Global datasets of hourly carbon and water fluxes simulated using a satellite-based process model with dynamic parameterizations

Responses to the comments on essd-2023-328 by Anonymous Referee #2

Jiye Leng et al.

The paragraphs in blue Italic indicate the corresponding revised paragraphs in the manuscript. The under-reviewed papers will be preprinted in the final manuscript before the publication.

1. Need particular explanation on the newly revised BEPS regarding how exactly the hourly GPP and ET were simulated but not from the previous version. If BEPS was designed to simulate the hourly products, were if just because the hourly inputs were not available before?

Thanks for the question. The newly revised BEPS v4.10 keeps the original structure and algorithms but with new standardized framework using the Doxygen format and Git version control. The name for BEPS was also revised from the former ‘Boreal Ecosystem Productivity Simulator’ to ‘Biosphere-atmosphere Exchange Process Simulator’. We also open-sourced the BEPS model after the code standardization. BEPS adopts hourly meteorological variables to simulate carbon and water fluxes using the same algorithms in Chen et al. (2012). Besides, BEPS has been comprehensively evaluated in several site-level, showing its capacity to simulate gross primary productivity (GPP) and evapotranspiration (ET) comparable to the eddy covariance measurements (Gonsamo et al., 2013; Luo et al., 2018; Luo et al., 2019).

Due to the computational capacity and data volume, the global hourly GPP and ET dataset based on BEPS has not yet been published before. In this study, compared to the previous research papers on BEPS, we adopted dynamic parameterizations to improve the accuracy of simulated carbon and water fluxes, and presented and analyzed the global GPP and ET in 2001-2020 at the hourly timescale for the first time.

References:
2. Need a better presentation and validation on the newly optimized key photosynthesis and stomatal conductance model parameters (i.e., \(V_{cmax}\) and \(m\)). Was it only revised for the flux sites (as presented in Figure 2)? How were they interpolated into the global scales and what are their uncertainties? How about the spatial and temporal variations of these parameters globally? Also, what will be the differences in the simulated GPP and ET between the new dynamic parameters and original fixed parameters in terms of accuracy and spatial pattern?

Thanks for the suggestion. As you are concerned, there are several problems that need to be addressed. This manuscript focuses on generating and sharing the reliable GPP and ET dataset. We included the global distributions of \(m\) and \(V_{cmax}\) in another paper which focuses on discussing the trend and spatial patterns of \(m\) and \(V_{cmax}\) (Leng et al., under review). However, we would add one chapter to include the spatial distributions, species distributions, and validations of \(m\) and \(V_{cmax}\) in this dataset manuscript to convince the readers on the accuracy and reliability of the dataset by Leng et al. (under review).

We included the validations of the Random Forest regressor for \(m\) and \(V_{cmax}\) estimations by courtesy of Leng et al. (under review), as shown in Error! Reference source not found.. The Random Forest regressor estimates \(m\) and \(V_{cmax}\) with good agreements to the optimized \(m\) and \(V_{cmax}\) from measured fluxes in both the training sets and the validation set. The Random Forest regressor can estimate \(m\) with \(R^2 = 0.95\) and RMSE = 1.414 in the training set (Figure 1a), and \(R^2 = 0.59\) and RMSE = 3.663 in the independent validation set (Figure 1b). The Random Forest regressor can estimate \(V_{cmax}\) with \(R^2 = 0.98\) and RMSE = 4.191 \(\mu\text{mol m}^{-2}\text{s}^{-1}\) (Figure 1c), and \(R^2 = 0.84\) and RMSE = 10.598 \(\mu\text{mol m}^{-2}\text{s}^{-1}\) in the independent validation set (Figure 1d). Most of the scatter points locate beside the 1:1 line in both the training and validation set, showing the good accuracy of the Random Forest regressor for \(m\) and \(V_{cmax}\) estimation. The Random Forest regressors build the bridge that links \(m\) and \(V_{cmax}\) derived from measured fluxes to the gridded data that can be expanded to global coverage and long timeseries.

![Figure 1. Comparisons of estimated \(m\) and \(V_{cmax25}\) from the Random Forest regressor and optimized \(m\) and \(V_{cmax25}\) from measured fluxes in the training set (a), (c) and in the independent validation set (b), (d) of the Random Forest Regressor, respectively. The color](image-url)
indicates the scatter density in each plot and the dotted lines indicate the 1:1 line in each plot. Courtesy of Leng et al. (under review).

The global distributions of retrieved $m$ and $V_{cmax}$ are shown in Figure 2a and Figure 3a, and PFT-dependent patterns of $m$ and $V_{cmax}$ are observed in Figure 2b and Figure 3b. We also included the monthly spatial patterns of global $m$ and $V_{cmax}$ during 2001-2020 in Figure 4 and Figure 5, respectively. Strong seasonal variations in $m$ and $V_{cmax}$ are observed in boreal regions while $m$ and $V_{cmax}$ in subtropical and tropical regions are fairly constant within a year.

![Figure 2](image.png)

*Figure 2. The spatial pattern of global $m$ (a) and the averaged $m$ in each PFT (b) during 2001-2020. Courtesy of Leng et al. (under review).*

![Figure 3](image.png)

*Figure 3. The spatial pattern of global $V_{cmax}$ (a) and the averaged $V_{cmax}$ in each PFT (b) during 2001-2020. Courtesy of Leng et al. (under review).*
To further validate the gridded global $m$ and $V_{\text{cmax}}$, we compared the global retrievals of $m$
and $V_{cmax}$ in this study with the $m$ census for various biomes from the review by Miner et al. (2017) and the $V_{cmax}$ field measurements collected from Smith et al. (2019), as shown in Figure 6. $m$ estimates in this study were compared with the mean and standard deviation in Miner et al. (2017) while $V_{cmax}$ observations with the timestamp of measurement were compared with the estimated $V_{cmax}$ in the corresponding time period in 2001-2020. Only 0.25° pixels with more than three $V_{cmax}$ measurements were selected in the comparison. The estimated $m$ in the global retrievals agrees well with the measured $m$, with $R^2 = 0.62$ ($P = 0.06$) and the estimated $V_{cmax}$ in the global retrievals agrees well with the measured $V_{cmax}$, with $R^2 = 0.59$ ($P < 0.001$).

Figure 6. Left: Comparison between the PFT-scale mean values of estimated $m$ from the Random Forest regressor and the measured $m$ values reported in the review by Miner et al. (2017). Each horizontal and vertical bar represents the mean $m \pm 1$ standard deviation in the literature values and the estimated values, respectively. The sample sizes of measured $m$ for each PFT are $n = 23$ (ENF), $n = 23$ (EBF), $n = 54$ (DBF), $n = 11$ (SH), $n = 5$ (GRA), and $n = 53$ (CRO). Right: Comparison between the PFT-scale mean values of predicted $V_{cmax}$ from the Random Forest regressor and the measured $V_{cmax}$ values reported in the review by Smith et al. (2019). Courtesy of Leng et al. (under review).

References:
3. Figure 1. The bottom-up order is quite counter-intuitive to readers. Suggest using top-down order.

Thanks for the suggestion. We have revised the Figure 7 as the top-down order.

![Methodological Steps](image)

**Methodological Steps**
- Flux data processing
- Flux data processing
- Residuals optimization @ monthly
- Feature selection, Training & Cross-validation
- Prediction @ monthly @ 0.25°
- BEPS-DP simulation @ hourly @ 0.25°

**BEPS-DP Ensembles of Global Two-leaf Hourly Carbon and Water Fluxes**
- Flux tower meteorological data
  - SW, TA, RH, P, WS
- RS-based products
  - LAI
- Eddy covariance flux tower data
  - GPP, LE

**Bayesian Parameter Optimization with carbon-water coupling cost function**
- ERAS Meteorological Reanalysis
  - SW, TA, RH, P, WS
- Optimized site-level
  - \( \cdot \frac{\text{LAI}}{\text{max}} \)
- RS-based products
  - LAI, LCC

**Global datasets**
- Random Forest Regressor
- Carbon fluxes: sunlit GPP, shaded GPP
- Water fluxes: sunlit T, shaded T, E @ hourly @ 0.25°

**Figure 7. Schematic overview of the methodology and data products of the BEPS model with dynamic parameterizations (BEPS-DP). The flow diagrams show the methodological steps (left) and the details (right) for the BEPS-DP datasets of global hourly two-leaf carbon and water fluxes. SW (shortwave radiation, W m\(^{-2}\)), TA (air temperature, °C), RH (relative humidity, %), P (precipitation, mm h\(^{-1}\)), WS (wind speed, m s\(^{-1}\)), GPP (gross primary productivity, g C m\(^{-2}\) h\(^{-1}\)), LE (latent heat, W m\(^{-2}\)).**

4. Figure 4, suggest adding the label of 1 in the slope subplots as a reference of good fitting.

Thanks for the suggestion. We have added the label of 1 in the slope and R\(^2\) subplots as a reference of good fitting, as shown in Figure 8.
Figure 8. Evaluation of modeled hourly and daily fluxes against the eddy covariance data in the independent validation set: site-year percentage of $R^2$ in (a) hourly GPP; (c) hourly ET; (i) daily GPP; (k) daily ET; site-year percentage of regression slopes in (b) hourly GPP; (d) hourly ET; (j) daily GPP; (l) daily ET; the mean and standard deviation (SD) of $R^2$ in each PFT in (e) hourly GPP; (g) hourly ET; (m) daily GPP; (o) daily ET; the mean and standard deviation of regression slopes in each PFT in (f) hourly GPP; (h) hourly ET; (n) daily GPP; (p) daily ET. The grey lines indicate 1.0 in $R^2$ and regression slopes as a reference of good fitting. The dashed grey lines in (e) – (h) and (m) – (p) indicate the mean of $R^2$ and regression slopes for all PFTs in GPP and ET.

5. Need better presentations on the diurnal patterns of GPP and ET for different vegetation function types. For example, providing hourly curves for different vegetation function types and validated against flux site observations. It is still unclear whether these products can capture the diurnal variations of GPP and ET.

Thanks for your suggestions. Since there are 20% of all the sites (809 site years) in the
independent validation set for the comparisons of modeled and measured GPP and ET, it would be too redundant to show all the diurnal patterns (i.e., the hourly curves) for different vegetation functional types in the manuscript. However, for better presentations of our GPP and ET product simulated based on BEPS with dynamic parameterizations, we randomly selected one site-year per each PFT and showed the simulated GPP and ET against flux site observations in three different stages (i.e., day of year 115-125, 195-205, and 275-285). The sites and site-year we selected in the presentations of diurnal curves were shown in the table below.

<table>
<thead>
<tr>
<th>Site Name</th>
<th>IGBP</th>
<th>Year</th>
<th>Lat</th>
<th>Lon</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH-Oe2</td>
<td>CRO</td>
<td>2013</td>
<td>47.2863</td>
<td>7.7343</td>
</tr>
<tr>
<td>DE-Hai</td>
<td>DBF</td>
<td>2004</td>
<td>51.0792</td>
<td>10.4530</td>
</tr>
<tr>
<td>IT-Cpz</td>
<td>EBF</td>
<td>2003</td>
<td>41.7053</td>
<td>12.3761</td>
</tr>
<tr>
<td>DE-Tha</td>
<td>ENF</td>
<td>2004</td>
<td>50.9624</td>
<td>13.5652</td>
</tr>
<tr>
<td>NL-Hor</td>
<td>GRA</td>
<td>2008</td>
<td>52.2404</td>
<td>5.0713</td>
</tr>
<tr>
<td>JP-SMF</td>
<td>MF</td>
<td>2005</td>
<td>35.2617</td>
<td>137.0788</td>
</tr>
<tr>
<td>AU-Dry</td>
<td>SAV</td>
<td>2010</td>
<td>-15.2588</td>
<td>132.3706</td>
</tr>
<tr>
<td>CN-Ha2</td>
<td>WET</td>
<td>2003</td>
<td>37.6086</td>
<td>101.3269</td>
</tr>
<tr>
<td>US-Ton</td>
<td>WSA</td>
<td>2011</td>
<td>38.4316</td>
<td>-120.9660</td>
</tr>
</tbody>
</table>

The diurnal variations of simulated GPP and ET against flux observations were shown below, in three different stages per each site year (the beginning, peak, and ending of the growing seasons). The $R^2$ for GPP and ET in different stages per each site year were shown on the left-top in each subplot.
6. Dataset links update

Besides, we updated the link of dataset thanks to the National Ecosystem Data Bank for providing the platform for sharing the dataset. The updated links include the hourly two-leaf
GPP and ET dataset (3.3 TB) and the accumulated daily two-leaf GPP and ET dataset (197 GB) in 2001-2020. The corresponding paragraphs (Abstract, Code and Data Availability) in the manuscript was also updated.

“Abstract
... The hourly and accumulated daily GPP and ET estimates are available at https://doi.org/10.57760/sciencedb.ecodb.00163 (Leng et al., 2023a) and https://doi.org/10.57760/sciencedb.ecodb.00165 (Leng et al., 2023b).”

“Code and Data Availability
The 0.25° × 0.25° global hourly two-leaf GPP and ET datasets for 2001-2020 are available at https://doi.org/10.57760/sciencedb.ecodb.00163 (Leng et al., 2023a). The datasets are provided in NetCDF4 format. The GPP datasets include two components, the hourly GPP of sunlit and shaded leaves. The ET datasets include three components, the hourly evapotranspiration, transpiration of sunlit and shaded leaves. Each hourly NetCDF4 file represents the GPP/ET in a year at an hourly scale (g C m⁻² h⁻¹/mm h⁻¹). The accumulated daily GPP and ET datasets for 2001-2020 are available at https://doi.org/10.57760/sciencedb.ecodb.00165 (Leng et al., 2023b). Each daily NetCDF4 file represents the GPP/ET in a year at a daily scale (g C m⁻² d⁻¹/mm d⁻¹) ...”

References: