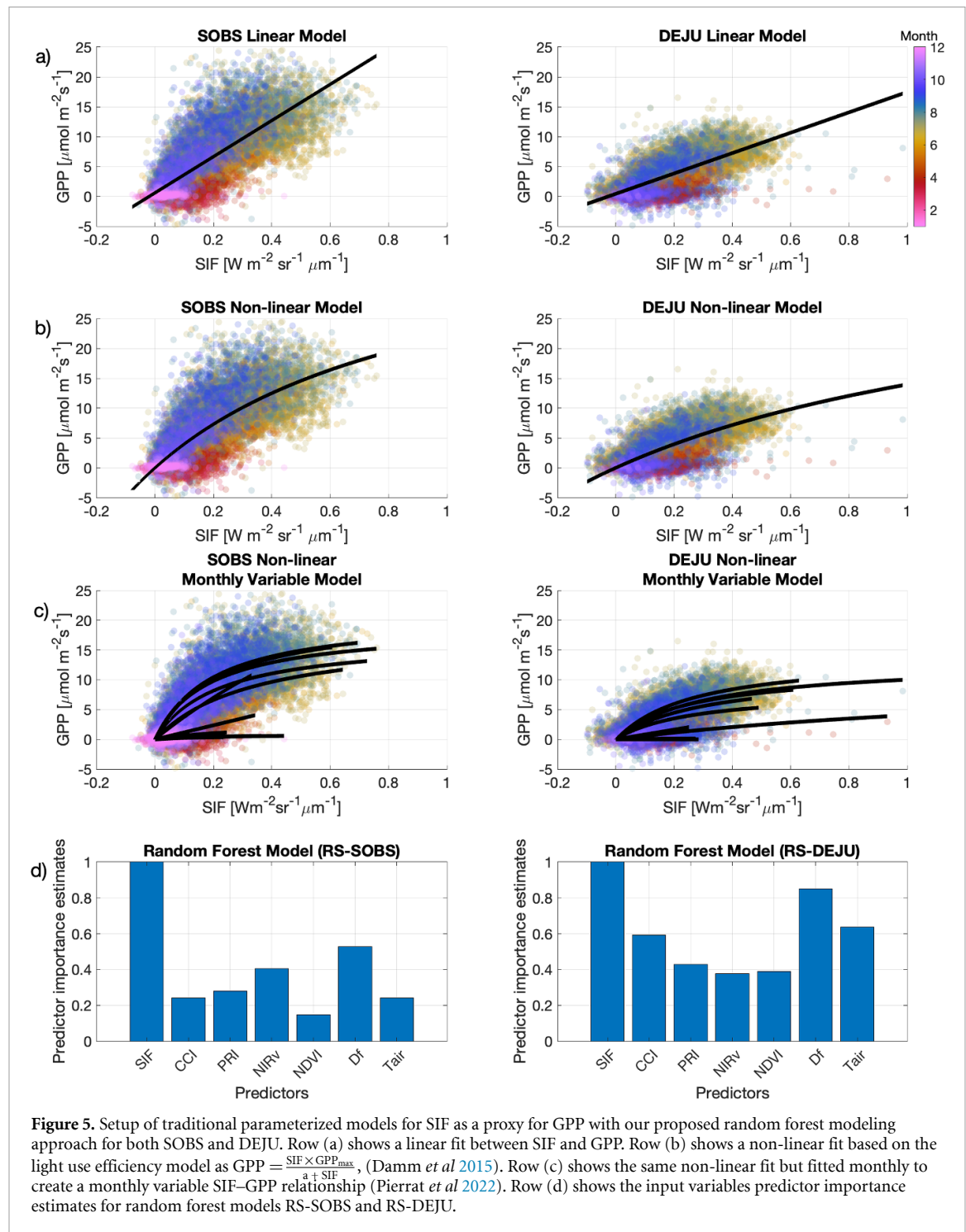
These results highlight the efficacy of random forest models for reproducing quantitative relationships and environmental dependencies for GPP and SIF without prescribing a parameterized model onto the data (Chen *et al* 2021). Because of this, and the additional information that can be provided by additional remote sensing metrics (section 3.1), we justify the use of random forest models for GPP prediction.

3.3. Random forest models compared with parameterized models for predicting GPP

We tested the ability of random forest models to improve GPP prediction by testing traditional

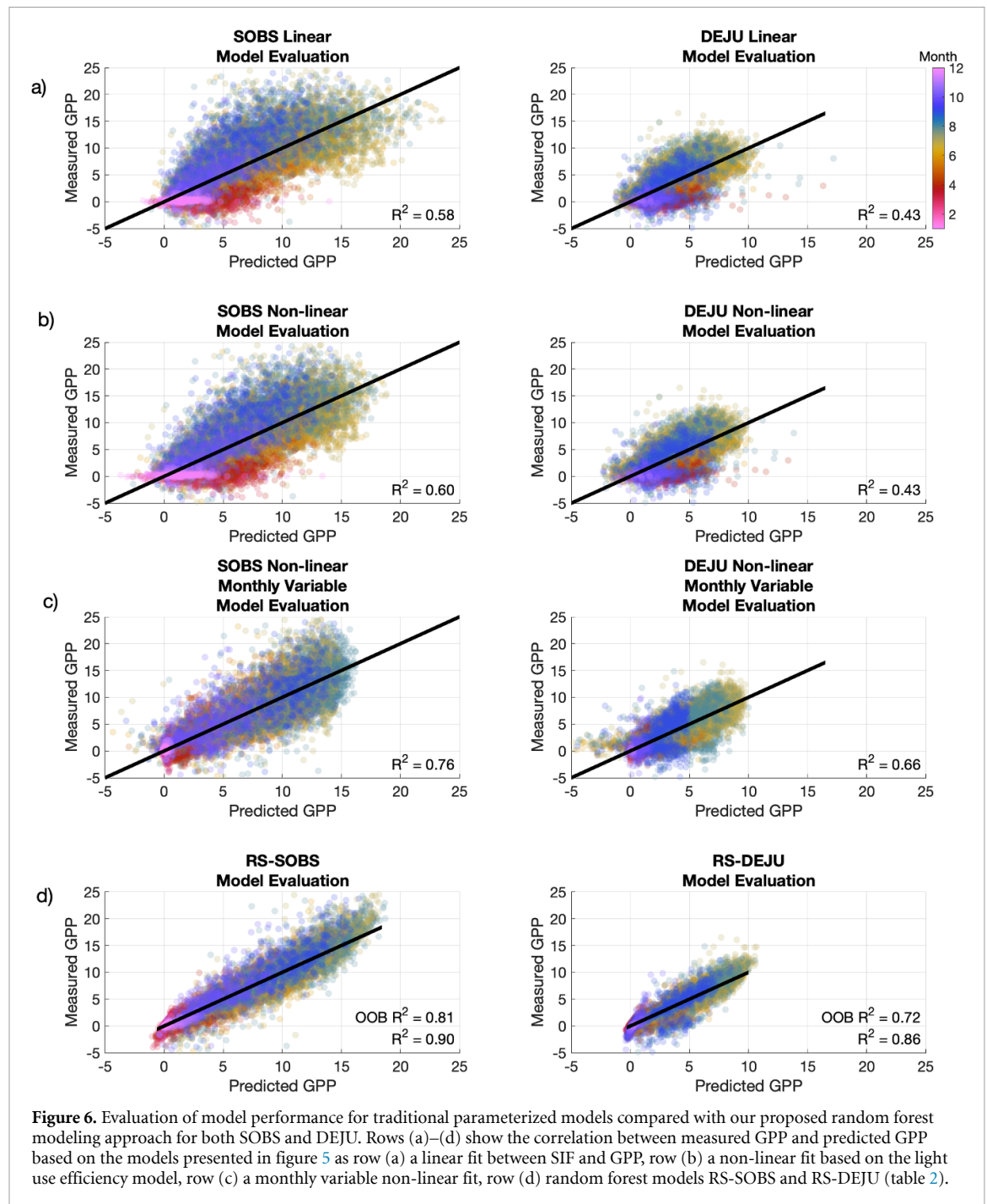
parameterized models (a linear SIF–GPP relationship, a non-linear SIF–GPP relationship that takes into account the light saturation of GPP (Monteith 1972) and a monthly variable non-linear SIF–GPP relationship (Damm *et al* 2015, Pierrat *et al* 2022) against our new random forest approach combining multiple remote sensing indices for predicting GPP (figures 5 and 6). The two random forest models used to test this (table 2, RS-SOBS and RS-DEJU) were driven by remote sensing and meteorological variables that reflect the physical and physiological controls on photosynthesis relevant at a half-hourly temporal resolution.

The linear fit between SIF and GPP (figure 5 row (a)) performs moderately-poorly with an $R^2 = 0.58$ at SOBS and $R^2 = 0.43$ at DEJU (figure 6 row (a)). The non-linear fit based on (Damm *et al* 2015) shows little to no improvement from the linear fit with $R^2 = 0.60$ at SOBS and $R^2 = 0.43$ at DEJU (figure 5 row (b)). Both models with a fixed SIF–GPP relationship overestimate GPP in winter due to a persistent winter light response of SIF. The monthly variable non-linear SIF–GPP relationship (figure 5 row (c)) shows a marked improvement in the predictability of GPP at both the SOBS and DEJU sites with $R^2 = 0.76$ and $R^2 = 0.66$ respectively (figure 6 row (c)). In addition to improved R^2 values, the monthly variable non-linear SIF–GPP relationship helps account for the persistent winter light response and the winter overestimation of GPP is no longer observed. The random forest models driven by remotely sensed products (figure 5 row (d)) improved the predictability of GPP compared to all parameterized models at both sites with $R^2 = 0.90$ and 0.86 and OOB $R^2 = 0.81$ and 0.73 at SOBS and DEJU respectively (figure 5 row (d)). The overestimation of GPP in winter is not observed using the random forest approach and residuals between predicted and measured GPP are more homoscedastic across seasons.



The predictive power of input variables is evaluated with the predictor importance estimates (figure 5 row (d)). SIF was the most important predictor at SOBS and second most important predictor at DEJU, highlighting the power of SIF for GPP prediction. DEJU showed a higher dependence on air temperature than SOBS which could point to stronger temperature limitations at this site. Df was an important predictor across both sites which highlights the dependence of GPP on diffuse vs. direct radiative conditions (Durand *et al* 2021, Pierrat

et al 2021b). NIRv was the third most important predictor at SOBS which could reflect the fact that SOBS is a mixed-species forest interspersed with deciduous trees and will thus have more dramatic changes in canopy structure over the course of the season than DEJU. CCI was moderately important at both SOBS and DEJU which highlights the relevance of CCI for capturing changes in sustained photoprotection over winter. NDVI was not a particularly important predictor for either SOBS or DEJU which could reflect the fact that NIRv is more effective at



capturing changes in canopy structure (Badgley *et al* 2017, 2019) than NDVI. PRI was also not particularly important for either SOBS or DEJU which suggests that rapidly reversible non-photochemical quenching dynamics may be more effectively captured in the SIF signal when a prescribed relationship between SIF and GPP is not used. Including PAR as an input variable does not improve model results but does decrease the predictor importance of SIF (figures S8 and S9). Substituting PAR for SIF as a predictor variable also does not change model performance (figures S10 and S11). This suggests that VIs effectively capture the structural ($fPAR$) and physiological (LUE_p) factors relevant for predicting GPP.

3.4. Random forest models for predicting GPP across boreal forest sites

Satellite remote sensing enables the approximation of GPP over a broader spatial range than tower-based measurements, making it essential for understanding regional GPP dynamics. However, relationships between remotely sensed products and GPP are often site specific and thus require a new model or set of parameters for each ecosystem or plant functional type. We tested the potential of our random forest modeling approach across sites for a ‘universal’ model for GPP based on remotely sensed products. We trained a model (RS-total) to predict GPP from data that are readily accessible or can be inferred

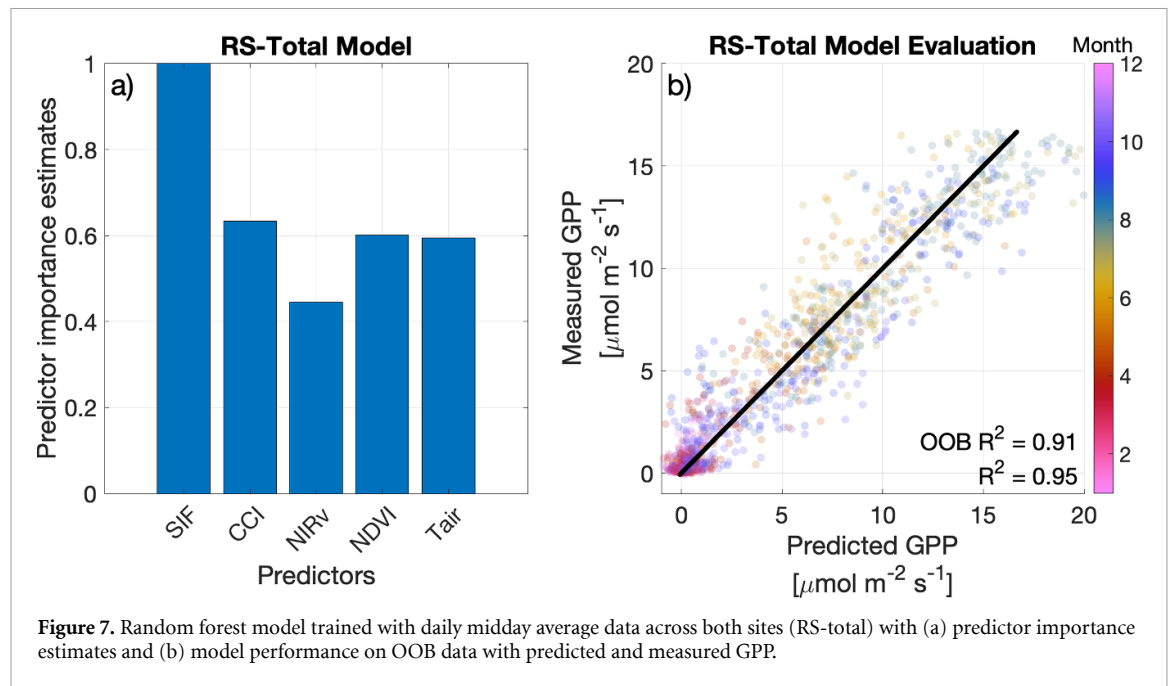


Figure 7. Random forest model trained with daily midday average data across both sites (RS-total) with (a) predictor importance estimates and (b) model performance on OOB data with predicted and measured GPP.

from satellite measurements (table 2, RS-total). We used data at a daily resolution averaged between 10:00 and 14:00 under all (figure 7) and only sunny (figure S12) sky conditions to replicate satellite observations across both the SOBS and DEJU sites together. SOBS and DEJU are at the latitudinal extremes of the boreal ecosystem, thus, testing the feasibility of interpolation across the boreal region.

Our results show excellent performance on the predictability of GPP based on remotely sensed metrics ($R^2 = 0.94$ and OOB $R^2 = 0.89$) (figure 7). Residuals between predicted and measured GPP are highly homoscedastic and GPP is not overestimated in winter. Predictor importance estimates show that SIF is the most valuable predictor for GPP, but air temperature, NDVI, CCI, and NIRv are all relevant predictors. To test the universality of this approach, we tested the same model but included a site flag as a predictor (either SOBS or DEJU) (figure S13). This model showed no improvement in performance ($R^2 = 0.94$ and OOB $R^2 = 0.89$) and had site flag as the least relevant predictor. This may be because importance estimates are biased towards predictors with many classes or different values (Loh and Shih 1997) and we only had two flags for the two sites. Alternatively, this, as well as the success of the model without the site flag, points to our random forest modeling approach being independent of site, and therefore potentially generalizable across the boreal biome at the satellite level (Sun et al 2017, Li et al 2018). This approach also works whether or not the data have been filtered for snow contamination (figure S4) and shows near identical results when only clear sky days are used (figure S12) which are both advantageous in the boreal. The combined random forest modeling approach also outperformed all other parameterized

models at a daily midday resolution for the two sites separately (figures S14 and S15) and worked well for the combined sites at a half-hourly resolution (figure S16). Substituting PAR for SIF also showed good model performance and similar predictor importance estimates (figure S17). This suggests that VIs are effective at capturing the physical and physiological effects relevant for predicting GPP. Our results support the use of this approach for improving the predictability of GPP from remote sensing observations because it can account for physical and physiological mechanisms impacting remotely sensed signals.

4. Conclusions

In this paper, we present a quantitative framework for predicting GPP using random forest models driven by a set of remotely sensed products and air temperature. Multiple years of tower-based remote sensing and GPP data across two field locations at the northern and southern ends of the North American boreal forest show that VIs and SIF contain valuable information on the physical and physiological drivers of GPP. Additionally, random forest models driven by environmental variables are able to reproduce light response curves of SIF and GPP and are thus able to capture quantitative relationships of plant physiology. These results justify the use of random forest models to predict GPP based on a set of random forest parameters. Random forest models outperform traditional parameterized models based on SIF alone for predicting GPP because they are able to incorporate physical and physiological information provided by additional remote sensing metrics without prescribing a parameterized model. This approach is not site specific and therefore has the potential to be

scaled across the boreal domain using satellite measurements. Finally, this approach has potential for use in other ecosystems where remote sensing data is available. Random forest models improve the utility of SIF and vegetation index data for scaling and understanding GPP across space and time and thus present an exciting opportunity to better understand vegetation's role in the global carbon cycle.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: [10.5281/zenodo.7231157](https://doi.org/10.5281/zenodo.7231157).

The meteorological, soil and eddy-covariance measurements at SOBS were made with support from the Global Institute for Water Security, University of Saskatchewan.

Data from DEJU were obtained from The National Ecological Observatory Network. The National Ecological Observatory Network is a program sponsored by the National Science Foundation and operated under cooperative agreement by Battelle. This material is based in part upon work supported by the National Science Foundation through the NEON Program. Some data used in this research were obtained through the NEON Assignable Assets program.

Acknowledgments

We would like to thank the reviewers and editor for their thoughtful comments which improved the quality of our manuscript. We would also like to acknowledge the support provided by NEON staff for their help in the collection of data used in this study.

Funding

This work was supported by NASA's Earth Science Division IDS (Awards 80NSSC17K0108 at UCLA, 80NSSC17K0110 at JPL) and ABoVE programs (Award 80NSSC19M0130). A portion of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. This material is also based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650604 and DGE-2034835. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. DRB and TSM were funded during this time by the Macrosystems Biology and NEON-Enabled Science program at NSF (Award 1926090). We acknowledge funding by the Canadian Space Agency and Natural Sciences and Engineering Research Council of Canada (NSERC).

ORCID iDs

Zoe Amie Pierrat  <https://orcid.org/0000-0002-6726-2406>

Jacob Bortnik  <https://orcid.org/0000-0001-8811-8836>

Bruce Johnson  <https://orcid.org/0000-0001-9013-5774>

Alan Barr  <https://orcid.org/0000-0003-2717-218X>

Troy Magney  <https://orcid.org/0000-0002-9033-0024>

David R Bowling  <https://orcid.org/0000-0002-3864-4042>

Nicholas Parazoo  <https://orcid.org/0000-0002-4424-7780>

Christian Frankenberg  <https://orcid.org/0000-0002-0546-5857>

Jochen Stutz  <https://orcid.org/0000-0001-6368-7629>

References

- Adams W W, Zarter C R, Ebbert V and Demmig-Adams B 2004 Photoprotective strategies of overwintering evergreens *BioScience* **54** 41–49
- Alton P B, North P R and Los S O 2007 The impact of diffuse sunlight on canopy light-use efficiency, gross photosynthetic product and net ecosystem exchange in three forest biomes *Glob. Change Biol.* **13** 776–87
- Anav A et al 2015 Spatiotemporal patterns of terrestrial gross primary production: A review *Rev. Geophys.* **53** 785–818
- Badgley G, Anderegg L D L, Berry J A and Field C B 2019 Terrestrial gross primary production: using NIRV to scale from site to globe *Glob. Change Biol.* **25** 3731–40
- Badgley G, Field C B and Berry J A 2017 Canopy near-infrared reflectance and terrestrial photosynthesis *Sci. Adv.* **3** e1602244
- Bai J, Zhang H, Sun R, Li X, Xiao J and Wang Y 2022 Estimation of global GPP from GOME-2 and OCO-2 SIF by considering the dynamic variations of GPP-SIF relationship *Agric. For. Meteorol.* **326** 109180
- Barr A G, Black T A, Hogg E H, Kljun N, Morgenstern K and Nesic Z 2004 Inter-annual variability in the leaf area index of a boreal aspen-hazelnut forest in relation to net ecosystem production *Agric. For. Meteorol.* **126** 237–55
- Barr A G, Morgenstern K, Black T A, McCaughey J H and Nesic Z 2006 Surface energy balance closure by the eddy-covariance method above three boreal forest stands and implications for the measurement of the CO₂ flux *Agric. For. Meteorol.* **140** 322–37
- Bonan G B 2008 Forests and climate change: forcings, feedbacks, and the climate benefits of forests *Science* **320** 1444–9
- Bowling D R, Logan B A, Hufkens K, Aubrecht D M, Richardson A D, Burns S P, Anderegg W R L, Blanken P D and Eiriksson D P 2018 Limitations to winter and spring photosynthesis of a Rocky Mountain subalpine forest *Agric. For. Meteorol.* **252** 241–55
- Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
- Chang C Y, Guanter L, Frankenberg C, Köhler P, Gu L, Magney T S, Grossmann K and Sun Y 2020 Systematic assessment of retrieval methods for canopy far-red solar-induced chlorophyll fluorescence using high-frequency automated field spectroscopy *J. Geophys. Res.* **125** e2019JG005533
- Chen A, Mao J, Ricciuto D, Xiao J, Frankenberg C, Li X, Thornton P E, Gu L and Knapp A K 2021 Moisture availability mediates the relationship between terrestrial

- gross primary production and solar-induced chlorophyll fluorescence: insights from global-scale variations *Glob. Change Biol.* **27** 1144–56
- Chen J M, Govind A, Sonnentag O, Zhang Y, Barr A and Amiro B 2006 Leaf area index measurements at Fluxnet-Canada forest sites *Agric. For. Meteorol.* **140** 257–68
- Cheng R et al 2020 Decomposing reflectance spectra to track gross primary production in a subalpine evergreen forest *Biogeosciences* **17** 4523–44
- Damm A, Guanter L, Paul-Limoges E, van der Tol C, Hueni A, Buchmann N, Eugster W, Ammann C and 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 Sens. Environ.* **166** 91–105
- Dechant B et al 2020 Canopy structure explains the relationship between photosynthesis and sun-induced chlorophyll fluorescence in crops *Remote Sens. Environ.* **241** 111733
- Demmig-Adams B and Adams W W 2006 Photoprotection in an ecological context: the remarkable complexity of thermal energy dissipation *New Phytol.* **172** 11–21
- Demmig-Adams B, Ebbert V, Zarter C R and Adams W W 2008 Characteristics and species-dependent employment of flexible versus sustained thermal dissipation and photoinhibition (*Advances in Photosynthesis and Respiration vol 21*) pp 39–48
- Dickinson Gibbons J and Chakraborti S 1985 *Nonparametric Statistical Inference* 4th edn (New York: Marcel Dekker, Inc.)
- Durand M, Murchie E H, Lindfors A V, Urban O, Aphalo P J and Robson T M 2021 Diffuse solar radiation and canopy photosynthesis in a changing environment *Agric. For. Meteorol.* **311** 108684
- Frankenberg C et al 2011 New global observations of the terrestrial carbon cycle from GOSAT: patterns of plant fluorescence with gross primary productivity *Geophys. Res. Lett.* **38** 1–6
- Friedlingstein P, Meinshausen M, Arora V K, Jones C D, Anav A, Liddicoat S K and Knutti R 2014 Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks *J. Clim.* **27** 511–26
- Gamon J A, Huemmrich K F, Stone R S and Tweedie C E 2013 Spatial and temporal variation in primary productivity (NDVI) of coastal Alaskan tundra: decreased vegetation growth following earlier snowmelt *Remote Sens. Environ.* **129** 144–53
- Gamon J A, Huemmrich K F, Wong C Y S, Ensminger I, Garrity S, Hollinger D Y, Noormets A and Peñuelas J 2016 A remotely sensed pigment index reveals photosynthetic phenology in evergreen conifers *Proc. Natl Acad. Sci. USA* **113** 13087–92
- Gamon J A, Serrano L and Surfus J S 1997 The photochemical reflectance index: an optical indicator of photosynthetic radiation use efficiency across species, functional types, and nutrient levels *Oecologia* **112** 492–501
- Garbulsky M F, Peñuelas J, Papale D, Ardö J, Goulden M L, Kiely G, Richardson A D, Rotenberg E, Veenendaal E M and Filella I 2010 Patterns and controls of the variability of radiation use efficiency and primary productivity across terrestrial ecosystems *Glob. Ecol. Biogeogr.* **19** 253–67
- Grossmann K, Frankenberg C, Magney T S, Hurlock S C, Seibt U and Stutz J 2018 PhotoSpec: a new instrument to measure spatially distributed red and far-red solar-induced chlorophyll fluorescence *Remote Sens. Environ.* **216** 311–27
- Gu L, Baldocchi D, Verma S B, Black T A, Vesala T, Falge E M and Dowty P R 2002 Advantages of diffuse radiation for terrestrial ecosystem productivity *J. Geophys. Res.* **107** ACL 2-1–23
- Hikosaka K and Tsujimoto K 2021 Linking remote sensing parameters to CO₂ assimilation rates at a leaf scale *J. Plant Res.* **134** 695–711
- Jarvis P G, Massheder J M, Hale S E, Moncrieff J B, Rayment M and Scott S L 1997 Seasonal variation of carbon dioxide, water vapor, and energy exchanges of a boreal black spruce forest *J. Geophys. Res.* **102** 28953–66
- Jiao W, Chang Q and Wang L 2019 The sensitivity of satellite solar-induced chlorophyll fluorescence to meteorological drought *Earth's Future* **7** 558–73
- Lasslop G, Reichstein M, Papale D, Richardson A, Arneth A, Barr A, Stoy P and Wohlfahrt G 2010 Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation *Glob. Change Biol.* **16** 187–208
- Li X, Xiao J and He B 2018 Chlorophyll fluorescence observed by OCO-2 is strongly related to gross primary productivity estimated from flux towers in temperate forests *Remote Sens. Environ.* **204** 659–71
- Liu X, Liu L, Hu J, Guo J and Du S 2020 Improving the potential of red SIF for estimating GPP by downscaling from the canopy level to the photosystem level *Agric. For. Meteorol.* **281** 107846
- Loh W-Y and Shih Y-S 1997 Split selection methods for classification trees *Stat. Sin.* **7** 815–40
- Lu X, Liu Z, Zhao F and 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
- Magney T S et al 2019 Mechanistic evidence for tracking the seasonality of photosynthesis with solar-induced fluorescence *Proc. Natl Acad. Sci. USA* **116** 11640–5
- Magney T S, Barnes M L and Yang X 2020 On the covariation of chlorophyll fluorescence and photosynthesis across scales *Geophys. Res. Lett.* **47** e2020GL091098
- Maguire A J et al 2020 On the functional relationship between fluorescence and photochemical yields in complex evergreen needleleaf canopies *Geophys. Res. Lett.* **47** e2020GL087858
- Marrs J K, Reblin J S, Logan B A, Allen D W, Reinmann A B, Bombard D M, Tabachnik D and Hutrya L R 2020 Solar-induced fluorescence does not track photosynthetic carbon assimilation following induced stomatal closure *Geophys. Res. Lett.* **47** e2020GL087956
- MATLAB (2019) 9 7.0.1190202 (Natick, MA: The MathWorks Inc.)
- Michaelis L and Menten M L 1913 Kinetik der Invertinwirkung *Biochem. Z* **49** 352
- Mohammed G H et al 2019 Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress *Remote Sens. Environ.* **231** 111177
- Monteith J L 1972 Solar radiation and productivity in tropical ecosystems *J. Appl. Ecol.* **9** 747
- National Ecological Observatory Network (NEON) 2022a Bundled data products—eddy covariance (DP4.00200.001) (<https://doi.org/10.48443/7CQP-3J73>)
- National Ecological Observatory Network (NEON) 2022b Relative humidity (DP1.00098.001) (<https://doi.org/10.48443/W9NF-K476>)
- National Ecological Observatory Network (NEON) 2022c Soil temperature (DP1.00041.001) (<https://doi.org/10.48443/9E56-PJ39>)
- National Ecological Observatory Network (NEON) 2022d Soil water content and water salinity (DP1.00094.001) (<https://doi.org/10.48443/GHRY-QW46>)
- Nelson P R et al 2022 Remote sensing of tundra ecosystems using high spectral resolution reflectance: opportunities and challenges *J. Geophys. Res.* **127** e2021JG006697
- Pappas C et al 2020 Aboveground tree growth is a minor and decoupled fraction of boreal forest carbon input *Agric. For. Meteorol.* **290** 108030
- Parazoo N C et al 2018 Spring photosynthetic onset and net CO₂ uptake in Alaska triggered by landscape thawing *Glob. Change Biol.* **24** 3416–35
- Pierrat Z et al 2021a Tower-based remote sensing reveals mechanisms behind a two-phased spring transition in a mixed-species boreal forest *J. Geophys. Res.* **126** 1–20
- Pierrat Z et al 2022 Diurnal and seasonal dynamics of solar-induced chlorophyll fluorescence, vegetation indices, and gross primary productivity in the boreal forest *J. Geophys. Res.* **127** e2021JG006588

- Pierrat Z, Norton A J, Baskin Monk L, Parazoo N, Maguire A, Grossmann K, Magney T, Barr A and Stutz J 2021b Radiative transfer and viewing geometry considerations for remote sensing as a proxy for carbon uptake in boreal ecosystems (<https://doi.org/10.1002/essoar.10509278.1>)
- Porcar-Castell A 2011 A high-resolution portrait of the annual dynamics of photochemical and non-photochemical quenching in needles of *Pinus sylvestris* *Physiol. Plant* **143** 139–53
- Porcar-Castell A *et al* 2021 Chlorophyll a fluorescence illuminates a path connecting plant molecular biology to Earth-system science *Nat. Plants* **7** 998–1009
- Porcar-Castell A, Tyystjärvi E, Atherton J, van der Tol C, Flexas J, Pfündel E E, Moreno J, Frankenberg C and Berry J A 2014 Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: mechanisms and challenges *J. Exp. Bot.* **65** 4065–95
- Raczka B *et al* 2019 Sustained nonphotochemical quenching shapes the seasonal pattern of solar-induced fluorescence at a high-elevation evergreen forest *J. Geophys. Res.* **124** 2005–20
- Richardson A D *et al* 2018 Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery *Sci. Data* **5** 180028
- Sims D A *et al* 2006 On the use of MODIS EVI to assess gross primary productivity of North American ecosystems *J. Geophys. Res.* **111** 4
- Starr G and Oberbauer S F 2003 Photosynthesis of arctic evergreens under snow: implications for tundra ecosystem carbon balance *Ecology* **84** 1415–20
- Stettz S G *et al* 2022 Resolving temperature limitation on spring productivity in an evergreen conifer forest using a model-data fusion framework *Biogeosciences* **19** 541–58
- Sun Y *et al* 2017 OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence *Science* **358** eaam5747
- Sun Y, Frankenberg C, Jung M, Joiner J, Guanter L, Köhler P and 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** 808–23
- Turner M *et al* 2014 Carbon stock and density of northern boreal and temperate forests *Glob. Ecol. Biogeogr.* **23** 297–310
- Tucker C J 1979 Red and photographic infrared linear combinations for monitoring vegetation *Remote Sens. Environ.* **8** 127–50
- van der Tol C, Berry J A, Campbell P K E and Rascher U 2014 Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence *J. Geophys. Res.* **119** 2312–27
- Verhoeven A 2014 Sustained energy dissipation in winter evergreens *New Phytol.* **201** 57–65
- Walter-McNeill A, Garcia M A, Logan B A, Bombard D M, Reblin J S, Lopez S, Southwick C D, Sparrow E L and Bowling D R 2021 Wide variation of winter-induced sustained thermal energy dissipation in conifers: a common-garden study *Oecologia* **197** 589–98
- Wang J A and Friedl M A 2019 The role of land cover change in Arctic-Boreal greening and browning trends *Environ. Res. Lett.* **14** 125007
- Wang X, Chen J M and Ju W 2020 Photochemical reflectance index (PRI) can be used to improve the relationship between gross primary productivity (GPP) and sun-induced chlorophyll fluorescence (SIF) *Remote Sens. Environ.* **246** 111888
- Wong C Y S and Gamon J A 2015a The photochemical reflectance index provides an optical indicator of spring photosynthetic activation in evergreen conifers *New Phytol.* **206** 196–208
- Wong C Y S and Gamon J A 2015b Three causes of variation in the photochemical reflectance index (PRI) in evergreen conifers *New Phytol.* **206** 187–95
- Wong C Y S, Mercado L M, Arain M A and Ensminger I 2022 Remotely sensed carotenoid dynamics improve modelling photosynthetic phenology in conifer and deciduous forests *Agric. For. Meteorol.* **321** 108977
- Wutzler T, Lucas-Moffat A, Migliavacca M, Knauer J, Sickel K, Šigut L, Menzer O and Reichstein M 2018 Basic and extensible post-processing of eddy covariance flux data with REddyProc *Biogeosciences* **15** 5015–30
- Yang H, Yang X, Zhang Y, Heskel M A, Lu X, Munger J W, Sun S and Tang J 2017 Chlorophyll fluorescence tracks seasonal variations of photosynthesis from leaf to canopy in a temperate forest *Glob. Change Biol.* **23** 2874–86
- Yang J C *et al* 2022 Gross primary production (GPP) and red solar induced fluorescence (SIF) respond differently to light and seasonal environmental conditions in a subalpine conifer forest *Agric. For. Meteorol.* **317** 108904
- Zeng Y, Badgley G, Dechant B, Ryu Y, Chen M and Berry J A 2019 A practical approach for estimating the escape ratio of near-infrared solar-induced chlorophyll fluorescence *Remote Sens. Environ.* **232** 111209