

Article

# Simultaneous Assimilation of Remotely Sensed Soil Moisture and FAPAR for Improving Terrestrial Carbon Fluxes at Multiple Sites Using CCDAS

Mousong Wu<sup>1,\*</sup>, Marko Scholze<sup>1</sup>, Michael Voßbeck<sup>2</sup>, Thomas Kaminski<sup>2</sup> and Georg Hoffmann<sup>2</sup>

- <sup>1</sup> Department of Physical Geography and Ecosystem Science, Lund University, 22362 Lund, Sweden; marko.scholze@nateko.lu.se
- <sup>2</sup> The Inversion Lab, 20251 Hamburg, Germany; Michael.Vossbeck@Inversion-Lab.com (M.V.); thomas.kaminski@inversion-lab.com (T.K.); georghoffmann64@gmail.com (G.H.)
- \* Correspondence: mousong.wu@nateko.lu.se; Tel.: +86-13570550392

Received: 26 September 2018; Accepted: 20 December 2018; Published: 25 December 2018



Abstract: The carbon cycle of the terrestrial biosphere plays a vital role in controlling the global carbon balance and, consequently, climate change. Reliably modeled CO<sub>2</sub> fluxes between the terrestrial biosphere and the atmosphere are necessary in projections of policy strategies aiming at constraining carbon emissions and of future climate change. In this study, SMOS (Soil Moisture and Ocean Salinity) L3 soil moisture and JRC-TIP FAPAR (Joint Research Centre—Two-stream Inversion Package Fraction of Absorbed Photosynthetically Active Radiation) data with respective original resolutions at 10 sites were used to constrain the process-based terrestrial biosphere model, BETHY (Biosphere, Energy Transfer and Hydrology), using the carbon cycle data assimilation system (CCDAS). We find that simultaneous assimilation of these two datasets jointly at all 10 sites yields a set of model parameters that achieve the best model performance in terms of independent observations of carbon fluxes as well as soil moisture. Assimilation in a single-site mode or using only a single dataset tends to over-adjust related parameters and deteriorates the model performance of a number of processes. The optimized parameter set derived from multi-site assimilation with soil moisture and FAPAR also improves, when applied at global scale simulations, the model-data fit against atmospheric CO<sub>2</sub>. This study demonstrates the potential of satellite-derived soil moisture and FAPAR when assimilated simultaneously in a model of the terrestrial carbon cycle to constrain terrestrial carbon fluxes. It furthermore shows that assimilation of soil moisture data helps to identity structural problems in the underlying model, i.e., missing management processes at sites covered by crops and grasslands.

**Keywords:** carbon cycle; SMOS soil moisture; JRC-TIP FAPAR; multi-site assimilation; uncertainty evaluation

# 1. Introduction

The terrestrial biosphere plays a vital role in the global carbon, water, and energy cycles [1,2]. Exchanges of carbon between the land surface and the atmosphere directly influence atmospheric carbon dioxide concentrations. The interactions of these three cycles result in a climate-carbon cycle feedback, which is a major source of uncertainty in climate change projections [2]. Large uncertainties exist in estimates of the terrestrial carbon balance both in observations and in model simulations [2–5]. Therefore, a better understanding of the response of the terrestrial biosphere to climate variability is both necessary and fundamental for improving projections of future climate change.



Several land data assimilation systems (LDASs) have been developed recently to assimilate various observations into land surface models, such as the carbon cycle data assimilation system (CCDAS) [6], the global land data assimilation system (GLDAS) [7], the coupled land vegetation LDAS (CLVLDAS) [8], and the LDAS-Monde [9]. These LDASs either optimize state variables or process parameters of the land surface model. CCDAS is an ideal tool to infer carbon exchanges both within the terrestrial biosphere and with the atmosphere by combining observations (remotely sensed or in-situ data) with process information as represented in models [6,10,11] to constrain model process parameters. A terrestrial biosphere model, in our case, the Biosphere Energy Transfer HYdrology (BETHY) model [12], parameterizes relevant processes in the terrestrial biosphere, from water and energy balance to photosynthesis. The actual values of the parameters used in these process-based models are often badly constrained by a limited set of observations taken on small scales with limited plant types and soil textures as well as climates. The typical spatial and temporal heterogeneity of various ecosystems can cause large uncertainties in modeling the global response of the terrestrial biosphere [2,13]. Also, the many nonlinear processes implemented in terrestrial biosphere models result in more complicated relationships between chosen parameters and modeled carbon fluxes at different temporal and spatial scales [3,14].

Assimilating observations into terrestrial biosphere models has been shown to be a powerful and efficient tool in constraining model parameters [10,11,15–19]. One of the observational datasets most used within the CCDAS, either as the unique observational constrain or jointly with others, is the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) [11,15–18,20,21]. Work has been done to assimilate FAPAR into the BETHY model, which was modified by an improved phenology scheme coupled to the model's two-flux representation of the radiative transfer in the canopy-soil system [16]. They found that most of the uncertainty reduction could be seen in parameters related to leaf phenology. The simultaneous assimilation of FAPAR with other datasets has also been shown to be promising. In a study by [22], FAPAR together with atmospheric  $CO_2$  concentrations were assimilated into BETHY, and the model's hydrology could be significantly constrained in a global simulation. Another work was done by assimilating latent heat flux and FAPAR for a woodland site in Botswana [18]. The study demonstrated that the model parameter controlling the maximum water uptake was efficiently constrained, but its optimal value was sensitive to the choice of the observational data used in the assimilation procedure. This leads to the probable conclusion that different observational constraints are connected to different processes in the terrestrial biosphere model. Recently, a comparison of assimilations with two FAPAR products using the ORCHIDEE model (Organising Carbon and Hydrology In Dynamic Ecosystems) [23] in an assimilation system was performed, with a simplified observation operator that employs the Lambert-Beer law to simulate FAPAR as a function of LAI (Leaf Area Index) [20]. They found that different FAPAR products could result in different model performances and in reducing uncertainties of different parameters. The combination with other datasets (e.g., eddy covariance fluxes) would improve the fit to independent observations.

Remotely sensed soil moisture has been widely used for assimilation in hydrological models [24–26], but also for assimilation in the context of updating gross carbon fluxes on a regional scale [27,28]. Recently, the assimilation work has shown that remotely sensed soil moisture is able to provide, as a complementary dataset, an additional constraint on hydrological as well as carbon cycle processes in a terrestrial biosphere model. Soil moisture, especially plant-available soil moisture, plays a key role in controlling photosynthesis and transpiration in plants and therefore regulates  $CO_2$  uptake through leaf stomata. Moreover, drought, freeze/thaw and, again, surface soil moisture control plant carbon cycling [29]. Nowadays, surface soil moisture can be observed by microwave sensors and is therefore independent of cloud cover. Such data sets have the potential to improve the simulation of hydrological processes in terrestrial biosphere models, thus resulting in improved terrestrial carbon cycle simulations.

Soil moisture from the European Space Agency's Soil Moisture and Ocean Salinity (SMOS) mission provides such a high-quality dataset for assimilation in hydrological and terrestrial biosphere models [19,30–32]. Assimilation of SMOS L3 soil moisture data together with atmospheric  $CO_2$  concentration data into the BETHY model to constrain the terrestrial carbon cycle on a global scale was done [19]. Assimilating SMOS soil moisture data into BETHY improved several aspects of the hydrological and carbon cycle [19]: (1) The agreement between both modeled and independently observed surface soil moisture (such as in-situ measurements of soil moisture from the International Soil Moisture Network (ISMN) [33]) improved significantly, and (2) the agreement between modeled and observed  $CO_2$  concentrations improved. A combined assimilation of FAPAR and soil moisture has been proposed by many studies as a potential improvement of the modeled carbon cycle [16,18,19].

In this study, we investigate the potential of such a simultaneous assimilation of remotely sensed surface soil moisture and FAPAR data to constrain the terrestrial carbon cycle both in single-site and in multi-site setups. The questions addressed are: What are the advantages of combining soil moisture and FAPAR data in constraining parameters of BETHY in a multiple sites approach? What do the posterior parameter values obtained by the assimilation experiments tell us about the mechanistic processes represented in the model? We also address the question whether an optimal parameter set obtained from site scale assimilation experiments improves global scale simulations.

## 2. Materials and Methods

#### 2.1. Study Site Descriptions

For this study, we chose 10 sites from the FLUXNET network as provided in the FLUXNET2015 data set (www.fluxdata.org, last accessed 6 July 2018). They were representative of different climate regimes, land cover types (in the model represented by Plant Functional Types (PFTs)), and soil textures, as depicted in Table 1. Detailed descriptions of all 10 sites can be found on the FLUXNET website (www.fluxdata.org). Site names used here are consistent with the naming and site codes from FLUXNET. The 10 sites contained 7 of the 13 PFTs used in BETHY, 4 of the 6 soil textures, and very different climate zones, such as a tropical site in northern Australia or a polar, tundra dominated site in Finland. Detailed PFT information is given in Table 2. At these sites, FLUXNET measurements were available from 2010 to 2014, corresponding to the SMOS L3 soil moisture data period used in this study.

Site Name	Latitude	Longitude	PFT_1	PFT_2	PFT_3	Texture
AU-DaS	-14.16	131.39	10 (0.70) *	7 (0.30)	/	Sandy loam
CA-Gro	48.22	-82.16	5 (0.85)	4 (0.10)	9 (0.05)	Loamy sand
CN-Sw2	41.79	111.90	9 (0.70)	10 (0.30)	/	Loam
DK-Sor	55.49	11.64	4 (0.95)	9 (0.05)	/	Loamy sand
ES-LMa	39.94	-5.77	9 (0.60)	3 (0.40)	/	Sandy loam
FI-Sod	67.36	26.64	5 (0.60)	4 (0.40)	/	Sandy loam
FR-Pue	43.74	3.60	3 (0.95)	9 (0.05)	/	Sandy clay loam
NL-Loo	52.17	5.74	5 (0.95)	9 (0.05)	/	Loamy sand
US-Kfs	39.05	-95.19	9 (0.95)	8 (0.05)	/	Loam
US-Wkg	31.74	-109.94	10 (0.90)	7 (0.10)	/	Loam

Table 1. Study site information, including latitude, longitude, PFTs, and textures.

\* The fraction of each PFT is given after PFT number.

#### 2.2. CCDAS Scheme

CCDAS is based on a terrestrial biosphere model (BETHY) and a gradient-based optimization method [10,16,34,35]. CCDAS can be run in a single-site, multi-site, or global-scale setup. BETHY simulates photosynthesis, autotrophic respiration, energy balance, and plant hydrological processes on an hourly scale, and phenology, soil hydrology, and heterotrophic respiration on a daily

scale. BETHY has 13 plant functional types (PFTs) to globally describe vegetation, as shown in Table A1. Each grid cell can be characterized by up to 3 PFTs, with a fractional coverage assigned for each PFT. Most PFTs use the same set of basic equations in modeling carbon cycle processes with different parameter values assigned. However, C3 and C4 photosynthesis and the growing season onset and termination (phenology) of each PFT are described by PFT-specific equations [16].

BETHY is forced by daily meteorological data, namely maximum and minimum air temperatures, precipitation, and incoming shortwave radiation. Here, we used meteorological data derived from the CRUNCEP V7 dataset (https://www.earthsystemgrid.org, last accessed 6 July 2018) as input data for running the model because the meteorological data measured at the sites are incomplete over the simulation period of 2010–2015. The original CRUNCEP data with a 6-hourly temporal resolution was converted to daily values. Meteorological data for each study site was extracted from the  $0.5^{\circ} \times 0.5^{\circ}$  gridded dataset. These data were then corrected by measured meteorological data from site meteorological stations to fill existing gaps in the site meteorology. The correction was done by using a linear regression method for air temperature and radiation. Precipitation was corrected by adjusting CRUNCEP gridded data to the total precipitation over the simulation period at the corresponding site. The model's hydrology was initialized by a spin-up run of the model using climatological forcing from 2009, which was sufficient for initializing the model as shown to be reasonable by [19].

CCDAS uses a probabilistic inversion method as described by [36]. A more detailed description of the optimization strategy can be found in [22]. In short, the system uses Gaussian Probability Density Functions (PDFs) to describe the state of the control vector, x (containing the model's process parameters), and the observations, d. The uncertainty covariance matrices of the model parameters and the observations are represented by  $C_x$  and  $C_d$ , respectively. The model, M(x), includes appropriate observation operators and maps the parameters to observables. The assimilation was done by minimizing the following cost function:

$$J_{total}(x) = \frac{1}{2} \begin{bmatrix} \sum (M_{sm}(x) - d_{sm})^T C_{d,sm}^{-1}(M_{sm}(x) - d_{sm}) \\ + \sum (M_{fapar}(x) - d_{fapar})^T C_{d,fapar}^{-1}(M_{fapar}(x) - d_{fapar}) + (x - x_0)^T C_0^{-1}(x - x_0) \end{bmatrix}$$
(1)

The summation in Equation (1) is over the different observations when assimilating the data simultaneously, in our case, soil moisture and FAPAR. The cost function,  $J_{total}$ , consists three parts, the parameter part,  $J_p$  (third term in Equation (1)), the soil moisture part,  $J_{sm}$  (first term in Equation (1)), and the FAPAR part,  $J_{fapar}$  (second term in Equation (1)). In practice, the minimization of  $J_{total}$  (=  $J_p + J_{sm} + J_{fapar}$ ) is performed iteratively by a gradient-based algorithm. The search direction is determined via the gradient of J with respect to the control vector, which, in this study, was derived from a finite difference approximation:

$$\frac{\partial J_{total}}{\partial x_i} \approx \frac{J_{total}(x_i + \varepsilon) - J_{total}(x_i)}{\varepsilon}$$
(2)

Within this approach, it is crucial to estimate a reasonable value for  $\varepsilon$  [37]. After a series of tests, we chose  $\varepsilon = 1.0 \times 10^{-2}$ . Figure 1 shows a flow chart of the calculation of the different terms in the cost function in CCDAS.



**Figure 1.** Flow chart of the calculation of the total cost function,  $J_{total} (= J_p + J_{sm} + J_{fapar})$  for simultaneous assimilation of soil moisture and FAPAR data as the sum of contributions from the parameter part ( $J_p$ ) and for the two data terms,  $J_{sm}$  and  $J_{fapar}$ . Ovals denote data and rectangular boxes denote processes (i.e., code modules).

#### 2.3. BETHY Soil Hydrology Scheme—Modified VIC 1L

To simulate surface soil moisture within BETHY in a way that can be compared to remotely sensed surface soil moisture, a modified version of the Variable Infiltration Capacity 1-layer (VIC-1 L) soil model by [38] was implemented in BETHY, for more details see [19]. The VIC model has the distinct property of representing within-grid cell spatial variability of infiltration capacity of the soil in a statistical fashion, something that is particularly suitable for simulations at the scale of the SMOS footprint of several hundred km<sup>2</sup> [39]. Therefore, a thin layer was added from which all soil evaporation occurs. The VIC-1 L evapotranspiration was also slightly modified, because in its original form, vegetation effects were not accounted for. This was needed to be consistent with transpiration and canopy throughfall of BETHY, and changed the base flow equation to prevent the soil from draining below field capacity. This new scheme is thus a 2-layer hydrology model with a surface and root-zone layer and the same 1-day time step as BETHY's original bucket model.

Soil moisture at saturation and at wilting point were computed from soil texture, total soil depth, and –for the root-zone layer—rooting depth. The values for rooting depth have been adapted from [12], taking account of the higher rooting depth of tropical trees and assigning a slightly higher rooting depth for temperate trees and shrubs and a lower one for grasses and tundra. A global soil depth and texture map and parameterization of soil porosity (saturation water content) and water potential was adapted from [40]:

$$\theta(\psi) = n(\psi/\psi_s)^{-1/b} \tag{3}$$

where  $\theta$  is the soil water content, *n* is the soil porosity,  $\psi$  is the soil water potential,  $\psi_s$  is the soil water potential at saturation, and b is the slope parameter of the retention curve. Details as well as validation of the modified VIC-1 L model are given in [19]. For the simulations described in this study, we used a physical depth of the surface soil layer of 40 mm to be comparable to the SMOS observations. *s<sub>n</sub>* and *s<sub>b</sub>* are scaling parameters for *n* and *b* included in the optimization. Additionally, soil hydrological parameters were set as fixed values derived from the VIC\_3L model [41].

#### 2.4. Parameters

In this study, we optimized a total of 100 parameters (Table A2) belonging to BETHY. Some of the parameters were global, some differentiated by PFT, and some by soil texture class. Parameters 72 to 100 were added to the version of CCDAS after implementation of a new phenology scheme [16]. For a detailed description of the parameters, 1 to 71, as well as the sensitivity we refer to [16].

The remaining parameters refer to stomatal control (72–74), energy balance (75), or soil water balance (76–100), and are described in detail in [19]. Rooting depth was differentiated by PFT. The soil hydrology parameters, i.e., the scaling factors for soil porosity and the slope parameter of the retention curve, were differentiated by soil texture class.

### 2.5. Assimilation and Validation Datasets

#### 2.5.1. SMOS Dataset

We used the reprocessed SMOS L3 (ver. 300) daily soil moisture product (L-band, 1.4 GHz) from CATDS (www.catds.fr, last accessed 6 July 2018) for the assimilation. The SMOS L3 soil moisture product represents soil moisture at a spatial resolution of ~25 km (EASE grid) within a thin surface layer (4 cm). The data were filtered considering radio frequency interference (RFI) according to [19]. Similar filtering was applied at grid points with a high probability of snow, frozen soil, water body, strong topography, and urban area.

We used the extended triple collocation (ETC) [42] analysis to calculate the uncertainty estimate for the SMOS L3 data. The other two datasets entered into the ETC calculation were the surface soil moisture computed by BETHY based on prior parameter values, and the Advanced SCATterometer (ASCAT) soil moisture observations (H109 Metop ASCAT DR2016 SSM time series 12.5 km sampling (see http://hsaf.meteoam.it) derived from an active microwave radar, C-band, 5.255GHz [43]). Then, the calculated uncertainty from ETC (model uncertainty) and uncertainty from data retrieval (data uncertainty) were formulated into the final SMOS soil moisture uncertainty, by:

$$\sigma_{SMOS} = \sqrt{\sigma_{data}^2 + \sigma_{model}^2} \tag{4}$$

where  $\sigma_{SMOS}$  is the SMOS soil moisture uncertainty used in the assimilation, and  $\sigma_{data}$  and  $\sigma_{model}$  are data and model uncertainty, respectively.

Additionally, we corrected the remotely sensed SMOS soil moisture to obviate the need for accurate soil texture data, and to retain as much useful information from the SMOS measurements as possible [19]. This was done by using the mean and standard deviation of the simulated surface soil moisture to correct the SMOS soil moisture:

$$S = \frac{\sigma_w(s - m_s)}{\sigma_s} + m_w \tag{5}$$

where *s* denotes the SMOS soil moisture (converted to mm),  $m_s$  is the long-term mean and  $\sigma_s$  its standard deviation, and *S* is the corrected SMOS soil moisture that is matched to the modeled surface soil moisture (in mm).  $m_w$  and  $\sigma_w$  are the long-term mean and standard deviation of the modeled soil moisture derived from the prior model parameter set.

#### 2.5.2. TIP-FAPAR Dataset

In the assimilation experiments, we used FAPAR data (16-day resolution) derived from the Joint Research Center with a Two-stream Inversion Package (JRC-TIP) [44]. The JRC-TIP FAPAR product was derived by the JRC-TIP algorithm from MODIS 0.01-degree albedo data [45]. A quality assurance was performed at two steps in the FAPAR retrieval algorithm. First, the MODIS broadband albedo product comes with a quality flag and low-quality input albedos were discarded before the JRC-TIP was applied. Second, dubious retrieval results were detected and if possible, improved in an iterative procedure [46]. Uncertainties for the FAPAR data were derived from the MODIS input albedo uncertainty by error propagation [47]. For more detailed information on the uncertainty propagation, we refer to the related papers [48].

We added an extra uncertainty of 0.05, taking into account very sparse observations (every 16-days). The total uncertainty of the FAPAR data was calculated as:

$$\sigma_{FAPAR} = \sqrt{\sigma_{data}^2 + 0.05^2} \tag{6}$$

where  $\sigma_{data}$  is the uncertainty of the product.  $\sigma_{FAPAR}$  is the FAPAR uncertainty used in the assimilation. A similar bias correction for the soil moisture observations (Equation (5)) was done for the JRC-TIP FAPAR product.

#### 2.5.3. FLUXNET Dataset

Estimates of daily net ecosystem exchange (NEE) were provided by FLUXNET2015 (www.fluxdata.org) for 8 of the 10 study sites, and for the two remaining sites by the European fluxes database (site ES-LMa; see www.europe-fluxdata.eu, last accessed 6 July 2018) and by the AmeriFlux data base (site US-Kfs; see https://ameriflux.lbl.gov, last accessed 6 July 2018), respectively. The NEE data covers the time period of interest of this study (2010–2014). In-situ daily soil moisture data covering 2010–2014 period from two sites, AU-DaS and US-Kfs, were also used for validation of the soil moisture simulations.

# 2.5.4. SIO CO<sub>2</sub> Concentrations

We used the same global setup as [19] in the  $8^{\circ} \times 10^{\circ}$  spatial resolution for validating our posterior net carbon flux results against atmospheric CO<sub>2</sub> exemplarity at two stations. The simulation period spanned from 2009 to 2015, with 2009 as an additional spin-up year. Monthly CO<sub>2</sub> concentrations were taken from the Scripps Institution of Oceanography (SIO) network (http://scrippsco2.ucsd.edu/, last accessed 6 July 2018). We used 2 stations (Table A3), for which both SIO and the Carbon Cycle Cooperative Global Air Sampling Network of the National Oceanic and Atmospheric Administration's (NOAA) Earth System Research Laboratory (ESRL) (ftp://aftp.cmdl.noaa.gov/data/trace\_gases/co2/ flask/surface/, last accessed 6 July 2018) CO<sub>2</sub> observations were available. As in [49], the uncertainties of the CO<sub>2</sub> observations were calculated based on the maximum of two values: The first one is the absolute difference between the SIO and the ESRL monthly mean values, and the second one is the standard deviation from a spline fit according to [50].

#### 2.6. Assimilation Experiments

We performed three groups of assimilation experiments for both single-site and multi-site cases. In the first experiment (denoted as 'sm'), we assimilated only SMOS L3 soil moisture data, in the second experiment (denoted as 'fapar'), we assimilated only JRC-TIP FAPAR data, and in the third experiment (denoted as 'sm+fapar'), we assimilated both SMOS L3 soil moisture and JRC-TIP FAPAR data. As mentioned, for two of the selected sites, DK-Sor and FI-Sod, no soil moisture data were available for assimilation due to the filtering of the SMOS L3 data set with high RFI. Therefore, only JRC-TIP FAPAR data were assimilated at these two sites. This resulted in 29 assimilation experiments (8 for the single-site 'sm', 10 for the single-site 'fapar', 8 for the single-site 'sm+fapar', 3 for the multi-site assimilation).

We note here the different spatial resolutions of the remote sensing products used in the assimilation (~25 km for the SMOS soil moisture and ~1 km for the JRC-TIP FAPAR as mentioned in the previous section) and the even smaller footprint (~few hundred meters) of the eddy covariance FLUXNET data used for validating our results. We also note the difference in the temporal resolutions of the SMOS soil moisture product (daily) and the JRC-TIP FAPAR product (16 daily). Hence, the number of soil moisture data points was higher than the FAPAR data points (6887 versus 1218 in the 'sm+fapar' multi-site case). However, our objective was to use all the observational information at the temporal and spatial scales available and not to achieve an equal balance between observational data streams in terms of numbers of observations. In any case, the weight of an individual observation in the cost

function calculation was a consequence of the specified uncertainty for this observation. In that sense, it was difficult to judge the constraint of an observational data streams on the model by the number of data points alone. The temporal resolution of a dataset determines the temporal scales that can be resolved by the data. The daily SMOS L3 product will be able to capture this temporal variability and constrain surface soil moisture and the associated hydrological processes on this time scale. However, phenological quantities (such as LAI) show less high temporal variability than soil moisture, therefore, the 16 daily FAPAR observations were of sufficient temporal resolution to constrain the phenology part of the model.

# 3. Results

#### 3.1. Optimization Performance

Table 2 shows a summary of the optimization results for the different experiments. For all three experiments, cost function values were largely reduced, with some exceptions in the 'fapar' experiment, where the initial cost functions were already very small (AU-DaS, 0% reduction in the cost function value, FR-Pue 1.38% reduction, and US-Wkg 7.74% reduction). In general, for all experiments, the optimization was successful, with gradients close to zero. In the 'sm+fapar' experiments, the cost function values were reduced by 16.37% to 69.28% in the case of single-site assimilation, and it was reduced by 49.49% in the case of multi-site assimilation.

**Table 2.** Summary of optimization results on cost function and gradient reduction for different assimilation experiments, 'sm' denotes soil moisture dataset, 'fapar' denotes FAPAR dataset, 'sm+fapar' denotes simultaneous assimilation soil moisture and FAPAR.

	sm	fapar	sm+fapar
Site	Cost Function Reduction (%)	Cost Function Reduction (%)	Cost Function Reduction (%)
AU-DaS	55.58	0	53.84
CA-Gro	22.39	26.83	21.78
CN-Sw2	65.61	57.82	69.28
DK-Sor	/	55.04	/
ES-LMa	40.92	34.46	35.52
FI-Sod	/	54.65	/
FR-Pue	16.88	1.38	32.59
NL-Loo	15.3	19.88	16.37
US-Kfs	51.88	86.12	52.79
US-Wkg	31.07	7.74	31.2
Multi-site *	35.27	55.06	49.49

\* 'Multi-site' denotes multi-site assimilation.

#### 3.2. Surface Soil Moisture and FAPAR

We calculated the cost function, *J*, over the simulation period both for the prior and the various posterior results to have a quantitative measure for the improvements obtained by the different assimilation experiments. We used the cost function as a performance metric for the assimilation because the cost function also considers the observation uncertainty (Equation (1)). Therefore, it was more suitable for quantifying the model performance with respect to the different variables.

Table 3 summarizes the cost function values for soil moisture (denoted as  $J_{sm}$  hereafter). Generally, the assimilation of the satellite datasets improved the model performance with respect to soil moisture. However, improvements differed between sites and depended sensitively on the assimilated dataset. As shown in Table 3, assimilation of soil moisture alone or simultaneously with FAPAR achieved the best performance in simulating soil moisture as indicated by the smallest  $J_{sm}$  values. Assimilating FAPAR only ('fapar') resulted in a larger  $J_{sm}$ , and thus inferior results for the simulated soil moisture than the other experiments for both single-site and multi-site cases.

Site	Prior	sm (Single) J <sub>sm</sub> /J <sub>p</sub>	sm (Multi) J <sub>sm</sub> /J <sub>p</sub>	fapar (Single) J <sub>sm</sub> /J <sub>p</sub>	fapar (Multi) J <sub>sm</sub> /J <sub>p</sub>	sm+fapar (Single) J <sub>sm</sub> /J <sub>p</sub>	sm+fapar (Multi) J <sub>sm</sub> /J <sub>p</sub>
AU-DaS	6427.59	1709.51/ 686.77	2327.26/ 420.94	$6427.50/\ 4.09 imes 10^{-7}$	5441.92/ 61.23	1995.35/ 103.94	2362.50/ 855.25
CA-Gro	2337.51	1710.34/ 71.84	2467.87/ 420.94	2232.91/ 1.65	2376.47/ 61.23	2788.61/ 73.95	2514.53/ 855.25
CN-Sw2	173.31	58.08/ 4.26	150.37/ 420.94	91.81/ 10.12	117.14/ 61.23	74.86/ 11.22	119.34/ 855.25
DK-Sor	/	/	/	/	/	/	/
ES-LMa	3250.70	1387.87/ 17.42	1531.16/ 420.94	2693.86/ 6.08	2652.02/ 61.23	1430.91/ 68.65	1587.30/ 855.25
FI-Sod	/	/	/	/	/	/	/
FR-Pue	78.90	30.02/ 46.75	30.49/ 420.94	79.42/ $1.78  imes 10^{-2}$	79.43/ 61.23	27.88/ 68.82	34.06/ 855.25
NL-Loo	1086.33	993.63/ 39.62	1481.05/ 420.94	1074.89/ 0.32	1073.81/ 61.23	1114.38/ 671.33	1463.87/ 855.25
US-Kfs	3860.16	2059.81/	2250.26/ 420.94	3124.62/	5556.66/ 61.23	2147.11/	2372.94/
US-Wkg	4500.40	2370.63/ 22.16	2443.57/ 420.94	4542.53/ $5.07 \times 10^{-2}$	4472.81/ 61.23	2371.58/ 21.13	2352.95/ 855.25

**Table 3.** Cost function values for soil moisture  $(J_{sm})$  and parameters  $(J_p)$  for prior and different assimilation experiments, 'sm' with soil moisture only, 'fapar' with FAPAR only, 'sm+fapar' with soil moisture and FAPAR data combined, 'multi' denotes multi-site assimilation, and 'single' denotes single-site assimilation.

In addition, we validated the posterior surface soil moisture values against in-situ soil moisture observations at eight sites from the FLUXNET2015 dataset and calculated the relative changes in  $R^2$  (determination coefficient) and RMSE (Root Mean Squared Error), shown in Table 4. For seven of the sites, either the  $R^2$  was improved (positive relative change values), or RMSE was improved (negative relative change values), or both  $R^2$  and RMSE were improved (four sites). As shown in Figure 2, at the site, US-Wkg, the RMSE for soil moisture reduced from 2.51 mm in the prior case to 1.92 mm in the posterior case. For the site, FI-Sod, located at high latitudes, with very limited observations available (427 observations during 2010–2015), both  $R^2$  and RMSE were not improved in the posterior case compared to the prior. The simultaneous assimilation of SMOS soil moisture and JRC-TIP FAPAR in the multi-site case was shown to improve not only carbon fluxes, but also, as expected, the soil moisture content in the surface layer.

**Table 4.** R<sup>2</sup> and RMSE relative changes (%) to prior simulation for site soil moisture calculated against in-situ observations when SMOS soil moisture and FAPAR were assimilated simultaneously in multi-site case, denoted as sm+fapar (multi).

C'I NI	Prior		sm+fap	oar (Multi)	Dete Desetter	01
Site Name	<b>R</b> <sup>2</sup>	RMSE (mm)	R <sup>2</sup> (%)	RMSE (%)	Data Duration	Observations
AU-DaS	0.58	6.02	1.72	-51.50	2010-2014	1826
CA-Gro	0.02	2.95	350.00	64.41	2010-2012	151
CN-Sw2	0.25	2.72	112.00	-25.37	2011	260
DK-Sor	0.001	1.79	7900.00	-6.15	2014	306
ES-LMa	0.51	1.93	-23.53	-22.80	2010-2011	418
FI-Sod	0.06	2.18	-16.67	6.42	2010-2014	427
FR-Pue	/	/	/	/	/	/
NL-Loo	0.16	3.31	106.25	105.14	2010-2014	1826
US-Kfs	/	/	/	/	/	/
US-Wkg	0.07	2.51	100.00	-23.51	2010-2014	1826



**Figure 2.** Comparison of simulated surface soil moisture (0–4 cm) at site US-Wkg against in-situ soil moisture observations. The blue dot-dash line denotes the simulation with the prior parameters, the red line denotes multi-site assimilation results with both soil moisture and FAPAR used for assimilation ('sm+fapar (multi)'), and the black dot denotes in-situ observations from the FLUXNET sites at the top layer.

Similarly,  $J_{fapar}$  values are shown in Table 5. In the 'sm' experiment,  $J_{fapar}$  values did not decrease much or even increase in comparison to the prior simulation. The smallest  $J_{fapar}$  values among the three experiments were achieved for the single-site assimilation, except for sites, FR-Pue, NL-Loo, and US-Wkg. For these sites, the cost function values for parameters ( $J_p$ ) were much smaller, which resulted in smaller total cost function values ( $J_{total}$ ) in the single-site optimizations. In the 'fapar' case, single-site assimilation achieved smaller  $J_{fapar}$  than multi-site assimilation. For the 'sm+fapar (multi)' case, the combination of SMOS soil moisture data with JRC-TIP FAPAR data resulted in a large cost function value for the parameters ( $J_p$  is 855.25). However, improvements in FAPAR in the 'sm+fapar(multi)' case were visible at most sites. For example, at site US-Kfs (Figure 3), when soil moisture and FAPAR were simultaneously assimilated at multiple sites, the FAPAR was constrained much better than the prior model simulation in matching remotely sensed FAPAR.

**Table 5.** Cost function values for FAPAR ( $J_{fapar}$ ) and parameters ( $J_p$ ) for prior and different assimilation experiments: 'Sm' with soil moisture only, 'fapar' with FAPAR only, 'sm+fapar' with soil moisture and FAPAR data combined, 'multi' denotes multi-site assimilation, and 'single' denotes single-site assimilation.

Site	Prior	sm (Single) J <sub>fapar</sub> /J <sub>p</sub>	sm (Multi) J <sub>fapar</sub> /J <sub>p</sub>	fapar (Single) J <sub>fapar</sub> /J <sub>p</sub>	fapar (Multi) J <sub>fapar</sub> /J <sub>p</sub>	sm+fapar (Single) J <sub>fapar</sub> /J <sub>p</sub>	sm+fapar (Multi) J <sub>fapar</sub> /J <sub>p</sub>
ALL-DaS	0.01	39.14/	0.08/	8.32/	71.56/	69.26/	27.77/
NO-Dao	0.01	101.43	218.50	5.47	92.26	103.94	855.25
CA- $Cro$	25.76	41.80/	39.71/	13.81/	28.31/	57.46/	18.61/
CA-GIU	25.70	51.88	218.50	4.83	92.26	73.95	855.25
CN Sur?	782 10	779.11/	3283.79/	13.28/	22.16/	293.45/	107.82/
CIN-3W2	782.10	8.32	218.50	6.52	92.26	25.65	855.25
DK-Sor	172 55	/	220.26/	21.88/	39.55/	/	36.13/
DR-501	172.55	/	218.50	9.98	92.26	/	855.25
ES I Ma	50 12	412.50/	127.60/	11.09/	18.33/	33.13/	18.66/
E3-Livia	39.12	31.25	218.50	12.61	92.26	68.65	855.25
EI Sod	274 72	/	709.31/	37.04/	38.60/	/	67.75/
11-30u	274.75	/	218.50	3.44	92.26	/	855.25
FR-Puo	2 90	2.96/	2.76/	15.85/	24.23/	11.59/	57.18/
TR-Tue	2.90	0.96	218.50	4.18	92.26	68.82	855.25
NI -Loo	1738	13.90/	14.17/	13.99/	31.05/	9.10/	33.07/
INL-LOO	17.30	24.80	218.50	4.09	92.26	671.33	855.25
LIC V 4 202 22	383 77	394.85/	438.06/	20.05/	71.26/	73.91/	37.10/
05-115	565.22	13.71	218.50	6.40	92.26	51.93	855.25
US Wha	1.86	2.99/	33.66/	24.79/	21.85/	25.39/	28.07/
0 <i>3</i> - <i>W</i> Kg	1.00	46.55	218.50	0.23	92.26	21.13	855.25



**Figure 3.** Comparison of simulated FAPAR at site US-Kfs against JRC-TIP FAPAR. The blue dot-dash line denotes the simulation with the prior parameters, the red line denotes multi-site assimilation results with both soil moisture and FAPAR used for assimilation ('sm+fapar (multi)'), and grey dots error bars denote JRC-TIP FAPAR with uncertainty.

# 3.3. Posterior Parameters

BETHY has 23 global parameters. In Figure 4, changes of these global parameters (aggregated as mean and standard deviation across the different experiments) are shown. Most of the global parameters did not change much compared to the prior.



**Figure 4.** Posterior (i.e., optimized) values for global parameters in different assimilation experiments (bars). Values were normalized by prior uncertainty for representation in one figure (**a**);  $\Lambda \sim$  is separately plotted in (**b**) due to its different range. Cross symbols denote the mean over all experiments and, error bars correspond to ±1 standard deviation.

Three parameters, however, showed a large spread in the different assimilation experiments:  $\xi$  (rate for initial linear leaf growth),  $fC_i$ ,  $C_3$  (the relation of maximum leaf internal CO<sub>2</sub> to ambient atmospheric CO<sub>2</sub> concentration for C3 plants), and *emis*0 (emissivity of the atmosphere). Large changes of these parameters were associated with the simultaneous assimilation experiments, that is 'sm+fapar'.

The largest changes in the initial leaf growth rate,  $\xi$ , were computed at the sites, CA-Gro and US-Wkg, in the 'smos+fapar' experiment. Sites dominated by C3 grass plant types (e.g., CN-Sw2 and US-Kfs) showed the largest changes for  $fC_i$ ,  $C_3$ .

## 4. Discussion

# 4.1. Posterior Parameters

In Figure 5, relative parameter changes, that is parameter changes from prior to posterior divided by their prior uncertainty, are shown. Some parameters related to phenology, water, and energy balance processes showed large changes (see parameters no. 57–75, and no. 89–97). This indicates that the observations of soil moisture and FAPAR have a major impact on the assimilation procedure and strongly constrain these processes in the model.



**Figure 5.** Parameter value changes relative to prior for all assimilation experiments (29 in total), with a cross symbol indicating the mean of changes, box for standard deviation, and whisker for max and min values, parameter numbers correspond to the numbers listed in Table A2.

A wide range of posterior values was computed for *emis*0 (Figure 5, Parameter no. 75). This parameter regulates the emissivity of the atmosphere and therefore has a strong impact on the surface water and energy fluxes through the radiation balance at the surface. At sites dominated by C3/C4 grasses (e.g., AU-DaS, ES-LMa, US-Kfs and US-Wkg), 'sm+fapar' experiments resulted in a remarkable range for *emis*0 across the respective sites. C3/C4 grasses had smaller root depths (in the model the prior values amount to 0.5 m) in comparison to trees (prior values of 1.5–3.0 m). Therefore, the growth conditions of grass in the model, but also in nature, depend strongly on soil moisture in the surface layer as well as on radiation. Moreover, surface conditions also differed from site to site due to roughness and surface cover variations, which potentially led to the large range of *emis*0 values.

Figure 6 shows the range of values for the two parameters that were spatially differentiated by soil texture. The parameters were related to the soil water retention curve, namely  $s_b$ , scalar for water retention curve shape parameter, and  $s_n$ , scalar for soil porosity. In Figure 6, the optimized values span a wide range across almost all soil textures. This indicates that there are large uncertainties in soil hydrology processes. In the BETHY model, soil hydrology was modeled by a bucket model with an overlying surface soil moisture layer [19]. The soil water retention curve cannot be robustly

constrained when soil moisture or FAPAR was assimilated as demonstrated by the large range of posterior values for the respective parameters.



**Figure 6.** Posterior (i.e., optimized) values for soil texture dependent parameters related to soil water processes, (**a**)  $s_b$  for different soil textures; (**b**)  $s_n$  for different soil textures. Normalization and symbols as in Figure 4.

When comparing the 'sm' experiment and the 'fapar' experiment, a sort of a "trade-off" between soil moisture and FAPAR becomes evident in the assimilation. The 'sm' assimilations resulted in relatively large changes in the parameter, *emis*0, that is the emissivity of the atmosphere, and the parameter,  $s_b$ , which controls the shape of the water retention curve (see [19]). To the contrary, both parameters hardly changed when assimilating FAPAR only ('fapar'). For 'fapar' experiments, parameters related to phenology experienced the largest changes relative to their prior values, e.g.,  $T_r$ (no. 62–63), which is the spatial range (essentially the with-in grid cell variability) of the phenology temperature trigger,  $k_L$  (no. 67–68), which denotes the inverse of leaf longevity, and  $\tau_W$  (no. 69–71), which parameterizes the length of dry spells before leaf shedding.

#### 4.2. Soil Moisture and FAPAR Controls on Soil Hydrology

Surface soil moisture and FAPAR data control the soil moisture process in different ways. FAPAR controls the photosynthetic processes and phenology as represented by the model. Both influence plant water uptake indirectly under water-limited conditions. In other assimilation studies, [15] and [18], the authors also noticed that the inclusion of FAPAR in the assimilation resulted in significant adjustments in the maximum plant-available soil moisture. The impact of FAPAR on the model's soil hydrology is not fully understood and obviously rather indirect compared to the assimilation of soil moisture data. Besides of FAPAR's role in controlling the model's phenology, the assimilation of FAPAR also constrains soil hydrology related parameters [15].

Vice versa, the assimilation of soil moisture observations constrains also the model's phenology and photosynthesis by influencing the water uptake and surface energy balance, as was demonstrated by [19]. In fact, assimilation of surface soil moisture corrects the process parameters in the model's hydrological scheme. Because surface and root zone soil moisture share common parameters, such as soil porosity (which depends on the soil texture in a given gridcell), to calculate soil water properties, the assimilation of surface soil moisture also constrains (through these parameters) plant-available soil moisture.

When assimilating soil moisture only ('sm'), single-site optimizations achieved better soil moisture results for most sites than multi-site assimilation (Table 3). For the FR-Pue site, the single-site optimization resulted in a larger  $J_{sm}$  than the single-site 'sm+fapar' optimization did. This was probably due to the relatively small number of SMOS soil moisture data points at FR-Pue (31) for the period from 2010 to 2015, which moreover showed nearly no temporal variability. An optimized parameter set for this site using these few and relatively constant observations was obviously difficult to compute in a single-site approach. Therefore, in the 'sm+fapar' case, the FAPAR observations provided a stronger constraint through common parameters on the soil moisture simulation than the actual observations at FR-Pue. We also noticed that the smaller  $J_{sm}$  at 'sm+fapar' for single-site optimization is compensated with a larger  $J_p$ . This indicates that due to limited soil moisture observations, model parameters were over-adjusted to achieve a better performance.

Better results for soil moisture were obtained when soil moisture and FAPAR were simultaneously assimilated ('sm+fapar'), with slight differences between the single-site and multi-site cases. This was due to specific site characteristics, which were not easily detectable in the multi-site assimilation and the processes embedded in the model. For example, the sites, AU-DaS, CN-Sw2, FR-Pue, and US-Kfs, are dominated by either grassland or cropland (or both). Water transport processes, however, were difficult to simulate at these sites by our optimization process since they are strongly affected by human management processes, such as irrigation, harvest, or grazing, which are not represented in the model. Another complicated case for our assimilation approach can be seen at the site, CA-Gro, which is located in North America and dominated by a cold climate. The soils at this site are subject to freezing/thawing conditions and snow cover for some periods of time. The associated cold region's soil hydrology is complicated and not fully represented in the model.

When soil moisture and FAPAR were simultaneously assimilated ('sm+fapar'), model parameters related to soil hydrology processes were constrained more than the ones related to phenology processes. There are two reasons for the stronger impact on the model's soil hydrology: (1) The assimilated soil moisture data control directly the surface water and energy balance, which then directly affect plant phenology changes. On the other hand, the FAPAR data have a direct control on the plant phenology, which then only indirectly influences the hydrological cycle. (2) The different temporal resolutions of soil water versus FAPAR data. The FAPAR data have a 16-day temporal resolution, while SMOS L3 data have a daily resolution; hence there were many more SMOS L3 data than FAPAR. In a simultaneous assimilation of both data sets, this favours a stronger constraint on soil moisture than FAPAR.

#### 4.3. Model Representation of In-situ Carbon Flux Observations

In a next step, we performed an independent analysis of the simulation results. In Figure 7, the Net Ecosystem Productivity (NEP) and Gross Primary Productivity (GPP) from site CA-Gro as simulated by the different assimilation experiments are compared with FLUXNET NEP and GPP data. For the whole 10 sites, results are shown in Figure A1. When comparing the eddy covariance measurements against our simulation results, we needed to keep in mind the difference in footprint size of the flux towers (typically a few hundred m<sup>2</sup>) to the spatial resolution of the products used in the assimilation (~600 km<sup>2</sup> for the SMOS soil moisture and ~1 km<sup>2</sup> for the JRC-TIP FAPAR). Therefore, we did not expect a perfect fit against the observations even with the posterior model. In fact, there were large differences of the various optimized NEP values. As shown in Figure 7, the assimilation of SMOS and FAPAR at multi-site achieved the best performance for NEP and GPP for site CA-Gro. For most of the sites, the optimized NEP and GPP captured the seasonality of observations very well, except for two sites, AU-DaS and site US-Wkg (Figures A1 and A2). As above, a major influence of land management processes not represented in the model seems a probable explanation for this deviation. Both sites are dominated by C4 grasses (C4Gr) and evergreen shrubs (EvShr) and are characterized

and grazing.



Figure 7. Simulated (a) net ecosystem productivity (NEP); (b) gross primary productivity (GPP) compared against NEP and GPP data derived from FLUXNET (black dots) at site CA-Gro. The blue line denotes the simulation with the prior parameters, and the red line denotes multi-site assimilation results with both soil moisture and FAPAR used for assimilation ('sm+fapar (multi)').

For the other sites, modelled NEP and GPP matched the fluxes for the non-growing season fluxes much better than for the growing season. For example, at the site CN-Sw2 (Figure A1c), the peak value of modeled NEP was much higher than the observations. For this site, only measurements for the year, 2011, were available. This site experienced a severe drought in 2011, causing a significant decrease in observed NEP during the growing season. Only the prior and soil moisture assimilation experiment in the single-site case detected the changes in July (Figure A1c). A similar decrease in observed NEP was detected at the site, US-Wkg (Figure A1j) between June and July. The discrepancy between modeled and measured NEP indicates again that the model was not able to capture the reduction in NEP in dry seasons.

Table 6 shows the relative changes of RMSE (%) to prior simulation for NEP (RMSE<sub>NEP</sub>) using the eddy covariance measurements at the sites and the simulated NEP from the various experiments. Generally, there was a reduction of the RMSE<sub>NEP</sub> values of up to 43% when compared to the prior simulation. However, especially when only one dataset was assimilated in a single-site experiment, the agreement with the NEP observations was occasionally degraded. In general, better results were achieved when soil moisture and FAPAR were simultaneously assimilated in the multi-site

Table 6. RMSE relative changes (%, positive values mean RMSE was degraded) for net ecosystem
productivity (NEP) in comparison with prior simulations, 'sm' with soil moisture solely, 'fapar' with
FAPAR solely, 'sm+fapar' with combination of soil moisture and FAPAR data, 'multi' denotes multi-site
assimilation, and 'single' denotes single-site assimilation.

Site	Prior	sm (Single)	sm (Multi)	fapar (Single)	fapar (Multi)	sm+fapar (Single)	sm+fapar (Multi)
AU-DaS	1.6232	0.87	2.26	0	3.64	0.97	4.71
CA-Gro	1.0033	5.94	18.01	5.4	18.25	-4.55	-7.24
CN-Sw2	2.6919	-7.25	-29.11	-27.6	-32.85	-20.79	-43.10
DK-Sor	2.3623	/	15.21	-8.24	19.82	-8.24	-0.52
ES-LMa	2.2769	-3.31	3.80	-1.61	12.68	-12.60	16.71
FI-Sod	1.4224	/	-26.36	-2.38	-29.62	-2.38	2.14
FR-Pue	1.6263	18.32	-14.54	-0.76	-17.32	49.89	-0.12
NL-Loo	2.0675	-0.65	6.37	-0.65	6.56	-0.46	-3.90
US-Kfs	1.7244	55.53	-12.35	10.96	3.93	32.82	-0.64
US-Wkg	0.6243	0.27	1.55	-0.02	0	0.21	0.82

Recently, assimilation experiments were conducted at a site-scale using latent heat and FAPAR observations by [18]. They noticed that the combination of two different datasets could lead to a higher uncertainty reduction in model parameters and in other model variables of interest than in single data set cases. In another study, single-site and multi-site assimilation experiments were compared with the CCDAS [51]. They also demonstrated a comparable model performance for multi-site assimilation to single-site assimilation. Also, the multi-site assimilation was found to result in a lower cost function than the single site optimizations [52]. We notice in Table 2 that for single-site and single dataset assimilations, our optimizations could not achieve as low final gradient values as in the multi-site, multi-dataset optimizations.

#### 4.4. Model Representation of Atmospheric CO<sub>2</sub>

We used atmospheric  $CO_2$  concentration observations to corroborate our assimilation results. We used the posterior values from the multi-site, multi-dataset optimization in a global model set-up forced by the global CRUNCEP meteorological driving data, i.e., the terrestrial model simulated global carbon fluxes. The simulated net  $CO_2$  exchange fluxes were then used as input together with a set of background CO<sub>2</sub> fluxes (ocean-atmosphere exchange fluxes, fossil fuel emissions, and land use change emissions) for the global tracer transport model, TM2, to simulate atmospheric  $CO_2$  concentrations for the years, 2010 to 2015. This was essentially the model set-up as described by [19], but here it was only used in a forward modelling sense. Table 7 summarizes the various performance metrics (R<sup>2</sup> and RMSE for the CO<sub>2</sub> time series, and mean error (ME) in the annual growth rates calculated as the method for calculating the annual growth rate in ESRL, https://www.esrl.noaa.gov/gmd/ccgg/mbl/crvfit/crvfit.html) for the two stations, Cape Kumukahi (KUM) and South Pole (SPO, see Table A4 for details on the stations). Obviously, the optimized parameter set improved simulated  $CO_2$  concentrations (Figure A3), as shown by the reduction in the average RMSE values in CO<sub>2</sub> concentrations over the whole time period: At station KUM, RMSE for CO<sub>2</sub> was reduced by 10.4%, and at station SPO, it was reduced by 20.6%. Similarly, the ME in the annual growth rates was improved by 5.5% for KUM and 11.3% for SPO, respectively. This demonstrates that the optimized parameters from multi-site assimilation using SMOS soil moisture and JRC-TIP FAPAR observation can provide a more reasonable parameter set for global scale model experiments. We note, however, that the R<sup>2</sup> values slightly worsened in the 'sm+fapar (multi)' experiment compared to the 'prior', by about 6% at the station KUM. This was mainly due to a reduction of the amplitude

in the seasonal cycle in atmospheric  $CO_2$  for the optimized case since, in contrast to [19], we did not assimilate atmospheric  $CO_2$  observations in this study.

	KUM			SPO		
Experiments	CO <sub>2</sub> Concentrations		Annual CO <sub>2</sub> Growth Rates CO <sub>2</sub>		ncentrations	Annual CO <sub>2</sub> Growth Rates
	R <sup>2</sup>	RMSE (ppm)	ME <sub>GR</sub> (ppm/yr)	R <sup>2</sup>	RMSE (ppm)	ME <sub>GR</sub> (ppm/yr)
prior	0.83	8.75	-2.17	0.99	5.60	-1.68
sm+fapar (multi)	0.78	7.84	-2.05	0.98	4.73	-1.49

**Table 7.** Summary of performance metrics for CO<sub>2</sub> concentrations and CO<sub>2</sub> annual growth rates at stations, KUM and SPO.

# 5. Conclusions

We performed data assimilation experiments at multiple and single sites with two sources of satellite datasets, namely SMOS L3 soil moisture and JRC-TIP FAPAR. These data sets were assimilated in the global terrestrial biosphere model, BETHY, over a time period from 2009 to 2015. The sites for the multiple and single site approach were 10 stations from the FLUXNET network. Model output was also analyzed and compared to independent NEP, GPP, and in-situ soil moisture observations at different sites. This is the first report on experiments that SMOS L3 soil moisture data were used in combination with remotely sensed FAPAR data within site-scale assimilation experiments in order to constrain model parameters.

In all cases, the cost function was reduced significantly, whether a single data set was assimilated or both simultaneously. However, the simultaneous assimilation of both data sets reduced the cost functions for the 10 considered FLUXNET sites to a larger degree (between 16.37% to 69.28%) than the experiments assimilating only one dataset. Parameters related to phenology, soil water, and energy balance processes showed the largest relative changes.

Generally, assimilation of FAPAR observations mainly changed parameters related to phenological processes in the BETHY model, and thus mainly improved simulated FAPAR. SMOS soil moisture, to the contrary, constrained efficiently the soil water and energy balance as computed by the model as indicated by larger changes in the related parameters, such as the emissivity of the atmosphere (*emis*0). In general, the combination of SMOS L3 soil moisture and FAPAR in a multi-site assimilation performed best in fitting both simulated soil moisture and FAPAR. Considering the range of optimal parameters computed in the different experiments, we conclude that a multi-site assimilation with various sources of well-processed observational data streams achieved the best overall results.

Soil moisture has shown to be important in controlling ecosystem production. Adding more detailed processes related to soil water and heat transfer in soil profiles to the model might therefore be key in improving the model response to extremes, such as drought, flood, or frost.

Posterior parameters were also analyzed by comparing the results from the different assimilation experiments. For single sites and single data set assimilation, some global and PFT dependent parameters showed a large range of posterior values. This was probably due to problems with sparse observational data, specific processes associated to a cold/arid climate, and more complicated phenology processes in certain PFT (e.g., crops and grasses) as imposed by management practices. In the assimilation experiments, the emissivity of the atmosphere (*emis*0) covered a wide range of posterior values, in particular considering that *emis*0 is a global parameter. Additional independent data on the radiative balance at the surface, such as outgoing longwave radiation, would help to constrain this parameter.

Comparison of simulated NEP with FLUXNET NEP observations as well as simulated soil moisture with FLUXNET in-situ soil moisture observations gave further evidence for the robustness of

an assimilation approach using both SMOS and FAPAR to constrain the carbon cycle simulation in the terrestrial biosphere model. When comparing simulation results against the in-situ observations, one needs to keep in mind the differences in the spatial representation of the assimilated products with the in-situ observations as mentioned above (see Section 4.3). Nevertheless, the fit against the NEP observations was improved at six of the 10 sites. For soil moisture, we see a reduction of the RMSE error against the in-situ observations of 51.5% and 23.4% for two sites, respectively. Our results from the multi-dataset assimilation (simultaneous assimilation of soil moisture and FAPAR) are consistent with previous studies recommending a combination of these two datasets in constraining parameters in terrestrial biosphere models [18,19]. Also, the evaluation of the model performance on a global scale against atmospheric  $CO_2$  indicated that optimized parameters result in a better fit as demonstrated by, for instance, a reduction of the RMSE of 10.4% against data from Cape Kumukahi and 20.6% against data from South Pole. However, the deterioration of the seasonal cycle amplitude, as mentioned in Section 4.4 is a sign for still unresolved sub-spaces in the total parameter space, i.e., parts of the biosphere model cannot be constrained by soil moisture and FAPAR data alone. By adding atmospheric CO<sub>2</sub> observations in the assimilation, we expect to constrain also these unresolved sub-spaces, at least partly. In general, assimilating further independent and complementary observations will help to constrain such unresolved sub-spaces in the total parameter space.

We demonstrated here that a careful definition of prior parameters and their uncertainties is of importance for an efficient optimization of these parameters. The analysis of the assimilation results, however, revealed that for some sites, the model was not significantly improved. This was especially true for sites covered by grassland and crops, i.e., land cover types that are influenced by management processes, which are not included in the model. Besides a more detailed representation of water and energy balance processes, special attention should be paid to include more sites marked by an even wider spread in climate zones, climate variability, and ecosystem types. Such a wider spread and the addition of additional observations might help to constrain still 'unobserved' (i.e., unobserved by the assimilated soil moisture and FAPAR data) processes in the model as mentioned above. Using remote sensing data has the advantage of providing global data products for the assimilation, which will be further explored in future studies.

**Author Contributions:** M.W. and M.S. conceived and designed this study. M.W. performed the experiments. G.H., M.V. and T.K. contributed to the analysis of the results and all authors co-wrote the manuscript.

**Funding:** This research was funded by the SNSB project, grant number Rymdstyrelsen 102/14, and the ESA project, contract number 4000117645/16/NL/SW. The APC was partly funded by Rymdstyrelsen 102/14 and partly by Lund University.

Acknowledgments: We thank three anonymous reviewers for improving the manuscript. We also thank Johan Nord from Department of Physical Geography and Ecosystem Science at Lund University for processing the CRUNCEP v7 meteorological data, Guillaume Monteil from Department of Physical Geography and Ecosystem Science at Lund University for providing Python code for processing and re-gridding remotely sensed soil moisture data and for fruitful discussions. We acknowledge the FLUXNET2015 (www.fluxdata.org) working group for providing the fluxnet dataset used in our analysis. This work used eddy covariance data acquired and shared by the FLUXNET community, including these networks: AmeriFlux, AfriFlux, AsiaFlux, CarboAfrica, CarboEuropeIP, CarboItaly, CarboMont, ChinaFlux, Fluxnet-Canada, GreenGrass, ICOS, KoFlux, LBA, NECC, OzFlux-TERN, TCOS-Siberia, and USCCC. The ERA-Interim reanalysis data are provided by ECMWF and processed by LSCE. The FLUXNET eddy covariance data processing and harmonization was carried out by the European Fluxes Database Cluster, AmeriFlux Management Project, and Fluxdata project of FLUXNET, with the support of CDIAC and ICOS Ecosystem Thematic Center, and the OzFlux, ChinaFlux and AsiaFlux offices.

Conflicts of Interest: The authors declare no conflict of interest.

# Appendix A

PFT No. Description		
1	Tropical broadleaved evergreen tree (TrEv)	
2	Tropical broadleaved deciduous tree (TrDec)	
3	Temperate broadleaved evergreen tree (TmpEv)	
4	Temperate broadleaved deciduous tree (TmpDec)	
5	Evergreen coniferous tree (EvCn)	
6	Deciduous coniferous tree (DecCn)	
7	Evergreen shrub (EvShr)	
8	Deciduous shrub (DecShr)	
9	C3 grass (C3Gr)	
10	C4 grass (C4Gr)	
11	Tundra vegetation (Tund)	
12	Swamp vegetation (Wetl)	
13	Crops (Crop)	
Soil texture No.	Description	
1	Loamy sand	
2	Sandy loam	
3	Loam	
4	Sandy clay loam	
5	Clay loam	
6	Loam	

Table A1. PFT and soil texture descriptions in BETHY model.

**Table A2.** Parameters to be calibrated in CCDAS, with prior values and uncertainties, parameters could be classified as PFT specified, soil texture specified, and global parameters that all sites share one value, as denoted after each parameter.

No.	Parameter	Description	Prior Value	Prior Uncertainty
1			$60  imes 10^{-6}$	$12  imes 10^{-6}$
2			$90  imes 10^{-6}$	$18  imes 10^{-6}$
3			$41 \times 10^{-6}$	$8.2  imes 10^{-6}$
4			$35  imes 10^{-6}$	$7 \times 10^{-6}$
5			$29  imes 10^{-6}$	$5.8 imes10^{-6}$
6			$53  imes 10^{-6}$	$10.6  imes 10^{-6}$
7	$V_{max}^{25}/k$ (PFT)	Maximum carboxylation rate (C3/C4)	$52  imes 10^{-6}$	$10.4  imes 10^{-6}$
8			$160  imes 10^{-6}$	$32  imes 10^{-6}$
9			$42  imes 10^{-6}$	$8.4 imes10^{-6}$
10			$8 imes 10^{-6}$	$1.6 imes10^{-6}$
11			$20  imes 10^{-6}$	$4 imes 10^{-6}$
12			$10  imes 10^{-6}$	$4 imes 10^{-6}$
13			$117  imes 10^{-6}$	$23.4  imes 10^{-6}$
14			1.96	0.098
15			1.99	0.0995
16			2.00	0.1
17			2.00	0.1
18			1.79	0.0895
19		Ratio V to I (maximum electron	1.79	0.0895
20	$a_{J,V}$ (PFT)	transport rate	1.96	0.098
21		unsport fue	1.66	0.083
22			1.90	0.095
23			$140  imes 10^{-6}$	$28  imes 10^{-6}$
24			1.85	0.0925
25			1.85	0.0925
26			1.88	0.094

No.	Parameter	Description	Prior Value	Prior Uncertainty
27	$f_{r,leaf}$ (Global)	Leaf respiration ratio	0.40	0.1
28	$f_{r  growth}$ (Global)	Growth respiration ratio	0.2	0.01
29	$O_{10 f}$ (Global)	Soil respiration temperature factor, fast pool	1.5	0.75
30	$O_{10,j}$ (Global)	Soil respiration temperature factor slow pool	1.5	0.75
31	$\tau_{\rm Clobal}$	Fast pool soil carbon turnover time	0.66	0.4
22	r (Clobal)	Soil maisture exponential for soil respiration	1.0	1.0
32	$f_{\rm c}$ (Global)		1.0	1.0
33	$f_S$ (Global)		0.2	0.2
34	$E_{R_d}$ (Global)	Activation energy, dark respiration	45,000	2250
35	$E_{V_{max}}$ (Global)	Activation energy, carboxylation rate (C3)	58,520	2926
36	$E_{K_O}$ (Global)	Activation energy, O <sub>2</sub>	35,948	1797
37	$E_{K_{\rm C}}$ (Global)	Activation energy, $CO_2$	59,356	2967
38	$E_k$ (Global)	Activation energy, carboxylation rate (C4)	50,967	2548
39	$\alpha_q$ (Global)	Photon capture efficiency (C3)	0.28	0.014
40	$\alpha_i$ (Global)	Quantum efficiency (C4)	0.04	0.002
41	$K_C^{25}$ (Global)	Michaelis-Menten constant CO <sub>2</sub>	$460  imes 10^{-6}$	$23  imes 10^{-6}$
42	$K_o^{25}$ (Global)	Michaelis-Menten constant O <sub>2</sub>	0.33	0.0165
43	$a_{\Gamma,T}$ (Global)	Temperature slope CO <sub>2</sub> compensation point	$1.6 imes10^{-6}$	$8.5 imes10^{-8}$
44			1.0	0.25
45			1.0	0.25
46 47			1.0	0.25
47			1.0	0.25
49			1.0	0.25
50	$\beta$ (PFT)	Net $CO_2$ sink/source factor	1.0	0.25
51			1.0	0.25
52			1.0	0.25
53			1.0	0.25
54			1.0	0.25
55			1.0	0.25
56			1.0	0.25
57	$\Lambda \sim$ (Global)	Maximum Leaf Area Index	4.0	0.1
58 59			10.0	1.0
60	$T_{\phi}$ (PFT)	Phenology temperature	2.0	2.0
61			15.0	2.0
62	T (DET)	Spatial range of phanalegy temperature trigger	2.0	1.0
63	$I_r$ (PF1)	Spatial range of phenology temperature trigger	2.0	1.0
64	$t_c$ (PFT)	Day length at leaf shedding	10.5	1.0
65	$t_r$ (PFT)	Spatial range of day length at leaf shedding	0.5	0.25
66	$\xi$ (Global)	Initial linear leaf growth	0.5	0.5
67 68	$k_L$ (PFT)	Inverse of leaf longevity	$\begin{matrix} 0.1 \\ 5\times 10^{-3} \end{matrix}$	$\begin{array}{c} 0.1\\5\times10^{-3}\end{array}$
69			30	30
70	$ au_W$ (PFT)	Length of dry spell before leaf shedding	30	30
71			30	30
72	$fC_i, C_3$ (Global)	Stomata internal to atmospheric $CO_2$ (C3)	0.65	0.065
73	$fC_i, C_4$ (Global)	Stomata internal to atmospheric CO <sub>2</sub> (C4)	0.37	0.037
74	C <sub>W0</sub> (Global)	Ratio of maximum water supply rate	0.5	0.5
75	emis0 (Global)	Emissivity of the atmosphere	0.64	0.032

# Table A2. Cont.

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

Parameter

 $s_{rdepth}$  (PFT)

 $s_b$  (Soil texture)

 $s_n$  (Soil texture)

Table A2. Cont.					
Description	Prior Value	Prior Uncertainty			
	1.0	0.1			
	1.0	0.1			
	1.0	0.1			
	1.0	0.1			
	1.0	0.1			
	1.0	0.1			

1.0

1.0

1.0

1.0

1.0

1.0

1.0

1.0

1.0

1.0

1.0

1.0

1.0

1.0

1.0

1.0

1.0

1.0

1.0

**Table A3.** RMSE reduction (%, positive value means RMSE was reduced) for gross primary productivity (GPP) simulation with prior and different assimilation methods, 'sm' with soil moisture solely, 'fapar' with FAPAR solely, 'sm+fapar' with combination of soil moisture and FAPAR data, 'multi' denotes multi-site assimilation, and 'single' denotes single-site assimilation.

Rooting depth scalar

Shape parameter retention curve scalar

Soil porosity scalar

Site	Prior	fapar (Multi)	fapar (Single)	sm (Multi)	sm (Single)	sm+fapar (Multi)	sm+fapar (Single)
AU-DaS	2.5496	63.01	19.43	35.99	26.78	19.43	0.00
CA-Gro	1.1403	63.49	34.96	34.21	50.49	15.79	13.36
CN-Sw2	3.8561	-61.80	-39.90	1.10	-17.86	-45.03	-41.82
DK-Sor *	4.6821	-17.34	-14.23	4.64	/	32.21	-14.23
ES-LMa	1.7476	43.52	-12.89	-9.72	30.43	7.91	-4.72
FI-Sod *	1.5081	22.76	-6.70	-25.74	/	-32.97	-6.70
FR-Pue	4.3716	-37.53	7.62	-46.73	3.38	-35.25	-0.36
NL-Loo	2.1479	-20.50	-8.23	33.75	-7.97	24.87	-5.68
US-Kfs	3.5734	-15.22	2.31	-42.75	-11.02	-10.28	-22.20
US-Wkg	1.3939	-0.91	-0.50	-0.81	-0.48	0.00	0.00

\* no soil moisture data available.

Table A4. CO<sub>2</sub> observation stations and standard error of fit from Keeling et al. (2001).

Station Name	Station Code	Latitude	Longitude	Elevation (m)	Standard Error of Fit (ppm)
Cape Kumukahi, Hawaii	KUM	19.5°N	154.8°W	3	0.7818
South PoleChristmas Island	SPO	90.0°S	157.3°W	28,102	0.1469

0.1

0.1

0.1

0.1

0.1

0.1

0.1

0.1

0.1

0.1

0.1

0.1

0.1

0.1

0.1

0.1

0.1

0.1

0.1

-10



**Figure A1.** Multi-year mean of simulated daily net ecosystem productivity (NEP) and observed data (dot) from FLUXNET. (**a**–**j**) corresponds to 10 sites listed in Table 6. The black line denotes simulation with prior parameter values, the blue line denotes multi-site assimilation result with soil moisture and FAPAR, the green line denotes multi-site assimilation result with soil moisture, the red line denotes multi-site assimilation result with soil moisture, the red line denotes multi-site and FAPAR from each site grid, the magenta line denotes single-site assimilation result with soil moisture from each site grid, and the dark yellow line denotes single-site assimilation results with FAPAR from each site grid.

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

-10

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec



**Figure A2.** Multi-year mean of simulated daily gross primary productivity (GPP) and observed data (dot) from FLUXNET. (**a**–**j**) corresponds to 10 sites listed in Table A3. The black line denotes simulation with prior parameter values, the blue line denotes multi-site assimilation result with soil moisture and FAPAR, the green line denotes multi-site assimilation result with soil moisture, the red line denotes multi-site assimilation result with soil moisture, the red line denotes multi-site assimilation result with soil moisture and FAPAR, the green and FAPAR, the cyan line denotes single-site assimilation result with soil moisture and result with form each site grid, the magenta line denotes single-site assimilation result with soil moisture from each site grid, and the dark yellow line denotes single-site assimilation results with FAPAR from each site grid.



**Figure A3.** Comparison of observed CO<sub>2</sub> concentrations from two sites from the Scripps Institution of Oceanography network with simulations from different experiments, (**a**) for site KUM; (**b**) for site SPO. 'Prior' represents simulations with prior parameter values, 'sm+fapar (multi)' represents simulations with optimized parameter values from multi-site assimilation using SMOS soil moisture and FAPAR.

#### References

- Pachauri, R.K.; Allen, M.R.; Barros, V.R.; Broome, J.; Cramer, W.; Christ, R.; Church, J.A.; Clarke, L.; Dahe, Q.; Dasgupta, P.; et al. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*; IPCC: Geneva, Switzerland, 2014.
- Friedlingstein, P.; Cox, P.; Betts, R.; Bopp, L.; von Bloh, W.; Brovkin, V.; Cadule, P.; Doney, S.; Eby, M.; Fung, I.; et al. Climate–carbon cycle feedback analysis: Results from the C4MIP model intercomparison. *J. Clim.* 2006, *19*, 3337–3353. [CrossRef]
- 3. Field, C.B.; Jackson, R.B.; Mooney, H.A. Stomatal responses to increased CO<sub>2</sub>: Implications from the plant to the global scale. *Plant Cell Environ*. **1995**, *18*, 1214–1225. [CrossRef]
- Wullschleger, S.D.; Epstein, H.E.; Box, E.O.; Euskirchen, E.S.; Goswami, S.; Iversen, C.M.; Kattge, J.; Norby, R.J.; van Bodegom, P.M.; Xu, X. Plant functional types in Earth system models: Past experiences and future directions for application of dynamic vegetation models in high-latitude ecosystems. *Ann. Bot.* 2014, 114, 1–16. [CrossRef] [PubMed]
- Le Quéré, C.; Moriarty, R.; Andrew, R.M.; Canadell, J.G.; Sitch, S.; Korsbakken, J.I.; Friedlingstein, P.; Peters, G.P.; Andres, R.J.; Boden, T.A.; et al. Global carbon budget 2015. *Earth Syst. Sci. Data* 2015, 7, 349–396. [CrossRef]
- 6. Kaminski, T.; Knorr, W.; Rayner, P.; Heimann, M. Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle. *Glob. Biogeochem. Cycles* **2002**, *16*, 14-1–14-16. [CrossRef]
- Rodell, M.; Houser, P.R.; Jambor, U.E.A.; Gottschalck, J.; Mitchell, K.; Meng, C.J.; Arsenault, K.; Cosgrove, B.; Radakovich, J.; Bosilovich, M.; et al. The global land data assimilation system. *Bull. Am. Meteorol. Soc.* 2004, 85, 381–394. [CrossRef]
- 8. Sawada, Y.; Koike, T.; Walker, J.P. A land data assimilation system for simultaneous simulation of soil moisture and vegetation dynamics. *J. Geophys. Res. Atmos.* **2015**, *120*, 5910–5930. [CrossRef]
- Albergel, C.; Munier, S.; Leroux, D.J.; Dewaele, H.; Fairbairn, D.; Barbu, A.L.; Gelati, E.; Dorigo, W.; Faroux, S.; Meurey, C.; et al. Sequential assimilation of satellite-derived vegetation and soil moisture products using SURFEX\_v8. 0: LDAS-Monde assessment over the Euro-Mediterranean area. *Geosci. Model Dev.* 2017, 10, 3889–3912. [CrossRef]
- 10. Rayner, P.; Scholze, M.; Knorr, W.; Kaminski, T.; Giering, R.; Widmann, H. Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS). *Glob. Biogeochem. Cycles* **2005**, *19*. [CrossRef]
- 11. Scholze, M.; Kaminski, T.; Rayner, P.; Knorr, W.; Giering, R. Propagating uncertainty through prognostic carbon cycle data assimilation system simulations. *J. Geophys. Res. Atmos.* **2007**, *112*. [CrossRef]
- 12. Knorr, W.; Heimann, M. Uncertainties in global terrestrial biosphere modeling: 1. A comprehensive sensitivity analysis with a new photosynthesis and energy balance scheme. *Glob. Biogeochem. Cycles* **2001**, *15*, 207–225. [CrossRef]
- 13. Solomon, S. Climate Change 2007—The Physical Science Basis: Working Group I Contribution to the Fourth Assessment Report of the IPCC; Cambridge University Press: Cambridge, UK, 2007; Volume 4.
- 14. Santaren, D.; Peylin, P.; Bacour, C.; Ciais, P.; Longdoz, B. Ecosystem model optimization using in situ flux observations: Benefit of Monte Carlo versus variational schemes and analyses of the year-to-year model performances. *Biogeosciences* **2014**, *11*, 7137–7158. [CrossRef]

- 15. Knorr, W.; Schulz, J.-P. Using satellite data assimilation to infer global soil moisture status and vegetation feedback to climate. *Remote Sens. Clim. Model. Synerg. Limit. Adv. Glob. Chang. Res. Ser.* **2001**, *7*, 273–306.
- 16. Knorr, W.; Kaminski, T.; Scholze, M.; Gobron, N.; Pinty, B.; Giering, R.; Mathieu, P.P. Carbon cycle data assimilation with a generic phenology model. *J. Geophys. Res. Biogeosci.* **2010**, *115*. [CrossRef]
- 17. Ziehn, T.; Kattge, J.; Knorr, W.; Scholze, M. Improving the predictability of global CO<sub>2</sub> assimilation rates under climate change. *Geophys. Res. Lett.* **2011**, *38*. [CrossRef]
- 18. Kato, T.; Knorr, W.; Scholze, M.; Veenendaal, E.; Kaminski, T.; Kattge, J.; Gobron, N. Simultaneous assimilation of satellite and eddy covariance data for improving terrestrial water and carbon simulations at a semi-arid woodland site in Botswana. *Biogeosciences* **2013**, *10*, 789–802. [CrossRef]
- Scholze, M.; Kaminski, T.; Knorr, W.; Blessing, S.; Vossbeck, M.; Grant, J.; Scipal, K. Simultaneous assimilation of SMOS soil moisture and atmospheric CO<sub>2</sub> in-situ observations to constrain the global terrestrial carbon cycle. *Remote Sens. Environ.* 2016, 180, 334–345. [CrossRef]
- 20. Bacour, C.; Peylin, P.; MacBean, N.; Rayner, P.J.; Delage, F.; Chevallier, F.; Weiss, M.; Demarty, J.; Santaren, D.; Baret, F.; et al. Joint assimilation of eddy covariance flux measurements and FAPAR products over temperate forests within a process-oriented biosphere model. *J. Geophys. Res. Biogeosci.* **2015**, *120*, 1839–1857. [CrossRef]
- Schürmann, G.; Kaminski, T.; Köstler, C.; Carvalhais, N.; Voßbeck, M.; Kattge, J.; Giering, R.; Rödenbeck, C.; Heimann, M.; Zaehle, S. Constraining a land-surface model with multiple observations by application of the MPI-Carbon Cycle Data Assimilation System V1.0. *Geosci. Model Dev.* 2016, *9*, 2999–3026. [CrossRef]
- 22. Kaminski, T.; Knorr, W.; Scholze, M.; Gobron, N.; Pinty, B.; Giering, R.; Mathieu, P.-P. Consistent assimilation of MERIS FAPAR and atmospheric CO<sub>2</sub> into a terrestrial vegetation model and interactive mission benefit analysis. *Biogeosciences* **2012**, *9*, 3173–3184. [CrossRef]
- Krinner, G.; Viovy, N.; de Noblet-Ducoudré, N.; Ogée, J.; Polcher, J.; Friedlingstein, P.; Ciais, P.; Sitch, S.; Prentice, I.C. A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Glob. Biogeochem. Cycles* 2005, 19. [CrossRef]
- 24. Lievens, H.; Tomer, S.K.; Al Bitar, A.; De Lannoy, G.J.; Drusch, M.; Dumedah, G.; Franssen, H.J.H.; Kerr, Y.H.; Martens, B.; Pan, M.; et al. SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia. *Remote Sens. Environ.* **2015**, *168*, 146–162. [CrossRef]
- Li, Y.; Grimaldi, S.; Pauwels, V.R.; Walker, J.P. Hydrologic model calibration using remotely sensed soil moisture and discharge measurements: The impact on predictions at gauged and ungauged locations. *J. Hydrol.* 2018, 557, 897–909. [CrossRef]
- 26. Pinnington, E.; Quaife, T.; Black, E. Impact of remotely sensed soil moisture and precipitation on soil moisture prediction in a data assimilation system with the JULES land surface model. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 2575–2588. [CrossRef]
- 27. Van der Molen, M.K.; De Jeu, R.A.M.; Wagner, W.; Van Der Velde, I.R.; Kolari, P.; Kurbatova, J.; Varlagin, A.; Maximov, T.C.; Kononov, A.V.; Ohta, T.; et al. The effect of assimilating satellite-derived soil moisture data in SiBCASA on simulated carbon fluxes in Boreal Eurasia. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 605–624. [CrossRef]
- 28. Barbu, A.; Calvet, J.-C.; Mahfouf, J.-F.; Lafont, S. Integrating ASCAT surface soil moisture and GEOV1 leaf area index into the SURFEX modelling platform: A land data assimilation application over France. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 173–192. [CrossRef]
- 29. Nemani, R.R.; Keeling, C.D.; Hashimoto, H.; Jolly, W.M.; Piper, S.C.; Tucker, C.J.; Myneni, R.B.; Running, S.W. Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science* 2003, 300, 1560–1563. [CrossRef]
- 30. Dumedah, G.; Walker, J.P. Intercomparison of the JULES and CABLE land surface models through assimilation of remotely sensed soil moisture in southeast Australia. *Adv. Water Resour.* **2014**, *74*, 231–244. [CrossRef]
- 31. Ridler, M.E.; Madsen, H.; Stisen, S.; Bircher, S.; Fensholt, R. Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark. *Water Resour. Res.* **2014**, *50*, 8962–8981. [CrossRef]
- 32. Wanders, N.; Karssenberg, D.; de Roo, A.; De Jong, S.; Bierkens, M. The suitability of remotely sensed soil moisture for improving operational flood forecasting. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 2343–2357. [CrossRef]
- Bartalis, Z.; Wagner, W.; Naeimi, V.; Hasenauer, S.; Scipal, K.; Bonekamp, H.; Figa, J.; Anderson, C. Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). *Geophys. Res. Lett.* 2007, 34. [CrossRef]

- 34. Scholze, M. Model Studies on the Response of the Terrestrial Carbon Cycle to Climate Change and Variability. Ph.D. Thesis, Universität Hamburg, Hamburg, Germany, 2003.
- 35. Kaminski, T.; Knorr, W.; Schürmann, G.; Scholze, M.; Rayner, P.J.; Zaehle, S.; Blessing, S.; Dorigo, W.; Gayler, V.; Giering, R.; et al. The BETHY/JSBACH carbon cycle data assimilation system: Experiences and challenges. *J. Geophys. Res. Biogeosci.* **2013**, *118*, 1414–1426. [CrossRef]
- 36. Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation;* SIAM: Philadelphia, PA, USA, 2005.
- 37. Santaren, D.; Peylin, P.; Viovy, N.; Ciais, P. Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France. *Glob. Biogeochem. Cycles* **2007**, *21*. [CrossRef]
- 38. Wood, E.F.; Lettenmaier, D.P.; Zartarian, V.G. A land-surface hydrology parameterization with subgrid variability for general circulation models. *J. Geophys. Res. Atmos.* **1992**, *97*, 2717–2728. [CrossRef]
- 39. Kerr, Y.H.; Waldteufel, P.; Wigneron, J.P.; Delwart, S.; Cabot, F.; Boutin, J.; Escorihuela, M.J.; Font, J.; Reul, N.; Gruhier, C.; et al. The SMOS mission: New tool for monitoring key elements of the global water cycle. *Proc. AEEE* **2010**, *98*, 666–687. [CrossRef]
- Clapp, R.B.; Hornberger, G.M. Empirical equations for some soil hydraulic properties. *Water Resour. Res.* 1978, 14, 601–604. [CrossRef]
- 41. Nijssen, B.; O'Donnell, G.M.; Lettenmaier, D.P.; Lohmann, D.; Wood, E.F. Predicting the discharge of global rivers. *J. Clim.* **2001**, *14*, 3307–3323. [CrossRef]
- 42. McColl, K.A.; Vogelzang, J.; Konings, A.G.; Entekhabi, D.; Piles, M.; Stoffelen, A. Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophys. Res. Lett.* **2014**, *41*, 6229–6236. [CrossRef]
- 43. Dorigo, W.A.; Wagner, W.; Hohensinn, R.; Hahn, S.; Paulik, C.; Xaver, A.; Gruber, A.; Drusch, M.; Mecklenburg, S.; Oevelen, P.V.; et al. The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 1675–1698. [CrossRef]
- 44. Pinty, B.; Lavergne, T.; Voßbeck, M.; Kaminski, T.; Aussedat, O.; Giering, R.; Gobron, N.; Taberner, M.; Verstraete, M.M.; Widlowski, J.L.; et al. Retrieving surface parameters for climate models from Moderate Resolution Imaging Spectroradiometer (MODIS)-Multiangle Imaging Spectroradiometer (MISR) albedo products. J. Geophys. Res. Atmos. 2007, 112. [CrossRef]
- Pinty, B.; Gobron, N.; Widlowski, J.L.; Gerstl, S.A.; Verstraete, M.M.; Antunes, M.; Bacour, C.; Gascon, F.; Gastellu, J.P.; Goel, N.; et al. Radiation transfer model intercomparison (RAMI) exercise. *J. Geophys. Res. Atmos.* 2001, 106, 11937–11956. [CrossRef]
- 46. Kaminski, T.; Pinty, B.; Voßbeck, M.; Lopatka, M.; Gobron, N.; Robustelli, M. Consistent retrieval of land surface radiation products from EO, including traceable uncertainty estimates. *Biogeosciences* **2017**, *14*, 2527–2541. [CrossRef]
- 47. Voßbeck, M.; Clerici, M.; Kaminski, T.; Lavergne, T.; Pinty, B.; Giering, R. An inverse radiative transfer model of the vegetation canopy based on automatic differentiation. *Inverse Probl.* **2010**, *26*, 095003. [CrossRef]
- 48. Pinty, B.; Andredakis, I.; Clerici, M.; Kaminski, T.; Taberner, M.; Verstraete, M.M.; Gobron, N.; Plummer, S.; Widlowski, J.L. Exploiting the MODIS albedos with the Two-stream Inversion Package (JRC-TIP): 1. Effective leaf area index, vegetation, and soil properties. *J. Geophys. Res. Atmos.* **2011**, *116*. [CrossRef]
- 49. Kaminski, T.; Scholze, M.; Vossbeck, M.; Knorr, W.; Buchwitz, M.; Reuter, M. Constraining a terrestrial biosphere model with remotely sensed atmospheric carbon dioxide. *Remote Sens. Environ.* **2017**, 203, 109–124. [CrossRef]
- 50. Keeling, C.D.; Piper, S.C.; Bacastow, R.B.; Wahlen, M.; Whorf, T.P.; Heimann, M.; Meijer, H.A. *Exchanges* of Atmospheric CO<sub>2</sub> and 13CO<sub>2</sub> with the Terrestrial Biosphere and Oceans from 1978 to 2000. I. Global Aspects; Scripps Institution of Oceanography: San Diego, CA, USA, 2001.
- Kuppel, S.; Peylin, P.; Maignan, F.; Chevallier, F.; Kiely, G.; Montagnani, L.; Cescatti, A. Model-data fusion across ecosystems: From multisite optimizations to global simulations. *Geosci. Model Dev.* 2014, 7, 2581–2597. [CrossRef]
- 52. Kuppel, S.; Peylin, P.; Chevallier, F.; Bacour, C.; Maignan, F.; Richardson, A.D. Constraining a global ecosystem model with multi-site eddy-covariance data. *Biogeosciences* **2012**, *9*, 3757–3776. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).