Title
Assessment of SMAP soil moisture for global simulation of gross primary production

Journal
Journal of Geophysical Research Biogeosciences, 122(7)

ISSN
2169-8953

Authors
He, Liming
Chen, Jing M
Liu, Jane
et al.

Publication Date
2017-07-01

DOI
10.1002/2016jg003603

Peer reviewed

Assessment of SMAP soil moisture for global simulation of gross primary production

Liming He

Jing M. Chen

Jane Liu

Stéphane Bélair

Xiangzhong Luo
First published: 16 June 2017
https://doi.org/10.1002/2016JG003603
Cited by: 4
UC-eLinks

Abstract

In this study, high-quality soil moisture data derived from the Soil Moisture Active Passive (SMAP) satellite measurements are evaluated from a perspective of improving the estimation of the global gross primary production (GPP) using a process-based ecosystem model, namely, the Boreal Ecosystem Productivity Simulator (BEPS). The SMAP soil moisture data are assimilated into BEPS using an ensemble Kalman filter. The correlation coefficient (r) between simulated GPP from the sunlit leaves and Sun-induced chlorophyll fluorescence (SIF) measured by Global Ozone Monitoring Experiment-2 is used as an indicator to evaluate the performance of the GPP simulation. Areas with SMAP data in low quality (i.e., forests), or with SIF in low magnitude (e.g., deserts), or both are excluded from the analysis. With the assimilated SMAP data, the r value is enhanced for Africa, Asia, and North America by 0.016, 0.013, and 0.013, respectively ($p < 0.05$). Significant improvement in r appears in single-cropping agricultural land where the irrigation is not considered in the model but well captured by SMAP (e.g., 0.09 in North America, $p < 0.05$). With the assimilation of SMAP, areas with weak model performances are identified in double or triple cropping cropland (e.g., part of North China Plain) and/or mountainous area (e.g., Spain and Turkey). The correlation coefficient is enhanced by 0.01 in global average for shrub, grass, and cropland. This enhancement is small and insignificant because nonwater-stressed areas are included.

1 Introduction

Gross primary production (GPP) is an important component in the global carbon cycle [*Gitelson et al.*, 2006; *Le Quere et al.*, 2015]. Quantitative estimates of the spatial and temporal distributions of GPP on regional to global scales are critical to the understanding of climate-carbon cycle feedback [*Xia et al.*, 2015]. GPP is widely modeled using process-based ecosystem models [*Morales et al.*, 2005]. Soil moisture plays a crucial role in vegetative processes and links

the carbon, water, and energy cycles in the physical climate system and biogeochemical cycles [*Seneviratne et al.*, 2010]. Therefore, soil moisture is one of the key variables in ecosystem models that constrains stomatal conductance and controls both plant water use and carbon uptake [*G. B. Bonan et al.*, 2014; *Liming He et al.*, 2014; *Mu et al.*, 2007]. However, a large uncertainty is associated with simulating soil moisture in these ecosystem models in applications on global and continental scales, because the simulation relies on the quality of input precipitation data, which often have large errors. Therefore, the estimate of terrestrial GPP is often biased due to errors in meteorological data sets [*Barman et al.*, 2014].

Soil moisture varies greatly in space and time so it is challenging to accurately map its variations. Recently, the Soil Moisture and Ocean Salinity (SMOS) and Soil Moisture Active Passive (SMAP) missions have been providing global maps of surface soil moisture with a target accuracy of 0.04 $m^3\ m^{-3}$ based on L-band radiometric measurements [*Entekhabi et al.*, 2010; *Font et al.*, 2001; *Kerr et al.*, 2012]. SMAP and SMOS are designed to map global soil moisture in a temporal resolution of every 2–3 days to improve our understanding of the coupling energy, water, and carbon cycles. It has been shown that assimilation of such remotely sensed surface soil moisture [*Kerr et al.*, 2012] can improve the modeling of various land surface processes [*Han et al.*, 2014; *Wanders et al.*, 2014].

Up to now, the potential capability of new SMAP soil moisture data to improve the performance of ecosystem models has not yet been well explored. Recently, it has been shown that spaceborne Sun-induced chlorophyll fluorescence (SIF) observations offer new possibilities for monitoring photosynthesis from space [*Frankenberg et al.*, 2011; *L. Guanter et al.*, 2014]. SIF captures the seasonal variation of GPP and can be used as a reliable proxy for validating the trend of GPP [*Lee et al.*, 2013; *Wagle et al.*, 2016]. The objective of this study is to evaluate the improvement of the global GPP simulation with the assimilated SMAP soil moisture in Boreal Ecosystem Productivity Simulator (BEPS) against SIF observations in the period of April to December 2015. This paper is organized in five sections starting with Introduction as section 1. Section 2 describes the ecosystem model, data assimilation system, and materials. The experiments for evaluating model performance are detailed in section 3. Section 4 provides the results and discussions. Finally, the conclusions are drawn in section 5.

2 Model, Method, and Materials

2.1 The BEPS Model

The Boreal Ecosystems Productivity Simulator (BEPS) is a process-based ecosystem model including water, energy, and carbon budgets and soil thermal transfer modules [*Chen et al.*, 2007; *Chen et al.*, 1999; *Chen et al.*, 2012]. In this model, GPP is modeled by scaling Farquhar's leaf-level biochemical model [*Farquhar et al.*, 1980] up to the canopy level using a "two-leaf" approach [*Chen et al.*, 1999; *Norman*, 1982]. The bulk stomatal conductances of the sunlit and shaded leaves for water vapor and CO_2 are calculated using a modified Ball-Woodrow-Berry (BWB) stomatal model [*Ball et al.*, 1987]. The Penman-Monteith equation [*Monteith*, 1965] is used to calculate the evaporation of intercepted water from the canopy and the ground surface, and canopy transpiration from sunlit and shaded leaves is computed following *Wang and Leuning* [1998]. The soil water dynamics is governed by the Richards equation [*Chen et al.*, 2007]. The soil profile is stratified in five layers with depths of 0.05 m, 0.10 m, 0.20 m, 0.40 m, and 1.2 m from top layer to bottom layer.

In BEPS, the influence of soil water on GPP is modeled through the modified BWB equation following *G. B. Bonan* [1995] and *Ju et al.* [2006]:

$$g = f_w\left(m\frac{Ah_s}{C_s}\right) + b \quad (1)$$

where g is the stomatal conductance (μmol m^{-2} s^{-1}) in leaf level; A is the net photosynthesis rate (μmol m^{-2} s^{-1}); m is a plant species dependent coefficient; h_s and C_s are the relative humidity and CO_2 concentration at the leaf surface, respectively; b is the residual conductance; and f_w is a soil water stress factor to adjust the slope of BWB equation and modeled as [*Chen et al.*, 2012; *Ju et al.*, 2006]:

$$f_w = \sum_{i=1}^{n} f_{w,i} w_i \quad (2)$$

where $f_{w,i}$ is the soil water stress factor in layer i and calculated as

$$f_{w,i} = \frac{1.0}{f_i(\psi_i) f_i\left(T_{s,i}\right)} \quad (3)$$

where $f_i(\psi_i)$ is a function of matrix suction $\psi_i(m)$ [*Zierl*, 2001]:

$$f_i(\psi_i) = \begin{cases} 1.0 + \left[\dfrac{\psi_i - 10.0}{10.0}\right]^{\alpha} & \psi_i > 10 \\ 1.0 & \text{else} \end{cases} \quad (4)$$

where α is suggested to be a function of plant type [*Chen et al.*, 2012]. The effect of soil temperature on soil water uptake is described as follows [*Gordon B. Bonan*, 1991]:

$$f_i\left(T_{s,i}\right) = \begin{cases} \dfrac{1.0}{1 - \exp\left(t_1 T_{s,i}^{t2}\right)} & T_{s,i} > 0 \\ \infty & else \end{cases} \quad (5)$$

where t_1 and t_2 are parameters determining the sensitivity of root water uptake to soil temperature and $T_{s,i}$ is the soil temperature at the ith layer. In BEPS, $t_1 = -0.02$ and $t_2 = 2.0$. To consider the variable soil water potential at different depths, the weight w_i is calculated as

$$w_i = \frac{R_i f_{w,i}}{\sum_{i=1}^{n} R_i f_{w,i}} \quad (6)$$

where R_i is the root fraction in layer i. The f_w is updated in an hourly step in BEPS.

2.2 The Data Assimilation of SMAP Soil Moisture

Data Assimilation of soil Moisture in Parallel (DAMP) is a package primarily developed to run at High Performance Computer in parallel mode for assimilating satellite-observed surface soil moisture into ecosystem models for this study. In this study, the SMAP soil moisture is assimilated into BEPS using DAMP in an hourly interval. In DAMP, BEPS is run on the Earth-fixed, global, cylindrical 36 km Equal-Area Scalable Earth Grid, Version 2.0 (EASE-Grid 2.0, 964 columns and 406 lines) [*Brodzik et al.*, 2014] in 36 km resolution. When SMAP data are available, an ensemble Kalman filter (EnKF) is used to adjust soil moisture states:

$$X_k^a = X_k^f + K\left(Y_k - H\left(X_k^f\right)\right) \quad (7)$$

where X_k^a and X_k^f are the analyzed and forecast soil moisture, respectively, in five layers at a time instant k; Y_k refers to the vector of corrected surface soil moisture observations from satellite using cumulative distribution functions (CDF) matching approach [*Reichle and Koster*, 2004]; and H is the measurement operator that maps the model state X_k, e.g., the predicted soil moisture, to Y_k, e.g., the observations. K is the Kalman gain matrix:

$$K = P_k^f H^T \left(H P_k^f H^T + R_k\right)^{-1} \quad (8)$$

where R is a diagonal matrix with each element indicating the soil moisture error variance and $P_k^f H^T$ is the cross covariance between any given state and prediction $H\left(X_k^f\right)$,

$$P_k^f H^T = \frac{\left(X_k^f - \overline{X}_k^f\right)}{u-1} q_k^T \quad (9)$$

$$q_k = H\left(X_k^f - \overline{X}_k^f\right) = \left(y_k^f - \overline{y}_k^f\right) \quad (10)$$

$H P_k^f H^T$ is the error covariance matrix of the prediction:

$$H P_k^f H^T = \frac{q_k q_k^T}{u-1} \quad (11)$$

and,

$$P_k^f = \frac{\left(X_k^f - \overline{X}_k^f\right)\left(X_k^f - \overline{X}_k^f\right)^T}{u-1} \quad (12)$$

u is the number of members in the ensemble.

One of the large limitations for EnKF-based DA systems is the resource-limited ensemble size. In our DA system, the model state vector size is on the order of one million (964*406*6*30% of the Earth's surface area) even for such a coarse spatial resolution (36 km). A localization method, by artificially reducing the spatial domain of influence of observations during the update, is usually used to reduce the necessary ensemble size by a few orders [*Sakov and Bertino*, 2011]. Another reason for using the localization method in EnKF is to avoid "spurious covariances," referring to covariances between distant or physically not connected state vector elements. In this study, the retrieved soil moisture from satellite measurement can be safely treated as local observations; i.e., the soil moisture value for one pixel is not correlated to its neighbor's values. With this assumption, the ensemble size in our DA system is reduced to the order of 10,000.

In this study, we used the covariance localization method, which modifies the update equation by replacing the state error covariance by its element-wise (Schur or Hadamard) product with a distance-based correlation matrix ρ [*Hamill et al.*, 2001; *Houtekamer and Mitchell*, 2001]:

$$P \longrightarrow \rho \circ P \quad (13)$$

In our DA system, the localized state error covariance matrix is not calculated explicitly. Instead, the terms $\left(\rho \circ P_k^f\right)H^T$ and $H\left(\rho \circ P_k^f\right)H^T$ in equations 9 and 11 are approximated as follows:

$$\left(\rho \circ P_k^f\right)H^T = \rho_1 \circ \left(P_k^f H^T\right) \quad (14)$$

where the elements of ρ_1 is one when the observation location is the same as the location of soil moisture status vector (soil moisture in five layers). Otherwise, its element values are zeros.

$$H\left(\rho \circ P_k^f\right)H^T = \rho_2 \circ \left(HP_k^f H^T\right) \quad (15)$$

where the ρ_2 is a square identity matrix with the same size of $HP_k^f H^T$.

By conducting each assimilation, the soil moisture status vector is updated and the new vector is used to calculate f_w for GPP modeling. When SMAP observation is unavailable for a grid at a specific hour, their ensemble members (soil moisture in the five layers for the grid) evolve based on Richards law in the BEPS soil moisture module and are not adjusted by the data assimilation. However, the accuracies of soil moisture can gradually deteriorate due to accumulation of uncertainties in forcing data and model structures, and therefore, the accuracy of GPP simulation may decrease depending on quality of forcing data.

2.3 Materials

The MERRA (Modern-Era Retrospective Analysis for Research and Applications) data (Version 2) [*Rienecker et al.*, 2011] from Goddard Space Flight Center, National Aeronautics and Space Administration (NASA), are used to drive BEPS. The data have a spatial resolution of 0.625° (longitude) by 0.5° (latitude) and a temporal resolution of 1 h. To drive BEPS, relative humidity, wind speed, and air temperature at 2 m above the surface, surface atmosphere pressure and incoming solar shortwave flux, and total precipitation at the surface level are spatially interpolated into 36 km grids. The precipitation data from MERRA are corrected by global gauge-based NOAA Climate Prediction Center "Unified" (CPCU) precipitation product.

The reprocessed Moderate Resolution Imaging Spectroradiometer (MODIS) leaf area index (LAI) product in 1 km resolution for 2015 is upscaled to 36 km resolution and used in BEPS [*H Yuan et al.*, 2011]. The fraction map of C4 plants is derived from "NACP Model Driver Data: Global Grassland C3 and C4 maps 0.5-Degree" (http://webmap.ornl.gov/). To separate sunlit and shaded leaves in the canopy, we use the clumping index map derived from MODIS bidirectional reflectance distribution function product [*Liming He et al.*, 2012]. The land cover map (classification system of plant function types (PFTs)) is from the MODIS land group [*Friedl et al.*, 2002]. The key BEPS parameters, such as V_{cmax} (the maximum rate of carboxylation) [*Zheng et al.*, 2017] and m (slope of the BWB equation), are mapped according to the PFTs and C4 fraction maps. The soil texture map from the SMAP team is used in BEPS [*O'Neill et al.*, 2015].

The "SMAP Level-2 Radiometer Half-Orbit 36 km EASE-Grid Soil Moisture" product from April to December 2015 is downloaded from the NASA National Snow and Ice Data Center Distributed Active Archive Center [*O'Neill et al.*, 2015]. This L2 soil moisture product provides estimates of global land surface conditions retrieved by SMAP passive microwave radiometer during the 6:00 A.M. descending half-orbit passes. SMAP L-band brightness temperatures are used to derive the surface soil moisture (0–5 cm) in unit of cm^3/cm^3, which are then resampled to the EASE-Grid 2.0 in 36 km resolution. The SMAP soil moisture is retrieved using brightness temperature in the V-polarization. Retrievals with "high-quality" flag are only located in semiarid and arid areas where vegetation is sparse. Unfortunately, there is no high-quality soil moisture retrieval in dense forest areas (Figure S1 in the supporting information).

Before data assimilation, the SMAP soil moisture in high quality is rescaled by the CDF matching approach for bias removal [*Reichle and Koster*, 2004]. This is done by matching the SMAP soil moisture CDF to BEPS simulated soil moisture CDF in the top layer for each pixel.

In this study, the Global Ozone Monitoring Experiment-2 (GOME-2) version 26 (V26) 740 nm terrestrial chlorophyll fluorescence data [*Joiner et al.*, 2013] are used to evaluate BEPS

simulated GPP. The GOME-2 L3 product contains the monthly mean SIF data at latitude and longitude gridded file. The spatial resolution for the L3 data in 2015 is 40 km by 40 km, which is compatible to our GPP simulation in 36 km by 36 km.

3 The Design of Experiment

All data sets are reprojected to the global 36 km EASE-Grid 2.0. Starting with one third total available water holding capacity in the top layer and higher soil moisture in deeper layers for each pixel, BEPS is run for at least five times, with each run initialized following the status of the previous run, to spin-up using forcing data in whole 2015. The last BEPS run is referred to as a “reference-run” or a “single-run,” representing a simulation of GPP without assimilating observations. In contrast, the data assimilation of SMAP observations in DAMP is referred to as an “EnKF-run.”

In DAMP, an EnKF-run is initialized with an exact starting status of the single-run to assimilate the rescaled SMAP soil moisture data. In the EnKF-run, the uncertainty of SMAP soil moisture is set to 0.04 $m^3\ m^{-3}$, which is very close to the recent assessment (0.038 $m^3\ m^{-3}$) [*Chan et al.*, 2016]. A white noise in 0.015 $m^3\ m^{-3}$ is added to the soil moisture state vector for covariance inflation in daily intervals. The ensemble average of 100 BEPS replicates is used to compare with results of single-run.

BEPS outputs sunlit and shaded GPP and evapotranspiration (ET) in hourly step. Both the single-run and EnKF results are summed to daily, monthly, and annual intervals for further analysis. For the monthly values, the temporal correlation (correlation coefficient, r) between simulated GPP (sunlit, shaded, and total) and GOME-2 SIF is used as an indicator of the performance of the ecosystem model in GPP simulation. Areas with SMAP data in low quality (i.e., forests as shown in Figure S1) or with SIF in low values (e.g., deserts) are excluded from the analysis. Since some monthly SIF data are missing, pixels with less than five valid SIF values in a year are also excluded. The difference in the temporal correlation between EnKF and single-run (Δr) is used to indicate if the model performance is improved after data assimilation (DA).

4 Result

4.1 Global Pattern of GPP and ET

Our ensemble-estimated global GPP from the EnKF-run in 2015 is 139.1, 82.7, and 56.4 Pg C yr^{-1} from the total, sunlit, and shaded leaves, respectively, which is similar to the estimate from the single-run of BEPS (138.3, 82.0, and 56.3 Pg C yr^{-1}). Shaded leaves contribute

41% of the total GPP in average globally. Our estimate of global GPP is slightly above the average value from literatures (Table 1). Large GPP values are found in humid and semihumid areas of the tropical and temperate rainforests (Figure 1). GPP values decrease along with deceasing solar radiation and/or available water.

Table 1. Global GPP Estimates Documented and in This Study

Period (Year)	GPP (Pg C yr^{-1})	Reference
2000–2003	108	*Y. J. Zhang et al.* [2009]
2000–2003	111	*W. P. Yuan et al.* [2010]
2000–2003	113	*Zhao et al.* [2005], *W. P. Yuan et al.* [2010]
1998–2005	123 ± 8	*Beer et al.* [2010]
2003	132 ± 22	*Chen et al.* [2012]
2000–2010	128 ± 1.5	*Yan et al.* [2015]
1990–2009	112 to 169 from 10 models	*Anav et al.* [2015]
2015	139	This study

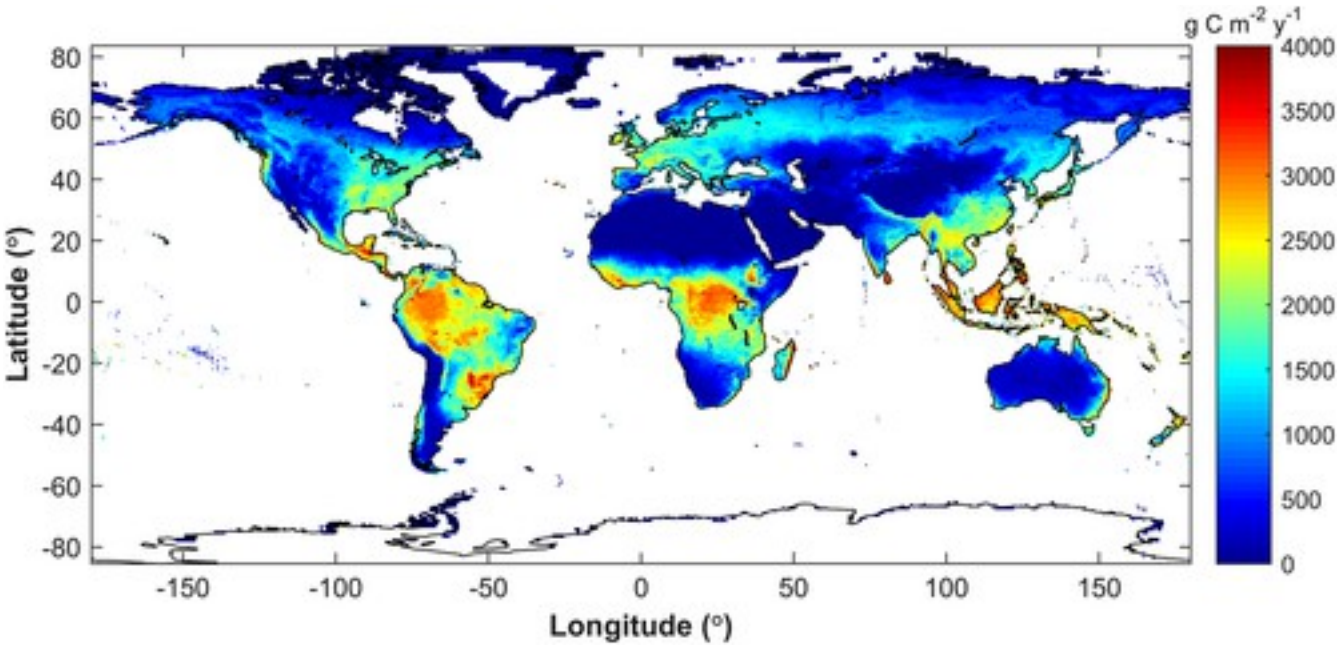

Figure 1

Open in figure viewerPowerPoint

The distribution of the global annual GPP (the ensemble mean) in 2015 estimated by ensemble Kalman filter.

Caption

Though the total GPP estimates from EnKF and single-run are close, their spatial distributions of GPP are quite different. The difference can be as large as ±300 g C m^{-2} yr^{-1} in some patches, suggesting that the carbon flux strongly responds to soil moisture changes in the model (Figure 2). In the Northern Hemisphere, these patches of positive differences are associated with irrigated cropland, which will be described in details in the next section. As expected, there is no difference in GPP estimates in the dense forest area where there are no SMAP soil moisture retrievals with high quality.

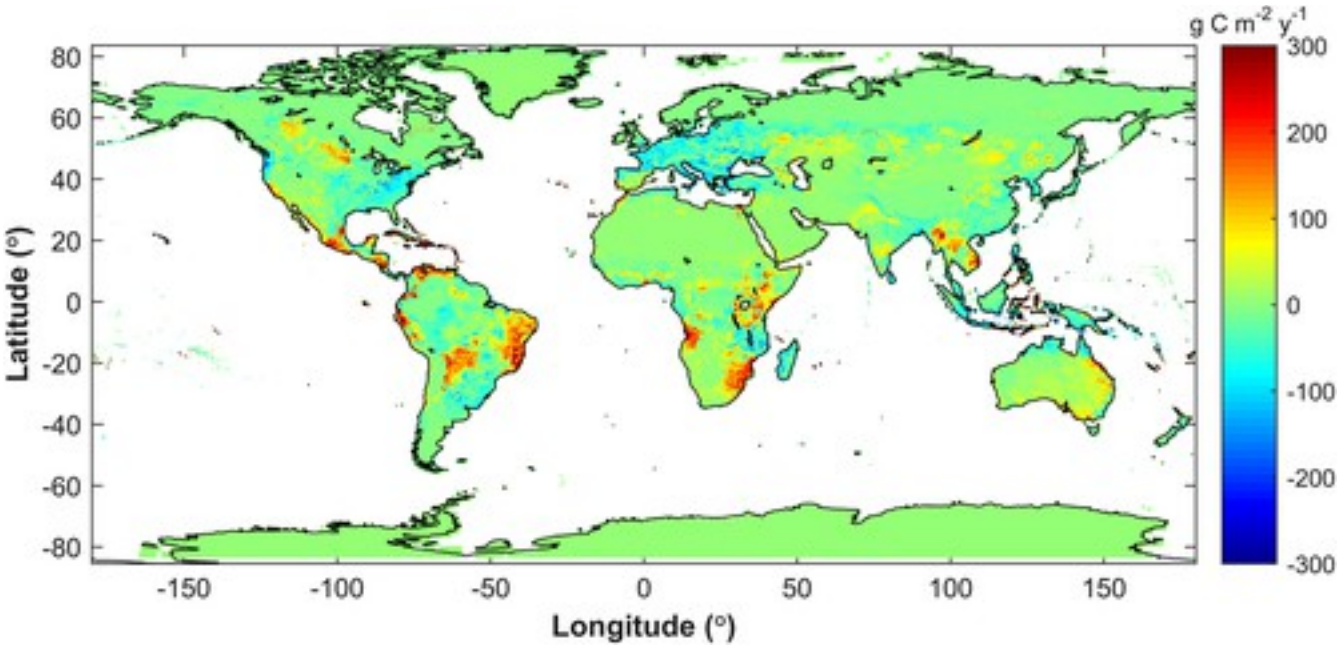

Figure 2

Open in figure viewerPowerPoint

The difference in the annual global GPP estimates between the two approaches (EnKF-mean minus single-run).

Caption

The GPP estimates from both EnKF and single-run highly correlate to GOME-2 SIF for most of the land area with the Pearson correlation coefficient (*r*) larger than 0.8 (Figure 3). High *r* is located at temperate and boreal forests where the seasonal variations of SIF and GPP are high. Negative correlations appear in the tropical rain forest. These negative correlations are due to the errors in the seasonal variation of LAI/SIF in the tropical area where cloud fractions are high all year round. In these areas, only few LAI data are available for a pixel during a year, making the temporal interpolation of LAI impossible. In the desert and arid areas, both the LAI and SIF values are too small to produce reliable temporal correlations. Thus, we exclude the tropical rainforests, desert, and arid areas in the following analysis.

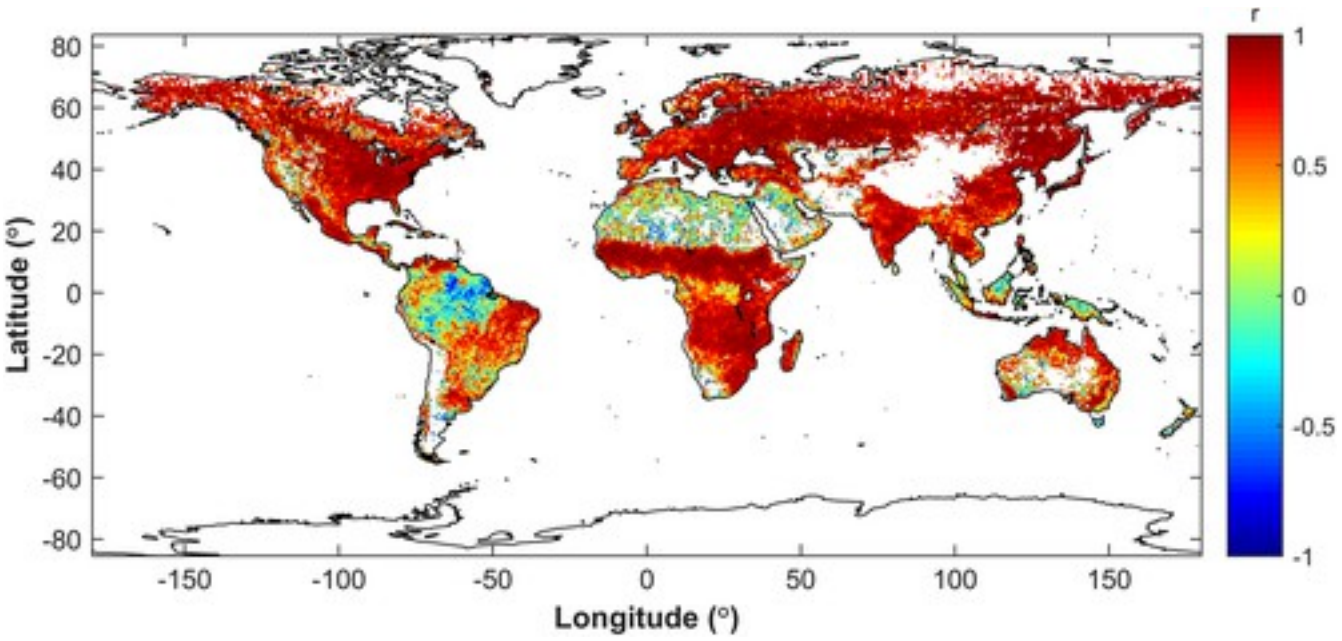

Figure 3

Open in figure viewerPowerPoint

The correlation coefficient (*r*) between SIF and EnKF estimated GPP. In tropical rainforest, the numbers of leaf area index (LAI) and SIF values are rather limited due to cloud contamination. In desert areas, both the values of LAI and SIF are too small to produce reasonable correlation. Therefore, data in desert and tropical rainforest are excluded from our analysis.

Caption

The Δ*r* maps for the total, sunlit, and shaped GPP components are produced. The three maps reveal same trend of the difference in *r*, although their magnitudes are different. Significant changes of Δ*r* are found for sunlit GPP (Figure 4). The absolute values of Δ*r* for shaded GPP are smaller than sunlit GPP. Though the same water stress factor is applied to both sunlit and shaded leaves in equation 1, the net photosynthesis rate for shaded leaves is lower than that of the sunlit leaves due to their lower solar irradiance. Therefore, shaded leaves demand less water than sunlit leaves, resulting in smaller|Δ*r*|. Thus, Δ*r* for the total GPP is between the Δ*r*values for sunlit and shaded GPP. In the following, we will use Δ*r* for sunlit GPP to assess the improvement of GPP.

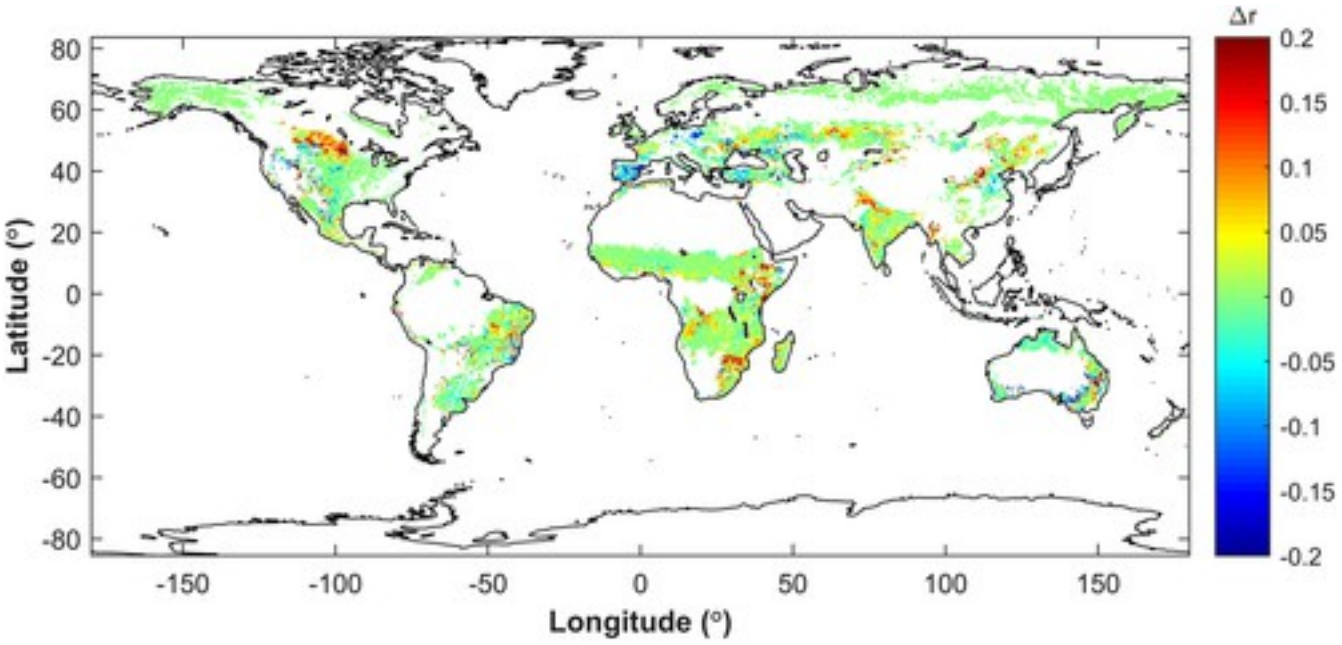

Figure 4

Open in figure viewerPowerPoint

The difference in the correlation coefficient (SIF versus sunlit GPP) between EnKF and single-run. Positive values indicate that the simulation of sunlit GPP is improved owe to SMAP.

Caption

Based on Figure 4, there are vast areas with Δ*r* close to zero in America, Asia, and Africa. At high latitudes, water is rarely a stressing factor for photosynthesis. Therefore, the value of |Δ*r*| is expected to be close to zero. In low and middle latitudes, these low |Δ*r*| values suggest that the performance of BEPS is rarely improved by assimilating SMAP soil moisture from the single-run which uses the gauge-corrected precipitation product.

The land area in Figure 4 mainly includes shrub, grass, and cropland. With the assimilation of SMAP data, the *r* for these land covers is improved by only 0.01 since nonwater-stressed area is included [*Vorosmarty et al.*, 2010] (http://12.000.scripts.mit.edu/mission2017/wp-content/uploads/2013/11/Figure-2.jpg). Forests are rarely included in our analysis except for Africa (Table 2). The degree of improvement of *r* varies among continents and land cover types (Table 2). The *r* for Africa, Asia, and North America are improved by 0.016, 0.013, and 0.013, respectively ($p < 0.05$). In Europe, the simulation of GPP shows a negative effect as *r* is reduced by 1.2%.

Table 2. (a) The Change of Correlation Coefficient (*r*) Between SIF and BEPS Simulation of Sunlit GPP After Using SMAP Soil Moisture for Data Assimilation for Various Plant Functional Typesa

	Asia	North America	Europe	Africa	South America	Australia	All Continents
Evergreen needleleaf trees	0.0*	−0.009	−0.002				−0.006
Evergreen broadleaf trees	0.036**	−0.004		0.016*	0.006		0.015*
Deciduous needleleaf trees	0.001*						0.001*
Deciduous broadleaf trees	0.035**	0.009*	0.011	0.02**	−0.006	−0.007	0.016**
Shrub	0.0	0.0	−0.019	0.02**	0.012*	0.008	0.005*
Grass	0.013**	0.001	−0.007	0.017**	0.004*	−0.006	0.009*
Cereal crops	0.022**	0.09**	−0.008	0.003*	−0.007	−0.002	0.015**
Broadleaf crops	0.023**	0.008*	−0.003	0.011*	−0.009		0.01*
All land cover	0.013*	0.013**	−0.012	0.016*	0.003*	0.001	0.01*

	Asia	North America	Europe	Africa	South America	Australia	All Continents
Evergreen needleleaf trees	0.0*	−0.009	−0.002				−0.006
Evergreen broadleaf trees	0.036**	−0.004		0.016*	0.006		0.015*
	*			*			

(b) The Number of Pixels Used in Table 2a for Statistics[b]

	Asia	North America	Europe	Africa	South America	Australia	Total
Evergreen needleleaf trees	21	147	45	0	0	0	213
Evergreen broadleaf trees	14	14	0	28	6	0	62
Deciduous needleleaf trees	102	0	0	0	0	0	102
Deciduous broadleaf trees	91	94	6	716	135	11	1053
Shrub	1350	704	146	713	267	447	3627

	Asia	North America	Europe	Africa	South America	Australia	All Continents
Evergreen needleleaf trees	0.0*	−0.009	−0.002				−0.006
Evergreen broadleaf trees	0.036**	−0.004		0.016*	0.006		0.015*
Grass	1547	1296	163	3508	2473	562	9549
Cereal crops	1776	471	1337	680	96	247	4607
Broadleaf crops	590	580	262	41	222	0	1695
Total	7701	4239	2737	9048	4577	1459	29761

- a Positive value indicates that the simulation of GPP is improved. Note: that the mark “*” (or “**”) indicates that *t* test is in favor of the alternate hypothesis that the data come from a population with a mean greater than 0 (or 0.01), at the 5% significance level.
- b Note that each pixel has an area of ~1296 km^2. Only pixels having a dominating land cover type and occupying more than 50% of pixel area are included in lines 1 to 8 for analysis. Line 9 shows the total pixel number for each continent. The sum of numbers in lines 1 to 8 may not be exactly equal to the number of pixels in the last line (total).

4.2 Cropland

Significant improvement of *r* is achieved in single-cropping agricultural land, especially in North America.

After assimilating SMAP data in BEPS, significant improvement of GPP simulation is found in the cereal cropland along the Canada-U.S. boundary, with *r* being improved by 9% on average. The MERRA-2 precipitation data are corrected by global gauge-based NOAA Climate Prediction Center “Unified” (CPCU) precipitation product to enhance the data's accuracy. However, the information on agricultural irrigation is not available. Our results indicate that SMAP can capture the positive impacts of agricultural irrigation on GPP.

In Canada, the information on irrigation is well documented; the irrigation volume varies by provinces, crop types, and available precipitation annually (http://www.statcan.gc.ca/pub/16-402-x/2011001/part-partie1-eng.htm). For example, more than 60% of irrigation water was used in the South Saskatchewan drainage region in 2010. From our analysis, it is clearly shown that the significant improvement of GPP simulation in Canada overlaps with the North and South Saskatchewen, and Assiniboine-Red drainage regions of 2006 Agricultural ecumene (http://www.statcan.gc.ca/pub/16-402-x/2011001/m002-eng.htm). How SMAP has captured the irrigation effect in Canada is shown in Figures 5-8. In Figure 5, the single-run shows dry summer and fall for a pixel. However, the EnKF shows a wetter summer as corrected by the SMAP data. Whenever the soil moisture estimate from single-run deviates from the SMAP-CDF observation, the soil moisture estimate from EnKF shows a clear tendency toward the SMAP-CDF observation, suggesting that the EnKF system works well as expected. As a result, the EnKF shows almost no water stress from July to October in contrast to the single-run case, which often shows significant water stress (Figure 6). Correspondingly, the EnKF predicts higher sunlit GPP from July to September than the single-run (Figure 7). The increase of GPP predicted by EnKF is supported by the SIF data (Figure 8). In the spring, both the single-run and EnKF overestimate GPP relative to the SIF data. This is because BEPS uses a constant V_{cmax}(maximum value of the growth season), but in fact the V_{cmax} increases from a very low value in the early growth season to its maximum following the accumulation of chlorophyll [*Liming He et al.*, 2014].

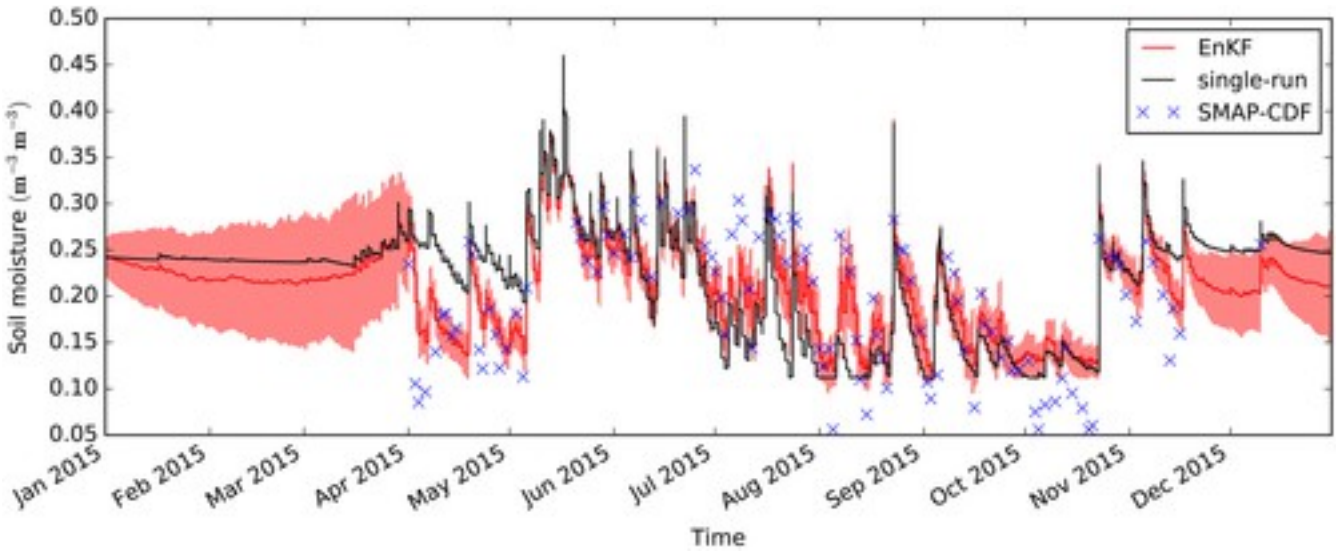

Figure 5

Open in figure viewerPowerPoint

The simulation of soil moisture in top layer (5 cm) by EnKF and a single-run at a pixel in Canada (47.3233°N, 98.0290°W). "SMAP-CDF" indicates the SMAP soil moisture that has been adjusted using the CDF matching approach. The red background color indicates one standard deviation of soil moisture simulated from the EnKF.

Caption

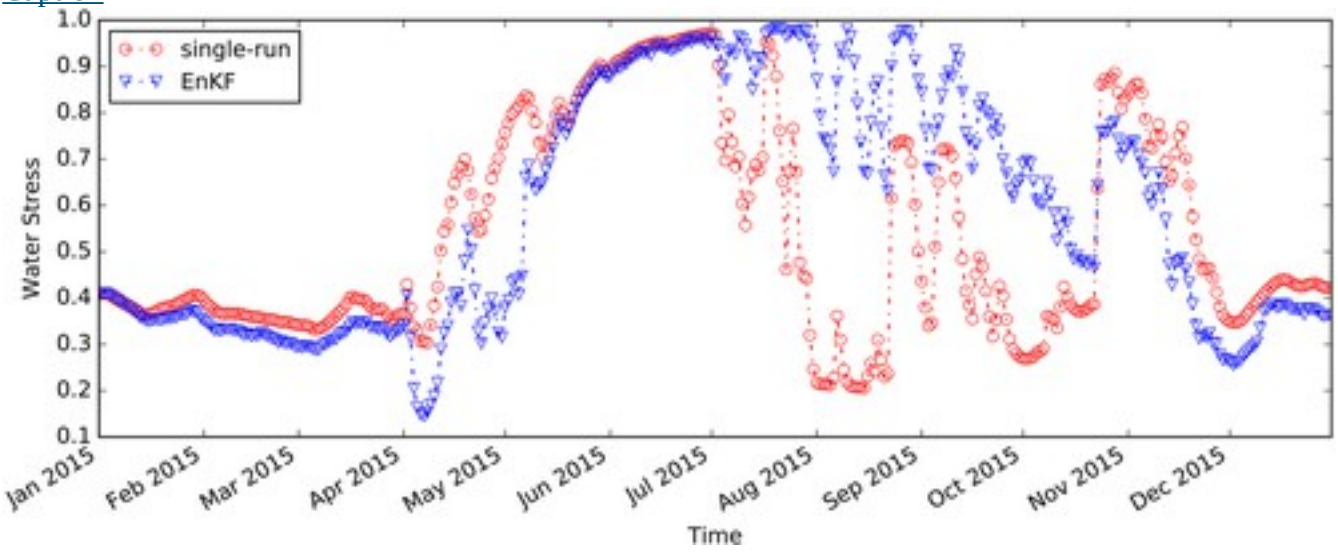

Figure 6

Open in figure viewerPowerPoint

Soil water stress factors (f_w) calculated by EnKF and single-run at a pixel in Canada (47.3233°N, 98.0290°W).

Caption

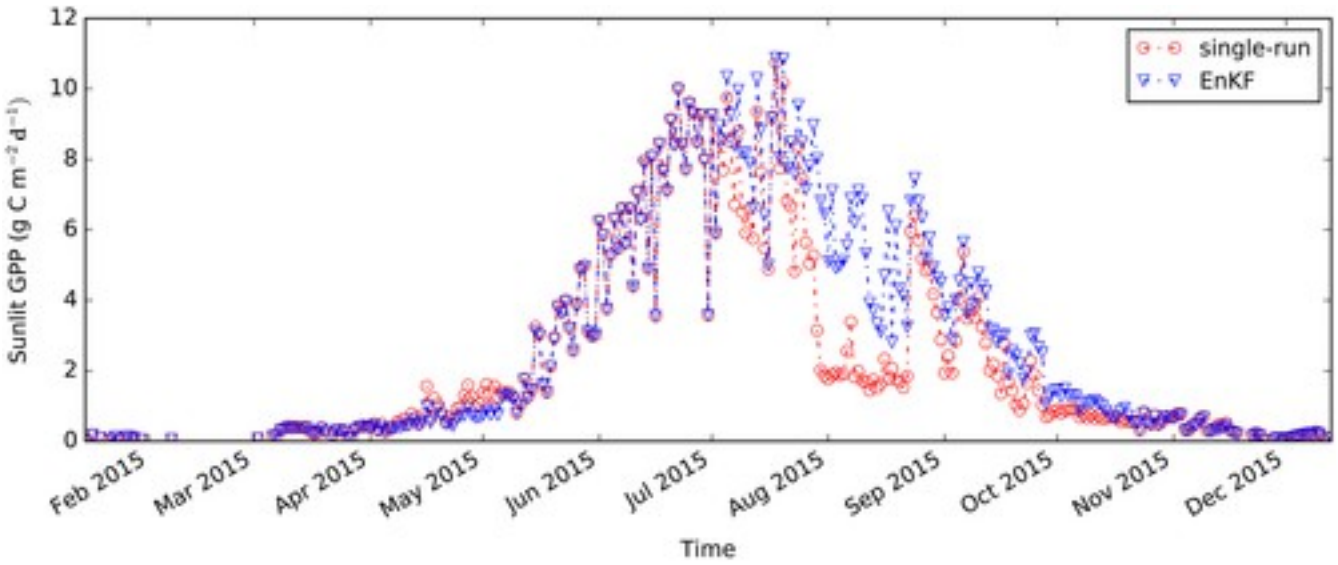

Figure 7

Open in figure viewerPowerPoint

Time series of sunlit GPP estimates by EnKF and single-run at a pixel in Canada (47.3233°N, 98.0290°W).

Caption

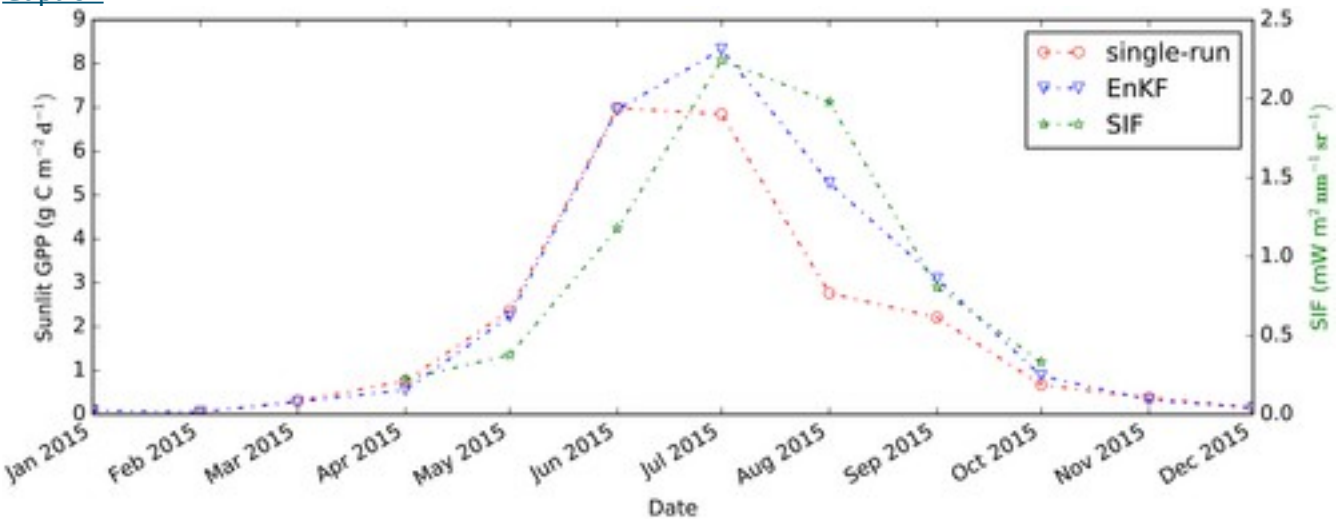

Figure 8

Open in figure viewerPowerPoint

Monthly time series of SIF, sunlit GPP estimates by EnKF, and single-run at a pixel in Canada (47.3233°N, 98.0290°W). The left *y* axis indicates the monthly averaged daily GPP estimation. The overestimation of Sun-GPP in early growth season may be attributed to a consistent V_{cmax} value used in the model.

Caption

Similar improvement of GPP simulation is also seen along the cropland belt (~50°N) going through Ukraine, Russia, Kazakhstan (most of its land is farm), in an oasis in Xinjiang province, and Northeast China where irrigation practice is common. In an Asian irrigation map (2000–2010) (http://waterdata.iwmi.org/applications/irri_area/), Kazakhstan was highlighted for the vast

rain-fed single cropping area. In Figure 4, the improvement of GPP simulation suggests that irrigation activities are captured by SMAP [*Saltanat et al.*, 2015]. India has the largest irrigated areas in the world. The belt area in northern India closely overlaps with an independent mapping of irrigation areas from remote sensing [*Thenkabail et al.*, 2009].

Figure 4 can also confirm the irrigation areas in South Africa [*Altchenko and Villholth*, 2015]. The improvement of GPP in other African areas may be related to recent development of irrigation systems in developing African countries.

Another example is given in Figure S2 to demonstrate the benefit of using SMAP data to improve GPP simulation. The pixel in Figure S2 is located in a cropland for maize in Northeast China, where the maize is sowed in spring and harvested in autumn. The single-run predicts low soil moisture (Figure S2a) and high water stress (Figure S2b) in August and September in contrast to the EnKF-run which adjusts the values of soil moisture according to SMAP observations. Correspondingly, the EnKF-run predicts high GPP in August and September (Figure S2c) and that is confirmed by the SIF data (Figure S2d).

5 Discussion

The SMAP data can be used for many studies on weather, climate, droughts, floods, wildfires landslides, human health, national security, and the global vegetation productivity. This study offers an enhanced understanding of assimilating SMAP soil moisture data into ecosystem models for global applications. The usefulness of SMAP data have not been fully exploited possibly due to some issues in data quality and limits. We focus our discussion on the data uncertainties and some inadequacy in data assimilation with details described in the following sections.

5.1 Uncertainty of SIF-GPP Relationship

As chlorophyll pigments can absorb photons to power photosynthesis, monitoring chlorophyll pigments from space has become a promising way to derive global photosynthetic productivity in recent years [*Cogliati et al.*, 2015]. SIF occurs during photosynthesis in leaves under excessive radiation [*Verrelst et al.*, 2016*]*. Measurements of SIF from vegetation by Greenhouse gases Observing Satellite [*Frankenberg et al.*, 2011; *L. Guanter et al.*, 2012; *Joiner et al.*, 2011] and GOME-2 satellite sensors [*Joiner et al.*, 2013; *Joiner et al.*, 2016] are well correlated with GPP simulations [*Duveiller and Cescatti*, 2016; *Frankenberg et al.*, 2011; *Joiner et al.*, 2014; *X. Yang et al.*, 2015; *Yoshida et al.*, 2015; *Yao Zhang et al.*, 2016b]. Over the last few years, research regarding to the SIF-GPP relationship has been rapidly advanced. Significant

correlations between SIF and GPP have been found in croplands [*Guan et al.*, 2016; *Wagle et al.*, 2016], tundra [*Luus et al.*, 2017], and forests [*Walther et al.*, 2016; *H. Yang et al.*, 2016], with *r* values over 0.9 for land cover types that have large SIF variations [*Verrelst et al.*, 2016; *Yongguang Zhang et al.*, 2016a; *Yao Zhang et al.*, 2016b]. It suggests that SIF can be a powerful tool to track photosynthetic rates on the canopy and ecosystem scales [*Liming He et al.*, 2017b]. Although the slope between GPP and SIF is higher for C4 crops than for C3 vegetation [*Liu et al.*, 2017; *Wood et al.*, 2017] and environmental conditions also affect the relationship [Verma et al., 2017], the correlations between SIF and GPP appear to be robust.

Satellite sensors receive SIF signals from both sunlit and shaded leaves. Since shaded leaves are less stressed in the absence of direct sunlight, their photoprotective mechanisms, like chlorophyll fluorescence, are less active. Therefore, sunlit foliage is responsible for most of the canopy far-red SIF emission received by satellite sensors. As a result, SIF is better correlated to GPP in sunlit leaves [*Z Wang*, 2014] and is used as a proxy of GPP in sunlit leaves in this study.

Large uncertainties of GOME-2 SIF occur over bright areas associated with high photon noise, e.g., shrub lands, deserts, and regions with snow/ice [*Kohler et al.*, 2015]. Because these areas often have low GPP and small ranges of GPP values, the SIF-GPP correlations are also low, which do not contribute much to statistic gains of data assimilation.

5.2 Accuracy of SMAP Data

Recent validations show that the accuracy of SMAP soil moisture products has achieved the mission's goal (0.04 $m^3\ m^{-3}$) [*Chan et al.*, 2016; *Colliander et al.*, 2017]. However, there are still several issues with the soil moisture retrieval from the L-band radiometer [*Wigneron et al.*, 2017]. For example, the soil moisture from SMAP data is found to dry more rapidly than in situ observations after rainfall events *Shellito et al.* [2016a], SMAP appears to be overly sensitive to summer precipitation [*Cai et al.*, 2017], the radio-frequency interference in SMAP is not fully solved yet [*Aksoy et al.*, 2016], and Sun-glint might cause dry bias in soil moisture, especially in mountain areas [*L. He et al.*, 2017a].

These mentioned issues suggest that the uncertainties of SMAP soil moisture vary spatiotemporally and information on these uncertainties is unavailable yet. A uniformed root-mean-square error (0.04 $m^3\ m^{-3}$) is not adequate to address the variation of uncertainty in data assimilation. The CDF matching approach is a way to remove biases between the model simulation and satellite retrieval of soil moisture so that the soil moisture observations can be used efficiently. This approach is particularly useful when there are no accurate soil texture data

available for global applications, as in this study. The sensitivity depth of L band also changes with soil moisture itself, suggesting that the CDF approach is necessary [*Shellito et al.*, 2016a]. Typically, the approach relies on a long time series (e.g., a few years) of satellite data and model simulations. In this study, only SMAP data in 8 months are available for the CDF matching. Once extreme events such as drought or flood happen during this period, the SMAP data adjusted to match a CDF could not represent the usual local variation of the time series of soil moisture. For example, Europe experienced a severe drought in 2015 summer. Relying on the input of precipitation data, the single-run of BEPS predicts a dry summer and low soil moisture. So in the CDF matching, the SMAP data in Europe, even containing the irrigation information, would be forced to match an unusual CDF from BEPS simulation which contains numerous low soil moisture values. As a result, the DA of SMAP in these areas hardly produced improvement in GPP simulation.

Large and highly variable slopes over mountainous areas will adversely affect the soil moisture retrievals as suggested by the SMAP algorithm. Nevertheless, soil moisture retrievals are still made over mountainous terrains. The soil moisture data in mountainous areas with moderate slopes, e.g., Spain and Turkey, are still marked with high quality in the SMAP product and used in our data assimilation. However, the BEPS performance over the regions becomes worse after incorporating the simulated SMAP data, suggesting that more attention should be paid on soil moisture retrievals over mountainous areas.

5.3 Cropping System

For global soil moisture simulation, there is little attention paid to the crop diversity in agriculture. Mono-cropping is the agricultural practice of growing a single crop year after year on the same land, in the absence of rotation through other crops or growing multiple crops on the same land. Multiple (double or triple) cropping is a practice of growing two or more cropsin the same piece of land during a single growth season [*Gallaher*, 2009]. Improvement with the assimilation of SMAP data is found to be weak in double or triple cropping cropland (e.g., part of North China Plain) and/or mountainous area (e.g., Spain and Turkey) (Figure 4).

When the crop rotation is in double or triple cropping system, the simulation of the seasonality of GPP becomes problematic, such as in the North China Plain (e.g., around Shandong, Henan, and An'hui provinces in China) [*Li et al.*, 2014; *Zabel et al.*, 2014] and parts of Europe [*Wu et al.*, 2015]. Without spatially explicit information of the cropping system, BEPS assumes a single cropping system, and therefore, the transition from a crop to the other and the gaps between two crops are not considered, causing unrealistic simulations of GPP of cropland with multiple

cropping systems. A typical double cropping system is the winter wheat and maize rotation in the North China Plain where the winter wheat is sowed in the later autumn after the harvest of maize and harvested in early summer in the next year, and the maize is sowed a few weeks before the harvest of winter wheat. Wheat is a C3 plant with a V_{cmax} value of 85 μmol m^{-2} $s^{-1}$and *m* of 8 (*m* is a plant species dependent coefficient, see equation 1). While maize is a C4 plant with a V_{cmax} value of 50 μmol m^{-2} s^{-1} and *m* of 4 [*Mo et al.*, 2012]. The winter wheat is replaced gradually in June since the cropland is managed by individuals. From the C4 map, we know that the fraction C4 plant is ~1/3 while we do not know when the cropping is rotated. So in the BEPS simulation, constant V_{cmax} and *m* values weighted by C4 fraction are assigned for each pixel for the entire growth season. In these areas, an improvement of GPP with SMAP data is not always expected.

5.4 Other Issues

In this study, BEPS is driven by the MERRA-2 reanalysis data in an off-line mode, for which the precipitation product is corrected by global gauge measurement so that soil moisture simulation is with an enhanced certainty. Some irrigation practice in cropland can be well modeled based on historical climate data. However, when SMAP soil moisture data are used for operational purpose, such as for crop growth monitoring and prediction, such high-accuracy precipitation product may be unavailable. Therefore, it is more beneficial to assimilate the real-time SMAP data.

Table 3 summarizes the degree to which improvement is made from data assimilation of SMAP soil moisture data for estimating GPP of various plant functional types.

Table 3. Summary of the Improvements From Assimilation of SMAP Soil Moisture Data for GPP Modeling

Regions	Significances of Improvement (Δ*r*)	Reasons
Global land surface	Nonsignificant improvement (0.01).	Use of gauge-corrected precipitation in reanalysis forcing data; bias in SMAP data and short data period for CDF correction; large areas without water stress
Cropland in single-	Significant improvement (0.09 in	Strong SIF variation for validation; explanation of irrigation practices

Regions	Significances of Improvement (Δr)	Reasons
cropping system	North America, 0.02 in Asia)	in SMAP product
Cropland in multicropping system	Non or negative improvement	Constant V_{cmax} value across seasons
Shrubland	Minor improvement (0.02 in Africa, 0.01 in South America)	Low SIF variation; use of gauge-corrected precipitation
Grassland	Minor improvement (0.02 in Africa, 0.01 in Asia)	Low SIF variation; use of gauge-corrected precipitation
Mountain	Non or negative improvement	Low soil moisture accuracy but not masked in the product
Tundra	No improvement	Water is usually not a stress factor in these area.
Forest	No evaluation except some sparse forests in Africa and mixed forests in Asia	Limited number of soil moisture in high quality due to high vegetation water content

6 Conclusion

Soil moisture is an essential climate variable controlling hydrological and ecosystem processes. However, there is large uncertainty in the metrological input data (mainly precipitation), which can prevent ecosystem models from accurately simulating soil moisture. In order to improve the simulation of the global GPP, it is desirable to make full usage of the remotely sensed soil moisture and to assimilate the soil moisture into these models. In this study, we used a data

assimilation (DA) system called DAMP to assimilate SMAP-observed soil moisture into an ecosystem model named BEPS, where the water, energy, and carbon exchanges between vegetation and the atmosphere are tightly coupled.

We found that the benefit of assimilating SMAP soil moisture for GPP simulation varies by regions and land cover types. Significant improvement is found in the single-cropping agricultural land. This is because the SMAP captured the agricultural irrigation activities that are not part of the meteorological inputs. Without the SMAP data, the GPP in cropland is underestimated. SMAP also reveals that the previous rain-fed croplands are turned into irrigation areas in many developing countries. In general, there is minor improvement of GPP simulation in Africa, Asia, and North America where there are vast areas with no water stress. In Europe, the GPP simulation becomes slightly worse after assimilating SMAP data, especially in mountainous Spain and Turkey. Our study also suggests that the current soil moisture product derived from SMAP data is already useful for improving ecosystem carbon flux estimation on the global scale, especially for irrigated cropland. However, the usefulness of the product for tropical, temperate, and boreal forests is still limited because high-quality retrievals of soil moisture over these ecosystems are limited.

Acknowledgments

This study was supported by the Canadian Space Agency grant (14SUSMAPTO). The MERRA-2 data are downloaded from https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/. The SIF data are from https://avdc.gsfc.nasa.gov/pub/data/satellite/MetOp/GOME_F/. The SMAP soil moisture is downloaded from https://nsidc.org/data/smap/smap-data.html. The LAI data are from http://globalchange.bnu.edu.cn/research/lai. The model codes that were written in C and Octave (MATLAB) are hosted at Github and can be granted for access by requesting Jing M. Chen at jing.chen@utoronto.ca.