Report 1

We appreciate the reviewer's positive evaluation of our manuscript.

Report 2

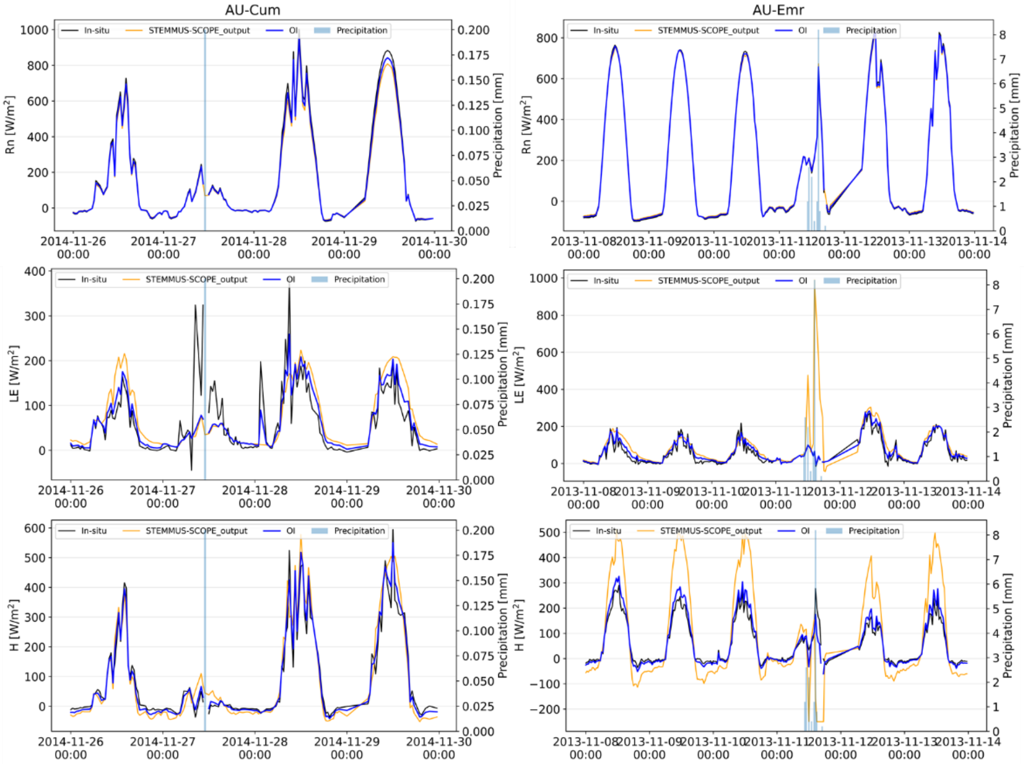
Surface water, energy, and carbon fluxes are critical to the study of the Earth's climate system, yet high-temporal-resolution global datasets remain relatively scarce. This study utilizes hourly-scale data from 2000 to 2020, combining the STEMMUS-SCOPE model with PLUMBER2 site observations, satellite remote sensing, and gridded meteorological data. Machine learning methods are employed to generate the global FluxHourly dataset at a 9 km resolution. The research integrates multiple site observation data with the SCOPE model, addresses data instability at sites through optimal interpolation, and extends the data globally using random forests. This work fills the gap in high-resolution flux data and provides an important tool for studying land-atmosphere interactions and ecosystem responses. The methodology is innovative, and the scientific value is significant. It is recommended for minor revisions with addressing the following issues.

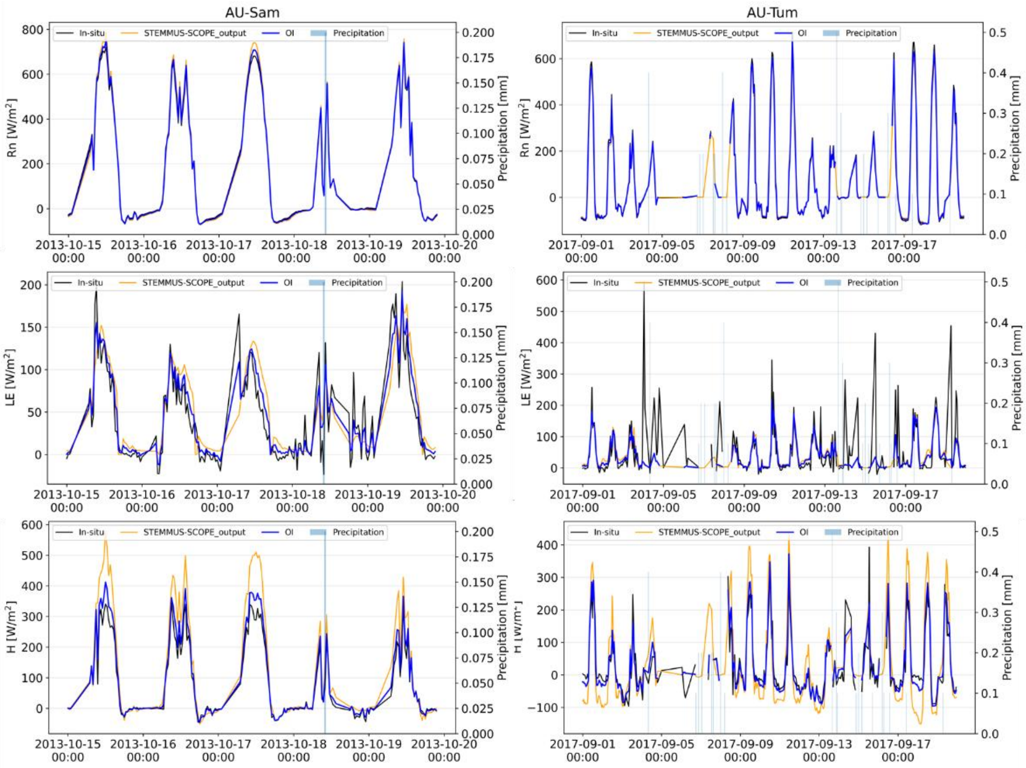
1. OI appears to enhance site-level data quality by integrating STEMMUS-SCOPE simulations and PLUMBER2 observations. However, specific comparisons of results before and after OI processing are not provided, which would help better understand its actual effectiveness. Additionally, OI is only applied at the site level, while the global product relies on RF extrapolation. This may lead to error propagation or bias adaptation due to differences in data scales. Improvements in site-level data quality may not fully reflect at the global scale. It is recommended to supplement the analysis with comparative results at both site and global levels before and after OI to further validate the effectiveness and applicability of the OI method.

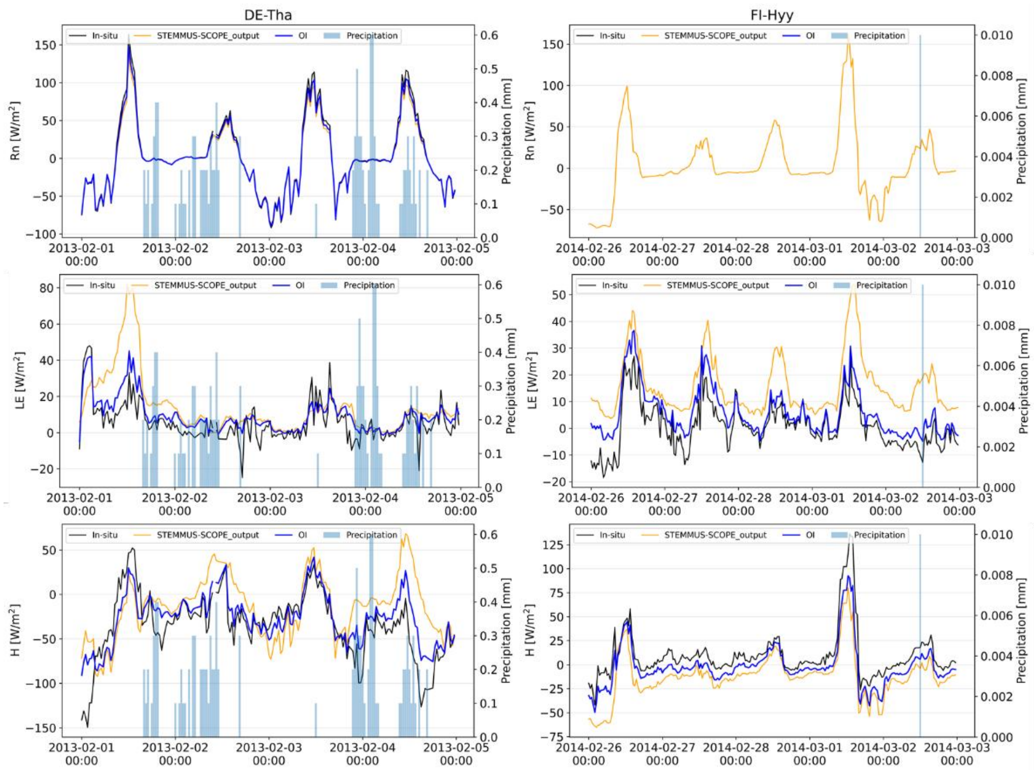
Reply: We thank the reviewer for this constructive suggestion. The main goal of applying OI in this study was to improve the physical consistency and continuity of site-level measurements, including latent heat (LE), sensible heat (H), and net radiation (Rn), by filtering out unrealistic values and short-term spikes. For example, OI helps remove abnormal LE during nighttime and precipitation periods while preserving the general temporal dynamics.

To illustrate this, we have added visual comparisons of time series before and after OI (see new Figure S2). In-situ flux measurements (black) provide the most direct observations but often contain short-term noise and unrealistic fluctuations, particularly during precipitation. The STEMMUS-SCOPE simulations (orange) are smoother and physically consistent but tend to exhibit systematic biases in magnitude. The OI result (blue) dynamically integrates these two sources. As a result, OI follows the observed diurnal dynamics where observations are reliable (e.g., station names in the below figure), while aligning more closely with the model during periods of degraded data quality (e.g., station names in the below figure). In the 6 example sites, OI reduces random variability, avoids unrealistic peaks, and enhances the physical consistency of Rn, LE, and H. This confirms that OI effectively reduces noise and enhances the reliability of site-level inputs used for RF training.

Regarding the reviewer's concern about potential scale differences, we acknowledge that the global product is derived through Random Forest (RF) extrapolation, while OI was applied only at the site level. The performance of the RF extrapolation was evaluated through spatial test, confirming that the model can capture large-scale spatial patterns. Nevertheless, we recognize that further improvement is possible. Future work will focus on enhancing the extrapolation performance by integrating advanced deep learning architectures that can better model nonlinear spatial–temporal dependencies.


AU-Cum
AU-Emr
In-situ
STEMMUS-SCOPE_output
OI
Precipitation
Rn [W/m2]
LE [W/m2]
H [W/m2]
Precipitation [mm]


AU-Sam
AU-Tum
In-situ
STEMMUS-SCOPE_output
OI
Precipitation
Rn [W/m2]
LE [W/m2]
H [W/m2]
Precipitation [mm]



2. Feature importance analysis indicates that incident shortwave radiation is the most critical predictor variable. Why is shortwave radiation so important, and is it related to specific physical mechanisms? Further elaboration on the underlying reasons is recommended. Additionally, inherent relationships exist among variables (e.g., between SIF and GPP), which could potentially improve the accuracy of certain outputs. However, the article does not delve into whether these relationships were considered to optimize inputs. It is suggested that the authors explore this potential to enhance data accuracy.

Reply: We thank the reviewer for this valuable comment. The high importance of incident shortwave radiation (Rin) identified by the Random Forest model can be explained by its fundamental physical role in driving land–atmosphere exchanges. This has been added to lines 189-194 in section 4.1.1.

"Rin provides the primary energy input for surface heating, evapotranspiration, and photosynthesis, thereby exerting strong control over both energy and carbon fluxes. For the energy components (Rn, H, LE, G), Rin determines the available energy that can be partitioned into different fluxes (Peng et al. 2021). For the carbon-related processes (GPP and SIF), Rin governs photosynthetically active radiation (PAR), which supplies the energy for carbon assimilation (Nemani et al. 2003; Running et al. 2004). These findings are therefore physically consistent with the observed dominance of Rin in the feature importance analysis."

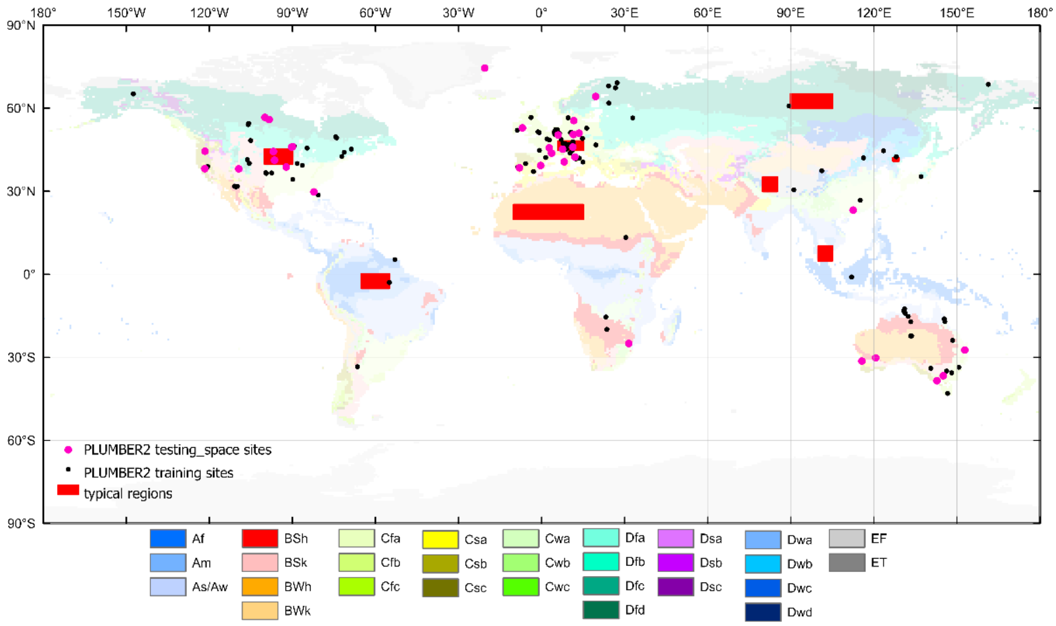
We thank reviewer's comments on the inherent relationships between SIF and GPP. It is to note that SIF and GPP are outputs of our STEMMUS-SCOPE model emulator. The mechanistic link between SIF and GPP has been described in the STEMMUS-SCOPE model (Wang et al. 2021), and therefore implicitly considered in our emulator.

3. In Fig. 1, the longitude label in the upper right should be 180° instead of 90°. It is recommended to add statistics on the number of sites in each climate zone and mark the spatial locations of the 34 testing_space sites. If these sites are uniformly distributed globally, it would further enhance the persuasiveness of the validation. Additionally, although the selection of the eight typical regions is explained later, it is suggested to clarify this in the title or main text of Fig. 1 to improve reader clarity.

Reply: Thanks for your helpful suggestions. We have corrected the longitude label from 90° to 180° in Figure 1. In addition, we added statistics on the number of sites in each climate zone. In total, the 170 sites cover 17 climate zones, and this information has been added in lines 79-81 in section 2.1.

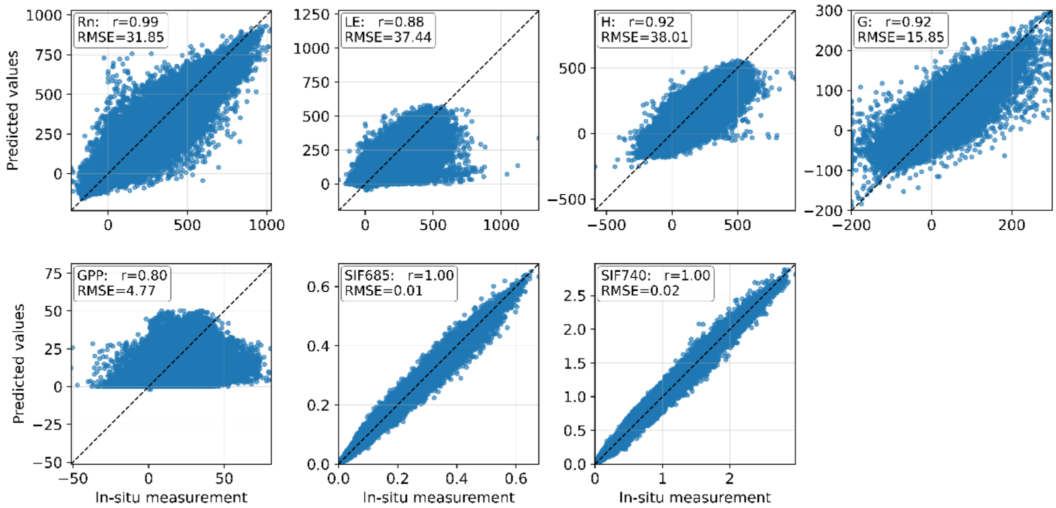
“The 170 sites cover 17 Köppen climate zones, with the following distribution: Af (2), Am (3), Aw (6), BSh (8), BSk (10), BWk (1), Cfa (22), Cfb (42), Csa (20), Csb (6), Dfa (5), Dfb (13), Dfc (23), Dwa (1), Dwb (2), DWc (1), and ET (5). ”

We also marked the spatial locations of the 34 testing_space sites in Figure 1, which were selected based on their IGBP types. Furthermore, the selection of the eight typical regions has been clarified in the caption of Figure 1 (line 85) to improve readability and clarity.



4. The simulated r-values are high (multiple 0.99 in Table 2), but r-values alone are insufficient to convince readers. It is recommended to use scatter plots to display the match between predicted values and actual observations for more intuitive validation of the results.

Reply: Thank you for your constructive suggestion. We have added scatter plots comparing the predicted and observed values for all flux variables to provide a more intuitive validation of the model performance (see new Figure S16) in the Supplementary Material. These plots clearly illustrate the consistency between the predicted and observed fluxes, complementing the statistical results shown in Table 2.



5. Lines 335–336 discuss uncertainty assessment but do not explicitly reference Fig. 9. It is recommended to directly cite the figure to improve clarity. Furthermore, Fig. 9 mentions that "the scale mismatch in predictor variables could lead to underestimation of fluxes." If the purpose is to explain the underestimation of fluxes such as LE, it is suggested to add a dedicated paragraph analyzing the potential mechanisms behind the underestimation. For example, grid-scale mismatches may lead to bias propagation in Rin, thereby affecting fluxes like LE. Supplementing such analyses would deepen the discussion and enhance the credibility of the results.

Reply: I have referred to Figure 9 in line 349. We appreciate the reviewer's insightful suggestion. Following the comment, we have added a paragraph in the discussion (lines 361-366). The new text reads as follows:

"Specifically, coarse-resolution gridded data tend to smooth spatial heterogeneity in surface radiation. When local high-radiation areas are averaged with surrounding low-value regions, the resulting grid-mean Rin becomes smaller than the true value at flux tower locations. Because Rin is the dominant driver of energy partitioning, this underestimation propagates through the energy balance and results in lower simulated LE. In other words, grid-scale aggregation dampens local extremes and introduces bias propagation from Rin to surface fluxes, explaining part of the systematic underestimation observed in our results."

6. The supplementary materials contain a considerable amount of disorganized information. It is recommended to reorganize them into a clear structure to improve readability.

Reply: We have reorganized the supplementary materials. The pages number became 25 from 47.

Report 3

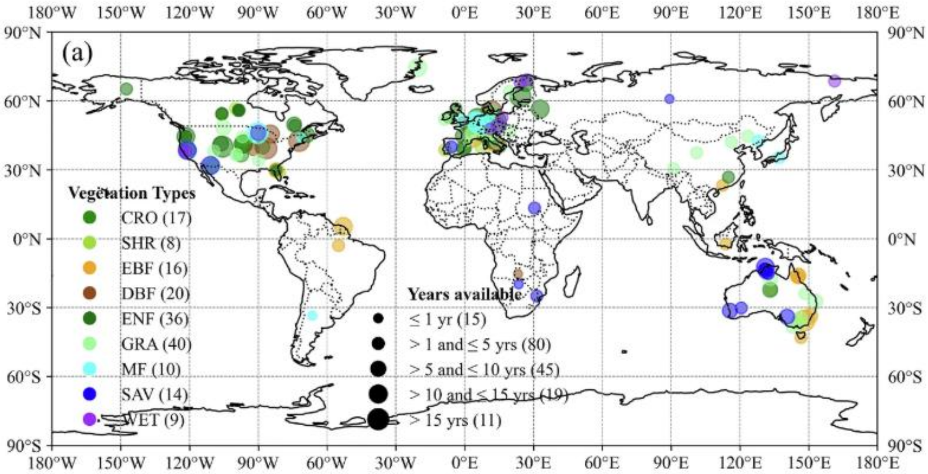
I do not have further comments, except two very minor ones:

1.It would be great to show years for each site covered and the number of site-year, or site-month or site-hour observations;

Reply: We appreciate the reviewer's suggestion. The temporal coverage of each flux tower site has been comprehensively summarized in our previous publication (Wang et al. 2025), where Figure 2 illustrates the years available for all sites (see below). To avoid redundancy, we have added a citation to that figure in the current manuscript (Section 2.1, lines 77-79).

Fig. 2

From: A physically consistent dataset of water-energy-carbon fluxes across the Soil-Plant-Atmosphere Continuum



2.Further reasons why including evergreen broadleaf forests (IT-Lav), mixed forests (BE-Vie), and permanent wetlands (US-Los) for validation should be further highlighted. After revise the minor issues, the manuscript can be accepted for publication.

Reply: We appreciate the reviewer's suggestion. The three validation sites (IT-Lav, BE-Vie, and US-Los) were intentionally selected to represent distinct ecosystem types and climatic conditions. IT-Lav represents evergreen broadleaf forests with high productivity and strong canopy–atmosphere coupling; BE-Vie represents mixed forests with transitional canopy structures and variable moisture regimes; and US-Los represents permanent wetlands with high water availability and dominant latent heat fluxes. These contrasting sites thus encompass a broad range of vegetation structures and energy–water exchange characteristics, providing a robust validation of our approach across different ecosystems. This rationale has been added to the revised manuscript (Section 4.1.2, lines 231–233).

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, 300*, 1560-1563, 2003

Peng, L., Wei, Z., Zeng, Z., Lin, P., Wood, E.F., & Sheffield, J. Reducing solar radiation forcing uncertainty and its impact on surface energy and water fluxes. *J. Hydrometeorol., 22*, 813-829, 2021

Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M., & Hashimoto, H. A continuous satellite-derived measure of global terrestrial primary production. *Bioscience, 54*, 547-560, 2004

Wang, Y., Zeng, Y., Alidoost, F., Schilperoort, B., Song, Z., Yu, D., Tang, E., Han, Q., Liu, Z., & Peng, X. A physically consistent dataset of water-energy-carbon fluxes across the Soil-Plant-Atmosphere Continuum. *Sci. Data, 12*, 1146, 2025

Wang, Y., Zeng, Y., Yu, L., Yang, P., Van der Tol, C., Yu, Q., Lü, X., Cai, H., & Su, Z. Integrated modeling of canopy photosynthesis, fluorescence, and the transfer of energy, mass, and momentum in the soil–plant–atmosphere continuum (STEMMUS–SCOPE v1. 0.0). *Geoscientific Model Development, 14*, 1379-1407, 2021