Remote Sensing of Environment 225 (2019) 441–457

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Estimating crop primary productivity with Sentinel-2 and Landsat 8 using machine learning methods trained with radiative transfer simulations

Aleksandra Wolanin[a,*], Gustau Camps-Valls[b], Luis Gómez-Chova[b], Gonzalo Mateo-García[b], Christiaan van der Tol[c], Yongguang Zhang[d], Luis Guanter[a]

[a] Section 1.4 Remote Sensing, GFZ German Research Centre for Geosciences, Helmholtz-Centre, Potsdam, Germany
[b] Image Processing Laboratory, Universitat de València, València, Spain
[c] Faculty of ITC, University of Twente, Enschede, the Netherlands
[d] Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, International Institute for Earth System Sciences, Nanjing University, 210023 Nanjing, China

ARTICLE INFO

Keywords:
Gross primary productivity (GPP)
Sentinel-2 (S2)
Landsat 8
Machine learning (ML)
Neural networks (NN)
Radiative transfer modeling (RTM)
Hybrid approach
Soil-canopy-observation of photosynthesis and the energy balance (SCOPE)
C3 crops

ABSTRACT

Satellite remote sensing has been widely used in the last decades for agricultural applications, both for assessing vegetation condition and for subsequent yield prediction. Existing remote sensing-based methods to estimate gross primary productivity (GPP), which is an important variable to indicate crop photosynthetic function and stress, typically rely on empirical or semi-empirical approaches, which tend to over-simplify photosynthetic mechanisms. In this work, we take advantage of all parallel developments in mechanistic photosynthesis modeling and satellite data availability for an advanced monitoring of crop productivity. In particular, we combine process-based modeling with the soil-canopy energy balance radiative transfer model (SCOPE) with Sentinel-2 and Landsat 8 optical remote sensing data and machine learning methods in order to estimate crop GPP. With this approach, we by-pass the need for an intermediate step to retrieve the set of vegetation biophysical parameters needed to accurately model photosynthesis, while still accounting for the complex processes of the original physically-based model. Several implementations of the machine learning models are tested and validated using simulated and flux tower-based GPP data. Our final neural network model is able to estimate GPP at the tested flux tower sites with r^2 of 0.92 and RMSE of 1.38 $gC\,d^{-1}\,m^{-2}$, which outperforms empirical models based on vegetation indices. The first test of applicability of this model to Landsat 8 data showed good results (r^2 of 0.82 and RMSE of 1.97 $gC\,d^{-1}\,m^{-2}$), which suggests that our approach can be further applied to other sensors. Modeling and testing is restricted to C3 crops in this study, but can be extended to C4 crops by producing a new training dataset with SCOPE that accounts for the different photosynthetic pathways. Our model successfully estimates GPP across a variety of C3 crop types and environmental conditions even though it does not use any local information from the corresponding sites. This highlights its potential to map crop productivity from new satellite sensors at a global scale with the help of current Earth observation cloud computing platforms.

1. Introduction

Monitoring spatio-temporal changes in the photosynthetic functioning of agricultural lands is of paramount importance for many societal, environmental and economical challenges within the current scenario of increasing demands of biofuels and food. In particular, the accurate estimation of the gross primary productivity (GPP, amount of carbon fixed by plants through photosynthesis) of agricultural lands is key for monitoring, understanding and forecasting crop's status and potential yields. GPP at various spatio-temporal scales (field, region, the globe) can be applied in order to compare the impact of different management practices (e.g., tillage or crop rotation) and spatio-temporal variations in geographic and meteorological conditions on crop photosynthesis (Baker and Griffis, 2005; Falge et al., 2002; Reeves et al., 2005).

Remote sensing provides consistent and systematic observations of the Earth surface and has therefore remarkably contributed to crop

* Corresponding author.
E-mail addresses: ola@gfz-potsdam.de (A. Wolanin), Gustau.Camps@uv.es (G. Camps-Valls), luis.gomez-chova@uv.es (L. Gómez-Chova), gonzalo.mateo-garcia@uv.es (G. Mateo-García), c.vandertol@utwente.nl (C. van der Tol), yongguang_zhang@nju.edu.cn (Y. Zhang), guanter@gfz-potsdam.de (L. Guanter).

https://doi.org/10.1016/j.rse.2019.03.002
Received 18 December 2018; Received in revised form 19 February 2019; Accepted 3 March 2019
Available online 01 April 2019
0034-4257/ © 2019 Published by Elsevier Inc.

monitoring on large scales. Satellite observations of crops have been applied for crop vegetation monitoring, crop yield forecasting and management decisions optimization by agriculture companies and sectoral organizations (e.g., Mulla, 2013; Pinter et al., 2003; Pulwarty and Sivakumar, 2014; Strachan et al., 2002; Wu et al., 2014). Over the last decade, both quantity and quality (including spectral and spatial resolution) of remote sensing data have been steadily increasing (Belward and Skien, 2015). For example, the Sentinel-2 mission of the European Copernicus program provides observations at a spatial resolution of 10–20 m, at multiple spectral bands in visible to shortwave infrared wavelengths with a 5-day revisit time, a long-term operation commitment and a free and open data policy (Drusch et al., 2012), which constitutes a great improvement as compared to other previous and current missions in terms of agricultural application.

GPP is typically modeled with three different approaches: process-based models (PBMs), semi-empirical light use efficiency (LUE) models (e.g., Zhang et al., 2012), and data-driven statistical models (Jung et al., 2011; Tramontana et al., 2016). PBMs are based on the mechanistic description of photosynthetic biochemical processes, usually as described in the Farquhar's photosynthesis model (Farquhar et al., 1980). GPP is first computed at the leaf level and then scaled-up to the whole canopy. In LUE models, GPP is explicitly decoupled into two terms: the amount of absorbed photosynthetically active radiation (APAR) and the LUE, the latter accounting for the effect of environmental conditions on photosynthesis (Monteith, 1972). Usually biome-specific relationships are established from empirical observations of GPP and APAR (e.g., Running et al., 2004), but Zhang et al. (2018) found that the expression of LUE based on PAR absorption by canopy chlorophyll tends to converge across biome types.

PBMs rely on more rigorous formulations than LUE models (e.g., Zhang et al., 2012), but they have the disadvantage of complexity and uncertainty of their parametrization. Although these input parameters for PBMs are interpreted as being more physical and biologically meaningful, many of them may be unavailable or highly uncertain. On the other hand, the fundamental assumptions underlying LUE models –that plant canopies behave like a big single-leaf, and their LUE is independent of the directional nature of solar radiation and vegetation structure– have been widely questioned already by Pury and Farquhar (1997) and continue to be discussed with support of flux data measurements (Gu et al., 2002; Propastin et al., 2012; Zhang et al., 2011). Furthermore, it is unclear how well these empirical relationships hold for spatial and temporal scales beyond those used to derive them, and how they might change under altering environmental conditions (e.g., Xin et al., 2015). The most widely used LUE model is applied in the Moderate Resolution Imaging Spectroradiometer (MODIS) GPP product, MOD17 (Running et al., 2004), currently available (Collection 6) globally at 8-day and 500 m resolution (Running et al., 2015). Despite general good performance of the model, evaluation of MOD17 for crops showed that it usually underestimates GPP for certain crops, for example soybeans and maize (e.g., Peng and Gitelson, 2012; Turner et al., 2005). This can be partly explained by neglecting the high heterogeneity of different crop types and a coarse spatial resolution, which does not allow separating observations of individual fields, as well as different irrigation and fertilization practices that are important for crop performance (Zhang et al., 2012). The VPM GPPV20 dataset –a more recent global GPP product that utilized MODIS datasets together with a reanalysis climate dataset and a land cover classification– was based on an improved LUE theory that uses the energy absorbed by chlorophyll (Zhang et al., 2017), and it's overall accuracy was relatively high, though it also underestimated cropland GPP (by ~15%).

A third approach to GPP estimation from remote sensing data is based on linking GPP fluxes at flux tower locations with observations of large spatial fields from satellites adopting advanced statistical and machine learning (ML) algorithms that use input variables from climate reanalysis and satellite data products (Jung et al., 2009; Jung et al., 2011; Tramontana et al., 2016; Xiao et al., 2008), such as the Max Planck Institute for Biogeochemistry (MPI-BCG) GPP product (Jung et al., 2011). Such methods are powerful in application, but being essentially a statistical approach, they share with more simple empirical LUE models the disadvantage of lacking the capacity to extrapolate to different conditions (Beer et al., 2010). In addition, the dataset needed to train such ML approaches should be sufficiently representative and cover a wide range of conditions, which is difficult in general and especially at the start of new satellite missions when the collected data is limited.

GPP was also previously estimated at a finer spatial resolution; e.g. Gitelson et al. (2012) assessed crop GPP with the Landsat data (spatial resolution of 30 m). They used the concept of total crop chlorophyll content, based on evaluation of performances of twelve vegetation indices (VIs) for estimating GPP using ground-based measurements (Peng and Gitelson, 2012). However, as these approaches use only simple VIs, the increased number of bands in the Sentinel-2 satellites and ongoing advancements in vegetation and GPP models motivate a more sophisticated application of available reflectance bands and the development of more flexible and powerful GPP algorithms.

In this work, we propose a hybrid approach for GPP estimation (Camps-Valls et al., 2011; Verrelst et al., 2016) based on the combination of process-based radiative transfer models (RTMs) with Sentinel-2 spectral reflectance data through ML algorithms. Rather than retrieving the biophysical parameters accounting for the impact of canopy structure and leaf pigments on the harvest of light, we convert spectral reflectance and meteorological information into GPP directly using statistical ML methods, such as random forests and neural networks. It is important to emphasize that the training is performed on the modeled data, rather than flux tower GPP, which allows us to simulate a broad range of conditions. This, as well as the use of all reflectance data (instead of derived products), makes the study different from purely data driven ML algorithms such as that of Jung et al. (2011) and LUE models like Gitelson et al. (2012). By adapting the same modeled data for different spectral characteristics of various instruments, our approach can remain consistent among multiple past and future satellites and still make the use of all available bands. Furthermore, it also can be applied across the range of spatial dimensions, independently of the footprints of the reference data used for the training of empirical models.

For the RTM, we use the soil-canopy energy balance radiative transfer model SCOPE to simulate the reflectance spectra, the light distribution in the vegetation, and the GPP as a function of the vegetation structure. The SCOPE model incorporates leaf model Fluspect (Vilfan et al., 2016) and canopy RTM 4SAIL (Verhoef et al., 2007), which can also be used for retrieving vegetation variables, e.g., leaf area index (LAI) and chlorophyll-*a* and *b* content (C_{ab}), as well as absorbed fraction of photosynthetically active radiation (fPAR) that is used as input to a number of LUE models (e.g., Weiss and Baret, 2016). We focus our investigation on soybeans and other C3 crops. However, the same approach can be used for C4 crops, after running simulations with appropriate biochemical settings (e.g., photochemical pathway, maximum carboxylation capacity, and temperature response).

This paper is structured as follows: in Section 2 we first introduce the SCOPE model, the satellite and meteorological data, flux tower sites for which our approach was tested, as well as other GPP models. We also introduce the methods applied for the analysis of the simulated data and the ML methods used. In Section 3, we analyze relationships between vegetation parameters relevant for GPP modeling, as well as the relationships between the components of the various LUE models, based on broad SCOPE simulations. We also discuss the performance of machine learning algorithms for different vegetation parameters, and explain how and why we chose to model GPP. Afterwards, we compare the results of our ML models applied to the satellite data with other GPP models and flux tower estimates. Finally, in Section 4, we conclude our findings and give an outlook for future work.

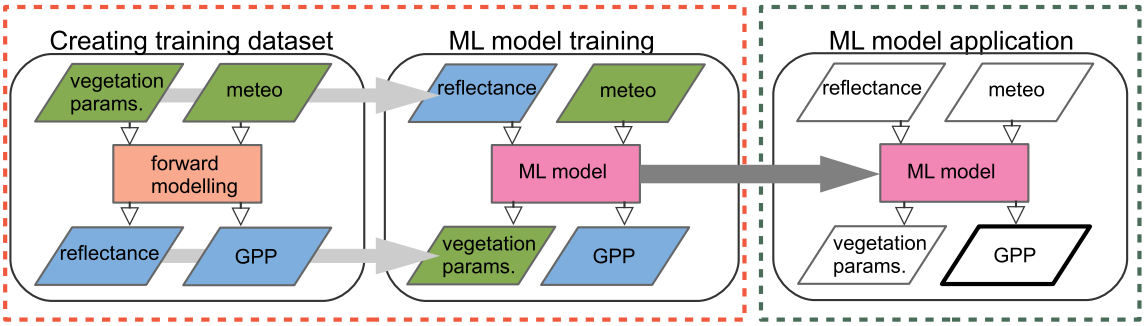

Fig. 1. Flow chart of the processing chain applied in this work. The ML model is trained on the dataset created from SCOPE simulations, afterwards the ML model is applied to the satellite and meteorological data.

2. Materials and methods

The overall process of creating and applying ML models for GPP estimation is schematically shown in Fig. 1. For creating the synthetic dataset we use the SCOPE model (Section 2.1). Afterwards, the model is applied to the reflectance data of Sentinel-2 and Landsat 8 data (Section 2.2 and Section 2.3, respectively), and meteorological dataset GLDAS 2.1 (Section 2.4). Initially, we considered three different workflows to estimate GPP (Fig. 2):

- retrieving vegetation parameters from satellite data, then running the SCOPE model in a forward mode. In this case some vegetation parameters are estimated, while others have to be set a-priori;
- retrieving fPAR from satellite data, then applying LUE model;
- estimate GPP directly from satellite and meteorological data.

To analyze these approaches and finally decide on the most suitable method, we analyzed the synthetic dataset created with SCOPE. We performed global sensitivity analyses (Section 2.7), examined the relationships between fPAR, LUE and GPP, as well as tested ML algorithms for retrieving various parameters (Section 2.8) using the modeled dataset. Eventually, we applied the ML model of GPP directly to satellite and meteorological data. We use data from flux tower sites (Section 2.5) for a feasibility test, and simple GPP models based on vegetation indices (Section 2.6) for comparison with our ML model.

2.1. The SCOPE model

The SCOPE model is a vertical (1-D) integrated radiative transfer and energy balance model (van der Tol et al., 2009). SCOPE calculates radiance spectra in the visible to thermal infrared range (0.4 to 50 μm) as observed above the canopy, as well as the fluxes of water, heat and carbon dioxide. SCOPE is continually updated, and recent significant improvements were introduced to link chlorophyll-*a* vegetation fluorescence to the photosynthesis processes within the framework of the *Photosynthesis Study* for the ESA FLEX mission (Mohammed et al., 2014).

SCOPE integrates radiative transfer and energy balance calculations at the level of individual leaves, as well as at the canopy level. The spectral transmittance and reflectance of the leaves are calculated with the Fluspect model (Vilfan et al., 2016). Radiative transfer within the canopy is based on the 4SAIL model (Verhoef et al., 2007). The leaf biochemical processes are based on Collatz et al. (1991) and Collatz et al. (1992) for C3 and C4 plants, respectively. The geometry of the

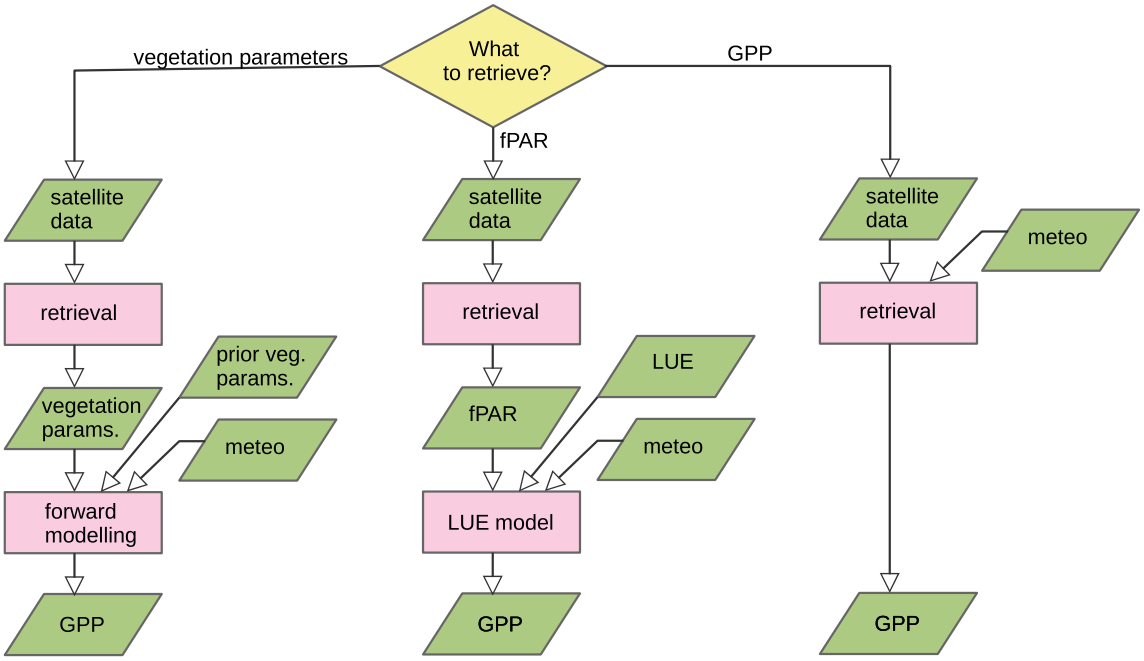

Fig. 2. Workflows of three considered approaches to estimate GPP using data from SCOPE.

Table 1
List of varied input parameters used in SCOPE model simulations. In this study we assumed uniform distribution of the input variables.

Symbol	Parameter	Unit	Min	Max
Leaf optical				
C_{ab}	Chlorophyll-*a* and *b* content	$\mu g\,cm^{-2}$	11	90
C_{ca}	Carotenoid content	$\mu g\,cm^{-2}$	0	40
C_{ant}	Anthocyanins content	$\mu g\,cm^{-2}$	0	40
C_{dm}	Dry matter content	$g\,cm^{-2}$	0.0	0.05
C_w	Leaf water equivalent layer	cm	0.0	0.1
C_s	Senescent material fraction	Fraction	0	0.9
N	Leaf thickness parameter	–	1	2.5
Canopy				
LAI	Leaf area index	$m^2\,m^{-2}$	0	9
h_c	Vegetation height	m	0.1	2
$LIDF_a$	Leaf inclination	–	−1	1
$LIDF_b$	Variation in leaf inclination	–	−1	1
Soil				
SMC	Volumetric soil moisture content in the root zone	–	0.01	0.7
BSMBrightness	BSM model parameter for soil brightness	–	0.01	0.9
BSMlat	BSM model parameter "lat"	–	20	40
BSMlon	BSM model parameter "long"	–	45	65
Geometry				
SZA	Solar zenith angle	Degree	0	85
Meteorology				
R_{in}	Broadband incoming shortwave radiation (0.4–2.5 μm)	$W\,m^{-2}$	0	1400
R_{li}	Broadband incoming longwave radiation (2.5–50 μm)	$W\,m^{-2}$	0	400
T_a	Air temperature	°C	−10	50
p	Air pressure	hPa	500	1030
e_a	Atmospheric vapour pressure	hPa	0	125
u	Wind speed	$m\,s^{-1}$	0	25

vegetation is treated in a stochastic way, where a probability of a leaf viewing in solar direction depends on the canopy parameters, and subsequently the different biophysical processes for sunlit and shaded components are considered. To simulate photosynthesis, SCOPE requires inputs of meteorological forcing, vegetation structure parameters, leaf biophysical parameters, and optical and plant physiological parameters. In the comparison of the simulated GPP (using Landsat data and locally measured weather data) to flux tower measurements, Bayat et al. (2018) found a typical root-mean-square error of 1.7 $\mu mol\,s^{-1}\,m^{-2}$ (for GPP of about 8 $\mu mol\,s^{-1}\,m^{-2}$, so about 20% error), with an r^2 of 0.65 during a drought episode. Their relatively low r^2 was mainly due to the overestimation of GPP during the drought. They also showed that the accuracy can be improved by including thermal information. More details on the SCOPE model can be found in van der Tol et al. (2009).

Here, we used the most recent SCOPE model release (version 1.70), in which optical coefficients used by the leaf model are consistent with the latest PROSPECT-D model (Féret et al., 2017). In addition, a new soil spectral reflectance Brightness-Shape-Moisture model (BSM) (Verhoef et al., 2018), which is based on the Global Soil Vectors of Jiang and Fang, (2012), has been added as an alternative to providing an input soil spectrum. In the BSM model, dry soil spectra are approximated using the soil brightness (B), and "lat" and "long" parameters that define spectral shape effects, while SMC parameter accounts for the soil moisture impact on the dry soil reflectance spectrum (see Table 1). In addition, a biochemical routine has been updated so that the internal CO_2 concentration in the leaf is calculated iteratively.

We adapted the model to work in parallel computing within the Matlab environment, and customized the input and output of the model as follows:

1. Added GPP to the output data, since the default output of the model covers only net canopy photosynthesis (GPP minus leaf dark respiration).
2. Added an option to calculate leaf maximum carboxylation capacity (V_{cmax}) at 25 °C (V_{cmax}^{25}), as a function of chlorophyll concentration C_{ab}, following (Houborg et al., 2013):

$$V_{cmax}^{25} = 2.5294C_{ab} - 27.34, \quad (1)$$

where V_{cmax}^{25} is in [$\mu mol\,m^{-2}\,s^{-1}$] and C_{ab} in [$\mu g\,cm^{-2}$].

2.2. Sentinel-2 data

Sentinel-2 is a wide-swath, high-resolution, and multi-spectral imaging mission, supporting Copernicus Land Monitoring, including the monitoring of vegetation, soil covers and water bodies, as well as observation of inland waterways and coastal areas. The Sentinel-2 Multispectral Instrument (MSI) samples 13 spectral bands spanning from the visible and the near infrared to the shortwave infrared (Fig. 3), including two new spectral bands in the so-called red edge region (at 705 nm and 740 nm), which are very important for retrieval of chlorophyll content (Clevers and Gitelson, 2013). The spatial resolution varies from 10 m to 60 m depending on the spectral band with a 290 km field of view (Drusch et al., 2012). Three bands at 60 m are mainly dedicated for atmospheric correction and cloud screening, which leaves ten bands aimed at *land* surface observations. Currently, there are two Sentinel-2 satellites operating in tandem: Sentinel-2A was launched in June 2015, and Sentinel-2B launched in March 2017, which enables a revisit time of < 5 days.

We visually chose relatively cloud-free images over fields of interest (see Section 2.5) for the years 2016–2017. These images were atmospherically corrected using the Sen2Cor (version 2.4) algorithm, converting top-of-atmosphere (TOA) reflectance into top-of-canopy

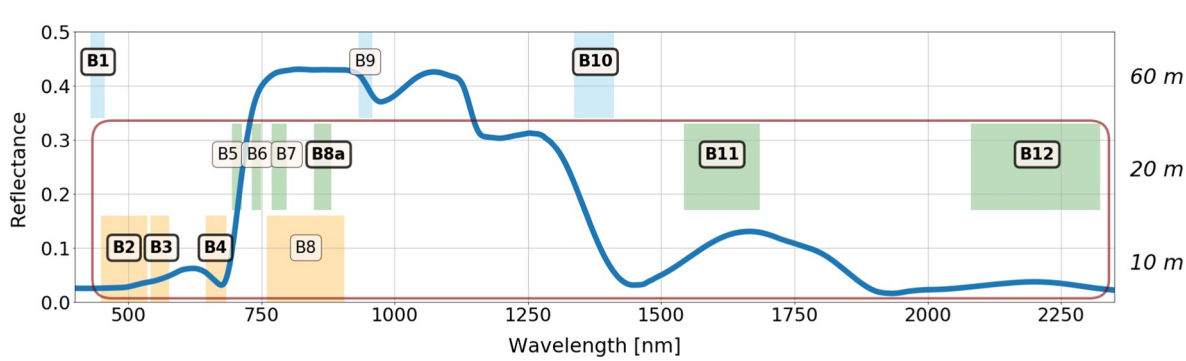

Fig. 3. Band settings of Sentinel-2A with respect to a typical vegetation reflectance spectrum. The bands in bold are those that overlap with Landsat 8 bands. The red rectangle encloses the bands used in this study. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(TOC) reflectance (Louis et al., 2016). In addition, Sen2Cor delivered information on pixel quality (cloud, cloud shadow, etc.). We used this information subsequently to evaluate more precisely if the images were cloudy. In few cases, when available, we used atmospherically corrected TOC data directly. Obtained TOC bands (B2–B8, B8a, B11, B12) were re-sampled to a common resolution of 20 m using the SNAP toolbox. We do not consider the effects of the resampling procedure, as we eventually calculate a mean value of the GPP over the whole fields.

2.3. Landsat 8 data

Landsat 8, a NASA and USGS collaboration, is the latest of the Landsat series and was launched in February of 2013. Operational Land Imager (OLI), an instrument onboard the Landsat 8 satellite, has overall similar spectral coverage to Sentinel-2 (sharing six common bands, see Fig. 3), but unfortunately does not cover as densely the vegetation red edge bands. The images of the Earth are collected with a 16-day repeat cycle, with a resolution of 30 m for bands of our interest (Storey et al., 2016).

We used in this study atmospherically corrected surface reflectance from Landsat 8/OLI (USGS Landsat 8 Surface Reflectance Tier 1) from the Google Earth Engine (GEE) platform (Gorelick et al., 2017). These data have been atmospherically corrected using LaSRC (Vermote et al., 2016) and include a cloud, shadow, water and snow mask produced using CFMASK (Foga et al., 2017), as well as a per-pixel saturation mask. We only used pixels which were marked *clear* by pixel quality attributes generated from the CFMASK algorithm.

2.4. Meteorological data

We used the meteorological data from Global Land Data Assimilation System (GLDAS) 2.1 that ingests satellite and ground-based observational data products to generate optimal fields of land surface states and fluxes (Rodell et al., 2004). GLDAS has been developed jointly by National Aeronautics and Space Administration (NASA) Goddard Space Flight Center (GSFC), and the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction (NCEP). It extends from the year 2000 to present with about one month latency and is updated monthly. Choosing this dataset as meteorological input was also motivated by its availability on GEE, which we plan to use in future for applying our GPP model globally. We used 3-hourly GLDAS-2.1 land surface model data available through GEE with a resolution of 0.25°. The data was exported for the dates of interest (availability of Sentinel-2 data) for the years 2016–2017. For Landsat 8 application, the GLDAS-2.1 data was used directly in GEE to estimate GPP. We did not perform any spatial interpolation, and used the meteorological data directly from the grid cells covering the chosen fields.

2.5. Flux tower sites

We used data from four flux tower sites located in the USA (US-Ro1, US-Ro2, US-Ro5, US-Ro6) and one site in Germany, DE-RuS (SE_EC_001 in the TERENO data portal http://teodoor.icg.kfa-juelich.de), for the feasibility test of our GPP models. Information on the location of the sites and crop types grown there can be found in Table 2. The sites were chosen based on the type of crop (soybeans and other C3 crops). The data was acquired for the dates of available Sentinel-2 observations for the years 2016 and 2017. We obtained GPP data directly from the Principle Investigators of the sites, and integrated half-hourly data to daily GPP values, which were then used as the reference value for the validation of our GPP model.

2.6. GPP estimated with vegetation indices

LUE models making use of VIs and incident photosynthetically active radiation (PAR_{in}) to estimate GPP for crops, were also applied for a comparison with our model. In previous studies different VIs were tested using ground-based in situ reflectance measurements (Peng and Gitelson, 2012), as well as Landsat (Gitelson et al., 2012) and MODIS data (Peng et al., 2013). These studies suggested different equations for GPP models using various VIs as input. The ones that showed the best performance in these studies were also tested here (cf. Table 3). We applied these approaches because they can be relatively easily adapted for our case (Sentinel-2 data, daily values), as compared to studies using MODIS data at 8-day temporal resolution and minimum 500 m spatial resolution (e.g. Wagle et al., 2015; Yuan et al., 2015; Q. Zhang et al., 2014). For VIs using a red edge band, we tested both red edge Sentinel-2 bands (B5 and B6), and eventually we chose B5, which led to higher correlation with flux tower GPP than using B6. We calculated 45% of daily integrated R_{in} values (following Running and Zhao, 2015) from GLDAS 2.1 to obtain PAR_{in} (even though the original equations were sometimes developed for PAR_{in} and sometimes for potential PAR_{in}). In addition, since none of these equations was actually designed for the band setting of Sentinel-2, we also established a linear function of red edge NDVI (reNDVI) and PAR_{in} using the flux tower validation dataset and calibrating the function directly on this data (cf. Table 3).

2.7. Global sensitivity analysis

Global sensitivity analysis (GSA) refers to a set of mathematical techniques aimed to analyze how the variation in the output of a numerical model can be attributed to variations of its inputs. Among others, GSA can be applied to evaluate the relative importance of each input variable in a model and can be used to identify the most influential variables affecting model outputs (Pianosi et al., 2015).

Here, we used the PAWN method (Pianosi and Wagener, 2015, the name derived from these authors names), which employs the entire model output distribution (cumulative distribution function, CDF) to quantify the sensitivity of the parameters and therefore it is applicable independently of the shape of the distribution. This is in contrast to variance-based sensitivity analysis (VBSA) that uses only the output variance, which might be not sufficient if the output distribution is multi-modal or highly skewed (see Pianosi et al., 2015 for more details). In addition, PAWN can be tailored to focus on particular ranges of the output, for instance extreme values (Pianosi et al., 2015). In the PAWN method, the sensitivity of the model output to the parameters due to direct and interaction effects is estimated with a PAWN total sensitivity index (T_i). The PAWN index has a range of variation between 0 and 1, with larger values reflecting higher importance. An input can

Table 2
Details about the flux tower sites used in this study (Griffis et al., 2004; Ney and Graf, 2017).

Site ID	Lon (°W)	Lat (°N)	Period	Crops
US-Ro1	−93.0898	44.7143	2016	Soybeans
US-Ro2	−93.0888	44.7288	2016	Kura Clover only
US-Ro5	−93.0576	44.6910	2017	Soybean
US-Ro6	−93.0578	44.6946	2017	Wheat/Kura Clover
DE-RuS	6.4472	50.8659	2016 & 2017	Winter barley in spring 2016, a catch crop mixture in fall 2016 and sugar beet in 2017

Table 3
Summary of vegetation indices used in this study. ρ_{green}, ρ_{red}, $\rho_{red\ edge}$ and ρ_{NIR} are reflectance in bands of green, red, red edge and near-infrared spectral regions and they refer to Sentinel-2 bands B3 (560 nm), B4 (665 nm), B5 (705 nm) and B8 (842 nm), respectively.

Vegetation index (VI)	VI abbreviation	VI formula	GPP ($x = VI \times PAR_{in}$)	Reference
Red edge chlorophyll index	$CI_{red\ edge}$	$\rho_{NIR}/\rho_{red\ edge} - 1$	$4.80\,ln(x) - 37.93$	Peng and Gitelson, 2012
Green chlorophyll index	$CI_{red\ edge}$	$\rho_{NIR}/\rho_{green} - 1$	$5.13\,ln(x) - 46.92$	Peng and Gitelson, 2012
Normalized difference vegetation index	NDVI	$(\rho_{NIR} - \rho_{red})/(\rho_{NIR} + \rho_{red})$	$2.07x - 6.19$	Gitelson et al., 2012
Green normalized difference vegetation index	greenNDVI	$(\rho_{NIR} - \rho_{green})/(\rho_{green} + \rho_{green})$	$2.86x - 11.9$	Gitelson et al., 2012
Enhanced vegetation index	EVI	$2.5(\rho_{NIR} - \rho_{red})/(\rho_{NIR} + 6\rho_{red} - 7.5\rho_{blue} + 1)$	$2.26x - 3.73$	Peng et al., 2013
Red edge normalized difference vegetation index	reNDVI	$(\rho_{NIR} - \rho_{red\ edge})/(\rho_{NIR} + \rho_{red\ edge})$	$1.61x - 1.75$	This study

be concluded to be non-influential, when T_i is below a threshold that depends on the chosen confidence level and the size of sample. The parameter space was sampled using Latin Hypercube Sampling (LHS) (McKay et al., 1979). In total, $N_u + N_c M_{PAWN} n_{PAWN}$ model runs are needed to approximate the total sensitivity index of all M_{PAWN} parameters, where N_u and N_c are the sample sizes of unconditional and conditional CDFs, respectively, and n_{PAWN} is the number of conditioning values of the model input. GSA was performed using the SAFE Toolbox (Pianosi et al., 2015).

2.8. Machine learning models

ML techniques map the relationship between the input (e.g., reflectances) and output (e.g., GPP) by fitting a flexible model directly to the data. Unlike parametric models that define an input-output mapping function, whose definition depends on a fixed set of parameters, the function in machine learning is typically non-parametric, nonlinear and very flexible. The weights of the model are fitted by using a training dataset (here provided by the forward modeling using SCOPE) in such a way that the model should perform well (i.e. provide accurate predictions) in a hold-out set, typically called validation or test dataset. Verrelst et al. (2012) compared four ML regression algorithms as candidates for biophysical parameter retrieval for Sentinel-2 and -3 and showed that Gaussian Process (GP) regression gave the most promising results. However, the main limitation of GP regression is the high computational cost for training and testing, as each test example has to be compared to all training samples (Quionero-Candela and Rasmussen, 2005).

Therefore, having in mind effective and global application of the developed models, we decided to eventually apply more efficient methods in terms of computational cost, such as neural networks (NNs) and random forests (RFs). Neural networks learn a relationship between input and output variables by establishing a set of nonlinear units (nodes with non-linear activation functions) organized in layers and connected by weights and biases that are equivalent to the regression parameters of classical parametric models (Bishop, 1995). They are a popular tool in the analysis of remotely sensed data (e.g., Mas and Flores, 2008), and have been already implemented in operational retrieval chains, including processing of Sentinel-2 data in the biophysical processor of the Sentinel Application Platform SNAP (Weiss and Baret, 2016). Random forests (RFs) are ensemble methods, which mean that a RF generates multiple estimators and aggregates their results. RFs can model complex interactions among input variables and are relatively robust with regard to outliers. They also have less parameters compared with NNs (Breiman, 2001) and recently were successfully applied in remote sensing applications (Tramontana et al., 2016; L. Wang et al., 2016). We trained these ML algorithms for the retrieval of vegetation parameters and modeling GPP using the data originated from the SCOPE model only.

Two training setups were engineered for this purpose:

- Case 1: Retrieving canopy and leaf parameters as well as retrieving fPAR. The input information is the reflectance data and the solar zenith angle (SZA) of the satellite observation.
- Case 2: Directly retrieving GPP. Here the input information is the reflectance data, SZA of observation, meteorological conditions, and SZA of a given modeling time step (which changes during the day, as opposite to the SZA of observation).

Multiple models were trained to estimate GPP, using reflectance data at Sentinel-2 resolution with all ten spectral *land* bands, but also a subset of bands that are common with Landsat 8 (i.e. B2, B3, B4, B8a, B11, B12, cf. Fig. 3). The GPP models were eventually applied to satellite data and GLDAS 2.1 meteorological data (Rodell et al., 2004) at 20 m spatial resolution (of the satellite data) and 3 h temporal resolution (temporal resolution of the meteorological dataset), for four temporal points per day (when incoming shortwave radiation was above zero). These values were then integrated to obtain daily values, which is a typical scale at which GPP from remote sensing data is evaluated. To compare data with flux tower measurements, we calculated daily GPP for the fields by taking average of Sentinel-2 (or Landsat 8) pixels within each field.

The ML regressions were applied in Python and were built using the scikit-learn toolkit (Pedregosa et al., 2011). If not mentioned otherwise, the settings of the ML models were set to *default* and the random state to 1.

2.9. Modeling set-up

We run multiple sets of SCOPE simulations to perform GSA (Section 2.7), and train ML algorithms (Section 2.8). Since the SCOPE model has a large number of input parameters, we tried to limit the number of considered parameters by building up on a recent study by Verrelst et al. (2015), in which the driving parameters for reflectance and SIF were investigated.

First of all, we focused on a number of vegetation and soil parameters that could be potentially retrieved from Sentinel-2 data. We varied 15 leaf, canopy and soil parameters (see Table 1) assuming a uniform distribution, while keeping other parameters constant. Furthermore, the variables were considered independent. We acknowledge that other distributions and assumptions might have led to a different performance of our ML models. However, since we wanted to focus on the concept of our approach, further optimization of ML algorithms is beyond the scope of this study, especially when considering a small validation dataset.

The value of V_{cmax}^{25} (leaf maximum carboxylation capacity at 25 °C) was set constant ($V_{cmax}^{25} = 100\ \mu mol\ m^{-2}\ s^{-1}$), or varied as a function of C_{ab}. The constant value of $100\ \mu mol\ m^{-2}\ s^{-1}$ was chosen following Y. Zhang et al. (2014), who applied it for SCOPE simulations for soybeans, and is commonly estimated for C3 crops (Kattge et al., 2009; Wullschleger, 1993). Since we focused on the vegetation and soil parameters first, the meteorological conditions were set to default SCOPE values ($R_{in} = 600\ W\ m^{-2}$, $T_a = 20\ °C$, $R_{li} = 300\ W\ m^{-2}$, $p = 970\ hPa$, $e_a = 15\ hPa$, $u = 2\ m\ s^{-1}$), as well as the SZA (30°). Regarding the geometry of observations, we used the constant values of *the observation zenith angle* (0°) and *the azimuthal difference between solar and observation angle* (90°), since both Sentinel-2 and Landsat 8 have a

relatively narrow field of view as well as quasi-nadir observations. We point out that $LIDF_a$ and $LIDF_b$ parameters were not independently sampled, but instead we used their sum ($LIDF_{a+b}$) and their difference ($LIDF_{a-b}$). This is motivated by the fact that their ($LIDF_a$ and $LIDF_b$) values must be chosen such that the sum of their absolute values equals to (or is smaller than) one, and therefore these parameters are not independent (Verhoef, 1998). As a solution we chose $LIDF_{a+b}$ and $LIDF_{a-b}$ to vary independently between −1 and 1, and based on their values we calculated $LIDF_a$ and $LIDF_b$. For these settings, we run the PAWN analyses, where N_u, N_c and n_{PAWN} were set to 1000, 400 and 30, respectively, which equaled to 181,000 simulations.

In order to train the ML GPP model, we additionally varied in simulations meteorological parameters and solar zenith angle (SZA). To cover a large range of vegetation and meteorological conditions, the variable ranges were based on previous studies that performed satellite retrievals and global sensitivity analysis of SCOPE data (Verrelst et al., 2015; Y. Zhang et al., 2014). All 22 parameters (see Table 1) were varied assuming uniform distribution (we run in total 177,000 simulations). These simulations were then re-run with different SZAs. This allows us to account for different SZAs at a given time step of GPP modeling (as opposed to the SZA during the Sentinel-2 observation). The SZAs in this scenario were chosen randomly between 0° and 85°. In this final training dataset, reflectance, meteorological data and SZA of observation are based on the original dataset, while GPP and SZA at a given time step are based on the re-run simulations.

For some combinations, the energy balance has not converged without adjusting the maximum iteration number or the maximum accepted error in the energy balance. These cases were not included in the training dataset for the ML model. This led to underrepresentation of cases with small LAI. To resolve this problem, we additionally performed a subset of calculations, where the value of LAI was set to 0.001, while all other variables varied as before (which led to a number of 10,700 additional simulations). For these cases, regardless whether simulations converged, we simply assigned the value of GPP to zero in all these scenarios. This subset was afterwards included in the training dataset to represent conditions of a very small (almost zero) LAI.

3. Results and discussion

This section gives empirical evidence of the performance of the proposed scheme for GPP estimation. We start the analysis by exploring the relative relevance of parameters using a sensitivity analysis approach. Then we examine the relationship between the components of the LUE model. After this analysis we provide quantitative results of GPP estimations using machine learning methods for the training and validation datasets. Results are then further validated for some selected flux towers and crops.

3.1. Predictor variables of GPP

We applied GSA to identify the most influential variables affecting GPP. We focused on the vegetation and soil parameters that can be potentially retrieved from the satellite data. Such an analysis should help to decide which vegetation parameters ought to be estimated in order to accurately estimate GPP with the SCOPE model. We used here a dataset, where V_{cmax}^{25} was varied as a function of C_{ab}, while meteorological conditions were constant. The results of PAWN indices computed over the whole output range are shown in Fig. 4a.

LAI was the most influential input (T_{LAI} of 0.97), followed by C_{ab} (T_{h_c} of 0.87), while other parameters were much less influential ($T_i < 0.5$). In addition, we performed the PAWN analysis in the following three GPP sub-ranges: small (< 5 μmol CO_2 m^{-2} s^{-1}), medium (5–20 μmol CO_2 m^{-2} s^{-1}) and high (> 20 μmol CO_2 m^{-2} s^{-1}). The influences of input parameters vary substantially for different sub-ranges. While LAI and C_{ab} turned out to be the most influential parameters predominantly for small and medium GPP, other variables (and especially h_c) are also influential for high GPP (Fig. 4b–d). LAI is the parameter that in general controls the presence and abundance of vegetation, and hence has a dominant role in determining GPP. The high influence of C_{ab} is due to its role in capturing light used for synthesis, but also because V_{cmax}^{25} is set as a function of C_{ab}; h_c is used in the model to calculate the roughness length for the momentum of the canopy displacement height, which in turn has an effect on the leaf temperatures and the gradients of water and CO_2 between the leaf surface and the atmosphere. As a result, we observe a high influence of h_c on GPP. Eventually, for high GPP, most of vegetation parameters become relevant to a certain degree (Fig. 4d). For example, other leaf pigments (content of carotenoid (C_{ca}) and content of anthocyanins (C_{ant})), dry matter content (C_{dm}) and canopy geometry parameters (*LIDF*) have all a stronger influence for medium and high GPP than for low GPP. Optically active leaf components compete with each other for light to absorb, while canopy geometry alters the relationship between sunlit and shaded leaves. The least influential vegetation variables are the leaf thickness parameter (N), leaf water equivalent layer (C_w) and senescent material fraction (C_s). Soil parameters have no influence in all cases. This is due to fact that the effect of soil properties on photosynthesis is

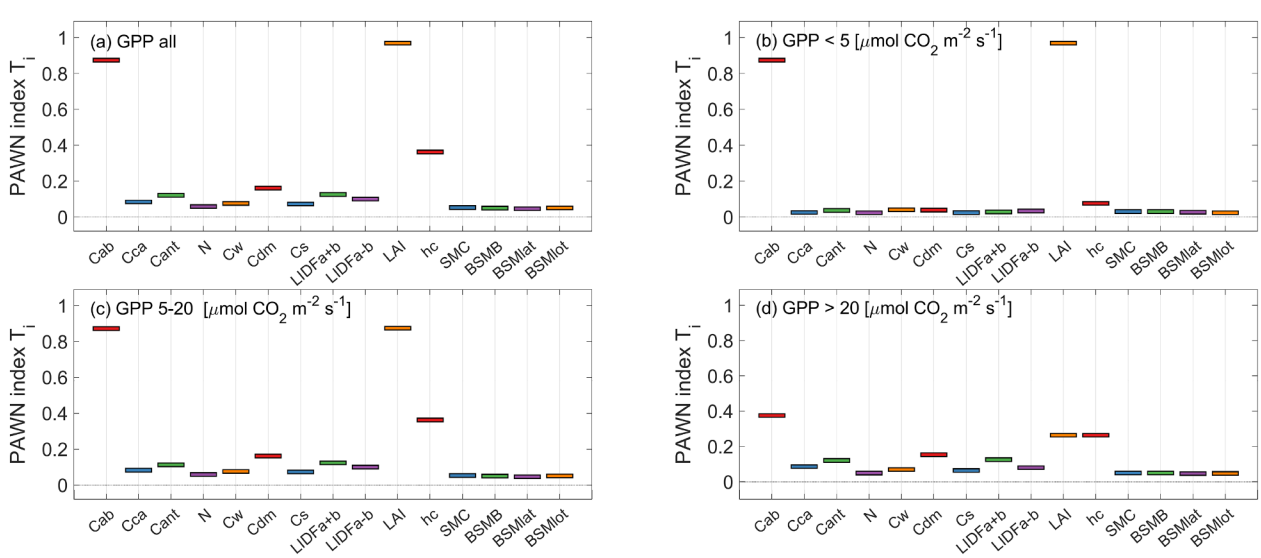

Fig. 4. PAWN indices for GPP for simulations with V_{cmax}^{25} dependent on C_{ab} for (a) all data, or three GPP sub-ranges: (b) small (< 5 μmol CO_2 m^{-2} s^{-1}); (c) medium (5–20 μmol CO_2 m^{-2} s^{-1}); and (d) high (> 20 μmol CO_2 m^{-2} s^{-1}). The boxes represent single values estimated for each input parameter.

not parameterized in the model.

The ranges of the input variables were chosen to be very broad in order to cover many various scenarios (see Table 1), but at any given time and place, their actual seasonal and daily variability can be very different. However, since it is very difficult to determine with certainty what is exactly probable for a global scale, we opted for including rather too much variability than too little. The description of the variability of each input parameter, including their range and distribution, can significantly affect the GSA results. For example, considering only a small range of pigment concentrations could decrease their influence on the output parameters as seen in the GSA. On the other hand, the parameter that has overall smaller influence can be sometimes the most important one if it changes more drastically than other parameters.

The vegetation parameters (LAI and Cab) that emerged as the most influential for estimating GPP, are exactly the ones that are often retrieved from remote-sensing data. However, more leaf and canopy parameters are important for precise calculation of GPP, especially for high GPP (which is often the case for crops, on which we focus in this study). Therefore, in case of using the SCOPE model, multiple further input variables should be estimated for running the model in the forward mode. Many of these variables are more difficult to estimate, and must be therefore assumed a priori, often with high uncertainties. However, if these assumption are not well constrained, it makes the global applications of such a complex model challenging.

3.2. Relationships between APAR and GPP

We also considered using SCOPE to estimate GPP by means of applying a LUE model, in which the only retrieved parameter would be fPAR, while LUE would be assumed constant and adjusted only by meteorological conditions (Fig. 2). In order to examine this application scheme, we analyzed the relationships between the components of the various LUE models (based on fPAR, fraction of PAR absorbed by chlorophyll, $fPAR_{Cab}$, or canopy chlorophyll content) as captured in SCOPE. The relationships between APAR, $APAR_{Cab}$ (PAR absorbed by chlorophyll), canopy chlorophyll content ($LAI{\cdot}C_{ab}$) and GPP were examined for a simple case of constant V_{cmax}^{25} and constant (default in SCOPE) meteorological conditions ($R_{in} = 600\,W\,m^{-2}$, $T_a = 20\,^{\circ}C$, $R_{li} = 300\,W\,m^{-2}$, $p = 970\,hPa$, $e_a = 15\,hPa$, $u = 2\,m\,s^{-1}$). This was done in order to dismiss the impact of biochemical variability on LUE.

Overall, obtained relationships are nonlinear, strongly scattered and heteroscedastic (Fig. 5), which does not agree with the assumptions made in the foreseen LUE model, where LUE would only vary with meteorological conditions. These simulations were performed for a constant irradiance and, therefore, the light response is not as saturating as typical light response curves. Thus, all variability is due to changes in leaf and canopy properties. However, it is important to keep in mind that this dataset includes strongly varied combinations. By limiting the data variability, for example by narrowing the relations between pigments and height of the canopy (e.g., C_{ca} 15–35% of C_{ab}, C_{ant} 30–60% of C_{ab}, h_c 1.2–1.6 m, chosen just as an example not to suggest any specific distributions), these relationships become much more linear, with larger scatter of data points only for high $APAR_{Cab}$, APAR and canopy chlorophyll content. This agrees with the GSA analysis, which also showed that the impact of more parameters is larger for high GPP. Since we use constant R_{in} ($600\,W\,m^{-2}$), this high variability towards larger values can be due to the photosynthetic efficiency being mitigated in the excessive light by other parameters (e.g., LIDFa that controls the angular distribution and therefore the ratio of sunlit and shaded leaves). As compared to APAR, $APAR_{Cab}$ appears to be a better parameter for estimating GPP. In the case of canopy chlorophyll content, the relationship saturates between 100 and 200 $\mu g\,cm^{-2}$. Gitelson et al. (2016) directly related GPP with canopy chlorophyll content and argued that GPP divided by incident PAR remained invariant, supporting the concept of an optimization of resource allocation (Field, 1991; Goetz and Prince, 1999). Since we used constant R_{in} in our simulations, we can directly compare the shapes of our curves with Gitelson et al. (2016). It is remarkable that Gitelson et al., 2016 also reported that GPP was very sensitive to canopy chlorophyll content up to 150 $\mu g\,cm^{-2}$, and not so much for canopy chlorophyll content above. Still, our results are more scattered and show a higher variability of GPP. However, since we used only a synthetic dataset, we can neither support nor refute their functional convergence hypothesis.

In general, these results suggest that LUE can change due to differences in leaf and canopy properties and that, by limiting the variability of the input variables, the SCOPE modeling results can converge into a more constant LUE. However, the confining of the input parameter distribution is not straightforward when all possible cases have to be accounted for. Nevertheless, as Zhang et al. (2018) found that maximum LUE (based on $APAR_{Cab}$) tends to converge across space and time, this approach could be further improved when the specific distributions of input parameters and their co-dependencies are investigated in detail.

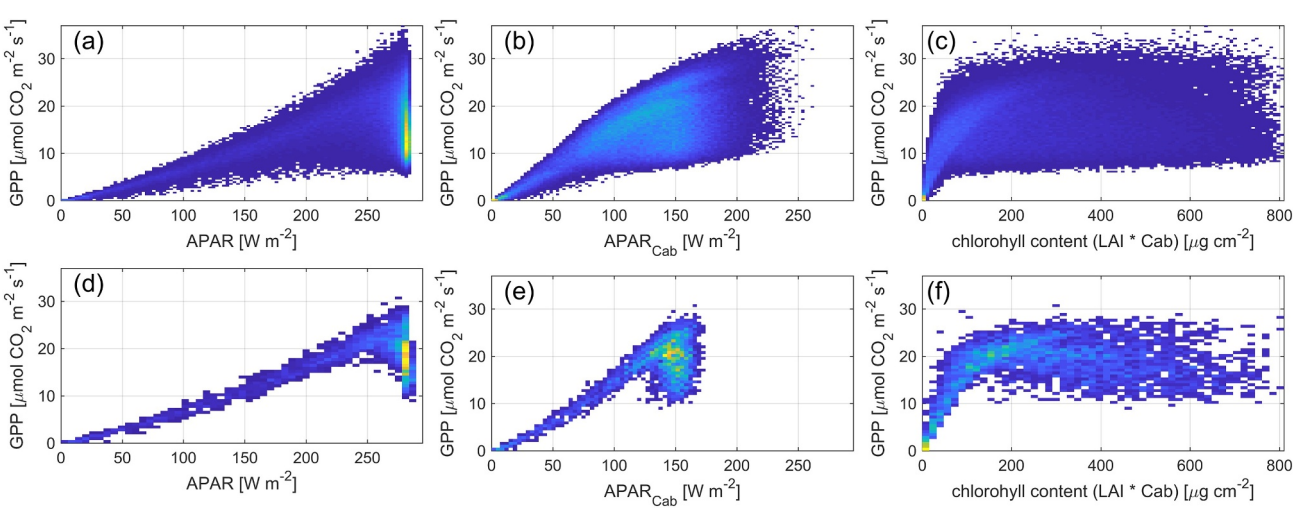

Fig. 5. Relationships between GPP and (a, d) APAR, (b, e) $APAR_{Cab}$ and (c, f) canopy chlorophyll content ($LAI{\cdot}C_{ab}$) for (top panel) the whole dataset and (bottom panel) a selected subset (C_{ca} 15–35% of C_{ab}, C_{ant} 30–60% of C_{ab}, h_c 1.2–1.6 m). The data were calculated with SCOPE, assuming constant V_{cmax}^{25} and constant meteorological conditions as default in SCOPE ($R_{in} = 600\,W\,m^{-2}$, $T_a = 20^{\circ}$ C, $R_{li} = 300\,W\,m^{-2}$, $p = 970\,hPa$, $e_a = 15\,hPa$, $u = 2\,m\,s^{-1}$).

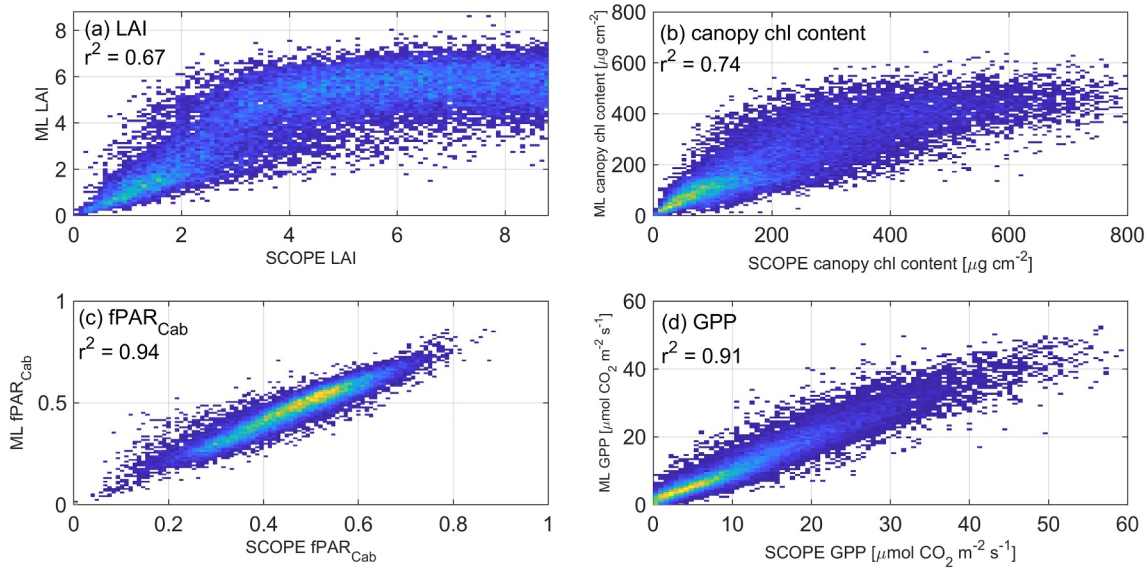

Fig. 6. Performances of NN models for (a) LAI, (b) canopy chlorophyll content (LAI·C_{ab}), (c) $fPAR_{Cab}$, and (d) GPP on the test dataset (the subset of SCOPE simulations).

3.3. Training of vegetation parameters and GPP models

Since before we were considering different workflows of estimating GPP using SCOPE (Fig. 2), we also tested the performance of various ML algorithms on the data modeled with SCOPE (dataset with V_{cmax}^{25} varying as a function of C_{ab} and changing meteorological conditions). In addition to GPP models, we compared performances of ML algorithms retrieving vegetation biophysical parameters using the SCOPE data. The GPP model was trained using satellite and meteorology data, while the other considered ML algorithms for LAI, canopy chlorophyll content and $fPAR_{Cab}$, used only satellite data. We used synthetic data to both train and validate ML algorithms. Simulated samples from SCOPE were hence divided into training (85%) and validation (15%) subsets. For LAI, canopy chlorophyll content and $fPAR_{Cab}$ retrievals, we used all ten (B2–B8, B8a, B11, B12) Sentinel-2 bands (reflectance output of SCOPE convolved to Sentinel-2 bands using spectral response functions) and the SZA of the observation. For models estimating GPP we used Sentinel-2 bands, both SZAs (of the observation and the modeling step), and the meteorological data.

We show results for NN, as RF gave similar results (data not shown). First, to compare the performance of the NNs for different parameters, we built a NN for each output separately. To minimize the effects of different atmospheric corrections and to at least partly harmonize the spectra across the sensors, for which we plan to apply our approach, we normalized the reflectance data to their spectral integral. In future, it would be optimal to use harmonized products, like e.g., Claverie et al. (2018). In addition, all inputs were normalized to fall in the range between zero and one (MinMax scaler). For this comparison, we used a simple NN: two hidden layers with 12 neurons with rectified linear unit function (ReLU). We also tested briefly other network structures, but results were similar and are not shown.

The performance of NN models for validation subsets is shown in Fig. 6 Our results show that retrieving vegetation parameters with a good accuracy is, in general, problematic. Even LAI, which is a commonly retrieved parameter, is difficult to estimate with a certain accuracy. For example, LAI has similar (even though much stronger) effects on the reflectance to LIDFa. Overall, retrieving multiple parameters from a limited number of Sentinel-2 bands is by nature an ill-posed inverse problem, where a set of possible solutions could lead to a match between the measured and the simulated reflectance data. Therefore, additional prior information can be helpful to improve the solution (Combal et al., 2003).

Our NN model of LAI showed a worse performance than the LAI retrieval that is implemented in SNAP, which is also using NN trained on the PROSAIL data (Weiss and Baret, 2016). Their results were validated with an independent test dataset (simulated using the same radiative transfer model), which also showed a minor LAI underestimation (but only around 6) and a much better performance overall. In addition, a comparison of LAI derived from Sentinel-2 data using this SNAP algorithm with a small set of non-destructive (optical) field reference measurements showed a very good agreement, with r^2 of 0.83 (Vuolo et al., 2016). There are several issues that might lead to the poor performance of our NN model. First of all, we used a more recent version of the SCOPE model with a higher number of input parameters, which can make it more difficult to train NNs. Furthermore, not only a number of input parameters, but also their distribution differs among this study and Vuolo et al. (2016), who used a Gaussian distribution. We used a uniform distribution and LHS in order to limit assumptions on underlying parameters as much as possible, but using a Gaussian distribution was shown to improve performance of the LAI retrieval in case of Verger et al. (2011). However, since GPP retrieval performed satisfactorily and it is the main focus in this study, we did not test adding additional a priori assumptions. We also applied a normalization of the satellite spectra to their sum across all bands, which might lead to the loss of some information from the magnitude. Overall, this is an ill posed problem, and an algorithm specifically designed to estimate LAI may impose a number of conditions to regularize the problem, as compared to our approach where we focus on GPP estimation.

NN performed better for canopy chlorophyll content and $fPAR_{Cab}$ than for LAI (Fig. 6c and d), which agrees with the study of (Verger et al. (2011). For canopy chlorophyll content, we obtained much better performance (r^2 of 0.74, as compared to r^2 of 0.67 for LAI), with only minor underestimation above 500 μg cm^{-2}. For $fPAR_{Cab}$, NN produced much better results (r^2 of 0.94), which shows that even though here we were not able to retrieve leaf and canopy properties accurately, more general characteristics of absorbed radiation can be retrieved well. Therefore, at least the fPAR element of the LUE model can be well observed using the Sentinel-2 data.

The GPP model performed much better than the LAI or canopy chlorophyll content retrievals, though similarly to $fPAR_{Cab}$, with a very small bias (mean error = 0.2 CO_2 s^{-1} m^{-2}) across the complete GPP range (Fig. 6). This suggests that it may be possible and actually easier to directly estimate GPP than to first retrieve other vegetation parameters, which would be then afterwards used to estimate GPP by

Table 4
Performance (r^2) of different NN models on training and test datasets (both with SCOPE), as well as on the validation (val.) data from the flux tower sites. For five different NN models we varied the number and size of hidden layers, as shown in the table.

	NN (hidden layers)				
	#1	#2	#3	#4	#5
	(12,12)	(20,20)	(20,12)	(12,12,12)	(40,20,12)
GPP r^2 train	0.92	0.93	0.92	0.93	0.96
GPP r^2 test	0.92	0.94	0.92	0.93	0.95
LAI r^2 test	0.58	0.62	0.62	0.59	0.68
GPP r^2 val.	0.86	0.88	0.92	0.89	0.91
GPP RMSE val.	1.72	1.66	1.38	1.51	1.41

running the original model in the forward mode. Good performance of the GPP and $fPAR_{Cab}$ models also suggests that the important information is already available in the Sentinel-2 bands, despite the fact that an accurate retrieval of leaf and canopy variables is very challenging (e.g., Fig. 6). However, using retrieved $fPAR_{Cab}$ in the LUE model is also not straightforward, as according to our previous analysis, LUE is not constant in SCOPE (Section 3.2).

Therefore, we chose to apply the ML model of GPP directly to the satellite data, instead of performing middle-step retrieval of vegetation parameters followed by re-running the original model in the forward mode. Our method makes the best use of the complexity of the process-based model (here SCOPE) in conditions of limited information about the system that we usually have, as it combines deep understanding of photosynthesis as implemented in the original model with the ML algorithms that are appropriate for the application to remote sensing data on a global scale. In addition, the algorithm design makes it easy to adjust or improve it, when the new version of the model is available - the data for training would have to be re-calculated and the algorithm re-trained, but it could be thereafter directly applied in an identical manner. Furthermore, ML models can be trained on data from the same model but with different spectral settings, which allows a global application across a range of different satellites while still being based on the same model.

Eventually, we also added LAI as an output to the ML models trained primarily to retrieve GPP, since LAI is the most influential parameter for GPP according to GSA. Although overall the performances of both architectures were similar, we obtained a small improvement in the performance of ML for GPP in case of our final model settings (of 0.01 for r^2 and of 0.2 $gC\,d^{-1}\,m^{-2}$ for RMSE the training and testing synthetic datasets). We considered different structures of the final ML models, and tested their performances on the training, testing and validation (from the flux tower) datasets (Tables 4 and 5). The results for both NN and RF were overall very similar. However, RF was in our case slightly affected by overfitting, since the determination coefficients were higher for the training than for the testing datasets. This was not the case for NNs, which had similar performance for training and testing datasets. Therefore, we focus here on NN models, while results for RF are shown in A.1. However, our approach in general does not rely on any specific ML method.

For NNs, we varied the number of hidden layers, the number of neurons per layers, the batch size (the number of training examples utilized in one iteration) as well as activation functions for hidden layers. After examination, we decided to use a batch size of 32, and ReLU as the activation function because of its good performance and low computational cost. Eventually, before finally applying ML algorithms to satellite data, we re-trained these ML models with the whole available SCOPE dataset. As for the numbers of hidden layers and neurons per layers, many different settings gave similarly good results for the synthetic dataset, so we ultimately compared the performances of five chosen architectures with the flux tower data (r^2 between 0.86 and 0.92, cf. Table 4). Eventually, we chose the NN model that performed best on the flux tower validation dataset. The best settings chosen for NN were afterwards applied to another ML model of GPP, for which we used only Sentinel-2 bands shared with Landsat 8.

3.4. Application to Sentinel-2

The GPP model was applied to the processed Sentinel-2 data. The comparison of resulting time series of GPP measured at flux towers and estimated GPP are shown in Fig. 7, and the overall results are compared in scatter plots in Fig. 8.

The NN model captures well the seasonal dynamics of the GPP, both in terms of the magnitude, as well as of the phenology. The good performance of the model is additionally confirmed by a strong linear relationship that was established for all seasons ($r^2 = 0.92$). The model successfully estimated GPP also outside of the growing season, and precisely tracked the emergence and senescence/harvest.

It must be stated that these results are for clear-sky data. Spectral reflectance data are affected by clouds, which is not accounted for in our statistical training and introduces errors in GPP models. In the first step of the selection of cloud-free days, Sentinel-2 cloud-free images were chosen visually. However, additional cloud check (based on the Sen2Cor classification of pixel into cloud, snow, shadow, etc.), showed that we included few days when the fields were covered by thin clouds, or partly by clouds and/or cloud shadows. These days show worse results (underestimation for most cases, cf. Fig. 12), which stresses the importance of atmospheric correction and scene classification.

Our ML algorithms performed better as compared to VI models (Fig. 8). However, the reNDVI model, which was fitted directly to the flux tower dataset, also yielded good results ($r^2 = 0.81$ and RMSE = 2.16 $gC\,d^{-1}\,m^{-2}$). We note that for VI models, using the red edge band in the reNDVI indeed improves GPP estimation as compared to EVI, and such LUE models are already widely used and show an overall good performance (e.g., Gitelson et al., 2012; Peng et al., 2013; Wagle et al., 2015; Yuan et al., 2015; Q. Zhang et al., 2014). However, we see the strength of our method not in directly outperforming empirical models, but in its potential to be used across a range of instruments with different spectral and spatial characteristics, and for a range of different conditions that might not be captured by empirical models, as a good performance of our model was obtained despite using no empirical information. First of all, it proves that the SCOPE model performs well and that it is a good and a reliable tool for coupled radiative transfer and biochemical modeling and therefore for relating reflectance data with GPP, also without having local information from the site. Accordingly, a possible future global application looks very promising.

The application of our approach to the Sentinel-2 data allows agriculture observations at a sub-field scale. The importance of the good spatial resolution is very clear when considering areas, which despite being homogeneously croplands, consist in fact of variable crop types. For example, as shown in Fig. 10, two observed fields (at flux-tower sites US-Ro1 and US-Ro2) demonstrate very different phenology. Kura Clover (which is a cover crop) was grown in 2016 at the site US-Ro2, and is photosynthetically active during the whole summer, while the phenology of soybeans is determined by planting and harvesting time, and its growing season is much shorter than for Kura Clover. Using Sentinel-2 images allows clear separation of different fields (not always the case for observations with coarser spatial resolution), which improves the differentiation of GPP among fields and therefore the estimation of the timing of crop phenology stage.

There are generally many issues that can hamper the model performance: the quality of the Sentinel-2 data and their atmospheric correction, the quality and the coarse spatio-temporal resolution of meteorological data, as well as the simplifications and assumptions used within the original model itself (especially lack of consideration of the soil moisture stress). However, even though there is no direct effect

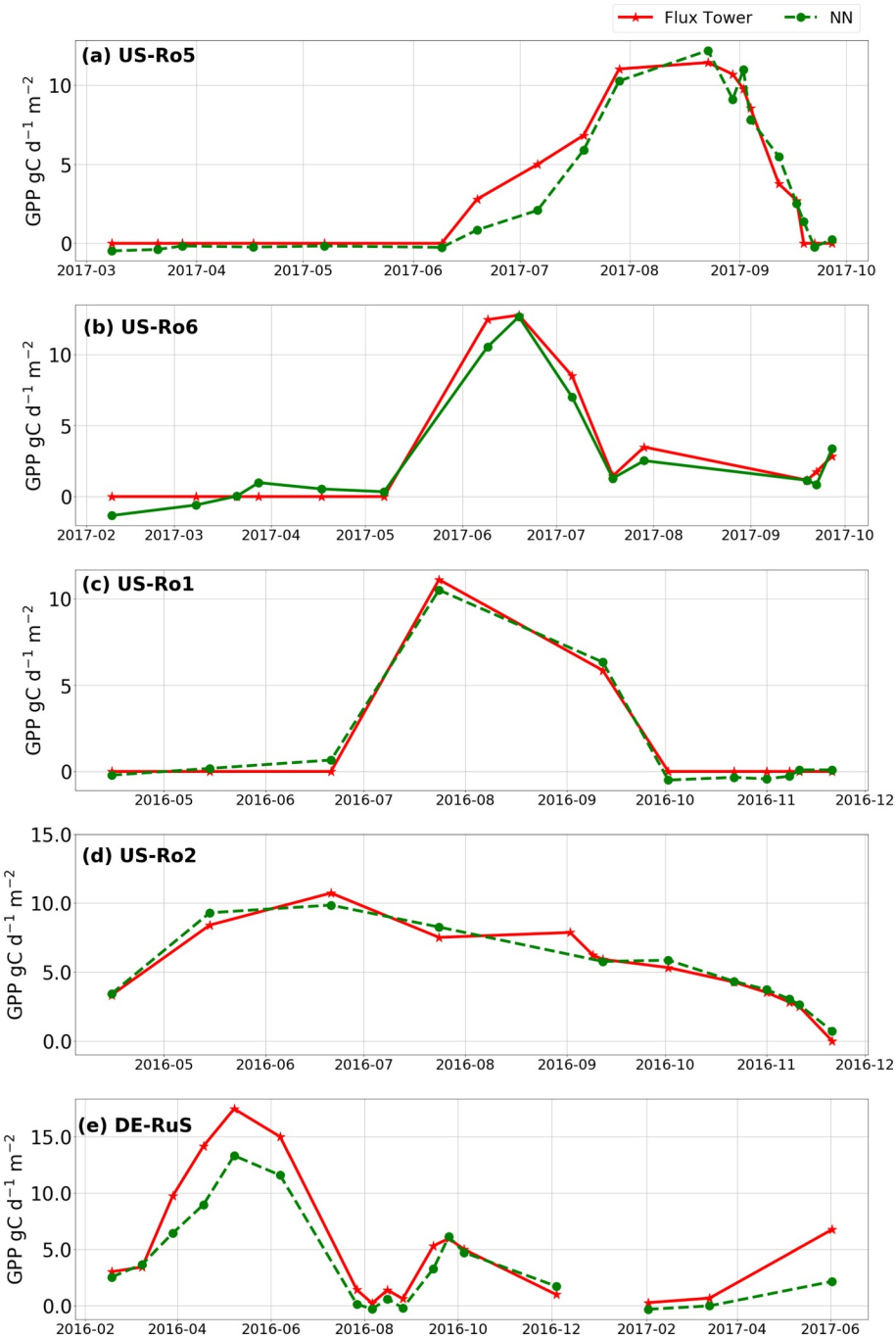

Fig. 7. Time series of GPP estimated from flux towers (red), and modeled with NN (green). Data points correspond to the days for which a clear Sentinel-2 image was available. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

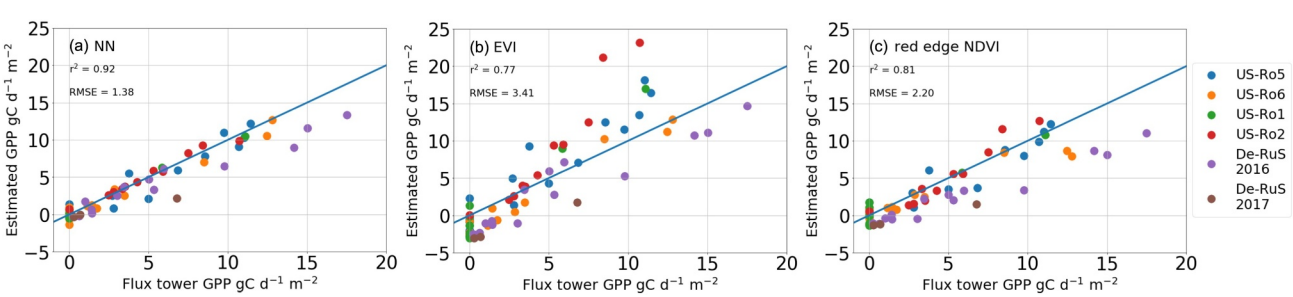

Fig. 8. Relationships between daily flux tower GPP and GPP estimated using (a) NN, (b) EVI and (c) reNDVI. Functions used for VIs are shown in Table 3. The straight line shows a 1:1 relationship.

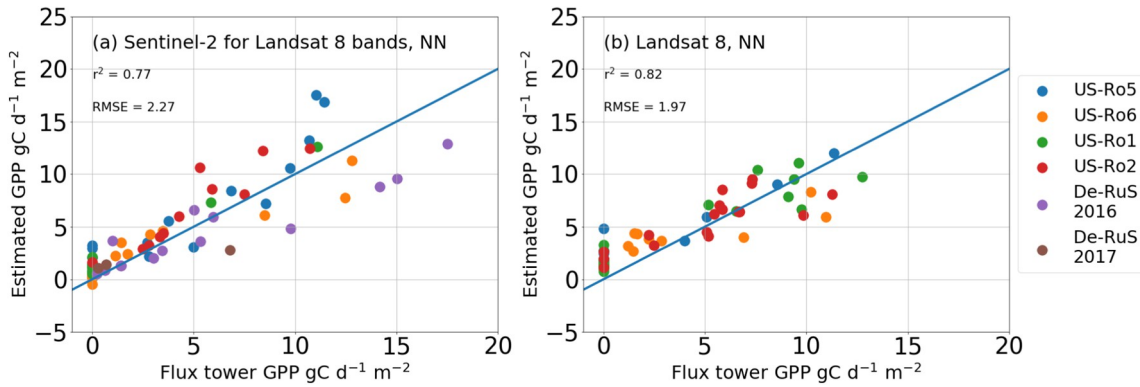

Fig. 9. Relationship between daily flux tower GPP and GPP estimated using (a) the subset of Sentinel-2 spectral bands that are also available in Landsat 8, (b) Landsat 8 data. The straight line shows a 1:1 relationship.

of soil moisture limitation within the model now, a prolonged stress is expected to have an effect on the canopy (e.g., through a reduction in C_{ab}, and eventually in LAI) that will be captured later by SCOPE and therefore by our ML model. We note that the applied flux tower dataset covers only the two years when no significant drought stress was affecting crops (even though these fields are not irrigated). The performance during drought episodes could be improved further by including thermal data (Bayat et al., 2018), which are however not available from Sentinel-2 data. Besides the limitations of the original model itself, ML algorithms are only a representation of the original model and do not exactly mirror its behavior. We tried to minimize the effects of atmospheric correction by normalizing the reflectance spectra. Nevertheless, the model was strongly impacted by the presence of clouds (Fig. 12).

In addition, footprints for the flux towers were not known and therefore we used a simple approach to calculate mean values over the whole fields. The footprints of flux towers for agricultural sites were estimated in previous studies to be up to 1–2 km (e.g., Chen et al., 2012; H. Wang et al., 2016), but they vary with wind speed and direction, turbulence intensity, surface roughness, measurement height, and atmospheric stability (Vesala et al., 2008), which can also lead to mismatches of flux tower and modeled estimates. Furthermore, we used rather coarse meteorological data (resolution of 0.25°) that do not capture finer spatial variability. We also do not explicitly account for the noise in the data, even though the performance of our models is affected by the uncertainties associated with the meteorological and satellite input data as well as the radiative transfer model itself.

Overall, the GPP model performed best for soybeans, for which the relationship between modeled and flux tower data was the most accurate (US-Ro1 and US-Ro5), and worst for the De-RuS site (Fig. 8). To apply our model to C4 crops, SCOPE simulations need to be redone accounting for the different photosynthetic pathways of the dark reaction of photosynthesis.

3.5. Application to Landsat 8

As the first attempt towards global application of our approach, we tested it on Landsat 8 data. We used the same NN settings as for Sentinel-2, but decreased the number of input satellite bands from ten to six. First, we tested the performance of the model on the Sentinel-2 data at bands shared with Landsat 8, which led to a decrease in model performance (r^2 = 0.77 and RMSE = 2.27 gC d^{-1} m^{-2}, see Fig. 9). The applied model had the same structure as the one chosen for the Sentinel-2 band setting, and therefore it could be expected to perform slightly better if specifically adjusted for Landsat 8.

Eventually we also applied NN model to the Landsat 8 data in GEE. The parameters of the model (scaling of the input parameters and the weights and intercepts of the neural network) were exported, and our final NN model was implemented in GEE for Landsat 8 data. The results were tested for the available flux tower data for the four sites in the USA (US-Ro1-2, US-Ro5-6), as the data for the DE-RuS site was only obtained for the Sentinel-2 overpasses. The overall performance of the model was good, but GPP was overestimated outside of the growing season (Fig. 9b).

The better performance of the model for Sentinel-2 bands suggests that the red edge bands do indeed improve ML modeling of GPP, which is also the case for the VI models. These bands have been shown to improve chlorophyll content estimation (Clevers and Gitelson, 2013), but in our case they seem to be especially useful for improving the

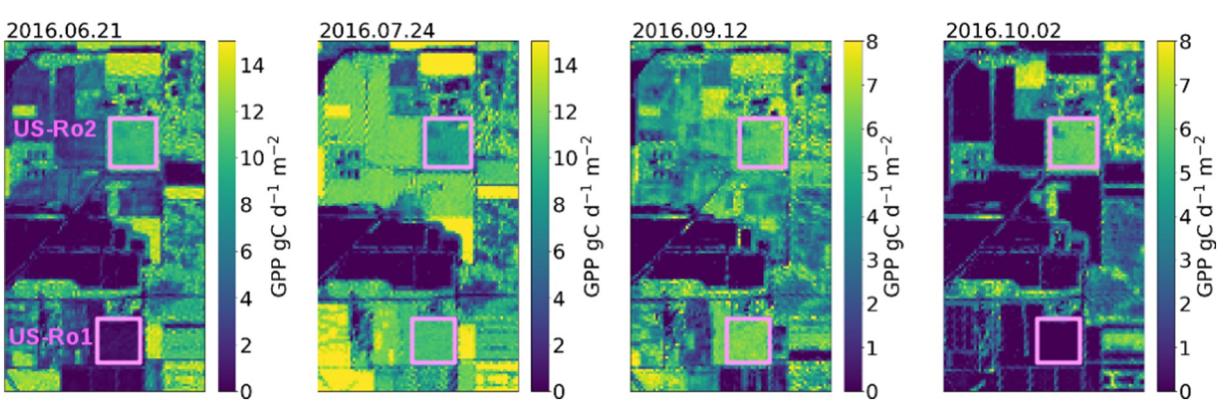

Fig. 10. Maps of estimated GPP using Sentinel-2 data over area neighboring the flux tower sites US-Ro1 and US-Ro2 for four days in summer 2016. The fields at these sites are highlighted. Soybeans and Kura Clover were grown at US-Ro1 and US-Ro2 in 2016, respectively.

model performance outside of the growing season (both for our ML models as well as for VIs).

Applying our ML models to both instruments simultaneously clearly increases the number of available data points for crop observations (Fig. 11). Interestingly, for the US-Ro1 site there are many more Landsat 8 observations available as compared to Sentinel-2, despite an overall better revisit time of Sentinel-2 (as there were fewer cloudy days during overpasses of Landsat 8). GPP values were quite similar among the sensors, especially during the growing season, which suggests a great potential to extend our approach to other instruments.

4. Conclusions and future work

Estimating photosynthesis of crops is crucial for the crop status monitoring and the forecasting of the agricultural production, and can be greatly supported by satellite remote sensing. Since recently, and

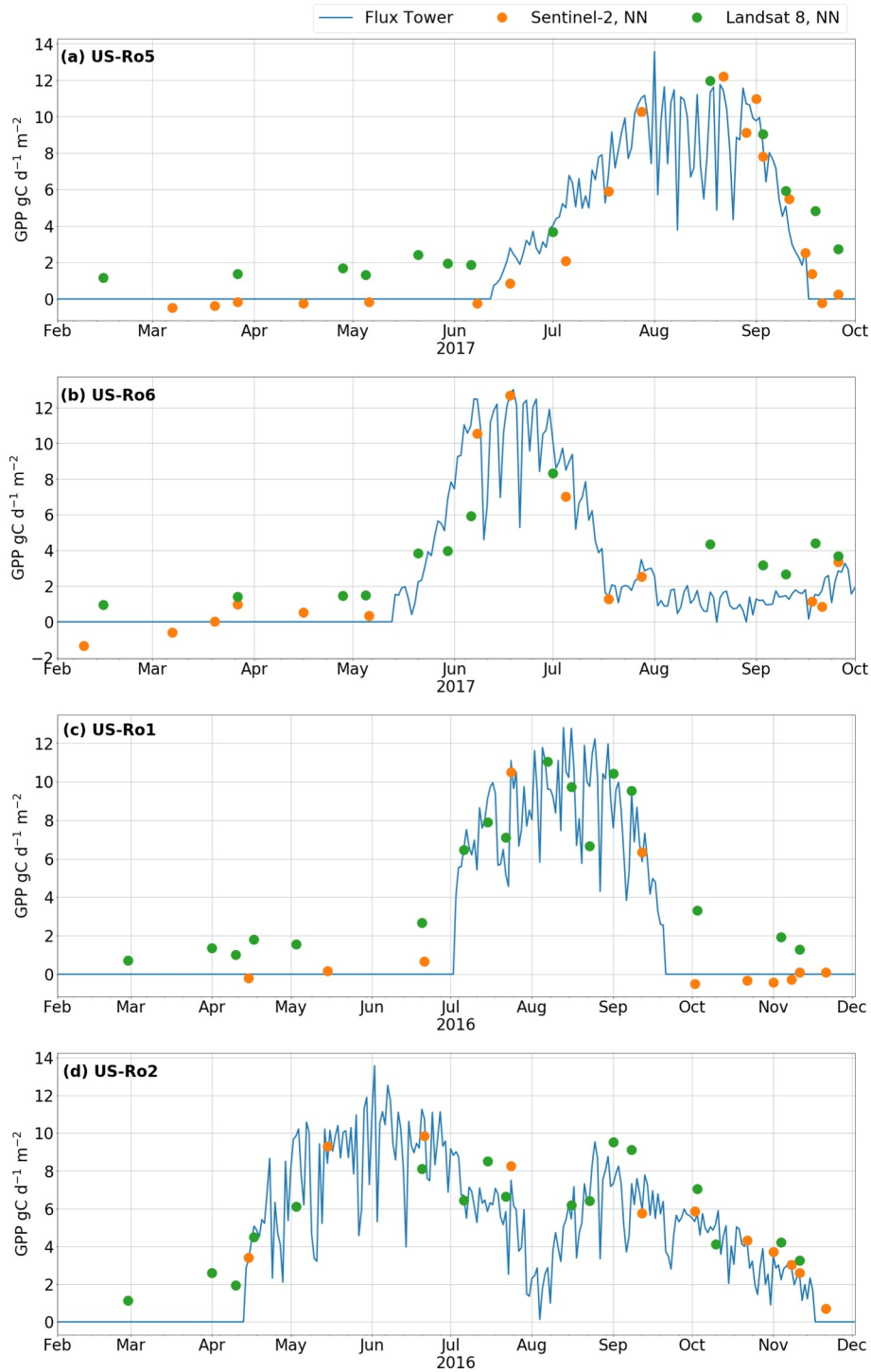

Fig. 11. Time series of GPP estimated from flux towers (blue), and modeled with NN using Sentinel-2 (orange) and Landsat 8 (green) data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

partly due to the advent of Sentinel-2 satellites, an unprecedented amount of data suitable for agriculture observations is available. Taking advantage of recent developments in satellite remote sensing technology, advances in machine learning and more complex and detailed models of photosynthetic processes, we developed a hybrid approach to model GPP with satellite data.

We have combined the process-based model SCOPE with ML algorithms to estimate GPP of C3 crops using satellite data and ancillary meteorological information. Several approaches were tested, and our final NN model estimated GPP at the tested flux towers very accurately (with r^2 of 0.92 and RMSE of 1.38 gC d^{-1} m^{-2}). ML models were more accurate than VI models, including the reNDVI model fitted directly into the flux tower dataset. Our proposed approach successfully estimated GPP across a variety of crop types and environmental conditions, also for time periods of no vegetation. This method was used for high spatio-temporal resolution monitoring of crops with Sentinel-2 and Landsat 8 data, but can be in fact further extended to other satellites. The results are promising and suggest a way to bridge process-based modeling for global application in an effective manner using a hybrid approach. Our model does not use any additional local information from the site, and therefore we plan to apply it globally using platforms providing cloud computing technology. Extending our approach to other sensors, including MODIS, will require additional accounting for spectral differences and the more complex geometry of observations. However, using data covering a longer time span will allow us to use a more extensive flux tower dataset for validation (e.g., FLUXNET2015 dataset), and therefore will provide a good opportunity for further model improvements. This will include testing a selection of the training dataset (e.g., selecting input distribution, assuming dependencies among parameters, adding noise to the data), model types and architectures, as well as procedures performed for harmonization of the datasets among satellites.

Acknowledgements

The work by AW, YZ and LG has been funded by the joint project of International Cooperation and Exchange Programs between NSFC and DFG (41761134082) and the Emmy Noether Programme (GlobFluo project) of the German Research Foundation. YZ was also financially supported by the General Program of National Science Foundation of China (41671421). The work by LGC and GMG has been funded by the Spanish Ministry of Economy and Competitiveness (TEC2016-77741-R, ERDF). We thank the PIs and Data Managers of the flux tower sites: Marius Schmidt (Forschungszentrum Jülich) and Alexander Graf (Forschungszentrum Jülich); John Baker (USDA-ARS), Timothy Griffis (University of Minnesota) and Cody Winker (USDA-ARS). We acknowledge the European Commission and the Copernicus Open Access Hub for the access to Sentinel-2 data. The GLDAS data used in this study were acquired as part of the mission of NASA's Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

Appendix A

A.1. RF

In the case of RF model of GPP, we tested settings including maximal depth, minimal samples leaf, as well as changing sample weight (in order to better represent scenarios with small LAI). Similarly to NN, the tests performed on synthetic dataset gave similarly good results, so five models were chosen to be compared with the flux tower data. The performances of these ML algorithms are shown in Table 5.

Table 5
Performance (r^2) of different ML models on training and test datasets (both with SCOPE), as well as on the validation (val.) data from the flux tower sites. For five different RF models, we varied maximum depth of the tree (MaxD), minimum number of samples required to be at a leaf node (MinLS), as well as sample weights (SW). In the case of the settings SW v.1, we increased the sample weight of data points with GPP below 2 μmol CO_2 m^{-2} s^{-1} to 2, and for the settings SW v.2 to 20.

	RF (settings)				
	#1	#2	#3	#4	#5
	Default	MaxD:20	MinLS:5	SW v.1	SW v.2
GPP r^2 train	0.98	0.98	0.96	0.98	0.98
GPP r^2 test	0.90	0.90	0.90	0.90	0.90
LAI r^2 test	0.51	0.51	0.52	0.51	0.51
GPP r^2 val.	0.84	0.84	0.85	0.87	0.89
GPP RMSE val.	1.98	1.99	1.93	1.70	1.58

The comparison of resulting time series of GPP measured at flux towers and GPP estimate with RF model (as well as NN model, including scenes covered by thin clouds) are shown in Fig. 12, and the overall results are compared in scatter plots in Fig. 13.

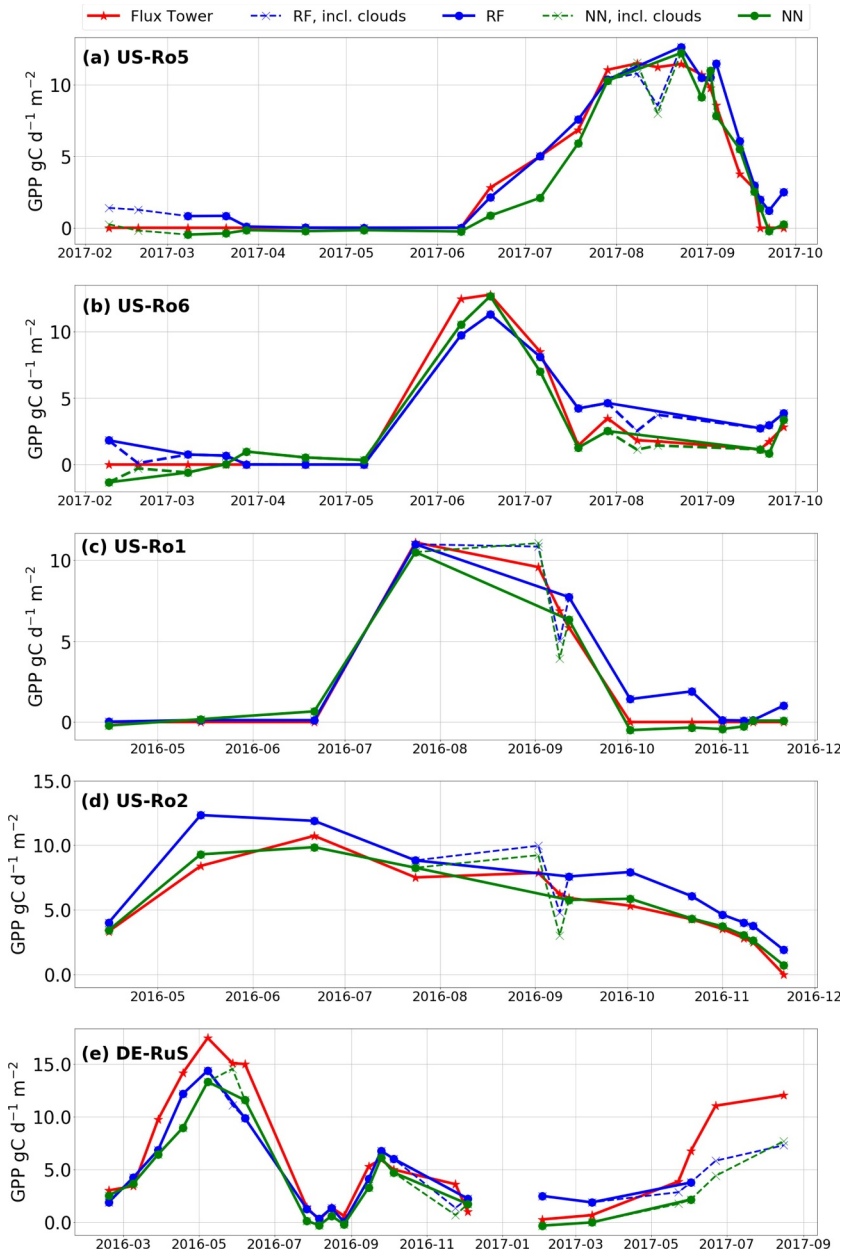

Fig. 12. Time series of GPP estimated from flux towers (red), and modeled with NN (green) and RF (blue). The straight line shows the GPP for only clear-sky dates, while dotted images show the dates when fields were covered by thin clouds (that were not removed directly in the visual check of the images). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

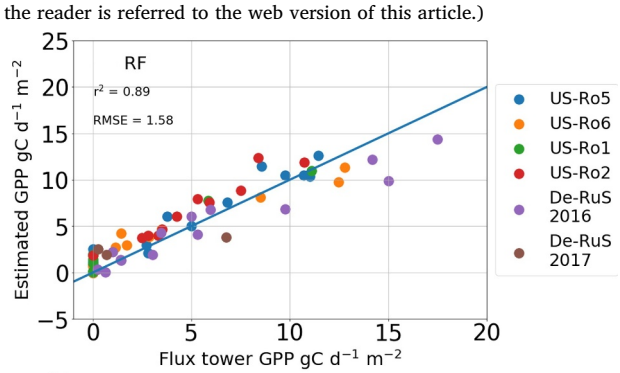

Fig. 13. Relationship between daily flux tower GPP and GPP estimated using the RF model. The straight line shows a 1:1 relationship.

References

Baker, J.M., Griffis, T.J., 2005. Examining strategies to improve the carbon balance of corn/soybean agriculture using eddy covariance and mass balance techniques. Agric. For. Meteorol. 128 (3), 163–177.

Bayat, B., van der Tol, C., Verhoef, W., 2018. Integrating satellite optical and thermal infrared observations for improving daily ecosystem functioning estimations during a drought episode. Remote Sens. Environ. 209, 375–394.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rdenbeck, C., Arain, M.A., Baldocchi, D., Bonan, G.B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K.W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F.I., Papale, D., 2010. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science 329 (5993), 834–838.

Belward, A.S., Skien, J.O., 2015. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. ISPRS J. Photogramm. Remote Sens. 103, 115–128.

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Camps-Valls, G., Tuia, D., Gómez-Chova, L., Jiménez, S., Malo, J., 2011. Remote Sensing Image Processing, 1st edition. Morgan & Claypool Publishers.

Chen, B., Coops, N.C., Fu, D., Margolis, H.A., Amiro, B.D., Black, T.A., Arain, M.A., Barr, A.G., Bourque, C.P.A., Flanagan, L.B., Lafleur, P.M., McCaughey, J.H., Wofsy, S.C., 2012. Characterizing spatial representativeness of flux tower eddy-covariance measurements across the Canadian Carbon Program Network using remote sensing and footprint analysis. Remote Sens. Environ. 124, 742–755.

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., Justice, C., 2018. The harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161.

Clevers, J., Gitelson, A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. Int. J. Appl. Earth Obs. Geoinf. 23, 344–351.

Collatz, G.J., Ball, J.T., Grivet, C., Berry, J.A., 1991. Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer. Agric. For. Meteorol. 54 (2), 107–136.

Collatz, G., Ribas-Carbo, M., Berry, J., 1992. Coupled photosynthesis-stomatal conductance model for leaves of C4 plants. Funct. Plant Biol. 19 (5), 519–538.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Mac, D., Pragnre, A., Myneni, R., Knyazikhin, Y., Wang, L., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance: using prior information to solve the ill-posed inverse problem. Remote Sens. Environ. 84 (1), 1–15.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., et al., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sens. Environ. 120, 25–36.

Falge, E., Baldocchi, D., Tenhunen, J., Aubinet, M., Bakwin, P., Berbigier, P., Bernhofer, C., Burba, G., Clement, R., Davis, K.J., Elbers, J.A., Goldstein, A.H., Grelle, A., Granier, A., Gumundsson, J., Hollinger, D., Kowalski, A.S., Katul, G., Law, B.E., Malhi, Y., Meyers, T., Monson, R.K., Munger, J.W., Oechel, W., Paw, U.K.T., Pilegaard, K., Rannik, L., Rebmann, C., Suyker, A., Valentini, R., Wilson, K., Wofsy, S., 2002. Seasonality of ecosystem respiration and gross primary production as derived from FLUXNET measurements. Agric. For. Meteorol. 113 (14), 53–74.

Farquhar, G.D., Caemmerer, S.v., Berry, J.A., 1980. A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. Planta 149 (1), 78–90.

Féret, J.B., Gitelson, A.A., Noble, S.D., Jacquemoud, S., 2017. PROSPECT-D: towards modeling leaf optical properties through a complete lifecycle. Remote Sens. Environ. 193, 204–215.

Field, C.B., 1991. Ecological scaling of carbon gain to stress and resource. In: Response of Plants to Multiple Stresses. Academic Press, San Diego, pp. 35–65.

Foga, S., Scaramuzza, P.L., Guo, S., Zhu, Z., Dilley, R.D., Beckmann, T., Schmidt, G.L., Dwyer, J.L., Hughes, M.J., Laue, B., Jun 2017. Cloud detection algorithm comparison and validation for operational Landsat data products. Remote Sens. Environ. 194, 379–390.

Gitelson, A.A., Peng, Y., Masek, J.G., Rundquist, D.C., Verma, S., Suyker, A., Baker, J.M., Hatfield, J.L., Meyers, T., 2012. Remote estimation of crop gross primary production with Landsat data. Remote Sens. Environ. 121, 404–414.

Gitelson, A.A., Peng, Y., Via, A., Arkebauer, T., Schepers, J.S., 2016. Efficiency of chlorophyll in gross primary productivity: a proof of concept and application in crops. J. Plant Physiol. 201, 101–110.

Goetz, S., Prince, S.D., 1999. Modelling terrestrial carbon exchange and storage: Evidence and implications of functional convergence in light-use efficiency. In: Advances in Ecological Research. vol. 28. Elsevier, pp. 57–92.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Griffis, T.J., Baker, J.M., Sargent, S.D., Tanner, B.D., Zhang, J., 2004. Measuring field-scale isotopic CO_2 fluxes with tunable diode laser absorption spectroscopy and micrometeorological techniques. Agric. For. Meteorol. 124 (1), 15–29.

Gu, L., Baldocchi, D., Verma, S.B., Black, T.A., Vesala, T., Falge, E.M., Dowty, P.R., 2002. Advantages of diffuse radiation for terrestrial ecosystem productivity. J. Geophys. Res. Atmos. 107 (D6), ACL 2–1.

Houborg, R., Cescatti, A., Migliavacca, M., Kustas, W.P., 2013. Satellite retrievals of leaf chlorophyll and photosynthetic capacity for improved modeling of GPP. Agric. For. Meteorol. 177, 10–23.

Jiang, C., Fang, H., 2012. Modeling soil reflectance using a global spectral library. In: AGU Fall Meeting, 2012, (San Francisco).

Jung, M., Reichstein, M., Bondeau, A., 2009. Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. Biogeosciences 6 (10), 2001–2013.

Jung, M., Reichstein, M., Margolis, H.A., Cescatti, A., Richardson, A.D., Arain, M.A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B.E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E.J., Papale, D., Sottocornola, M., Vaccari, F., Williams, C., 2011. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. J. Geophys. Res. 116.

Kattge, J., Knorr, W., Raddatz, T., Wirth, C., 2009. Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models. Glob. Chang. Biol. 15 (4), 976–991.

Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U., Cadau, E., Gascon, F., August 2016. Sentinel-2 Sen2Cor: L2A processor for users. In: Ouwehand, L. (Ed.), ESA Living Planet Symposium 2016. Vol. SP-740 of ESA Special Publications (on CD). Spacebooks Online, pp. 1–8.

Mas, J.F., Flores, J.J., 2008. The application of artificial neural networks to the analysis of remotely sensed data. Int. J. Remote Sens. 29 (3), 617–663.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239–245.

Mohammed, G., Ac, A., Daumard, F., Drusch, M., Galle, A., Goulas, Y., Magnani, F., Malenovsky, Z., Moreno, J., Olejnickova, J., Pernokis, D., Rivera, J., Verrelst, J., Rascher, C., van der Tol, C., Verhoef, W., Volta, A., 2014. FLEX/Sentinel-3 Tandem Mission Photosynthesis Study an Investigation of Steady-State Chlorophyll Fluorescence and Photosynthesis in Terrestrial Vegetation.

Monteith, J.L., 1972. Solar radiation and productivity in tropical ecosystems. J. Appl. Ecol. 9 (3), 747–766.

Mulla, D.J., 2013. Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. Biosyst. Eng. 114 (4), 358–371.

Ney, P., Graf, A., Oct 2017. High-resolution vertical profile measurements for carbon dioxide and water vapour concentrations within and above crop canopies. Bound.-Layer Meteorol. 166 (3), 449–473.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, D., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12 (Oct), 2825–2830.

Peng, Y., Gitelson, A.A., 2012. Remote estimation of gross primary productivity in soybean and maize based on total crop chlorophyll content. Remote Sens. Environ. 117, 440–448.

Peng, Y., Gitelson, A.A., Sakamoto, T., 2013. Remote estimation of gross primary productivity in crops using MODIS 250 m data. Remote Sens. Environ. 128, 186–196.

Pianosi, F., Wagener, T., 2015. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. Environ. Model Softw. 67, 1–11.

Pianosi, F., Sarrazin, F., Wagener, T., 2015. A Matlab toolbox for global sensitivity analysis. Environ. Model Softw. 70, 80–85.

Pinter, J., Hatfield, J.L., Schepers, J.S., Barnes, E.M., Moran, M.S., Daughtry, C.S.T., Upchurch, D.R., 2003. Remote Sensing for Crop Management. Photogramm. Eng. Remote Sens. 6, 647–664.

Propastin, P., Ibrom, A., Knohl, A., Erasmi, S., 2012. Effects of canopy photosynthesis saturation on the estimation of gross primary productivity from MODIS data in a tropical forest. Remote Sens. Environ. 121, 252–260.

Pulwarty, R.S., Sivakumar, M.V.K., 2014. Information systems in a changing climate: early warnings and drought risk management. Weather Clim. Extremes 3, 14–21.

Pury, D.d., Farquhar, G., 1997. Simple scaling of photosynthesis from leaves to canopies without the errors of big-leaf models. Plant Cell Environ. 20 (5), 537–557.

Quionero-Candela, J., Rasmussen, C.E., 2005. A unifying view of sparse approximate Gaussian process regression. J. Mach. Learn. Res. 6 (Dec), 1939–1959.

Reeves, M.C., Zhao, M., Running, S.W., 2005. Usefulness and limits of MODIS GPP for estimating wheat yield. Int. J. Remote Sens. 26 (7), 1403–1421.

Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin, J.K., Walker, J.P., Lohmann, D., Toll, D., 2004. The global land data assimilation system. Bull. Am. Meteorol. Soc. 85 (3), 381–394.

Running, S.W., Zhao, M., 2015. Daily GPP and annual NPP (MOD17A2/A3) products NASA earth observing system MODIS land algorithm. In: MOD17 Users Guide.

Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M., Hashimoto, H., 2004. A continuous satellite-derived measure of global terrestrial primary production. BioScience 54 (6), 547–560.

Running, S.W., Mu, Q., Zhao, M., 2015. MOD17A2H MODIS/Terra Gross Primary Productivity 8-Day L4 Global 500 m SIN Grid V006.

Storey, J., Roy, D.P., Masek, J., Gascon, F., Dwyer, J., Choate, M., 2016. A note on the temporary misregistration of Landsat-8 Operational Land Imager (OLI) and sentinel-2 multi spectral instrument (MSI) imagery. Remote Sens. Environ. 186, 121–122.

Strachan, I.B., Pattey, E., Boisvert, J.B., 2002. Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance. Remote Sens. Environ. 80 (2), 213–224.

Tramontana, G., Jung, M., Schwalm, C.R., Ichii, K., Camps-Valls, G., Ráduly, B., Reichstein, M., Arain, M.A., Cescatti, A., Kiely, G., Merbold, L., Serrano-Ortiz, P., Sickert, S., Wolf, S., Papale, D., 2016. Predicting carbon dioxide and energy fluxes across global fluxnet sites with regression algorithms. Biogeosciences 13 (14), 4291–4313.

Turner, D.P., Ritts, W.D., Cohen, W.B., Maeirsperger, T.K., Gower, S.T., Kirschbaum, A.A., Running, S.W., Zhao, M., Wofsy, S.C., Dunn, A.L., Law, B.E., Campbell, J.L., Oechel, W.C., Kwon, H.J., Meyers, T.P., Small, E.E., Kurc, S.A., Gamon, J.A., 2005. Site-level evaluation of satellite-based global terrestrial gross primary production and net

primary production monitoring. Glob. Chang. Biol. 11 (4), 666–684.
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 (12), 3109–3129.
Verger, A., Baret, F., Camacho, F., Feb 2011. Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: evaluation over an agricultural area with CHRIS/PROBA observations. Remote Sens. Environ. 115 (2), 415–426.
Verhoef, W., 1998. Theory of Radiative Transfer Models Applied in Optical Remote Sensing of Vegetation Canopies. (Ph.D. thesis).
Verhoef, W., Jia, L., Xiao, Q., Su, Z., 2007. Unified optical-thermal four-stream radiative transfer theory for homogeneous vegetation canopies. IEEE Trans. Geosci. Remote Sens. 45 (6), 1808–1822.
Verhoef, W., van der Tol, C., Middleton, E.M., 2018. Hyperspectral radiative transfer modeling to explore the combined retrieval of biophysical parameters and canopy fluorescence from FLEX–Sentinel-3 tandem mission multi-sensor data. Remote Sens. Environ. 204, 942–963.
Vermote, E., Justice, C., Claverie, M., Franch, B., Nov 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56.
Verrelst, J., Muoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., Moreno, J., 2012. Machine learning regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3. Remote Sens. Environ. 118, 127–139.
Verrelst, J., Rivera, J.P., van der Tol, C., Magnani, F., Mohammed, G., Moreno, J., 2015. Global sensitivity analysis of the SCOPE model: what drives simulated canopy-leaving sun-induced fluorescence? Remote Sens. Environ. 166, 8–21.
Verrelst, J., Pablo Rivera, J., Gitelson, A., Delegido, J., Moreno, J., Camps-Valls, G., 2016. Spectral band selection for vegetation properties retrieval using Gaussian processes regression. Int. J. Appl. Earth Obs. Geoinf. 52, 554–567.
Vesala, T., Kljun, N., Rannik, Rinne, J., Sogachev, A., Markkanen, T., Sabelfeld, K., Foken, T., Leclerc, M.Y., 2008. Flux and concentration footprint modelling: state of the art. Environ. Pollut. 152 (3), 653–666.
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.
Vuolo, F., Tak, M., Pipitone, C., Zappa, L., Wenng, H., Immitzer, M., Weiss, M., Baret, F., Atzberger, C., 2016. Data service platform for Sentinel-2 surface reflectance and value-added products: system use and examples. Remote Sens. 8 (11), 938.
Wagle, P., Xiao, X., Suyker, A.E., 2015. Estimation and analysis of gross primary production of soybean under various management practices and drought conditions. ISPRS J. Photogramm. Remote Sens. 99, 70–83.
Wang, H., Jia, G., Zhang, A., Miao, C., Wang, H., Jia, G., Zhang, A., Miao, C., 2016a. Assessment of spatial representativeness of eddy covariance flux data from flux tower to regional grid. Remote Sens. 8 (9), 742.
Wang, L., Zhou, X., Zhu, X., Dong, Z., Guo, W., 2016b. Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. Crop J. 4 (3), 212–219.
Weiss, M., Baret, F., 2016. S2ToolBox Level 2 Products: LAI, FAPAR, FCOVER, Version 1.1.
Wu, B., Meng, J., Li, Q., Yan, N., Du, X., Zhang, M., Feb. 2014. Remote sensing-based global crop monitoring: experiences with China's CropWatch system. Int. J. Digit. Earth 7 (2), 113–137.
Wullschleger, S.D., 1993. Biochemical limitations to carbon assimilation in C3 plants. A retrospective analysis of the a/ci curves from 109 species. J. Exp. Bot. 44 (5), 907–920.
Xiao, J., Zhuang, Q., Baldocchi, D.D., Law, B.E., Richardson, A.D., Chen, J., Oren, R., Starr, G., Noormets, A., Ma, S., Verma, S.B., Wharton, S., Wofsy, S.C., Bolstad, P.V., Burns, S.P., Cook, D.R., Curtis, P.S., Drake, B.G., Falk, M., Fischer, M.L., Foster, D.R., Gu, L., Hadley, J.L., Hollinger, D.Y., Katul, G.G., Litvak, M., Martin, T.A., Matamala, R., McNulty, S., Meyers, T.P., Monson, R.K., Munger, J.W., Oechel, W.C., Paw, U.K.T., Schmid, H.P., Scott, R.L., Sun, G., Suyker, A.E., Torn, M.S., 2008. Estimation of net ecosystem carbon exchange for the conterminous United States by combining MODIS and AmeriFlux data. Agric. For. Meteorol. 148 (11), 1827–1847.
Xin, Q., Broich, M., Suyker, A.E., Yu, L., Gong, P., 2015. Multi-scale evaluation of light use efficiency in MODIS gross primary productivity for croplands in the Midwestern United States. Agric. For. Meteorol. 201, 111–119.
Yuan, W., Cai, W., Nguy-Robertson, A.L., Fang, H., Suyker, A.E., Chen, Y., Dong, W., Liu, S., Zhang, H., 2015. Uncertainty in simulating gross primary production of cropland ecosystem from satellite-based models. Agric. For. Meteorol. 207, 48–57.
Zhang, M., Yu, G.-R., Zhuang, J., Gentry, R., Fu, Y.-L., Sun, X.-M., Zhang, L.-M., Wen, X.-F., Wang, Q.-F., Han, S.-J., Yan, J.-H., Zhang, Y.-P., Wang, Y.-F., Li, Y.-N., 2011. Effects of cloudiness change on net ecosystem exchange, light use efficiency, and water use efficiency in typical ecosystems of China. Agric. For. Meteorol. 151 (7), 803–816.
Zhang, F., Chen, J.M., Chen, J., Gough, C.M., Martin, T.A., Dragoni, D., 2012. Evaluating spatial and temporal patterns of MODIS GPP over the conterminous U.S. against flux measurements and a process model. Remote Sens. Environ. 124, 717–729.
Zhang, Q., Cheng, Y.-B., Lyapustin, A.I., Wang, Y., Xiao, X., Suyker, A., Verma, S., Tan, B., Middleton, E.M., 2014a. Estimation of crop gross primary production (GPP): I. Impact of MODIS observation footprint and impact of vegetation BRDF characteristics. Agric. For. Meteorol. 191, 51–63.
Zhang, Y., Guanter, L., Berry, J.A., Joiner, J., van der Tol, C., Huete, A., Gitelson, A., Voigt, M., Khler, P., 2014b. Estimation of vegetation photosynthetic capacity from space-based measurements of chlorophyll fluorescence for terrestrial biosphere models. Glob. Chang. Biol. 20 (12), 3727–3742.
Zhang, Y., Xiao, X., Wu, X., Zhou, S., Zhang, G., Qin, Y., Dong, J., 2017. A global moderate resolution dataset of gross primary production of vegetation for 2000–2016. Sci. Data 4, 170165.
Zhang, Y., Xiao, X., Wolf, S., Wu, J., Wu, X., Gioli, B., Wohlfahrt, G., Cescatti, A., van der Tol, C., Zhou, S., Gough, C.M., Gentine, P., Zhang, Y., Steinbrecher, R., Ardö, J., Apr 2018. Spatio-temporal convergence of maximum daily light-use efficiency based on radiation absorption by canopy chlorophyll. Geophys. Res. Lett. 45 (8), 3508–3519.