

# H2CM (v1.0): hybrid modeling of global water–carbon cycles constrained by atmospheric and land observations

Zavud Baghirov<sup>1,2</sup>, Markus Reichstein<sup>1,5</sup>, Basil Kraft<sup>4</sup>, Bernhard Ahrens<sup>1</sup>, Marco Körner<sup>2,3</sup>, and Martin Jung<sup>1</sup>

<sup>1</sup>Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, Germany

<sup>2</sup>Department of Aerospace and Geodesy, TUM School of Engineering and Design, Technical University of Munich (TUM), Germany

<sup>3</sup>Munich Data Science Institute, Technical University of Munich (TUM), Munich, Germany

<sup>4</sup>ETH Zurich, Environmental Systems Science, Zurich, Switzerland

<sup>5</sup>ELLIS Unit Jena at Michael-Stifel-Center Jena for Data-driven and Simulation Science, Jena, Germany

**Correspondence:** Zavud Baghirov (zbaghirov@bgc-jena.mpg.de)

**Abstract.** We present the Hybrid Hydrological Carbon Cycle Model (H2CM)—a global model that couples the terrestrial water and carbon cycles by integrating a process-informed deep learning approach with observational constraints for the water and carbon cycles. H2CM extends the hybrid hydrological model with vegetation (H2MV) to represent key terrestrial carbon fluxes, including gross primary productivity (GPP), autotrophic and heterotrophic respiration at daily resolution and 1-degree spatial scale. H2CM uses neural networks to learn and predict ecosystem properties governing water and carbon fluxes, such as carbon and water use efficiencies and basal respiration rate. H2CM ~~uniquely combines~~ provides a “reanalysis” of recent land water-carbon cycle variations by combining multiple observational constraints synergistically: on top of hydrological and vegetation data constraints on terrestrial water storage variations, snow water equivalent, evapotranspiration, runoff and fraction of photosynthetically active radiation, the carbon cycle is informed by an observation-based GPP product, and net ecosystem exchange (NEE) from satellite and in-situ based atmospheric  $CO_2$  inversion datasets. H2CM reproduces the seasonal and interannual dynamics of carbon fluxes well. H2CM outperforms both purely data-driven models as well as state-of-the-art process-based model ensembles in capturing NEE seasonality, especially in challenging regions such as the South American tropics and Southern Africa. Moreover, H2CM reveals emergent spatial patterns in precipitation use efficiency, light use efficiency, and water-carbon coupling, consistent with empirical ecological understanding. Notably, we show that H2CM learns to represent the rain pulse effect on respiration in dry regions, which is often not well reproduced by global models. H2CM represents a key step toward a new generation of data-driven diagnostic land surface models, with planned extensions to include the energy cycle.

## 1 Introduction

The water and carbon cycles are critical components of Earth’s ecosystems, significantly influencing our understanding of climate, water resources, and carbon dynamics. Previous studies have demonstrated that global water and carbon cycles are strongly interconnected (Jung et al., 2017; Humphrey et al., 2018).

Global water and carbon cycle processes are typically modeled using two main strategies: data-driven modeling and process-based modeling (PBM). Data-driven approaches often involve machine learning (ML) to estimate quantities (e.g., fluxes) related to the water or carbon cycle (Dou et al., 2018; Shi et al., 2024; Tian et al., 2023; Jung et al., 2019, 2020; Nelson et al., 25 2024). ML can learn from observational data with minimal prior knowledge, relaxing uncertain assumptions and potentially leading to new insights and accurate predictions. This becomes increasingly relevant as the volume of Earth observation data grows (Huntingford et al., 2019; Rolnick et al., 2022; Schneider et al., 2017; Eyring et al., 2024). However, a significant caveat of using ML to explain the Earth system is that these models are very difficult to interpret in terms of learned intermediate processes and mechanisms (e.g., they function as a “black box”). Additionally, they may suffer from extrapolation issues and 30 they do not guarantee adherence to well-established process-knowledge (Shen et al., 2023; Reichstein et al., 2019).

Unlike the ML strategy, PBMs represent process-understanding explicitly and adhere to fundamental laws, such as mass conservation (Le Quéré et al., 2012; Sitch et al., 2015, 2024). By design, PBMs’ simulations can output various diagnostic variables that are easy to interpret and help to understand drivers of water-carbon cycle variations in the model. However, PBMs abstract the complex processes governing water and carbon cycles and require numerous assumptions about processes 35 due to incomplete process knowledge. Thus, assumptions and modeling approaches vary across different PBMs, while this uncertainty is reflected in significant inter-model spread of simulations (O’Sullivan et al., 2022). Additionally, unlike ML approaches, PBMs are not designed for fully exploiting the growing volume of Earth observations (Nearing et al., 2021; Shen et al., 2018; Kraft et al., 2021).

A novel approach to modeling global water and carbon cycles—hybrid modeling—has recently emerged. This approach 40 combines ML and PBM within a single framework and aims for leveraging the advantages of both while mitigating their challenges. For instance, hybrid modeling can replace uncertain parameters or process representations of a PBM with ML estimations, while retaining established process knowledge (e.g., mass balance) in the PBM formulations. This can reduce the physical inconsistencies of ML, as ML predictions must pass through process formulations that constrain them to obey process knowledge and maintain physical units. Hybrid modeling also reduces the need for prior assumptions about PBM’s uncertain 45 components, if they can be learned from data. Consequently, hybrid modeling allows for leveraging the growing volume of Earth observations through its ML component while maintaining physical plausibility through its PBM component (Eyring et al., 2024; Reichstein et al., 2019; Shen et al., 2023). However, this remains a young and evolving field; global-scale hybrid modeling studies are at the proof-of-concept stage and extensive evaluation and assessment of the plausibility of these models is still needed (Baghirov et al., 2025; Kraft et al., 2021).

50 A major coupling mechanism between the global water and carbon cycles is related to the carbon-water trade-off during photosynthesis. As plants open their stomata to fix carbon dioxide from the atmosphere they are transpiring water back into the atmosphere (Katul et al., 2012; Heimann and Reichstein, 2008). Soil moisture is another nexus of the water and carbon cycles, as it regulates many processes such as photosynthesis (Humphrey et al., 2018) and heterotrophic respiration (Zhang et al., 2018).

55 In this study, we introduce the hybrid hydrological-carbon cycle model (H2CM), building on the hybrid hydrological model with vegetation (H2MV, Baghirov et al. (2025)). H2MV combines a hydrological model with deep learning (DL)

(LeCun et al., 2015), while being constrained by an array of complementary observational data streams on terrestrial water storage (TWS) variations, snow water equivalent (SWE), evapotranspiration (ET), runoff and fraction of photosynthetically active radiation (fAPAR). H2CM extends this approach to include the carbon cycle, where transpiration and gross primary productivity (GPP) are explicitly coupled through neural network (NN) predictions of water use efficiency (WUE). H2CM also represents carbon use efficiency (CUE, defined as the ratio of net to gross primary production) and basal heterotrophic respiration rate (R<sub>b</sub>) by NNs to simulate respiration processes. To improve plausibility and causality of NN predictions we couple separate neural networks, which receive individual sets of meaningful inputs for their task, rather than using one large network that receives all inputs to make all estimations. The model simulates spatio-temporal dynamics of GPP and net primary productivity (NPP), heterotrophic respiration (R<sub>h</sub>), and thus net ecosystem exchange (NEE) representing the net carbon exchange between land and atmosphere. H2CM's GPP estimations are directly constrained using observation-based data, while NEE estimations are constrained using atmospheric inversion products. Thus, H2CM aims at facilitating a "reanalysis" of carbon-water cycle variations by a joint interpretation of diverse observational data streams based on conceptual process-understanding incorporated into deep learning.

This study has the overall objective to describe and evaluate H2CM. To this end we 1) evaluate the model's performance against withheld data constraints; 2) compare the model's GPP, terrestrial ecosystem respiration (TER), and NEE simulations with state-of-the-art data-driven and process-based models; 3) assess the plausibility of the model's learned global patterns for precipitation, water, carbon, and light use efficiencies; 4) illustrate the model's potential to explain and understand carbon flux variations in the challenging region of Southern Africa.

## 2 Methods

### 2.1 ~~Datasets~~H2CM

~~Our model receives meteorological time series data for each grid-cell, including precipitation, net radiation, air temperature, vapor pressure deficit, and the atmospheric concentration of CO<sub>2</sub>, as inputs (Table 2). We use net radiation as a single forcing term rather than separately prescribing shortwave and longwave components. This simplification leverages the strong empirical correlation between shortwave and net radiation at daily scales (Jung et al., 2024). Additionally, we incorporate static inputs, which vary only between grid-cells that consist of soil properties, land cover fractions, elevation, and wetlands, as described in (Baghirov et al., 2025). In general, we selected meteorological and static datasets that are widely used in the community, quality checked, and offer the best compromise between spatial/temporal resolution and observational accuracy for each variable.~~

~~All meteorological forcing and data constraints were aggregated to a 1° spatial resolution. The spatial resolutions of static inputs were aggregated to 1/30°. Meteorological forcing data are maintained in their native daily temporal resolutions, while a monthly temporal resolution is applied to the data constraints.~~

We use the term “data constraints” to refer to the observation-based data used as target variables to optimize the model. These data guide the model’s behavior through its NN components. We utilize data constraints of TWS variations, SWE, ET, runoff, fAPAR, GPP, and NEE (Table 2).

We use the OCO-2 L4 NLGIS atmospheric inversion MIP (v10) ensemble median to constrain our estimations of NEE. H2CM consists of three primary modules: a static (fully connected) NN module, which generates spatially varying parameters from land cover, soil, and elevation inputs; a dynamic NN module, which produces spatially and temporally varying variables from meteorological forcing; and the process-based (water-carbon cycle) module, which receives the output of the static and dynamic modules, and where fluxes and states are calculated through process equations, which also ensure mass conservation. Note that H2CM’s static NN module outputs are at  $1^\circ$  spatial resolution for the period 2015–2020, which uses satellite column and in-situ  $\text{CO}_2$  observations jointly. Additionally, we use CarboScope (in-situ based NEE inversion product) to constrain the long-term spatial average monthly anomalies of NEE estimations. Both OCO-2 MIP and CarboScope estimate the carbon exchange, which includes contributions from fire emissions that we do not model. Therefore, we use fire emission data from GFED (v4) (Chen et al., 2023) to subtract fire emissions from OCO-2 and CarboScope inversions. We only use the mean seasonal cycle of GPP and ET from FLUXCOM-X-BASE, and runoff from GRUN to constrain the seasonality of our simulations (Table 2), due to caveats of representing interannual variability of these data products (Ghiggi et al., 2019; Jung et al., 2019, 2020; Nelson et al., 2024).

Additionally, we apply a soft constraint on the spatial distribution of mean annual CUE ( $\text{NPP} / \text{GPP}$ ), using the TRENDY (v11) ensemble median of global process-oriented ecosystem models (Sitch et al., 2024) as data constraint. This helps to constrain plausible relative contributions of autotrophic and heterotrophic respiration in H2CM. Note that this step only constrains the magnitude of mean CUE based on theoretical expectations, while temporal variations of CUE emerge from training H2CM, whereas the dynamic NN and process-based modules’ simulations are at  $1^\circ$  spatial and daily temporal resolution.

Overview of datasets: meteorological forcings, static inputs, and data/observational constraints. All temporally varying constraints were aggregated to monthly resolution. The Resolution column lists the original (native) resolution of each dataset.

## 2.2 H2CM

### 2.1.1 Hydrological cycle

We utilize the hydrological cycle model component of H2MV (Baghirov et al., 2025) within our model. This hydrological model estimates three primary water storages: snow, soil moisture, and groundwater. It estimates key water fluxes, including snowfall and snowmelt, soil and groundwater recharge, evapotranspiration (divided into transpiration, soil evaporation, and interception evaporation), and runoff (divided into fast and slow runoff) to update these water storages.

In the equations below, parameters with the superscript  $\langle s, t \rangle$  indicate variables that vary both spatially ( $s$ ) and temporally ( $t$ ), while those with the superscript  $\langle s \rangle$  refer solely to spatial variation. Parameters without any superscript represent

120 globally constant parameters (denoted by the Greek letter  $\beta$  or  $Q_{10}$ ) that do not vary either in space or time and are learned by the neural network.

The main coupling mechanism between water and carbon cycles in H2CM is through linking transpiration and GPP by modelling WUE.

Transpiration (T) is modeled as described by Baghirov et al. (2025):

$$125 \quad T^{<s,t>} = \text{fAPAR}^{<s,t>} \cdot ET_{pot}^{<s,t>} \cdot \alpha_T^{<s,t>} \quad (\text{in mm day}^{-1}) \quad , \quad (1)$$

where transpiration is a function of the predicted vegetation state (fAPAR), potential evapotranspiration, and a neural network-learned parameter  $\alpha_T$ , which represents effective conductance or stress. The predictions of  $\alpha_T$  depend on relative soil moisture, vapor pressure deficit, net radiation, and static variables (Table 1).

Parameters such as  $\alpha_T$  are explicit NN outputs. During training, the NN predicts  $\alpha_T$  from its inputs, and this value—together  
 130 with variables like fAPAR—is passed to the process-based component to estimate transpiration, which contributes to total ET. Because ET is constrained by observations, training minimizes the loss between predicted and observed ET, thereby adjusting  $\alpha_T$ , among other parameters, iteratively. Moreover,  $\alpha_T$  is also influenced indirectly through other observational constraints (e.g., TWS, runoff, GPP), so multiple datasets jointly influence its learning. Further details on the hydrological model are available in Baghirov et al. (2025).

## 135 2.1.2 Carbon cycle

We employ a simple, conceptual carbon cycle model to estimate carbon cycle fluxes (Fig. 1).

GPP is the amount of carbon dioxide absorbed by vegetation to produce organic matter through photosynthesis. We model GPP as:

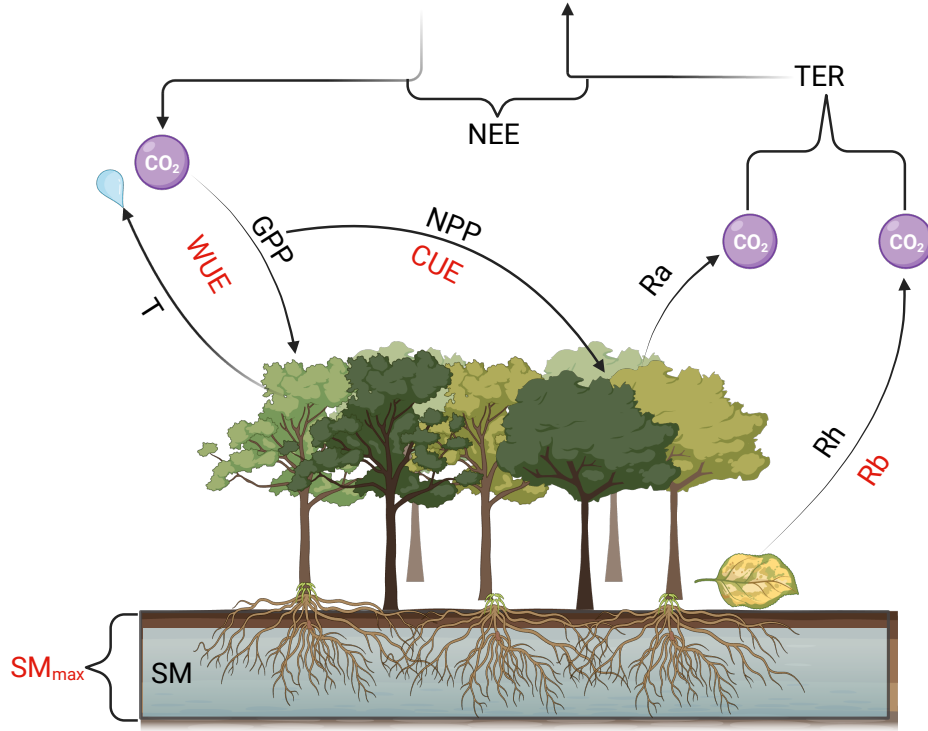
$$GPP^{<s,t>} = T^{<s,t>} \cdot \alpha_{WUE}^{<s,t>} \cdot CO_2^{<s,t>} \cdot \beta_{CO_2} \quad (\text{in gC m}^{-2} \text{ day}^{-1}) \quad (2)$$

140 where GPP is a function of transpiration  $T$ , atmospheric  $CO_2$  concentration, the global constant  $\beta_{CO_2}$  describing the  $CO_2$  fertilization effect and NN learned  $\alpha_{WUE}$  describing WUE without the  $CO_2$  fertilisation effect. The  $\beta_{CO_2}$  term is a globally constant, trainable parameter that regulates the strength of the  $CO_2$  fertilization effect on GPP. Although  $\beta_{CO_2}$  does not explicitly represent a dynamic  $CO_2$  fertilization effect, the model's linear dependence on atmospheric  $CO_2$  interacts with  $\alpha_{WUE}$  which is learned by the neural network as a nonlinear function of vapor pressure deficit, relative soil moisture, net  
 145 radiation, and static variables (Table 1).

NPP is the net carbon available after autotrophic plant respiration. NPP is modeled as:

$$NPP^{<s,t>} = GPP^{<s,t>} \cdot CUE^{<s,t>} \quad (\text{in gC m}^{-2} \text{ day}^{-1}) \quad (3)$$

where NPP is a function of GPP and CUE. CUE is directly learned by the NN which receives air temperature, vapor pressure deficit, net radiation, atmospheric  $CO_2$  concentration, relative soil moisture, fAPAR and NPP at previous time step, and static  
 150 variables (Table 1). Autotrophic respiration is simply calculated as the difference between GPP and NPP.



**Figure 1.** High level overview of the carbon cycle related processes in H2CM: Acronyms above or below the arrows shown in black (e.g., GPP) represent modeled parameters. Acronyms next to the arrows and modelled parameters shown in red (e.g., WUE), indicate parameters directly learned by neural networks to model the corresponding parameter next to the arrow. T: transpiration, WUE: water use efficiency, GPP: gross primary productivity, CUE: carbon use efficiency, NPP: net primary productivity, NEE: net ecosystem exchange, Ra: autotrophic respiration, Rh: heterotrophic respiration, Rb: basal respiration rate, TER: terrestrial ecosystem respiration,  $SM_{max}$ : maximum soil moisture capacity, SM: soil moisture. Created in BioRender (Baghirov, 2025a).

Heterotrophic respiration, Rh, refers to the process by which non-photosynthetic organisms (e.g., microorganisms) decompose organic matter, releasing carbon dioxide back to the atmosphere. We model Rh using a traditional  $Q_{10}$  function:

$$Rh^{<s, t>} = Q_{10}^{\frac{T_{air}^{<s, t>} - T_{ref}}{10}} \cdot Rb^{<s, t>} \quad (\text{in } \text{gC m}^{-2} \text{ day}^{-1}) \quad (4)$$

where  $Q_{10}$ , a global constant learned by the NN, describes the temperature sensitivity factor, indicating by which factor respiration increases with a  $10^\circ\text{C}$  rise in air temperature.  $T_{air}$  is the current air temperature at time  $t$ ,  $T_{ref}$  is the reference temperature set to  $15^\circ\text{C}$ , and Rb is the basal respiration rate essentially representing the availability of carbon for Rh and it is learned by a recurrent (stateful) NN as a function of precipitation, net radiation, fAPAR and NPP at previous time step, and static variables (Table 1). TER is the sum between autotrophic (Ra) and heterotrophic (Rh) respiration representing the total

amount of carbon dioxide being respired to the atmosphere. NEE is calculated as the difference between TER and GPP, where  
160 negative values indicate a land carbon sink.

### 2.1.3 ~~High-level overview~~ Overview of the hybrid architecture

H2CM consists of three primary modules: a static (fully connected) NN module, which generates spatially varying parameters  
from Figure 2 illustrates the overall architecture of H2CM:

(A) Static inputs and compression. Panel A1 represents the static environmental inputs, including land cover, soil ~~and~~  
165 ~~elevation inputs~~; a dynamic NNs module, which produces spatially and temporally varying variables from meteorological  
forcing; and the process-based (water-carbon cycle) module, which receives the output of the static and dynamic modules,  
~~and where fluxes are calculated through process equations, which also ensure mass conservation.~~ properties, wetland  
extent, and digital elevation. These features are passed through a fully connected neural network (FC-NN 1; A2) that  
170 compresses spatially varying but temporally invariant information into a latent vector. FC-NN 1 has two hidden layers  
with 150 and 12 units, respectively. The 12-unit latent vector is used as compressed static features and serves as a shared  
input to all dynamic sequence layers. The outputs of this static network (A3) include spatially varying parameters such  
as maximum soil moisture capacity ( $SM_{max}$ ) and  $\alpha_{Ei}$  controlling interception evaporation.

(B–E) Dynamic sequence models. H2CM ~~'s~~ static module consists of a FCNN (Fig. ??G) that receives static land properties  
as inputs (Fig. ??F) and compresses them. These compressed representations are connected to another FCNN (omitted in ??G  
175 for clarity), which further transforms them into physically interpretable static parameters (e.g.,  $SM_{max}$ , the maximum water  
holding capacity of the soil) (Fig. ??H). Additionally, these compressed representations of land properties are then sent as  
inputs to the the dynamic module (Fig. ??B) to facilitate interaction effects between meteorological forcing and land properties.

In the dynamic module, meteorological forcing data (inputs) at a single time step  $t$  (Fig. ??A) are processed through  
180 task-specific NNs, mostly a dynamic NN layer, specifically an LSTM (Hochreiter and Schmidhuber, 1997) (Fig. ??B). LSTM is  
a type of recurrent neural network (RNN) architecture designed to effectively capture long-range dependencies and patterns  
in sequential data by leveraging its gating mechanisms to learn memory effects (Staudemeyer and Morris, 2019). The LSTM  
transforms meteorological inputs at time step  $t$  into higher-dimensional parameters, which are then fed into includes three long  
short-term memory (LSTM) networks and an additional fully connected neural ~~networks~~ (FCNN) (omitted in Fig. ??B for  
185 clarity) to convert them network (FC-NN 2) to model time-varying processes. Each LSTM contains one hidden layer with 100  
units and is connected to a small fully connected layer that transforms hidden states into physically interpretable parameters like  
WUE (Fig. ??C), that are used in the process-based module. For this, the LSTM also receives estimated water and vegetation  
states (Fig. ??E) from the previous time step, besides its own learned internal hidden and cell states (representing long-term  
memory). These dynamic modules produce spatio-temporal predictions:

- 190 – LSTM 1 (B1–B3): Receives the compressed static features together with dynamic drivers such as net radiation, precipitation, relative soil moisture, snow, groundwater, and fAPAR at time  $t - 1$ . It predicts coefficients ( $\alpha_{rsoil}$ ,  $\alpha_{rgw}$ ,  $\alpha_{smelt}$ ) that control soil recharge, groundwater recharge, and snowmelt processes.
- LSTM 2 (C1–C3): Takes inputs including the static compressed vector, air temperature, vapor pressure deficit,  $CO_2$  concentration, relative soil moisture, and fAPAR and NPP at time  $t - 1$ . It estimates carbon use efficiency and fAPAR.
- 195 – LSTM 3 (D1–D3): Uses static compressed features, net radiation, precipitation, fAPAR, and NPP at time  $t - 1$  to predict the basal respiration rate and  $\alpha_{Es}$  parameter that controls soil evaporation.
- FC-NN 2 (E1–E3): A fully connected neural network with two hidden layers (150 and ~~compressed representations of static land inputs (Fig. ??F and G) to make estimations for the next time step. All estimations within the dynamic module vary in both space and time.~~ 12 units) that predicts water use efficiency and the  $\alpha_T$  parameter, which represents effective conductance or stress response.
- 200

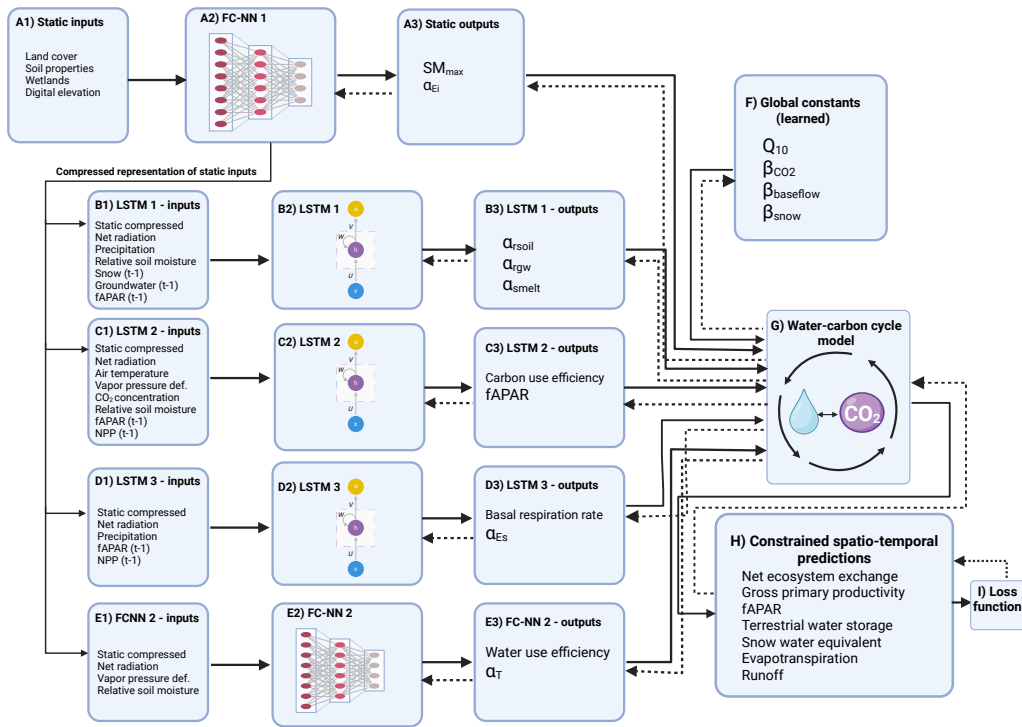
~~H2CM also includes globally constant learned parameters (Fig. ??I) to represent parameters that are fixed across both space and time.~~

~~In the process-based module, meteorological forcing (Fig. ??A)~~

- (F) Global constants (learned). A set of globally learned parameters ( $Q_{10}$ , ~~NN~~ $\beta_{CO_2}$ ,  $\beta_{baseflow}$ ,  $\beta_{snow}$ ) provide scaling relationships for temperature sensitivity of respiration,  $CO_2$  fertilization, baseflow recession, and correction of snowfall.
- 205
- (G) Coupled water-carbon cycle model. The ~~outputs from the dynamic (Fig. ??C), and static modules (Fig. ??H), and global constants (Fig. ??I) are input into a conceptual water-carbon cycle~~ static and dynamic subnetworks are passed to a differentiable, process-based ~~model (Fig. ??D) to estimate water and carbon fluxes, as well as water states (Fig. ??E).~~ water-carbon cycle model that enforces mass balance between fluxes. This model represents the physical coupling between hydrological and carbon processes, ensuring consistency between water storage, evapotranspiration, carbon assimilation, and respiration.
- 210
- (H) Constrained spatio-temporal predictions. During training, all components of the hybrid architecture are optimized jointly, including the static and dynamic neural subnetworks and the differentiable process-based water-carbon cycle model. Gradients are propagated through the full model from the loss function back to the trainable parameters, so that all trainable components are updated simultaneously. The neural network components generate spatially and spatio-temporally varying parameters that are passed into the process model, which produces simulated water fluxes and storages and carbon fluxes. Some of these ~~predictions are directly constrained using observations or observation-based data (Fig. ??E).~~ Note that the model is executed sequentially for each time step, starting from  $t = 0$  until the final time step. For a detailed description of the hybrid architecture, see ~~outputs (net ecosystem exchange, gross primary productivity, fAPAR, terrestrial water storage, snow water equivalent, evapotranspiration, and runoff) are compared against observational targets through a composite loss function (I). Because the process-based component is fully differentiable, gradients of~~
- 215
- 220

the loss propagate through the process equations back to the neural subnetworks via automatic differentiation (dashed arrows in Fig. 2 and Appendix A-). This enables the networks to learn physically consistent parameterizations that minimize discrepancies between modeled and observed dynamics.

225



**Figure 2. Simplified overview** of the H2CM framework. **Note:** NNs refers to neural networks. Static inputs (e.g., land cover, soil, elevation) are processed through a fully connected neural network (two hidden layers: 150 and FC-NN refers 12 units) to fully-connected neural networks generate compressed static features. Note that this figure omits the 'guided' nature of the model. These features feed into three LSTM networks (different groups each with one hidden layer of neural networks-100 units) and an additional fully connected network that receive different inputs for their task predict spatio-temporal parameters regulating hydrological and carbon processes, including  $\alpha_{rsoil}$ ,  $\alpha_{rgw}$ ,  $\alpha_{smelt}$ ,  $\alpha_{Es}$ ,  $\alpha_{Ei}$ , and  $\alpha_T$ . Globally learned constants ( $Q_{10}$ ,  $\beta_{CO2}$ ,  $\beta_{baseflow}$ ,  $\beta_{snow}$ ) control temperature sensitivity,  $CO_2$  fertilization, baseflow recession, and only shows very high-level overview snow corrections. Outputs from all subnetworks are coupled through a differentiable, mass-balanced water-carbon process model. The process layer produces water fluxes and storages and carbon fluxes, some of which are directly constrained by observations, including NEE, GPP, fAPAR, TWS, SWE, ET, and runoff. The dashed lines represent backpropagation used to train H2CM. Created in BioRender (Baghirov, 2026c) (Baghirov, 2026d).

When we integrate the carbon cycle with the hydrological cycle from H2MV, the number of NN predicted parameters and inputs to the model increases. This expansion raises the risk of the NN component making estimations based on unreasonable relationships, known as short-cut learning (Geirhos et al., 2020). This issue arises due to potential covariations among the

**Table 1.** Guided neural networks: The first column shows the estimated parameters, the second column lists the inputs used in the estimation process, the third column details the types of neural networks used for estimating these parameters, and the last column shows behaviour each neural network controls.  $\alpha_{WUE}$ : Water use efficiency without  $CO_2$  fertilisation, CUE: Carbon use efficiency, fAPAR: fraction of absorbed photosynthetically active radiation, Rb: Basal respiration rate, FCNN: Fully connected neural networks, LSTM: Long short-term memory, Rn: Net radiation,  $\frac{SM}{SM_{max}}$ : relative soil moisture, SWE: snow water equivalent, GW: Groundwater, VPD: vapor pressure deficit,  $CO_2$ : atmospheric carbon-dioxide concentration, NPP: Net primary productivity. The  $\alpha$  parameters that are used in hydrological cycle are introduced in Baghirov et al. (2025).

Parameter	Inputs	Neural networks	Neural network's role
$SM_{max}, \alpha_{Ei}$	Static	FCNN	Spatial parameters
$\alpha_{r_{soil}}, \alpha_{r_{gw}}, \alpha_{s_{melt}}$	Rn, Precipitation, $\frac{SM}{SM_{max}}, SWE^{t-1}, GW^{t-1}, fAPAR^{t-1}, Static$	LSTM	Most hydrological behaviour
$\alpha_{WUE}, \alpha_T$	Rn, VPD, $\frac{SM}{SM_{max}}, Static$	FCNN	Water-carbon coupling
CUE, fAPAR	Rn, $T_{air}, VPD, CO_2, \frac{SM}{SM_{max}}, fAPAR^{t-1}, NPP^{t-1}, static$	LSTM	Vegetation & its respiration
Rb, $\alpha_{Es}$	Rn, Precipitation, $fAPAR^{t-1}, NPP^{t-1}, Static$	LSTM	Upper soil sensitivity

inputs, where different inputs or their combinations might be used to estimate certain parameters, even if such associations are implausible based on established process knowledge.

To address this, we ‘guide’ NNs by ensuring they learn from plausible inputs when estimating specific parameters. Instead of employing a single large NN to predict all parameters simultaneously, which is a common approach, we divide the neural networks into groups. Each group is restricted to learning from plausible relationships, enhancing the reliability of the model’s predictions (Table 1 and Fig. 2).

## 2.2 Model optimization

### 2.2.1 Cross validation (CV)

To evaluate the generalizability of H2CM, we use a 10-fold cross-validation (CV) setup. ~~This process involves dividing the data~~ For this purpose, the spatial domain is divided into 10 subsets, or CV folds, and training 10 different folds composed of different grid cells. We then train 10 separate models, each time leaving out one fold as a of grid cells as the validation set and using the remaining folds for training. The training set is used to optimize the model, while the validation fold is exclusively for hyperparameter tuning, such as selecting the learning rate, and for early stopping. Furthermore, we keep a separate set of grid cells as a testing set, which is used only after training to assess the model’s performance and generalizability. Importantly, the testing set is never exposed to the models during the training phase (Fig. A1).

We divide the global data into training, validation, and testing sets using a spatial-only splitting method (Fig. A1). This approach ensures that different grid cells are included in each of the training, validation, and testing sets. To address spatial autocorrelation among grid cells, we implement a spatial blocking technique as recommended by Roberts et al. (2017). In this approach, “blocks” refer to spatially contiguous groups of  $5^\circ \times 5^\circ$  grid cells (25 grid cells in total) that are treated as single

sampling units. These blocks are randomly selected from the global data, ensuring that the majority of the data (approximately 80% of the grid cells) is allocated to the training set. The remaining ~20% of the data is reserved for validation and testing. In practice, each of the 10 cross-validation folds contains roughly 1/11 of the total data for validation, while ~1/11 of the data is held out as a fixed testing set for final evaluation. The random selection of blocks also ensures that the training, validation, and testing sets approximately represent grid cells from all continents.

This split is crucial for accurately testing the model’s generalizability on unseen testing grid cells and helps in understanding potential overfitting to the training data. Ideally, testing on both spatial and temporal dimensions, known as spatio-temporal splitting, would provide a more rigorous assessment. However, with the current set-up implementing proper time splitting for cross-validation is not feasible due to inconsistent and limited temporal coverage across different data constraints (Table 2). To assess the model’s generalizability across both space and time, we conducted an additional experiment in which the model was trained using data from 2001 to 2017 and evaluated on the subsequent two years (2018–2019; Appendix [D C](#)). The results show qualitative agreement with the spatial CV setup, thereby justifying its validity in this study.

### 2.2.2 Loss function

We use the mean squared error (MSE) as the loss function, following the approach in Baghirov et al. (2025). The loss function MSE

$$L(X, \Theta, \beta) = \frac{1}{N_c} \sum_{c=1}^C \sum_{i=1}^{N_c} (y_{c,i} - \hat{y}_{c,i})^2 \quad (5)$$

evaluates the model’s performance based on the inputs  $X$ , the neural network weights  $\Theta$ , and the global constants  $\beta$ . In this context,  $C$  denotes the number of data constraints,  $N_c$  is the number of data points for each constraint  $c$ , and  $y_{c,i}$  and  $\hat{y}_{c,i}$  represent the observed and predicted values for the constraint  $c$ , respectively. Throughout the training process,  $\Theta$  and  $\beta$  are adjusted to minimize the overall loss  $L$ . Note that we apply a Z-transformation to predictions and observations before computing the loss. This standardizes each variable by subtracting its mean and dividing by its standard deviation, removing the effect of different units and balancing the contributions of each data constraint to the total loss.

The loss function combines all data constraints into a single objective, where MSE is computed by concatenating the data across grid cells and monthly data for each variable. The only exception is for the data constraint on long-term NEE from CarboScope. In situ-based inversions such as CarboScope are generally more robust when averaged globally but become increasingly uncertain at the grid-cell level. Therefore, we first compute the spatial mean of both CarboScope and H2CM NEE monthly anomalies during training within each batch ([subset of training samples processed together in one optimization step](#)). This yields time series of spatial averages for which we calculate the loss.

### 2.2.3 Model training

We use the strategy outlined by Baghirov et al. (2025) to optimize H2CM for each CV fold separately, which includes utilizing the Adam optimizer (Kingma and Ba, 2014). We set the initial learning rate to 0.005 and used a step-wise learning-rate

scheduler (StepLR; Paszke et al. (2019)) that decays the rate by a fixed factor at fixed epoch intervals. Hyperparameters were  
280 tuned by training multiple model variants and selecting the configuration that achieved strong validation performance with  
stable training. During training, all inputs to the neural networks are standardized using Z-transformation (so deviations from  
the mean are preserved in standardized units). We also implement early stopping, a technique that monitors both validation  
and training losses. This approach halts further training if the model’s performance on the validation set either declines or  
remains unchanged, thereby preventing overfitting to the training data. Throughout training, we keep track of the validation  
285 loss and select the final model based on the smallest total loss observed on the validation set. This selected model is then used  
for making final predictions, such as on the testing set.

### 2.3 Datasets

Our model receives meteorological time series data for each grid-cell, including precipitation, net radiation, air temperature,  
vapor pressure deficit, and the atmospheric concentration of  $CO_2$ , as inputs (Table 2). We use net radiation as a single forcing  
290 term rather than separately prescribing shortwave and longwave components. This simplification leverages the strong empirical  
correlation between shortwave and net radiation at daily scales (Jung et al., 2024). Additionally, we incorporate static inputs,  
which vary only between grid-cells that consist of soil properties, land cover fractions, elevation, and wetlands, as described in  
(Baghirov et al., 2025). In general, we selected meteorological and static datasets that are widely used in the community,  
quality-checked, and offer the best compromise between spatial/temporal resolution and observational accuracy for each  
295 variable.

All meteorological forcing and data constraints were aggregated to a  $1^\circ$  spatial resolution. The spatial resolutions of static  
inputs were aggregated to  $1/30^\circ$ . Meteorological forcing data are maintained in their native daily temporal resolutions, while a  
monthly temporal resolution is applied to the data constraints.

Throughout, we use “data constraints” to denote data-driven constraints implemented as observation-based loss terms using  
300 TWS, SWE, ET, runoff, fAPAR, GPP, and NEE (Table 2). These terms guide the model’s predictions but are not physically  
defined hard bounds or conservation laws. Physical constraints (e.g., mass balance) are enforced by the process-oriented  
component of the model.

We use the OCO-2 LNLGIS atmospheric inversion MIP (v10) ensemble median to constrain our estimations of NEE at  $1^\circ$   
spatial resolution for the period 2015-2020, which uses satellite column and in-situ  $CO_2$  observations jointly. Additionally, we  
305 use CarboScope (in-situ based NEE inversion product) to constrain the long-term spatial average monthly anomalies of NEE  
estimations. Both OCO-2 MIP and CarboScope estimate the carbon exchange, which includes contributions from fire emissions  
that we do not model. Therefore, we use fire emission data from GFED (v4) (Chen et al., 2023) to subtract fire emissions  
from OCO-2 and CarboScope inversions. We only use the mean seasonal cycle of GPP and ET from FLUXCOM-X-BASE,  
and runoff from GRUN to constrain the seasonality of our simulations (Table 2), due to caveats of representing interannual  
310 variability of these data products (Ghiggi et al., 2019; Jung et al., 2019, 2020; Nelson et al., 2024).

Additionally, we apply a soft constraint on the spatial distribution of mean annual CUE (NPP / GPP), using the TRENDY  
(v11) ensemble median of global process-oriented ecosystem models (Sitch et al., 2024) as data constraint. This helps to

**Table 2.** [Overview of datasets: meteorological forcings, static inputs, and data/observational constraints. All temporally varying constraints were aggregated to monthly resolution. The Resolution column lists the original \(native\) resolution of each dataset.](#)

Name	Resolution		Temporal coverage	Data	Reference
	Spatial	Temporal			
<b>Meteorological forcing</b>					
Precipitation	1°	Daily	2001-2019	GPCP 1dd v1.2	Huffman et al. (2016)
Net radiation	1°	Daily	2001-2019	CERES SYN1deg Ed4A	Wielicki et al. (1996), Doelling (2017)
Air temperature	0.5°	Daily	2001-2019	CRUNCEP v8	Harris et al. (2014), Viovy (2018)
Vapor pressure deficit	1°	Daily	2001-2019	ERA5	Hersbach et al. (2020), Soci et al. (2024)
Atmospheric CO <sub>2</sub>	1°	Hourly	2001-2019	CAMS	Inness et al. (2019), Agustí-Panareda et al. (2023)
<b>Static data</b>					
Soil properties	1/120°	-	-	Soil grids v2	Hengl et al. (2017), Poggio et al. (2021)
Land cover fractions	1/360°	-	-	Globland30 v1	Chen et al. (2015)
Digital elevation model	1/120°	-	-	GTOPO	Center (1997)
Wetlands	1/240°	-	-	Tootchi	Tootchi et al. (2019)
<b>Data constraints</b>					
Terrestrial water storage	0.5°	Monthly	2001-2017	GRACE Tellus JPL RL06M v1	Watkins et al. (2015)
fAPAR	500m	8 daily	2001-2019	MOD15A2H	Myneni et al. (2015)
Snow water equivalent	0.25°	Monthly	2001-2018	GlobSnow v3 bias corrected	Luoju et al. (2014), Luoju et al. (2021)
Evapotranspiration	0.05°	Hourly	2001-2019	FLUXCOM-X-BASE	Jung et al. (2019), Nelson et al. (2024)
Runoff	0.5°	Monthly	2001-2019	GRUN v2	Ghiggi et al. (2021a), Ghiggi et al. (2021b)
Gross primary productivity	0.05°	Hourly	2001-2019	FLUXCOM-X-BASE	Jung et al. (2020), Nelson et al. (2024)
Net ecosystem exchange	1°	Monthly	2015-2019	OCO-2 v10 MIP	Byrne et al. (2022)
Net ecosystem exchange	1°	Monthly	2001-2019	Jena CarboScope	Rödenbeck et al. (2018)

[constrain plausible relative contributions of autotrophic and heterotrophic respiration in H2CM. Note that this step only constrains the magnitude of mean CUE based on theoretical expectations, while temporal variations of CUE emerge from training H2CM.](#)

315

## 2.4 Model evaluation

We assess our model's performance by utilizing the root mean square error (RMSE), Pearson's correlation coefficient ( $r$ ), and the standard deviation ratio (SDR), which compares the predicted standard deviation to the observed standard deviation. These metrics are also calculated based on a decomposition into the mean seasonal cycle (MSC) and monthly anomalies ( $A$ ).

320 We calculate MSC as:

$$MSC(m) = \frac{1}{Y} \sum_{y=1}^Y p_{m,y} \quad (6)$$

where  $p_{m,y}$  denotes the predicted or observed variable for month  $m$  and year  $y$ , and  $Y$  represents the total number of years.

**Table 3. Emerging global patterns: parameters** Carbon-water cycle metrics used in this study, their with abbreviations, ratios definitions, and respective units:

Parameter	Abbreviation	Ratio	Unit
Precipitation use efficiency	PUE	$\frac{NPP}{Prec}$	$gC\ kgH_2O^{-1}$
Water use efficiency	WUE	$\frac{GPP}{T}$	$gC\ kgH_2O^{-1}$
Carbon use efficiency	CUE	$\frac{NPP}{GPP}$	-
Light use efficiency	LUE	$\frac{GPP}{APAR}$	$gC\ MJ^{-1}$

The monthly anomaly ( $A(m, y)$ ) represents the deviation of a given month's value from its mean seasonal cycle:

$$A(m, y) = p_{m,y} - MSC(m) \quad (7)$$

325 where  $p_{m,y}$  represents the estimated or observed variable for a certain month  $m$  and year  $y$ , and  $MSC(m)$  denotes MSC for that specific month  $m$ .

To assess the plausibility of simulated carbon-water relations we assess emerging spatial patterns of key functional parameters detailed in Table 3.

### 3 Results and discussions

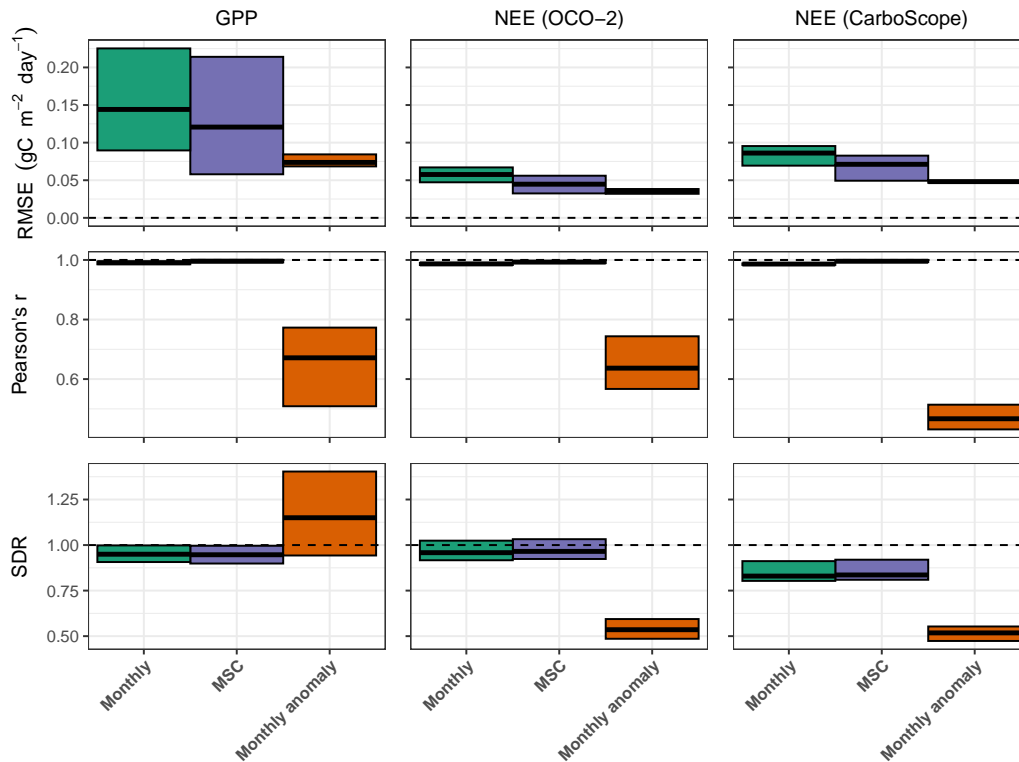
#### 330 3.1 Model evaluation

##### 3.1.1 Validation with independent test set

In this section, we evaluate the performance of H2CM in reproducing the monthly, MSC, and monthly anomalies of GPP and NEE across the testing set that was not exposed to the model during training (Fig. A1).

335 H2CM nearly perfectly reproduces the monthly and seasonal patterns of both GPP and NEE, with a Pearson's correlation coefficient ( $r$ ) close to 1 (Fig. 3, Fig. B1, Fig. B2 and Fig. B3). The monthly anomaly of GPP and NEE (based on OCO-2 data) has a Pearson's  $r$  of approximately 0.7 (mean across members), while the ~~long-term~~ monthly anomaly of NEE based on CarboScope data has a Pearson's  $r$  of approximately 0.5. ~~Note the small RMSE for NEE monthly anomaly, indicating that the reduced monthly anomaly variance contributes to lower correlations.~~ Uncertainties in the inversions of NEE monthly anomaly when locating in space and time also contribute to lowered performance metrics, as some of the variance in the inversions is  
340 noise, not signal.

The SDR for the monthly and seasonal patterns of GPP and NEE is close to 1, indicating the model's strong capability in accurately reproducing these patterns. The SDR for GPP monthly anomaly is greater than one, suggesting an overestimation of GPP monthly anomaly compared to FLUXCOM-X-BASE. This outcome is expected, as we do not explicitly constrain GPP monthly anomaly, and FLUXCOM-X-BASE is known to underestimate the GPP interannual variance (Nelson et al., 2024).



**Figure 3.** Model performance for carbon fluxes based on the spatial average over testing-set grid cells. Cross bars represent the minimum and maximum errors across the 10 cross-validation folds, and the central lines indicate the mean error. Rows correspond to performance metrics, and columns represent model constraints. RMSE denotes root mean squared error, and SDR is the standard deviation ratio (the ratio between predicted and observed standard deviation). Horizontal dashed lines denote ideal metric values, with  $R^2$  and SDR equal to 1 and RMSE equal to 0.

345 Conversely, H2CM underestimates NEE monthly anomaly in both short-term (OCO-2) and long-term (CarboScope) contexts, as indicated by a low SDR value.

In terms of RMSE, H2CM tends to exhibit higher errors for monthly and seasonal data compared to monthly ~~anomaly~~. ~~The higher RMSE in the monthly data reflects errors in reproducing both the amplitude and phase anomalies. This elevated RMSE primarily reflects the large magnitude of the seasonal cycle, such that even relatively small absolute deviations lead to~~  
 350 ~~substantial RMSE values. Despite this, the near-unity Pearson's r and SDR indicate that the phase and relative amplitude~~ of the seasonal cycle are well captured. In contrast, the monthly anomalies exclude ~~this the dominant~~ seasonal component, resulting in smaller RMSE due to their reduced variance, but typically lower correlations because only irregular year-to-year variations remain.

Globally, H2CM reproduces GPP patterns from FLUXCOM-X-BASE with RMSEs of 0.1, 0.09, and 0.05  $gC\ m^{-2}\ day^{-1}$  for monthly data, mean seasonal cycles (MSC), and monthly anomalies, respectively. For NEE from OCO-2 satellite inversions, RMSEs are 0.04, 0.03, and 0.02  $gC\ m^{-2}\ day^{-1}$  for the same categories. Compared to CarboScope in-situ inversions, H2CM yields RMSEs of 0.07, 0.06, and 0.03  $gC\ m^{-2}\ day^{-1}$ . Bias is negligible (close to 0) for GPP and NEE (OCO-2), and for NEE (CarboScope) is 0.01  $gC\ m^{-2}\ day^{-1}$  for monthly and MSC data, and zero for monthly anomalies (Table B1).

Note that, although the results discussed here are based on the spatial-only cross-validation setup, Appendix D-C demonstrates that the model also generalizes well to unseen years, at least over short-term future periods. [For example, when testing our model on unseen years for monthly GPP and NEE values, Pearson's r is close to 1 for GPP and 0.96 for NEE. These results hold for both the spatial split \(original model\) and spatio-temporal split experiments, with very similar ranges across cross-validation folds \(Fig. C1\).](#)

Model performance on water cycle-related data constraints is presented in Appendix E-B. We do not delve into the details here, as the performance is qualitatively similar to that of H2MV, which has been thoroughly discussed in Baghirov et al. (2025).

### 3.1.2 Global patterns

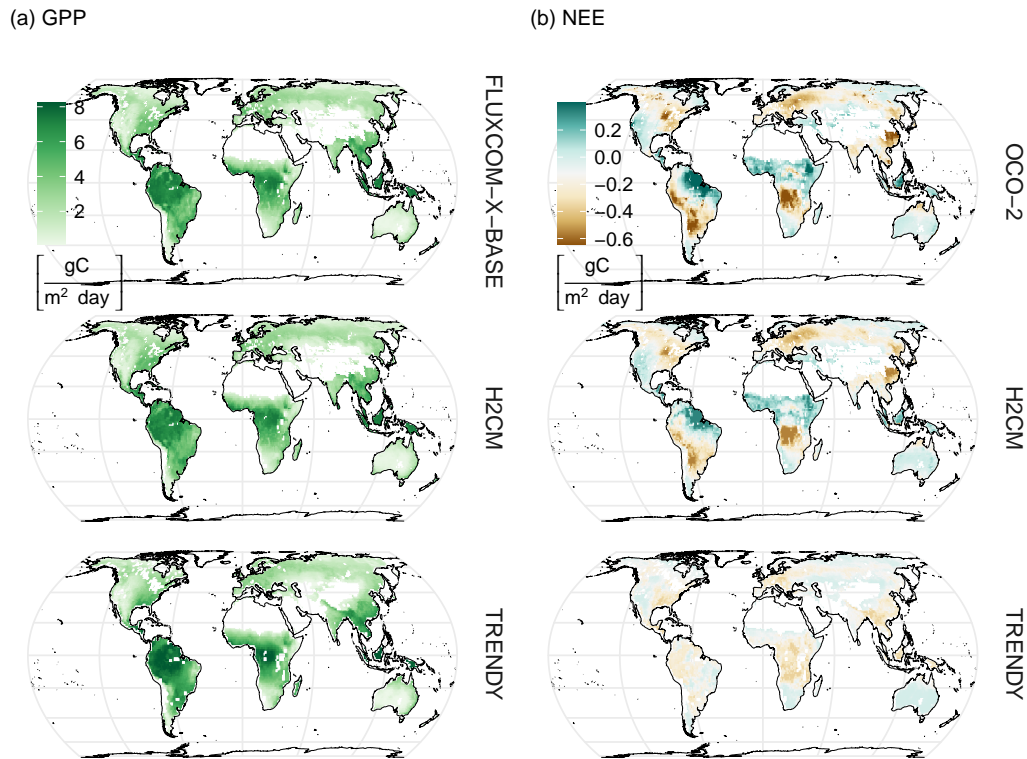
#### 3.1.3 Gross and net carbon fluxes

In this section, we evaluate H2CM's performance in terms of learned global mean spatial patterns of GPP and NEE. We compare H2CM's estimations to the data constraints provided by FLUXCOM-X-BASE and OCO-2 satellite inversions, as well as to the estimations from TRENDY.

FLUXCOM-X-BASE, H2CM, and TRENDY show very similar spatial patterns of mean GPP with largest values in wet tropical regions such as South America, Central Africa, and Southeast Asia (Fig. 4(a)). While mean annual GPP from H2CM and FLUXCOM-X-BASE are nearly identical, TRENDY tends to show somewhat larger GPP in central South America, central Africa, and Southeast Asia, while lower GPP across much of the Northern Hemisphere (Fig. B10(a)). H2CM reproduces the spatial patterns of mean NEE from the OCO-2 MIP, whereas TRENDY's estimations show little spatial gradients (Fig. 4(b) and (Fig. B10(b))). Specifically, both H2CM and OCO-2 inversions suggest the largest carbon sources in northeastern South America, tropical Asia, and in the Savannah belt south of the Sahara in Africa. In contrast, the eastern parts of North America and Europe, the southwest of South America, areas south of the Equator in Africa, and the eastern parts of China appear as carbon sinks in the period 2015-2019.

#### 3.1.4 Global spatial patterns of functional diagnostics

In this section, we qualitatively evaluate emerging global patterns in H2CM, focusing on key indicators such as precipitation, water, carbon, and light-use efficiency. These ratios, calculated based on mean annual values, help to diagnose and understand water-carbon cycle dynamics and coupling. Thus, this serves as credibility checks and a qualitative functional evaluation against patterns from existing studies in the literature.



**Figure 4.** Comparison of predicted versus target global GPP and NEE against TRENDY estimates. Maps show the median predictions across members: H2CM (10-fold cross-validation, with members corresponding to folds), OCO-2 (inversion ensemble members), and TRENDY (individual process-based models). Column (a) shows GPP, while column (b) shows NEE. Mean annual averages are calculated over 2001–2019 for GPP and 2015–2019 for NEE.

H2CM estimates largest PUE values for high latitudes, particularly in regions such as western Canada, parts of Europe, and the Eurasian boreal regions (Fig. 5(a)). These findings strongly align with the results reported by Chen et al. (2020) that estimated rain use efficiency based on a light use efficiency model forced by satellite data, both in terms of spatial patterns and the magnitude of the estimations. Additionally, another study by Liu et al. (2024b) employed a regression model based on meteorology and remote sensing, to estimate rain use efficiency, specifically focusing on Australia. Although they used GPP to estimate rain use efficiency instead of NPP as we did, their findings for Australia are also very consistent with our model's estimates for the region such as reduced PUE in Australia's interior. Overall, H2CM estimates low PUE in very wet and very dry regions, indicating that PUE peaks at intermediate moisture conditions. In dry areas, the low PUE pattern can be attributed to lower vegetation growth rates (Paruelo et al., 1999), enhanced losses by evaporation (lower T/ET, Nelson et al. (2024)), or water losses through runoff after rare but intense rainfall. In wet regions, a large fraction of precipitation is lost to runoff, and other factors such as light and nutrients may limit photosynthesis, which also can also cause a decline of PUE (Paruelo et al.,

1999; Huxman et al., 2004). We expect that PUE is very well constrained in H2CM because precipitation is a model input, and NPP is well constrained by data constraints on GPP and CUE.

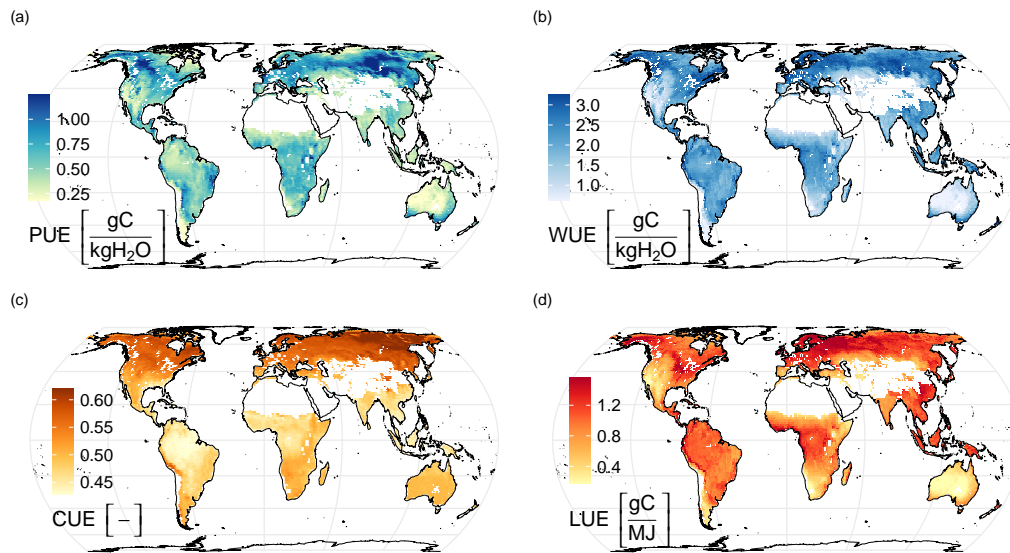
H2CM estimates high WUE in regions such as northern North America (specifically Alaska and Canada), Northern Europe, the Eurasian boreal regions, certain areas in South America, Central Africa, the southernmost parts of Australia, and New Zealand (Fig. 5(b)). In contrast, H2CM predicts lower WUE for most arid regions. Low WUE estimations for arid regions is consistent with theory and observations predicting declining WUE with increasing VPD (Boese et al., 2019; Li et al., 2023; Medlyn et al., 2011, Medlyn et al. (2017); Katul et al., 2009). The attenuated WUE in tropical forests might be related to larger costs of transporting water through tall trees (Prentice et al., 2013). Our study defines WUE with respect to transpiration, while empirical studies often define it as a function of total evapotranspiration. This can lead to significant differences due to spatial variations of transpiration versus evapotranspiration (Lawrence et al., 2007; Wei et al., 2017b). However, Ito and Inatomi (2012) provides global WUE estimations as a function of GPP and transpiration based on process-based model simulations, which align well with H2CM's WUE estimations in terms of both spatial patterns and magnitudes. H2CM's WUE is constrained by GPP and ET data, while mainly fAPAR (also constrained by observations) and the respective process formulations constrain the partitioning of ET into transpiration and evaporation components.

H2CM generally estimates higher CUE values in higher latitudes and lower CUE values in most of the Southern Hemisphere (Fig. 5(c)). The range of these estimations is relatively narrow, approximately between 0.45 and 0.6, indicating that roughly 45-60% of GPP is converted into net biomass. The patterns reflect the general expectation that the fraction of autotrophic respiration increases with temperature and biomass. This is expected because we have constrained the spatial pattern of mean annual CUE to align with predictions from the TRENDY ensemble of process models. Other studies have consistently reported similar findings in terms of magnitude and spatial patterns (He et al., 2018; Konings et al., 2019; Liu et al., 2019; Tao et al., 2023; Wang et al., 2022; Zhang et al., 2014).

H2CM estimates high effective LUE for regions such as the northwest parts of North America, the eastern states of Canada and the United States, Europe, the Eurasian boreal regions, South America, Central Africa, and Southeast Asia (Fig. 5(d)). Low LUE for dry regions are expected due to increased water limitations. A study by Liu et al. (2024a), using FLUXNET site data and satellite-derived proxies, found results closely matching ours, particularly in spatial patterns. Similarly, Wei et al. (2017a) used random forest to estimate LUE based on flux tower data, aligning mostly with our estimations, though they report lower LUE values for the northwest part of North America and the Eurasian boreal regions.

### 3.2 Estimated seasonality of carbon fluxes across major regions

In this section, we assess the estimated seasonality of NEE, GPP, and TER as predicted by H2CM. Conceptually, H2CM is intended to be a bridge between purely data-driven approaches and purely process-based models. Therefore, we compare our estimations with both a pure data-driven model (FLUXCOM-X-BASE, Nelson et al. (2024)) and state-of-the-art process-based models (TRENDY, Sitch et al. (2024)). Our evaluation focuses on global and four major TransCom regions covering different climates: Eurasian Boreal, North American Temperate, South American Tropical, and Southern Africa (Fig. E1).



**Figure 5.** Emerging global patterns in H2CM: a) Precipitation Use Efficiency ( $PUE = NPP / Prec$ ), b) Water Use Efficiency ( $WUE = GPP / T$ ), c) Carbon Use Efficiency ( $CUE = NPP / GPP$ ), and d) Light Use Efficiency ( $LUE = GPP / APAR$ ). The maps represent the median across 10 cross-validation folds. The mean annual average is calculated over the years 2001-2019.

430 Overall, H2CM outperforms both FLUXCOM-X-BASE (in terms of NEE) and TRENDY (in terms of NEE and GPP) globally and across all regions, as measured by RMSE and  $R^2$  (Fig. 6). H2CM, FLUXCOM-X-BASE, and TRENDY median all accurately capture most of the seasonal variation in global NEE and for the Eurasian Boreal region. However, the range among different TRENDY models is very large, and the ensemble median underestimates the peak of net carbon uptake in summer, apparently due to an underestimation of GPP (Fig. 6).

435 For the North American Temperate region, both H2CM and FLUXCOM-X-BASE effectively capture most of the seasonal variation in NEE, with  $R^2$  values of 1 and 0.98, respectively. In terms of TER, there is good agreement across all approaches (Fig. 6). Similar as for the Eurasian boreal region, TRENDY median underestimates the net carbon uptake peak in summer peak of NEE, and exhibits a slight phase shift of NEE seasonality. This issue seems attributable to TRENDY's GPP estimations, where similar problems are observed (underestimation of the peak and slight shift).

440 In the South American Tropical region, only H2CM captures 69% of the subtle seasonal variation in NEE, which shows a seasonal cycle characterized by a double peak of net carbon release. FLUXCOM-X-BASE captures the first carbon release peak qualitatively but not the second, while TRENDY's seasonality seems unrelated to the OCO-2 inversion estimates. Interestingly, the seasonality of GPP is similar between H2CM, FLUXCOM, and TRENDY, while seasonality of TER deviate across modeling approaches. The discrepancy of FLUXCOM and TRENDY with respect to NEE variations of H2CM is dominated  
 445 by TER, as H2CM estimates a positive TER peak in October, which TRENDY and FLUXCOM-X-BASE do not. The peak of

net carbon uptake suggested by OCO-2 inversions does not coincide with the GPP peak, which is qualitatively reproduced by H2CM. One possible explanation for the poor NEE estimations in the South American Tropical region is the potential omission of certain processes in the models. Given the role of respiration in shaping net carbon release patterns in the wet tropics and the apparent discrepancies related to TER among models, it seems that wetness effects on respiration remain to be a modelling issue. H2CM's neural network component seems effective to compensate for this to some extent by learning responses from the observational constraints.

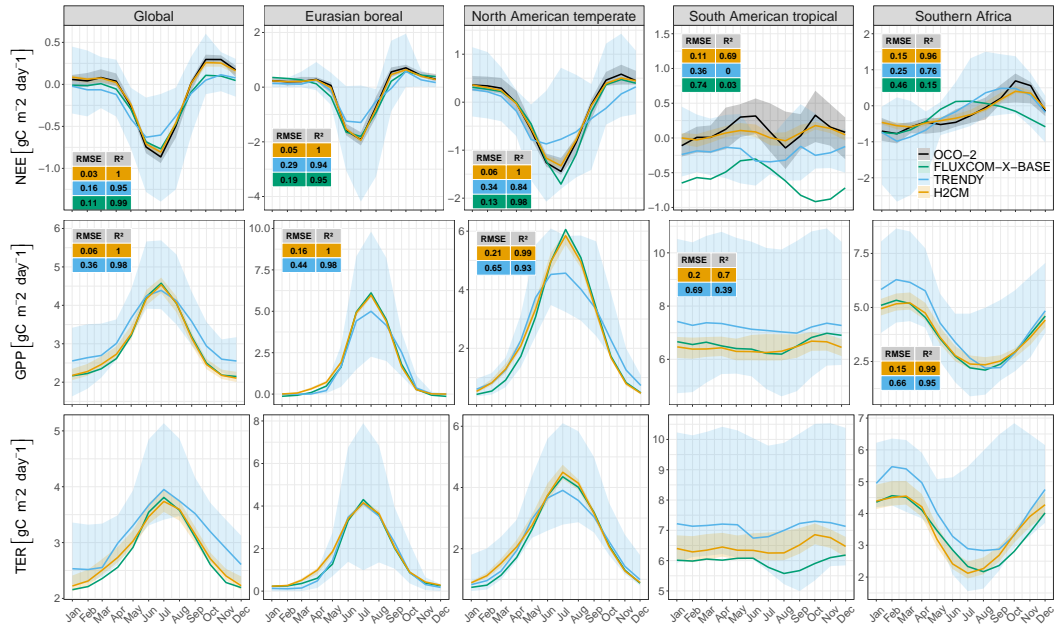
In Southern Africa, H2CM explains most of the variation in NEE (based on OCO-2 MIP) with an  $R^2$  of 0.96. TRENDY reasonably reproduces the patterns with an  $R^2$  of 0.76, while FLUXCOM-X-BASE performs poorly with an  $R^2$  of 0.16. Also in this subtropical dry region, NEE seasonality and discrepancies among models seem related to respiration as both H2CM and TRENDY capture most of the variation in GPP in terms of  $R^2$  (Fig. 6). H2CM accurately reproduces the prominent peak of net carbon release at the end of the year, whereas TRENDY captures it poorly, showing a slight shift in the peak estimation, and FLUXCOM-X-BASE completely misses it (Fig. 6). Several studies have highlighted the challenges of correctly capturing NEE seasonality in dry regions (Jung et al., 2020; Lee et al., 2025; Metz et al., 2025; Nelson et al., 2024; Bodesheim et al., 2018). The carbon release peak in Southern Africa was attributed to the rapid stimulation of microbial respiration due to re-wetting when transitioning from the dry to the rainy season (Metz et al., 2023, 2025). Also for dry subtropical regions FLUXCOM-X-BASE and TRENDY models struggle to accurately represent the differential responses of photosynthesis and respiration processes to water effects. These carbon-water interactions lead to complex seasonal variations of NEE, which are characterized by carbon release peaks in contrast to boreal and temperate regions dominated by a GPP driven carbon uptake peak. In the next section, we further diagnose the origin of the carbon release peak in Southern Africa by H2CM.

### 3.3 Net carbon release peak in Southern Africa

In this section, we focus on the drylands of Southern Africa—a region with large contribution to global NEE interannual variability but where FLUXCOM-X-BASE and TRENDY struggle to reproduce even seasonal dynamics. We analyze how H2CM represents the patterns and processes governing NEE seasonality in this region and show that our hybrid modeling approach provides insight into these complex dynamics.

To infer which regions contribute most to the end-of-year net carbon release peak in Southern Africa, we plotted the NEE difference between December and August (Fig. 7a). Interestingly, the largest amplitude occurs in the sub-humid region covering the Miombo woodlands, and not in the more-dry-drier southern part of Africa. Similarly, Metz et al. (2025) inferred that the more northern regions of Southern Africa contribute most to the peak based on an atmospheric inversion using GOSAT data, while those did not permit a more detailed spatial assessment.

To better understand the simulated carbon dynamics by H2CM, we plot daily variations of NEE, NPP, and heterotrophic respiration for a grid cell in the center of this region (Fig. 7a). The NEE variations from August to December are characterized by a gradual increase until about November, driven by a concomitant rise in Rh (Fig. 7a). With the onset of the rain, both Rh and NEE exhibit sharp increases. This phenomenon aligns with the “Birch effect” (Birch, 1964; Jarvis et al., 2007), where re-wetting causes a respiration pulse. Recent studies based on atmospheric inversions using GOSAT data have found that this

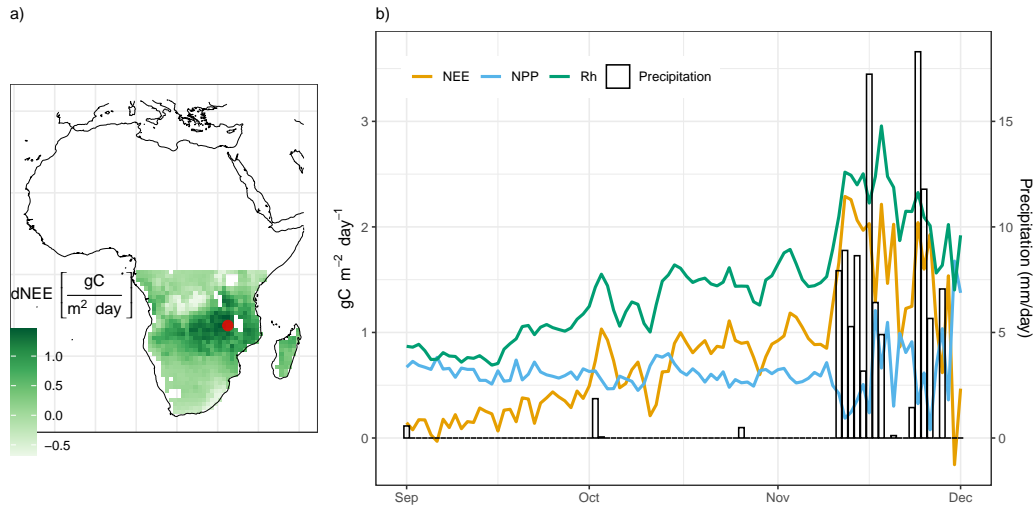


**Figure 6.** Predicted seasonality of carbon fluxes compared with data constraints, FLUXCOM-X-BASE, and TRENDY across four major Transcom regions. Shaded areas represent the range across model members (or 10 CV folds in H2CM), with lines indicating the median. Each column corresponds to a different region, and each row represents a distinct carbon flux. Subplots for NEE and GPP include tables with metrics for each region. NEE comparisons involve H2CM, FLUXCOM-X-BASE, and TRENDY against OCO-2 inversions, while GPP comparisons involve H2CM and TRENDY with FLUXCOM-X-BASE. Metrics for TER are omitted due to the absence of direct data constraints. Note varying y-axis scales for different regions.

480 effect shapes NEE variations in dry areas such as Australia (Metz et al., 2023), Southern Africa (Metz et al., 2025) and dry regions of South America (Vardag et al., 2025).

Interestingly, at this location, we also observe that NPP declines when it rains, which further amplifies the NEE response. This suggests that in H2CM, the region’s GPP gets limited by light rather than water, due to reduced radiation from cloudiness. In fact, previous studies indicate that the Miombo forests remain green throughout the dry season (Küçük et al., 2022), likely  
 485 due to water access through deep rooting (Stocker et al., 2023). Our findings based on H2CM suggest that light limitation of photosynthesis in the Miombo contributes to the end-of-season NEE peak in Southern Africa, in addition to the previously known re-wetting driven respiration response.

This example highlights H2CM’s ability to capture and explain complex carbon cycle phenomena. It suggests that H2CM learned plausible carbon flux responses to high-frequency (daily) variations in environmental conditions, even though only  
 490 low-frequency (monthly) data constraints were used during training.



**Figure 7.** a) The difference in NEE between the peak month (December) and the onset of the gradual increase (August). The black point on the map indicates the specific grid cell location from which the time-series data is derived. b) Response of  $CO_2$  fluxes to rainfall in Southern Africa. The left y-axis represents  $CO_2$  flux values, while the right y-axis corresponds to precipitation values. The time-series is based on daily flux predictions from September through December for a single year 2006 and median across 10-CV folds.

### 3.4 Discussion of the H2CM approach

#### 3.4.1 Scope and strength

H2CM is a process-aware hybrid framework that integrates machine learning with simplified process-based formulations, and can be positioned within the broader family of model–data integration strategies that aim to combine physics with machine learning. This family includes approaches such as physics-informed neural networks (Raissi et al., 2019, Tartakovsky et al. (2020), Wang et al. (2020)), physics-guided machine learning (Khandelwal et al., 2020, Pawar et al. (2021), Karpatne et al. (2017)), and differentiable modeling (Shen et al., 2023). Rather than reproducing the detailed parameterizations used in comprehensive land-surface models, H2CM employs conceptual process formulations to maintain interpretability and ensure parameter identifiability under the available observational constraints.

The primary objective of H2CM is to deliver a consistent, observation-constrained ‘reanalysis’ of coupled carbon–water states and fluxes over recent decades by fusing diverse data streams within a process-informed ML architecture. H2CM is not intended to replace land surface models coupled into Earth system models, nor to provide long-term projections; such applications would require more complete representations of processes and feedback. Instead, H2CM can be perceived as a bridge between empirical and process-based modeling which enables a coherent interpretation of only partially observed coupled carbon–water cycle variability and interactions. H2CM is a global model in the sense that it is applied consistently across grid cells covering the global land surface, while it remains local as it does not simulate lateral processes among grid

cells. Although the domain and grid are, in principle, arbitrary, the present design is tailored to comparatively coarse grid cells for improved consistency with spatially coarse observational constraints like TWS from GRACE or NEE inversions.

H2CM advances the field by combining previous approaches—such as fully data-driven models like FLUXCOM(-X-BASE) (Jung et al., 2019, Jung et al. (2020), Nelson et al. (2024)) and fully process-based models like TRENDY (Sitch et al., 2015, Sitch et al. (2024))—by aiming at combining their strengths: it learns directly from global observations through its machine learning component, while respecting conceptual process understanding and mass balance. Relative to process-based TRENDY models, which rely on numerous parameterizations yet struggle with regional and climate-specific patterns in observation-based carbon fluxes, H2CM better captures key spatial–temporal features in observations (Fig. 6). Likewise, compared with fully data-driven FLUXCOM approaches that depend primarily on site-level eddy covariance data and can underperform in tropical and subtropical regions with sparse coverage, H2CM leverages complementary data streams and process-aware structure to improve generalization (Fig. 6).

Both, the rich information provided by complementary Earth Observations to H2CM, and the flexibility of machine learning to exploit these, contribute to the good performances of the model. Clearly, a key factor for the improved performance of H2CM compared to FLUXCOM in capturing NEE variations inferred from atmospheric inversions is the integration of this data constraint in the hybrid modeling approach. While this might also explain H2CM’s improved performance compared to TRENDY, the machine learning component used to parameterize the conceptual process equations seems to contribute to the success significantly. A useful point of reference is the model-data-integration approach by Lee et al. (2025), which shares a similar model structure of conceptual process equations of carbon and water cycles, and nearly identical observational constraints. The key difference is the representation of parameters in process equations: the study by Lee et al. (2025) assumes globally constant parameters, whereas H2CM uses neural networks to predict parameters that vary spatially and/or temporally. Overall, both approaches, being strongly data-constrained, capture large-scale patterns well. However, H2CM achieves higher performance, in particular in dry and wet tropical regions with partly complex linkages of carbon and water cycles. This illustrates the added value of the adaptive machine learning module in H2CM for better model performance. Taken together, H2CM’s skill emerges from the synergy of conceptual process formulations, diverse observational constraints, and adaptive parameterization through machine learning.

### 3.4.2 Uncertainties and limitations

H2CM heavily depends on data-driven learning, incorporating both meteorological forcings and data constraints. The primary constraints on the carbon cycle include GPP seasonality from FLUXCOM-X-BASE, NEE from OCO-2, and global NEE monthly anomalies from CarboScope. Each of these data sources is subject to uncertainties, which could propagate to H2CM. Accounting for data uncertainties explicitly in H2CM, e.g. based on a ~~bayesian~~ Bayesian approach, would be desirable conceptually. However, the lack of adequate and comparable quantitative information on data uncertainties across different data streams limits this approach in practice.

We tried to mitigate known issues of the observational constraints by making the loss function sensitive to the more robust patterns in the data, e.g. by using only the mean seasonal cycles of GPP by FLUXCOM-X-BASE due to known limitations for

interannual variability (Jung et al., 2020; Nelson et al., 2024). However, uncertainty of the GPP mean seasonal cycle of X-BASE remain for example in the region of tropical South America. Thus, potential biases in the seasonal cycle of FLUXCOM-X-BASE GPP likely propagate to H2CM, which could then also compromise simulations of respiration due to the concomitant data constraint on NEE.

545 To constrain H2CM's NEE estimations, we utilize an atmospheric inversion ensemble using in-situ and OCO-2 satellite-based column  $XCO_2$  data. Despite significant efforts and advancements in  $XCO_2$  retrievals from OCO-2, the preferential sampling of clear-sky conditions and remaining retrieval biases may affect inferred NEE variations (Connor et al., 2016; Kulawik et al., 2019a, b; Worden et al., 2017). Additional uncertainties from various methodological choices, particularly in atmospheric transport modeling, result in poorly constrained spatial inversion results (Yun et al., 2025; Zhang et al., 2022; 550 Byrne et al., 2022). We used  $1^\circ$  OCO-2 MIP results to constrain NEE, which, despite the added value of extensive  $XCO_2$  data, may be overly optimistic. Therefore, some discrepancies between H2CM-simulated and OCO-2 MIP-based NEE at  $1^\circ$  resolution (Fig. B10(b)) are due to this data constraint uncertainty. Since the issue of loose spatial constraints is even more pronounced for atmospheric inversions using only the sparse in-situ  $CO_2$  network, we used only the spatial batch average as a data constraint from CarboScope.

555 Exploiting H2CM's explicit process interpretability for process understanding requires sufficiently strong observational constraints to overcome equifinality—i.e., different processes or pathways yielding similar outcomes when data are not informative enough. Examining the spread among 10 CV models trained on different cross-validation folds and with different random initializations indicates good qualitative robustness under varying training data (Baghirov et al., 2025). Nevertheless, the globally constant temperature sensitivity of respiration ( $Q_{10}$ ) and the scalar for the relative  $CO_2$  fertilization effect on GPP 560 ( $\beta_{CO_2}$ ) show signs of non-identifiability. Across CV folds, the median  $Q_{10}$  (unitless) is 1.25, with a range of 1.24–1.27, lower than literature values for heterotrophic/soil respiration (1.4–2; Zhou et al. (2009); Meyer et al. (2018); Niu et al. (2021)). For  $\beta_{CO_2}$  the median value learned across CV folds is 35.82 (% 100 ppm<sup>-1</sup>). The range of values extends from 26.88 (% 100 ppm<sup>-1</sup>) to 62.71 (% 100 ppm<sup>-1</sup>), which exceeds observational estimates of roughly 14–16 % 100 ppm<sup>-1</sup> (AINSWORTH and ROGERS, 2007; Wang et al., 2020; Ueyama et al., 2020; Zhan et al., 2024). To probe this, we imposed informative priors 565 on these global constants (Appendix ED). Under these priors, H2CM recovered  $\beta_{CO_2} = 15$  (% 100 ppm<sup>-1</sup>) and  $Q_{10} = 1.5$  — matching the priors exactly—while other model results and performance remained qualitatively unchanged. These findings indicate that with the current data constraints,  $Q_{10}$  and  $\beta_{CO_2}$  are not fully identifiable, underscoring the value of stronger observational constraints and the incorporation of process knowledge (e.g., via priors or loss-function regularization) for decadal-scale assessments with H2CM.

570 Deviations of H2CM simulations from the data constraints such as for NEE in tropical South America (Fig. 6) are useful indications for missing information or issues in the data. This is because H2CM cannot fit the data constraints if (1) the relevant signals are not present in the model forcing data; or (2) the model's process formulations do not permit it, e.g. due to a missing process representation that cannot be captured by the flexibility of the machine learning component; or (3) because there is a conflict among different observational constraints that cannot be satisfied simultaneously given the process formulations.

575 While incorporating process understanding in H2CM also provides theoretical constraints that can help to regularize the NNs (Reichstein et al., 2019; Shen et al., 2023), a trade-off between model complexity and parameter identifiability is obvious. Thus the simple nature of process representations in H2CM is justified by the limited information content of the used observations to disentangle subprocesses, and by minimum requirements on interpretability. For example, we use net radiation as a single forcing term rather than separately prescribing shortwave and longwave components for simplicity as shortwave and net radiation are strongly correlated at daily time scale (Jung et al., 2024). Currently, H2CM lacks several important components of the land system, which are relevant for the coupled water and carbon cycles. For example, it does not yet explicitly model vegetation and soil carbon pools, as the current focus has been on spatial variations of carbon fluxes from sub-seasonal to interannual time scales. Furthermore, H2CM does not incorporate the effects of disturbances such as fire, or other drivers including land-use change. Note that dynamic vegetation changes are not explicitly represented in H2CM; however, temporal variations in fAPAR provide an implicit representation of phenological changes. There are emerging opportunities from the observational side which could support the integration of these components in H2CM in the future.

#### 4 Conclusions

We introduced H2CM - a hybrid global carbon-water cycle model constrained by diverse observational data streams, which acts as a bridge between empirical and process-based modeling and enables a coherent interpretation of only partially observed coupled carbon-water cycle variability and interactions. H2CM integrates key hydrological data streams such as terrestrial water storage variations from GRACE, but also combines flux tower based bottom-up and atmospheric top-down data constraints of the carbon cycle. We demonstrated that H2CM successfully integrates the strengths of machine learning and process-based modelling, learning directly from observations while obeying to process understanding. H2CM reproduces major observed patterns of carbon flux variations and spatial patterns of functional carbon-water coupling properties. Our parallel evaluation of H2CM with the purely data-driven FLUXCOM and process-based TRENDY model ensemble further highlights the added value of our hybrid approach in modelling uncertain moisture effects on respiration, which shape NEE variations in dry and wet tropical regions significantly.

H2CM opens new avenues for studying the global carbon-water cycle and lays the groundwork for further developments toward a hybrid data-driven land-surface model. Advancing in this direction will require integrating the surface energy cycle, incorporating dynamic vegetation and explicit carbon pools, representing additional key processes and disturbances (e.g., permafrost, fire, and land use/management), and resolving sub-daily dynamics. These opportunities are forward-looking, while the grand challenge will be to achieve a robustly constrained model with causal functioning required to simulate conditions with no analogs in the contemporary observational era.

. The model simulations are accessible via <https://doi.org/10.5281/zenodo.18434085> (Baghirov, 2026b). The model code can be accessed via <https://doi.org/10.5281/zenodo.18400385> (Baghirov, 2026a). The training data can be accessed via <https://doi.org/10.5281/zenodo.16575309>

(Baghirov, 2025b). For the most current version of the code, please visit the public repository at <https://github.com/zavud/h2cm> (last access: 31 January 2026). We are open to sharing additional variables (that are not shared here) upon request.

## Appendix A: Detailed overview of the hybrid architecture in H2CM

Figure 2 illustrates the overall architecture of H2CM:-

610 1. **Static inputs and compression.** Panel A1 represents the static environmental inputs, including land cover, soil properties, wetland extent, and digital elevation. These features are passed through a fully connected neural network (FC-NN 1; A2) that compresses spatially varying but temporally invariant information into a latent vector. FC-NN 1 has two hidden layers with 150 and 12 units, respectively. The 12-unit latent vector is used as compressed static features and serves as a shared input to all dynamic sequence layers. The outputs of this static network (A3) include spatially varying parameters such as maximum soil moisture capacity ( $SM_{max}$ ) and  $\alpha_{E_i}$  controlling interception evaporation.

(B–E) Dynamic sequence models. H2CM includes three long short-term memory (LSTM) networks and an additional fully connected neural network (FC-NN 2) to model time-varying processes. Each LSTM contains one hidden layer with 100 units and is connected to a small fully connected layer that transforms hidden states into physically interpretable parameters. These dynamic modules produce spatio-temporal predictions:-

- 620
- LSTM 1 (B1–B3): Receives the compressed static features together with dynamic drivers such as net radiation, precipitation, relative soil moisture, snow, groundwater, and fAPAR at time  $t-1$ . It predicts coefficients ( $\alpha_{r_{soil}}$ ,  $\alpha_{rgw}$ ,  $\alpha_{smelt}$ ) that control soil recharge, groundwater recharge, and snowmelt processes.
  - LSTM 2 (C1–C3): Takes inputs including the static compressed vector, air temperature, vapor pressure deficit,  $CO_2$  concentration, relative soil moisture, and fAPAR and NPP at time  $t-1$ . It estimates carbon use efficiency and fAPAR.

625

  - LSTM 3 (D1–D3): Uses static compressed features, net radiation, precipitation, fAPAR, and NPP at time  $t-1$  to predict the basal respiration rate and  $\alpha_{E_s}$  parameter that controls soil evaporation.
  - FC-NN 2 (E1–E3): A fully connected neural network with two hidden layers (150 and 12 units) that predicts water use efficiency and the  $\alpha_T$  parameter, which represents effective conductance or stress response.

Global constants (learned). A set of globally learned parameters ( $Q_{10}$ ,  $\beta_{CO_2}$ ,  $\beta_{baseflow}$ ,  $\beta_{snow}$ ) provide scaling relationships for temperature sensitivity of respiration,  $CO_2$  fertilization, baseflow recession, and correction of snowfall. Coupled water-carbon cycle model. The outputs from the static and dynamic subnetworks are passed to a differentiable, process-based water-carbon cycle model that enforces mass balance between fluxes. This model represents the physical coupling between hydrological and carbon processes, ensuring consistency between water storage, evapotranspiration, carbon assimilation, and respiration. Constrained spatio-temporal predictions. During training, the entire hybrid architecture—including the static and dynamic neural subnetworks and the differentiable process water-carbon cycle model—is optimized end-to-end. The neural network

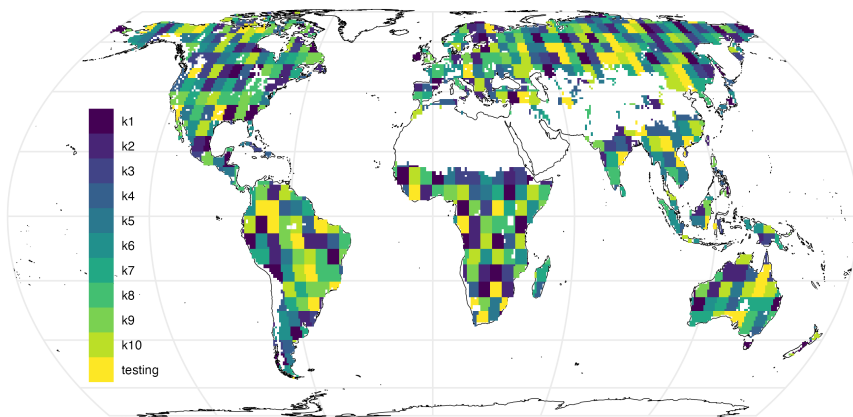
630

635

components generate spatially and spatio-temporally varying parameters that are passed into the process model, which produces simulated water fluxes and storages and carbon fluxes. Some of these outputs (net ecosystem exchange, gross primary productivity, fAPAR, terrestrial water storage, snow water equivalent, evapotranspiration, and runoff) are compared against observational targets through a composite loss function. Because the process-based component is fully differentiable, gradients of the loss propagate through the process equations back to the neural subnetworks via automatic differentiation. This enables the networks to learn physically consistent parameterizations that minimize discrepancies between modeled and observed dynamics.

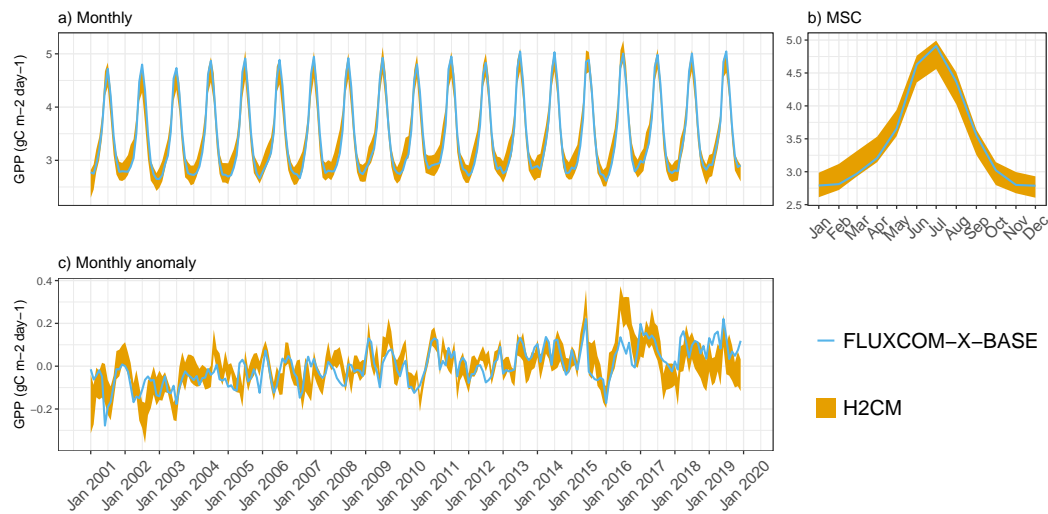
Detailed overview of the H2CM framework. Static inputs (e.g., land cover, soil, elevation) are processed through a fully connected neural network (two hidden layers: 150 and 12 units) to generate compressed static features. These features feed into three LSTM networks (each with one hidden layer of 100 units) and an additional fully connected network that predict spatio-temporal parameters regulating hydrological and carbon processes, including  $\alpha_{rsoil}$ ,  $\alpha_{rgw}$ ,  $\alpha_{smelt}$ ,  $\alpha_{Es}$ ,  $\alpha_{Ei}$ , and  $\alpha_T$ . Globally learned constants ( $Q_{10}$ ,  $\beta_{CO2}$ ,  $\beta_{baseflow}$ ,  $\beta_{snow}$ ) control temperature sensitivity,  $CO_2$  fertilization, baseflow recession, and snow corrections. Outputs from all subnetworks are coupled through a differentiable, mass-balanced water-carbon process model. The process layer produces water fluxes and storages and carbon fluxes, some of which are directly constrained by observations, including NEE, GPP, fAPAR, TWS, SWE, ET, and runoff. reated in BioRender (Baghirov, 2026d).

## 650 Appendix A: Cross validation

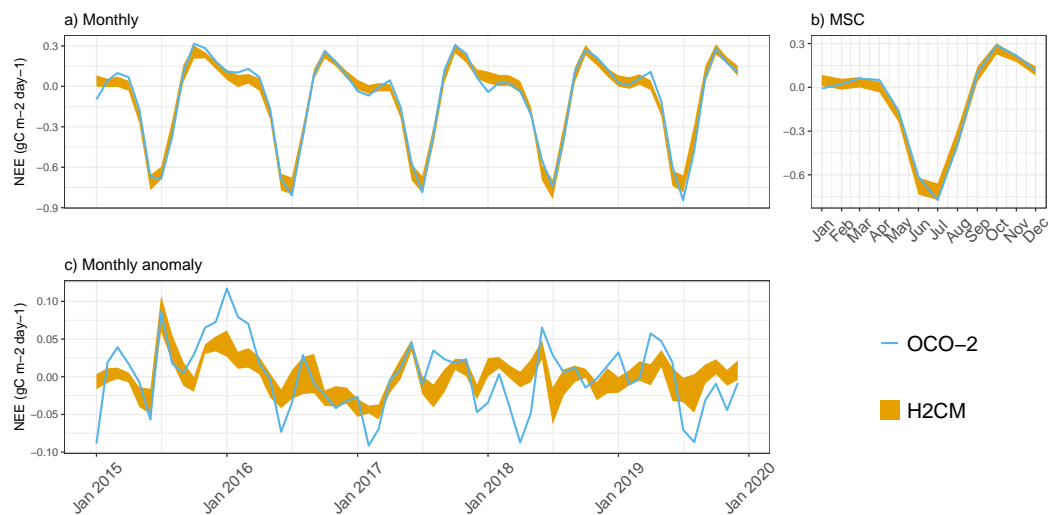


**Figure A1.** Validation sets for 10 different models and a fixed testing set. Note that, during training, each fold has a separate and unique validation set and all models were tested on the same testing set.

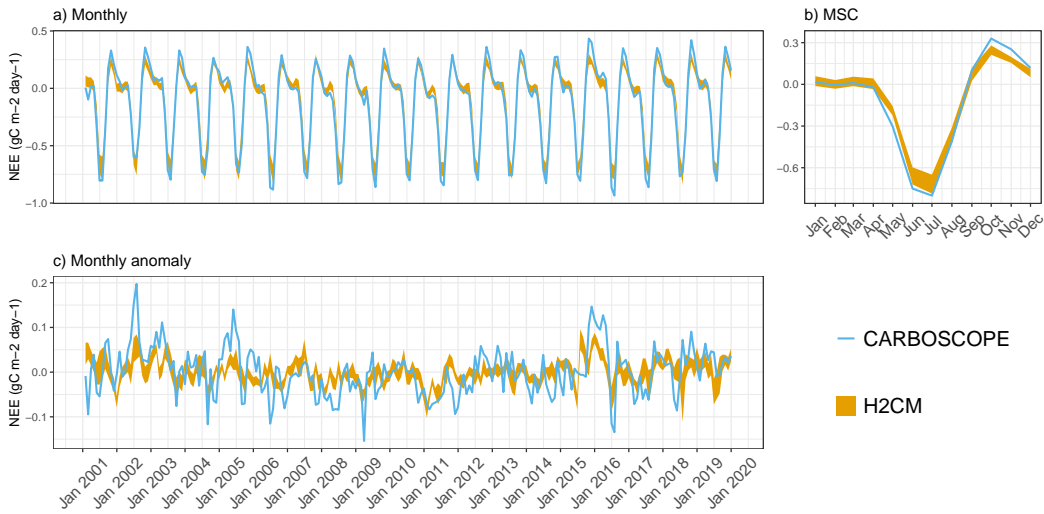
## Appendix B: Model evaluation



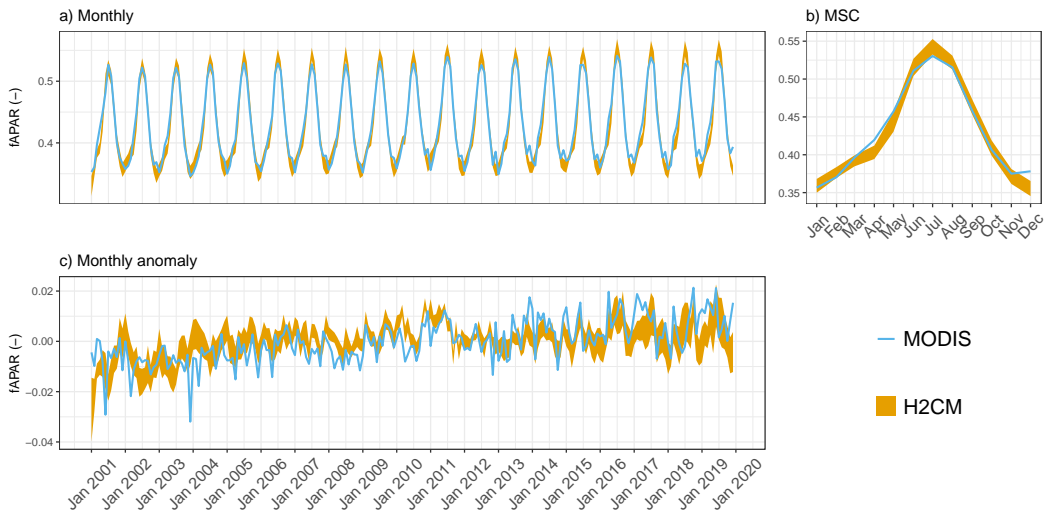
**Figure B1.** Predicted versus target GPP (FLUXCOM-X-BASE) over the testing set (spatial domain) across 10 CV folds: a) monthly, b) mean seasonal cycle over the entire period of data constraint and c) monthly anomaly



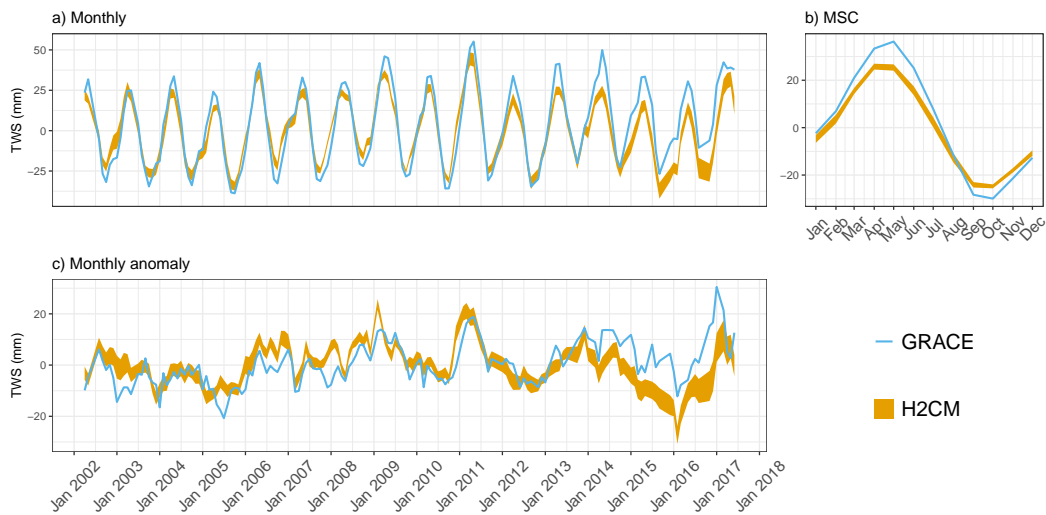
**Figure B2.** Predicted versus target NEE (OCO-2) over the testing set (spatial domain) across 10 CV folds: a) monthly, b) mean seasonal cycle over the entire period of data constraint and c) monthly anomaly



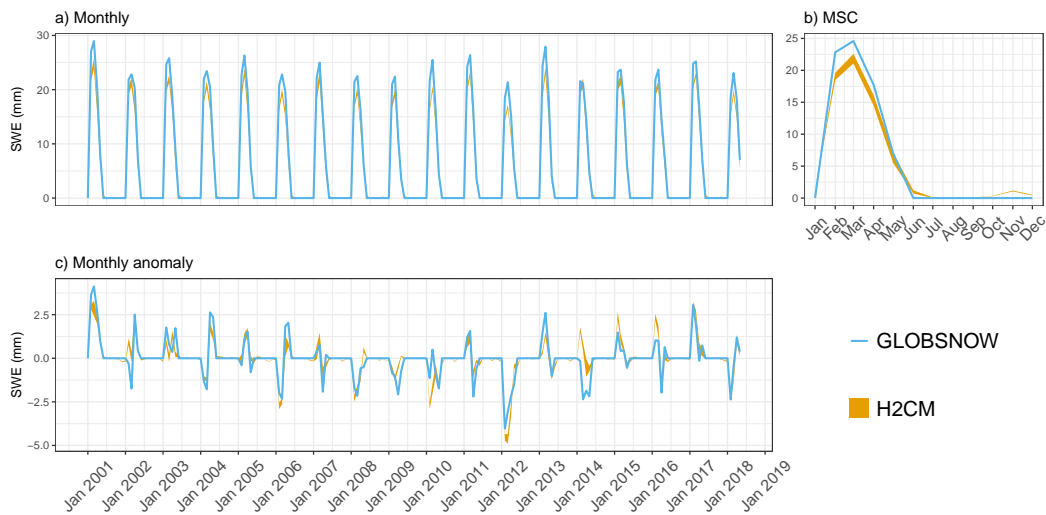
**Figure B3.** Predicted versus target NEE (CarboScope) over the testing set (spatial domain) across 10 CV folds: a) monthly, b) mean seasonal cycle over the entire period of data constraint and c) monthly anomaly



**Figure B4.** Predicted versus observed mean fAPAR over the testing set (spatial domain) across 10 CV folds: a) monthly, b) mean seasonal cycle over the entire period of data constraint and c) monthly anomaly.



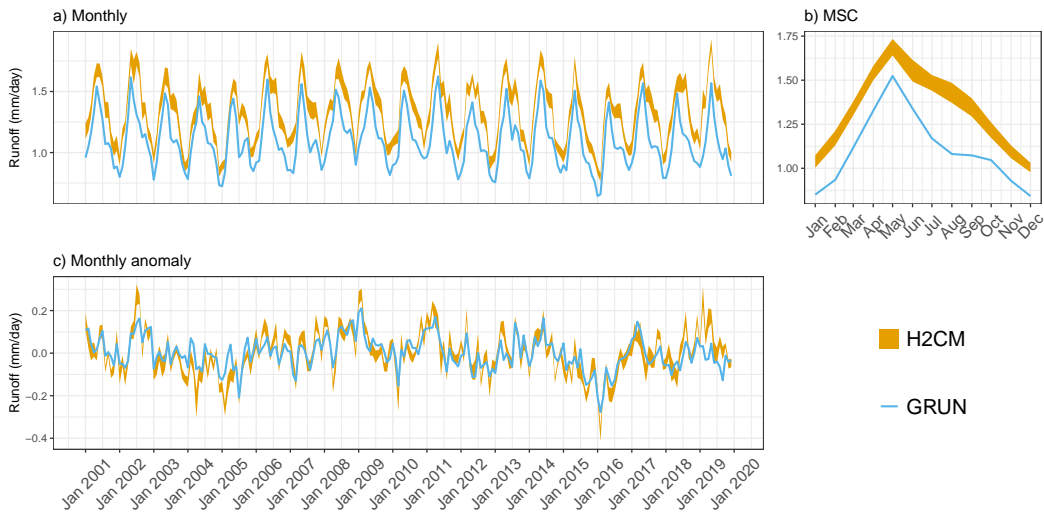
**Figure B5.** Predicted versus observed mean terrestrial water storage (TWS) (anomaly) over the testing set (spatial domain) across 10 CV folds: a) monthly, b) mean seasonal cycle over the entire period of data constraint and c) monthly anomaly.



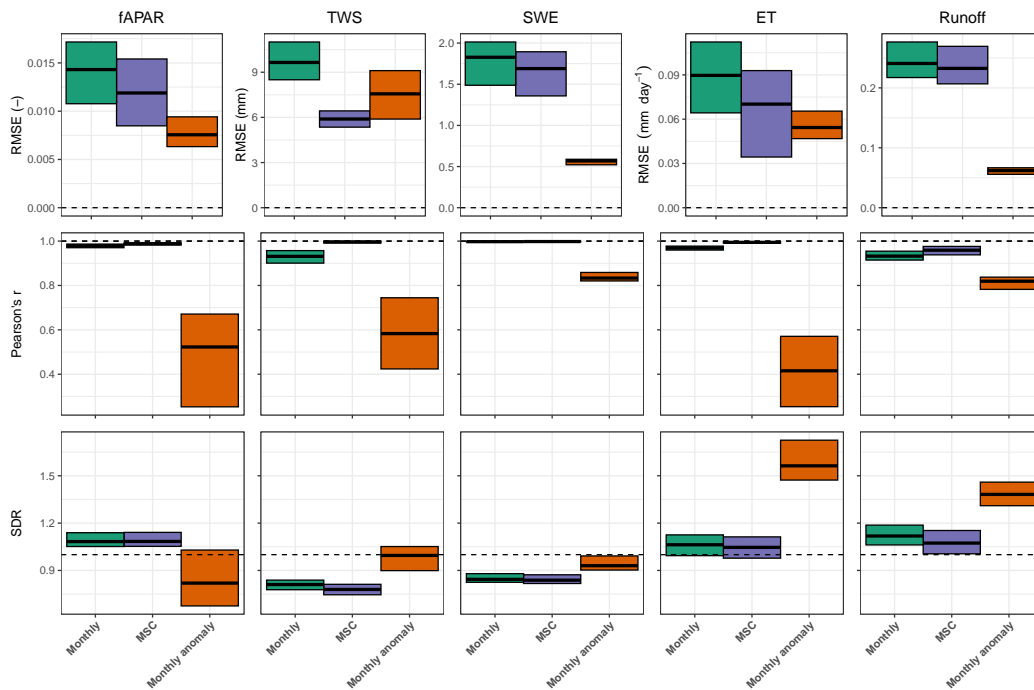
**Figure B6.** Predicted versus observed mean snow water equivalent (SWE) over the testing set (spatial domain) across 10 CV folds: a) monthly, b) mean seasonal cycle over the entire period of data constraint and c) monthly anomaly



**Figure B7.** Predicted versus target mean ET over the testing set (spatial domain) across 10 CV folds: a) monthly, b) mean seasonal cycle over the entire period of data constraint and c) monthly anomaly



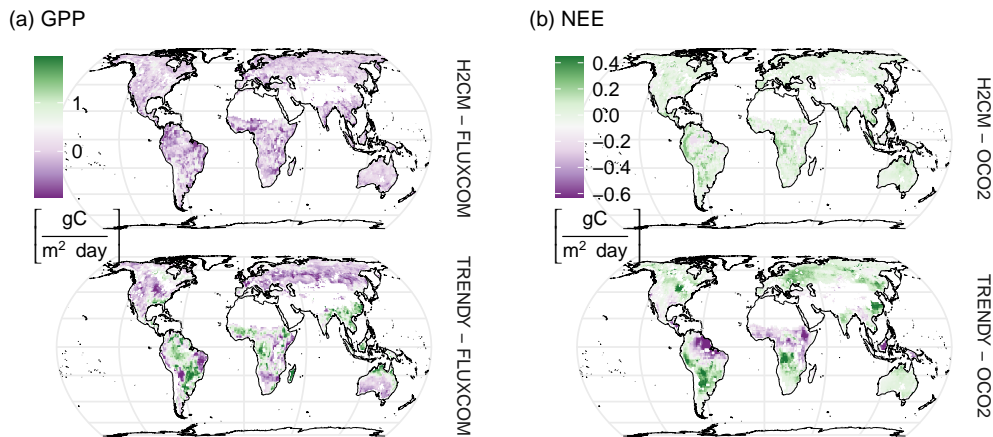
**Figure B8.** Predicted versus target mean runoff over the testing set (spatial domain) across 10 CV folds: a) monthly, b) mean seasonal cycle over the entire period of data constraint and c) monthly anomaly.



**Figure B9.** Model performance for vegetation, water storages and fluxes based on the spatial average over testing-set grid cells. Cross bars represent the minimum and maximum errors across the 10 cross-validation folds, and the central lines indicate the mean error. Rows correspond to performance metrics, and columns represent model constraints. RMSE denotes root mean squared error, and SDR is the standard deviation ratio (the ratio between predicted and observed standard deviation). [Horizontal dashed lines denote ideal metric values, with  \$R^2\$  and SDR equal to 1 and RMSE equal to 0.](#)

**Table B1.** Benchmarking metrics (RMSE and Bias) for modeled carbon fluxes against data constraints. Metrics are reported for gross primary productivity (GPP) and for net ecosystem exchange (NEE) benchmarked against OCO-2 and CarboScope inversion products. Each section (GPP, NEE (OCO-2), NEE (CarboScope)) summarizes results for Monthly, mean seasonal cycle (MSC), and Anomaly data. All metrics are based on globally averaged, area-weighted time series. Reported values represent the median across 10-fold cross-validation runs, with values in brackets indicating the minimum and maximum across folds.

Metric	Monthly	MSC	Anomaly
<b>GPP</b>			
RMSE [gC m <sup>-2</sup> day <sup>-1</sup> ]	0.1, [0.06, 0.22]	0.09, [0.05, 0.22]	0.05, [0.04, 0.06]
Bias [gC m <sup>-2</sup> day <sup>-1</sup> ]	0, [-0.2, 0.11]	0, [-0.2, 0.11]	0, [0, 0]
<b>NEE (OCO-2)</b>			
RMSE [gC m <sup>-2</sup> day <sup>-1</sup> ]	0.04, [0.03, 0.05]	0.03, [0.02, 0.04]	0.02, [0.02, 0.03]
Bias [gC m <sup>-2</sup> day <sup>-1</sup> ]	-0.01, [-0.03, 0.03]	-0.01, [-0.03, 0.03]	0, [0, 0]
<b>NEE (CarboScope)</b>			
RMSE [gC m <sup>-2</sup> day <sup>-1</sup> ]	0.07, [0.06, 0.08]	0.06, [0.04, 0.08]	0.03, [0.03, 0.04]
Bias [gC m <sup>-2</sup> day <sup>-1</sup> ]	0.01, [0, 0.04]	0.01, [0, 0.04]	0, [0, 0]

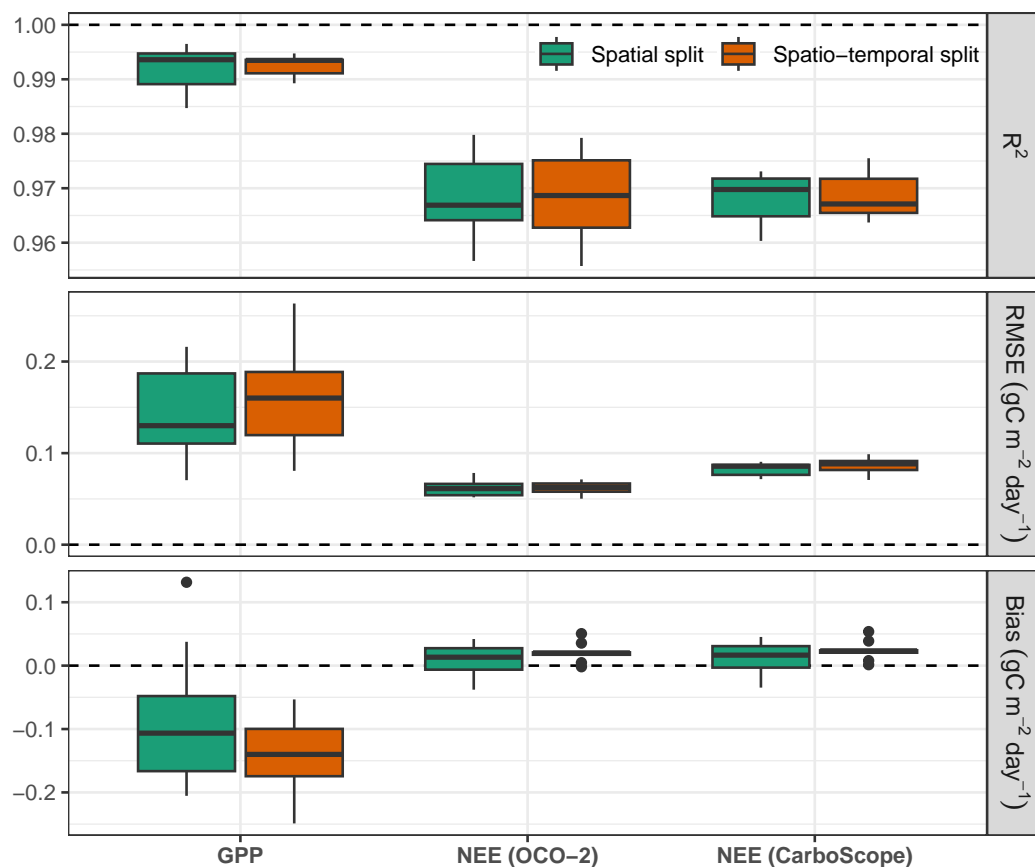


**Figure B10.** Spatial differences between modelled and target carbon fluxes for (a) gross primary production (GPP) and (b) net ecosystem exchange (NEE). The figure compares the median estimations from H2CM (10-fold cross-validation), OCO-2 (inversion ensemble), and TRENDY (process-based model ensemble). For GPP, H2CM and TRENDY are evaluated against FLUXCOM, averaged over 2001–2019; for NEE, both are evaluated against OCO-2 MIP, averaged over 2015–2019.

## Appendix C: Model generalization across space and time

Figure C1 shows the performance comparison between spatially split and spatio-temporally split cross-validation folds using post-2017 time-series data. In the spatial split experiment, the model was trained on the complete 2001–2019 time series while holding out specific spatial grid cells for testing. In contrast, the spatio-temporal split experiment was trained on data from 2001–2017, with all data after 2017 withheld for spatio-temporal testing.

Overall, the results suggest that the model maintains consistent performance when evaluated on the two unseen years, demonstrating generalisability not only across space but also over time—at least when tested on short-term future periods. In the spatio-temporal split experiment, neither the testing grid cells nor the final two years (2018–2019) were included in model training or validation, ensuring a fully unseen temporal and spatial domain during testing.



**Figure C1.** Model performance evaluated on the testing set comprising post-2017 time-series data, averaged over the testing grid cells. In the spatial split experiment, the model was trained on the full 2001–2019 dataset while holding out specific spatial grids for testing. In contrast, the spatio-temporal split experiment was trained only on data from 2001–2017, ensuring that no post-2017 information was included during training. Boxplots illustrate the distribution of errors across 10 cross-validation folds. [Horizontal dashed lines denote ideal metric values, with bias and RMSE equal to 0 and  \$R^2\$  equal to 1.](#)

## Appendix D: Regularization on global constants

To incorporate prior knowledge into H2CM's estimation of its global parameters, we add the following penalty term to the loss:

$$l_p = \frac{(\hat{\theta} - \theta)^2}{\sigma_\theta^2} \quad (D1)$$

665 Here,

- $\hat{\theta}$  is the estimated global constant,
- $\theta$  is the target (reference) value based on prior knowledge, and
- $\sigma_\theta$  controls the strength of the regularization (analogous to the expected uncertainty in  $\hat{\theta}$ ).

Note that formulation does not represent a Bayesian update but instead serves as a constraint that discourages large deviations  
670 from previously known reasonable values.

In our experiments:

- For  $\beta_{CO_2}$ , we set

$$\theta = 15\% 100 \text{ppm}^{-1}, \quad \sigma_\theta = 5\% 100 \text{ppm}^{-1}.$$

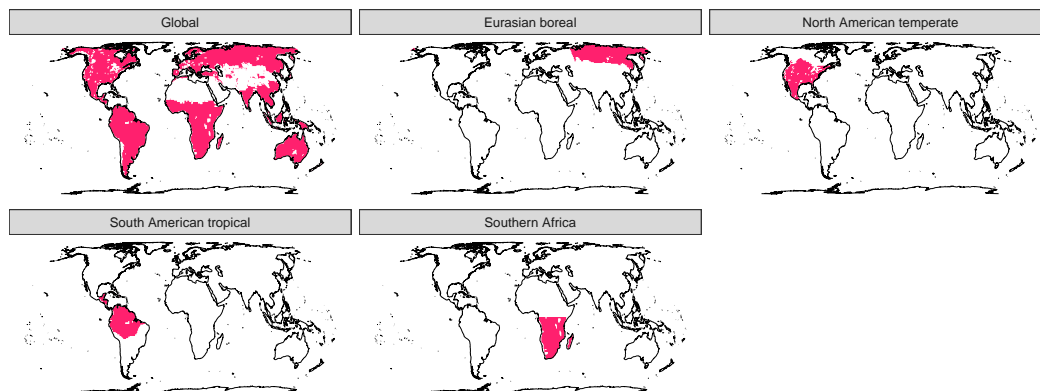
- For the temperature sensitivity,  $Q_{10}$ , we use

675  $\theta = 1.5, \quad \sigma_\theta = 0.3.$

Under this setup, H2CM estimates

$$\beta_{CO_2} = 15\% 100 \text{ppm}^{-1}, \quad Q_{10} = 1.5,$$

which coincide with their reference values. We note that the observed ~~equality between the estimations~~ agreement between estimates and reference values may partly ~~reflect a strong nudging~~ be driven by the strong influence of the prior (regularization) term in the loss function, which constrains the model and reduces its ability to deviate from the reference. At the same  
680 time, this behavior also reflects underdetermination: the observational data do not provide sufficient information to move the parameters substantially away from the prior. Without additional data or process constraints, H2CM's global constants remain underdetermined by the current observations and knowledge alone; therefore, including prior terms in the loss function is currently important.



**Figure E1.** Major Transport and Climate Monitoring (TransCom, Baker et al. (2006)) regions studied in this study

. ZB implemented the model, conducted the analysis, and drafted the manuscript. MJ designed main components of the water-carbon cycle model structure, with inputs from MR, BA, and ZB. All authors contributed intellectual input to the design, analysis, and writing.”

. The authors declare no competing interests.

. We gratefully acknowledge financial support through the German Aerospace Center (DLR) with funds provided by the Federal Ministry  
 690 for Economic Affairs and Climate Action (BMWK) due to an enactment of the German Bundestag. We further acknowledge the support by  
 the European Research Council (ERC) Synergy Grant “Understanding and Modelling the Earth System with Machine Learning (USMILE)”  
 under the Horizon 2020 research and innovation programme. Zavud Baghirov is supported by the International Max Planck Research School  
 for Global Biogeochemical Cycles (IMPRS-gBGC). Bernhard Ahrens was funded by the European Union Horizon Europe research and  
 innovation program under Grant agreement No. 101086179 (AI4SoilHealth). We gratefully acknowledge the financial support from the Max  
 695 Planck Society, which enabled us to publish this paper as open-access.

## References

- Agustí-Panareda, A., Barré, J., Massart, S., Inness, A., Aben, I., Ades, M., Baier, B. C., Balsamo, G., Borsdorff, T., Bousserez, N., et al.: The CAMS greenhouse gas reanalysis from 2003 to 2020, *Atmospheric Chemistry and Physics*, 23, 3829–3859, 2023.
- AINSWORTH, E. A. and ROGERS, A.: The response of photosynthesis and stomatal conductance to rising [CO<sub>2</sub>]: mechanisms and environmental interactions, *Plant, Cell & Environment*, 30, 258–270, <https://doi.org/10.1111/j.1365-3040.2007.01641.x>, 2007.
- 700 Baghirov, Z.: High level overview of the carbon cycle related processes in H2CM, <https://BioRender.com/k6gy7an>, created with BioRender. Accessed 1 July 2025, 2025a.
- Baghirov, Z.: H2CM - inputs and targets, <https://doi.org/10.5281/zenodo.16575309>, 2025b.
- Baghirov, Z.: zavud/h2cm: v1.0.1: Improve TWS anomaly estimation, <https://doi.org/10.5281/zenodo.18400385>, 2026a.
- 705 Baghirov, Z.: H2CM - model simulations, <https://doi.org/10.5281/zenodo.18434085>, 2026b.
- Baghirov, Z.: Simplified overview of the H2CM framework, <https://BioRender.com/k6gy7an>, created with BioRender. Accessed 31 January 2026, 2026c.
- Baghirov, Z.: Overview of the H2CM framework., <https://BioRender.com/0kf418a>, created with BioRender. Accessed 31 January 2026, 2026d.
- 710 Baghirov, Z., Jung, M., Reichstein, M., Körner, M., and Kraft, B.: H2MV (v1.0): global physically constrained deep learning water cycle model with vegetation, *Geoscientific Model Development*, 18, 2921–2943, <https://doi.org/10.5194/gmd-18-2921-2025>, 2025.
- Baker, D. F., Law, R. M., Gurney, K. R., Rayner, P., Peylin, P., Denning, A. S., Bousquet, P., Bruhwiler, L., Chen, Y., Ciais, P., Fung, I. Y., Heimann, M., John, J., Maki, T., Maksyutov, S., Masarie, K., Prather, M., Pak, B., Taguchi, S., and Zhu, Z.: TransCom 3 inversion intercomparison: Impact of transport model errors on the interannual variability of regional CO<sub>2</sub> fluxes, 1988–2003, *Global Biogeochemical Cycles*, 20, <https://doi.org/10.1029/2004gb002439>, 2006.
- 715 Birch, H.: Mineralisation of plant nitrogen following alternate wet and dry conditions, *Plant and soil*, pp. 43–49, 1964.
- Bodesheim, P., Jung, M., Gans, F., Mahecha, M. D., and Reichstein, M.: Upscaled diurnal cycles of land–atmosphere fluxes: a new global half-hourly data product, *Earth System Science Data*, 10, 1327–1365, 2018.
- Boese, S., Jung, M., Carvalhais, N., Teuling, A. J., and Reichstein, M.: Carbon–water flux coupling under progressive drought, *Biogeosciences*, 16, 2557–2572, <https://doi.org/10.5194/bg-16-2557-2019>, 2019.
- 720 Byrne, B., Baker, D. F., Basu, S., Bertolacci, M., Bowman, K. W., Carroll, D., Chatterjee, A., Chevallier, F., Ciais, P., Cressie, N., et al.: National CO<sub>2</sub> budgets (2015–2020) inferred from atmospheric CO<sub>2</sub> observations in support of the global stocktake, *Earth System Science Data Discussions*, 2022, 1–59, 2022.
- Center, U. E.: USGS 30 ARC-second Global Elevation Data, GTOPO30, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. URL: <https://rda.ucar.edu/datasets/ds758.0>, 1997.
- 725 Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., et al.: Global land cover mapping at 30 m resolution: A POK-based operational approach, *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 7–27, 2015.
- Chen, Y., Hall, J., Van Wees, D., Andela, N., Hantson, S., Giglio, L., Van Der Werf, G. R., Morton, D. C., and Randerson, J. T.: Multi-decadal trends and variability in burned area from the 5th version of the Global Fire Emissions Database (GFED5), *Earth System Science Data Discussions*, 2023, 1–52, 2023.
- 730 Chen, Z., Wang, W., Yu, Z., Xia, J., and Schwartz, F. W.: The collapse points of increasing trend of vegetation rain-use efficiency under droughts, *Environmental Research Letters*, 15, 104072, 2020.

- Connor, B., Bösch, H., McDuffie, J., Taylor, T., Fu, D., Frankenberg, C., O'Dell, C., Payne, V. H., Gunson, M., Pollock, R., et al.: Quantification of uncertainties in OCO-2 measurements of XCO<sub>2</sub>: Simulations and linear error analysis, *Atmospheric Measurement Techniques*, 9, 5227–5238, 2016.
- 735 Doelling, D.: CERES Level 3 SYN1DEG-DAYTerra+ Aqua HDF4 file–Edition 4A, NASA Langley Atmospheric Science Data Center DAAC, 2017.
- Dou, X., Yang, Y., and Luo, J.: Estimating forest carbon fluxes using machine learning techniques based on eddy covariance measurements, *Sustainability*, 10, 203, 2018.
- 740 Eyring, V., Gentine, P., Camps-Valls, G., Lawrence, D. M., and Reichstein, M.: AI-empowered next-generation multiscale climate modelling for mitigation and adaptation, *Nature Geoscience*, 17, 963–971, 2024.
- Geirhos, R., Jacobsen, J.-H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., and Wichmann, F. A.: Shortcut learning in deep neural networks, *Nature Machine Intelligence*, 2, 665–673, 2020.
- Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L.: GRUN: an observation-based global gridded runoff dataset from 1902 to 2014, *Earth System Science Data*, 11, 1655–1674, 2019.
- 745 Ghiggi, G., Humphrey, V., Gudmundsson, L., and Seneviratne, S. I.: G-RUN ENSEMBLE, <https://doi.org/10.6084/m9.figshare.12794075.v1>, 2021a.
- Ghiggi, G., Humphrey, V., Seneviratne, S. I., and Gudmundsson, L.: G-RUN ENSEMBLE: A multi-forcing observation-based global runoff reanalysis, *Water Resources Research*, 57, e2020WR028 787, 2021b.
- 750 Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations—the CRU TS3. 10 Dataset, *International journal of climatology*, 34, 623–642, 2014.
- He, Y., Piao, S., Li, X., Chen, A., and Qin, D.: Global patterns of vegetation carbon use efficiency and their climate drivers deduced from MODIS satellite data and process-based models, *Agricultural and Forest Meteorology*, 256, 150–158, 2018.
- Heimann, M. and Reichstein, M.: Terrestrial ecosystem carbon dynamics and climate feedbacks, *Nature*, 451, 289–292, 2008.
- 755 Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., et al.: SoilGrids250m: Global gridded soil information based on machine learning, *PLoS one*, 12, e0169 748, 2017.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, *Quarterly journal of the royal meteorological society*, 146, 1999–2049, 2020.
- 760 Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural computation*, 9, 1735–1780, 1997.
- Huffman, G., Bolvin, D., and Adler, R.: GPCP version 1.2 one-degree daily precipitation data set, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, 10, D6D50K46, 2016.
- Humphrey, V., Zscheischler, J., Ciais, P., Gudmundsson, L., Sitch, S., and Seneviratne, S. I.: Sensitivity of atmospheric CO<sub>2</sub> growth rate to observed changes in terrestrial water storage, *Nature*, 560, 628–631, 2018.
- 765 Huntingford, C., Jeffers, E. S., Bonsall, M. B., Christensen, H. M., Lees, T., and Yang, H.: Machine learning and artificial intelligence to aid climate change research and preparedness, *Environmental Research Letters*, 14, 124 007, 2019.
- Huxman, T. E., Smith, M. D., Fay, P. A., Knapp, A. K., Shaw, M. R., Loik, M. E., Smith, S. D., Tissue, D. T., Zak, J. C., Weltzin, J. F., Pockman, W. T., Sala, O. E., Haddad, B. M., Harte, J., Koch, G. W., Schwinning, S., Small, E. E., and Williams, D. G.: Convergence across biomes to a common rain-use efficiency, *Nature*, 429, 651–654, <https://doi.org/10.1038/nature02561>, 2004.

- 770 Inness, A., Ades, M., Agustí-Panareda, A., Barré, J., Benedictow, A., Blechschmidt, A.-M., Dominguez, J. J., Engelen, R., Eskes, H.,  
Flemming, J., et al.: The CAMS reanalysis of atmospheric composition, *Atmospheric Chemistry and Physics*, 19, 3515–3556, 2019.
- Ito, A. and Inatomi, M.: Water-use efficiency of the terrestrial biosphere: a model analysis focusing on interactions between the global carbon  
and water cycles, *Journal of Hydrometeorology*, 13, 681–694, 2012.
- Jarvis, P., Rey, A., Petsikos, C., Wingate, L., Rayment, M., Pereira, J., Banza, J., David, J., Miglietta, F., Borghetti, M., et al.: Drying and  
775 wetting of Mediterranean soils stimulates decomposition and carbon dioxide emission: the “Birch effect”, *Tree physiology*, 27, 929–940,  
2007.
- Jung, M., Reichstein, M., Schwalm, C. R., Huntingford, C., Sitch, S., Ahlström, A., Arneth, A., Camps-Valls, G., Ciais, P., Friedlingstein, P.,  
et al.: Compensatory water effects link yearly global land CO<sub>2</sub> sink changes to temperature, *Nature*, 541, 516–520, 2017.
- Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., and Reichstein, M.: The  
780 FLUXCOM ensemble of global land-atmosphere energy fluxes, *Scientific data*, 6, 74, 2019.
- Jung, M., Schwalm, C., Migliavacca, M., Walther, S., Camps-Valls, G., Koirala, S., Anthoni, P., Besnard, S., Bodesheim, P., Carvalhais, N.,  
et al.: Scaling carbon fluxes from eddy covariance sites to globe: synthesis and evaluation of the FLUXCOM approach, *Biogeosciences*,  
17, 1343–1365, 2020.
- Jung, M., Nelson, J., Migliavacca, M., El-Madany, T., Papale, D., Reichstein, M., Walther, S., and Wutzler, T.: Technical note: Flagging  
785 inconsistencies in flux tower data, *Biogeosciences*, 21, 1827–1846, <https://doi.org/10.5194/bg-21-1827-2024>, 2024.
- Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., and Kumar, V.: Theory-  
Guided Data Science: A New Paradigm for Scientific Discovery from Data, *IEEE Transactions on Knowledge and Data Engineering*, 29,  
2318–2331, <https://doi.org/10.1109/TKDE.2017.2720168>, 2017.
- Katul, G. G., Oren, R., Manzoni, S., Higgins, C., and Parlange, M. B.: Evapotranspiration: A process driving mass transport and energy  
790 exchange in the soil-plant-atmosphere-climate system, *Reviews of Geophysics*, 50, 2012.
- Katul, G. G., Palmroth, S., and Oren, R.: Leaf stomatal responses to vapour pressure deficit under current and CO<sub>2</sub>-enriched atmosphere  
explained by the economics of gas exchange, *Plant, Cell & Environment*, 32, 968–979, <https://doi.org/10.1111/j.1365-3040.2009.01977.x>,  
2009.
- Khandelwal, A., Xu, S., Li, X., Jia, X., Stienbach, M., Duffy, C., Nieber, J., and Kumar, V.: Physics Guided Machine Learning Methods for  
795 Hydrology, <https://doi.org/10.48550/ARXIV.2012.02854>, 2020.
- Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*, 2014.
- Konings, A. G., Bloom, A. A., Liu, J., Parazoo, N. C., Schimel, D. S., and Bowman, K. W.: Global satellite-driven estimates of heterotrophic  
respiration, *Biogeosciences*, 16, 2269–2284, 2019.
- Kraft, B., Jung, M., Körner, M., Koirala, S., and Reichstein, M.: Towards hybrid modeling of the global hydrological cycle, *Hydrology and  
800 Earth system sciences discussions*, 2021, 1–40, 2021.
- Küçük, u., Koirala, S., Carvalhais, N., Miralles, D. G., Reichstein, M., and Jung, M.: Characterizing the Response of Vegetation  
Cover to Water Limitation in Africa Using Geostationary Satellites, *Journal of Advances in Modeling Earth Systems*, 14,  
<https://doi.org/10.1029/2021ms002730>, 2022.
- Kulawik, S. S., Crowell, S., Baker, D., Liu, J., McKain, K., Sweeney, C., Biraud, S. C., Wofsy, S., O’Dell, C. W., Wennberg, P. O.,  
805 et al.: Characterization of OCO-2 and ACOS-GOSAT biases and errors for CO<sub>2</sub> flux estimates, *Atmospheric Measurement Techniques  
Discussions*, 2019, 1–61, 2019a.

- Kulawik, S. S., O'Dell, C., Nelson, R. R., and Taylor, T. E.: Validation of OCO-2 error analysis using simulated retrievals, *Atmospheric Measurement Techniques*, 12, 5317–5334, 2019b.
- Lawrence, D. M., Thornton, P. E., Oleson, K. W., and Bonan, G. B.: The Partitioning of Evapotranspiration into Transpiration, Soil  
810 Evaporation, and Canopy Evaporation in a GCM: Impacts on Land–Atmosphere Interaction, *Journal of Hydrometeorology*, 8, 862–880, <https://doi.org/10.1175/jhm596.1>, 2007.
- Le Quéré, C., Andres, R. J., Boden, T., Conway, T., Houghton, R. A., House, J. I., Marland, G., Peters, G. P., Van der Werf, G., Ahlström, A., et al.: The global carbon budget 1959–2011, *Earth System Science Data Discussions*, 5, 1107–1157, 2012.
- LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, *nature*, 521, 436–444, 2015.
- 815 Lee, H., Jung, M., Carvalhais, N., Reichstein, M., Forkel, M., Bloom, A. A., Pacheco-Labrador, J., and Koirala, S.: Spatial Attribution of Temporal Variability in Global Land–Atmosphere CO<sub>2</sub> Exchange Using a Model–Data Integration Framework, *Journal of Advances in Modeling Earth Systems*, 17, <https://doi.org/10.1029/2024ms004479>, 2025.
- Li, F., Xiao, J., Chen, J., Ballantyne, A., Jin, K., Li, B., Abraha, M., and John, R.: Global water use efficiency saturation due to increased vapor pressure deficit, *Science*, 381, 672–677, <https://doi.org/10.1126/science.adf5041>, 2023.
- 820 Liu, Y., Yang, Y., Wang, Q., Du, X., Li, J., Gang, C., Zhou, W., and Wang, Z.: Evaluating the responses of net primary productivity and carbon use efficiency of global grassland to climate variability along an aridity gradient, *Science of the Total Environment*, 652, 671–682, 2019.
- Liu, Z., He, C., Xu, J., Sun, H., Dai, X., Cui, E., Qiu, C., Xia, J., and Huang, K.: Observed increasing light-use efficiency of terrestrial gross primary productivity, *Agricultural and Forest Meteorology*, 359, 110–126, 2024a.
- 825 Liu, Z., Skrzypek, G., Batelaan, O., and Guan, H.: Rain use efficiency gradients across Australian ecosystems, *Science of The Total Environment*, 933, 173–181, 2024b.
- Luojus, K., Pulliainen, J., Takala, M., Lemmetyinen, J., Kangwa, M., Eskelinen, M., Metsämäki, S., Solberg, R., Salberg, A., Bippus, G., et al.: GlobSnow-2 Final Report, Global Snow Monitoring for Climate Research, European Space Agency, 2014.
- Luojus, K., Pulliainen, J., Takala, M., Lemmetyinen, J., Mortimer, C., Derksen, C., Mudryk, L., Moisander, M., Hiltunen, M., Smolander, T.,  
830 et al.: GlobSnow v3. 0 Northern Hemisphere snow water equivalent dataset, *Scientific Data*, 8, 163, 2021.
- Medlyn, B. E., DUURSMA, R. A., EAMUS, D., ELLSWORTH, D. S., PRENTICE, I. C., BARTON, C. V. M., CROUS, K. Y., DE ANGELIS, P., FREEMAN, M., and WINGATE, L.: Reconciling the optimal and empirical approaches to modelling stomatal conductance: RECONCILING OPTIMAL AND EMPIRICAL STOMATAL MODELS, *Global Change Biology*, 17, 2134–2144, <https://doi.org/10.1111/j.1365-2486.2010.02375.x>, 2011.
- 835 Medlyn, B. E., De Kauwe, M. G., Lin, Y., Knauer, J., Duursma, R. A., Williams, C. A., Arneth, A., Clement, R., Isaac, P., Limousin, J., Linderson, M., Meir, P., Martin-StPaul, N., and Wingate, L.: How do leaf and ecosystem measures of water-use efficiency compare?, *New Phytologist*, 216, 758–770, <https://doi.org/10.1111/nph.14626>, 2017.
- Metz, E.-M., Vardag, S. N., Basu, S., Jung, M., Ahrens, B., El-Madany, T., Sitch, S., Arora, V. K., Briggs, P. R., Friedlingstein, P., et al.: Soil respiration–driven CO<sub>2</sub> pulses dominate Australia’s flux variability, *Science*, 379, 1332–1335, 2023.
- 840 Metz, E.-M., Vardag, S. N., Basu, S., Jung, M., and Butz, A.: Seasonal and interannual variability in CO<sub>2</sub> fluxes in southern Africa seen by GOSAT, *Biogeosciences*, 22, 555–584, 2025.
- Meyer, N., Welp, G., and Amelung, W.: The Temperature Sensitivity (Q<sub>10</sub>) of Soil Respiration: Controlling Factors and Spatial Prediction at Regional Scale Based on Environmental Soil Classes, *Global Biogeochemical Cycles*, 32, 306–323, <https://doi.org/10.1002/2017gb005644>, 2018.

- 845 Myneni, R., Knyazikhin, Y., and Park, T.: MOD15A2H MODIS/Terra leaf area Index/FPAR 8-Day L4 global 500m SIN grid V006, NASA EOSDIS Land Processes DAAC, 2015.
- Nearing, G. S., Kratzert, F., Sampson, A. K., Pelissier, C. S., Klotz, D., Frame, J. M., Prieto, C., and Gupta, H. V.: What role does hydrological science play in the age of machine learning?, *Water Resources Research*, 57, e2020WR028 091, 2021.
- Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L.,  
850 et al.: X-BASE: the first terrestrial carbon and water flux products from an extended data-driven scaling framework, *FLUXCOM-X*, *Biogeosciences*, 21, 5079–5115, 2024.
- Niu, B., Zhang, X., Piao, S., Janssens, I. A., Fu, G., He, Y., Zhang, Y., Shi, P., Dai, E., Yu, C., Zhang, J., Yu, G., Xu, M., Wu, J., Zhu, L., Desai, A. R., Chen, J., Bohrer, G., Gough, C. M., Mammarella, I., Varlagin, A., Fares, S., Zhao, X., Li, Y., Wang, H., and Ouyang, Z.: Warming homogenizes apparent temperature sensitivity of ecosystem respiration, *Science Advances*, 7, <https://doi.org/10.1126/sciadv.abc7358>,  
855 2021.
- O’Sullivan, M., Friedlingstein, P., Sitch, S., Anthoni, P., Arnoeth, A., Arora, V. K., Bastrikov, V., Delire, C., Goll, D. S., Jain, A., Kato, E., Kennedy, D., Knauer, J., Lienert, S., Lombardozzi, D., McGuire, P. C., Melton, J. R., Nabel, J. E. M. S., Pongratz, J., Poulter, B., Séférian, R., Tian, H., Vuichard, N., Walker, A. P., Yuan, W., Yue, X., and Zaehle, S.: Process-oriented analysis of dominant sources of uncertainty in the land carbon sink, *Nature Communications*, 13, <https://doi.org/10.1038/s41467-022-32416-8>, 2022.
- 860 Paruelo, J. M., Lauenroth, W. K., Burke, I. C., and Sala, O. E.: Grassland precipitation-use efficiency varies across a resource gradient, *Ecosystems*, 2, 64–68, 1999.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, <https://doi.org/10.48550/ARXIV.1912.01703>, 2019.
- 865 Pawar, S., San, O., Aksoyulu, B., Rasheed, A., and Kvamsdal, T.: Physics guided machine learning using simplified theories, *Physics of Fluids*, 33, <https://doi.org/10.1063/5.0038929>, 2021.
- Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, *SOIL*, 7, 217–240, <https://doi.org/10.5194/soil-7-217-2021>, 2021.
- Prentice, I. C., Dong, N., Gleason, S. M., Maire, V., and Wright, I. J.: Balancing the costs of carbon gain and water transport: testing a new  
870 theoretical framework for plant functional ecology, *Ecology Letters*, 17, 82–91, <https://doi.org/10.1111/ele.12211>, 2013.
- Raissi, M., Perdikaris, P., and Karniadakis, G.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational Physics*, 378, 686–707, <https://doi.org/10.1016/j.jcp.2018.10.045>, 2019.
- Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat, F.: Deep learning and process understanding  
875 for data-driven Earth system science, *Nature*, 566, 195–204, 2019.
- Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillerá-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., et al.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, *Ecography*, 40, 913–929, 2017.
- Rödenbeck, C., Zaehle, S., Keeling, R., and Heimann, M.: How does the terrestrial carbon exchange respond to inter-annual climatic  
880 variations? A quantification based on atmospheric CO<sub>2</sub> data, *Biogeosciences*, 15, 2481–2498, 2018.
- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., et al.: Tackling climate change with machine learning, *ACM Computing Surveys (CSUR)*, 55, 1–96, 2022.

- Schneider, T., Lan, S., Stuart, A., and Teixeira, J.: Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations, *Geophysical Research Letters*, 44, 12–396, 2017.
- 885 Shen, C., Laloy, E., Elshorbagy, A., Albert, A., Bales, J., Chang, F.-J., Ganguly, S., Hsu, K.-L., Kifer, D., Fang, Z., et al.: HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community, *Hydrology and Earth System Sciences*, 22, 5639–5656, 2018.
- Shen, C., Appling, A. P., Gentine, P., Bandai, T., Gupta, H., Tartakovsky, A., Baity-Jesi, M., Fenicia, F., Kifer, D., Li, L., et al.: Differentiable modelling to unify machine learning and physical models for geosciences, *Nature Reviews Earth & Environment*, 4, 552–567, 2023.
- 890 Shi, H., Zhang, Y., Luo, G., Hellwich, O., Zhang, W., Xie, M., Gao, R., Kurban, A., De Maeyer, P., and Van de Voorde, T.: Machine learning-based investigation of forest evapotranspiration, net ecosystem productivity, water use efficiency and their climate controls at meteorological station level, *Journal of Hydrology*, 641, 131 811, 2024.
- Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., Doney, S. C., Graven, H., Heinze, C., Huntingford, C., et al.: Recent trends and drivers of regional sources and sinks of carbon dioxide, *Biogeosciences*, 12, 653–679, 2015.
- 895 Sitch, S., O’sullivan, M., Robertson, E., Friedlingstein, P., Albergel, C., Anthoni, P., Armeth, A., Arora, V. K., Bastos, A., Bastrikov, V., et al.: Trends and drivers of terrestrial sources and sinks of carbon dioxide: An overview of the TRENDY project, *Global Biogeochemical Cycles*, 38, e2024GB008 102, 2024.
- Soci, C., Hersbach, H., Simmons, A., Poli, P., Bell, B., Berrisford, P., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Radu, R., et al.: The ERA5 global reanalysis from 1940 to 2022, *Quarterly Journal of the Royal Meteorological Society*, 150, 4014–4048, 2024.
- 900 Staudemeyer, R. C. and Morris, E. R.: Understanding LSTM—a tutorial into long short-term memory recurrent neural networks, arXiv preprint arXiv:1909.09586, 2019.
- Stocker, B. D., Tumber-Dávila, S. J., Konings, A. G., Anderson, M. C., Hain, C., and Jackson, R. B.: Global patterns of water storage in the rooting zones of vegetation, *Nature Geoscience*, <https://doi.org/10.1038/s41561-023-01125-2>, 2023.
- Tao, F., Huang, Y., Hungate, B. A., Manzoni, S., Frey, S. D., Schmidt, M. W., Reichstein, M., Carvalhais, N., Ciais, P., Jiang, L., et al.: 905 Microbial carbon use efficiency promotes global soil carbon storage, *Nature*, 618, 981–985, 2023.
- Tartakovsky, A. M., Marrero, C. O., Perdikaris, P., Tartakovsky, G. D., and Barajas-Solano, D.: Physics-Informed Deep Neural Networks for Learning Parameters and Constitutive Relationships in Subsurface Flow Problems, *Water Resources Research*, 56, <https://doi.org/10.1029/2019wr026731>, 2020.
- Tian, Z., Yi, C., Fu, Y., Kutter, E., Krakauer, N. Y., Fang, W., Zhang, Q., and Luo, H.: Fusion of multiple models for improving gross 910 primary production estimation with eddy covariance data based on machine learning, *Journal of Geophysical Research: Biogeosciences*, 128, e2022JG007 122, 2023.
- Tootchi, A., Jost, A., and Ducharne, A.: Multi-source global wetland maps combining surface water imagery and groundwater constraints, *Earth System Science Data*, 11, 189–220, 2019.
- Ueyama, M., Ichii, K., Kobayashi, H., Kumagai, T., Beringer, J., Merbold, L., Euskirchen, E. S., Hirano, T., Marchesini, L. B., Baldocchi, 915 D., Saitoh, T. M., Mizoguchi, Y., Ono, K., Kim, J., Varlagin, A., Kang, M., Shimizu, T., Kosugi, Y., Bret-Harte, M. S., Machimura, T., Matsuura, Y., Ohta, T., Takagi, K., Takanashi, S., and Yasuda, Y.: Inferring CO<sub>2</sub> fertilization effect based on global monitoring land-atmosphere exchange with a theoretical model, *Environmental Research Letters*, 15, 084 009, <https://doi.org/10.1088/1748-9326/ab79e5>, 2020.
- Vardag, S. N., Metz, E., Artelt, L., Basu, S., and Butz, A.: CO<sub>2</sub> Release During Soil Rewetting Shapes the Seasonal Carbon Dynamics in 920 South American Temperate Region, *Geophysical Research Letters*, 52, <https://doi.org/10.1029/2024gl111725>, 2025.

- Viovy, N.: CRUNCEP version 7-atmospheric forcing data for the community land model, 2018.
- Wang, M., Zhao, J., and Wang, S.: Detection of Carbon Use Efficiency Extremes and Analysis of Their Forming Climatic Conditions on a Global Scale Using a Remote Sensing-Based Model, *Remote Sensing*, 14, 4873, 2022.
- 925 Wang, S., Zhang, Y., Ju, W., Chen, J. M., Ciaï, P., Cescatti, A., Sardans, J., Janssens, I. A., Wu, M., Berry, J. A., Campbell, E., Fernández-Martínez, M., Alkama, R., Sitch, S., Friedlingstein, P., Smith, W. K., Yuan, W., He, W., Lombardozzi, D., Kautz, M., Zhu, D., Lienert, S., Kato, E., Poulter, B., Sanders, T. G. M., Krüger, I., Wang, R., Zeng, N., Tian, H., Vuichard, N., Jain, A. K., Wiltshire, A., Haverd, V., Goll, D. S., and Peñuelas, J.: Recent global decline of CO<sub>2</sub> fertilization effects on vegetation photosynthesis, *Science*, 370, 1295–1300, <https://doi.org/10.1126/science.abb7772>, 2020.
- Watkins, M. M., Wiese, D. N., Yuan, D.-N., Boening, C., and Landerer, F. W.: Improved methods for observing Earth's time variable mass 930 distribution with GRACE using spherical cap mascons, *Journal of Geophysical Research: Solid Earth*, 120, 2648–2671, 2015.
- Wei, S., Yi, C., Fang, W., and Hendrey, G.: A global study of GPP focusing on light-use efficiency in a random forest regression model, *Ecosphere*, 8, e01724, 2017a.
- Wei, Z., Yoshimura, K., Wang, L., Miralles, D. G., Jasechko, S., and Lee, X.: Revisiting the contribution of transpiration to global terrestrial evapotranspiration, *Geophysical Research Letters*, 44, 2792–2801, <https://doi.org/10.1002/2016gl072235>, 2017b.
- 935 Wielicki, B. A., Barkstrom, B. R., Harrison, E. F., Lee III, R. B., Smith, G. L., and Cooper, J. E.: Clouds and the Earth's Radiant Energy System (CERES): An earth observing system experiment, *Bulletin of the American Meteorological Society*, 77, 853–868, 1996.
- Worden, J. R., Doran, G., Kulawik, S., Eldering, A., Crisp, D., Frankenberg, C., O'Dell, C., and Bowman, K.: Evaluation and attribution of OCO-2 XCO<sub>2</sub> uncertainties, *Atmospheric Measurement Techniques*, 10, 2759–2771, 2017.
- Yun, J., Liu, J., Byrne, B., Weir, B., Ott, L. E., McKain, K., Baier, B. C., Gatti, L. V., and Biraud, S. C.: Quantification of regional net CO<sub>2</sub> flux 940 errors in the Orbiting Carbon Observatory-2 (OCO-2) v10 model intercomparison project (MIP) ensemble using airborne measurements, *Atmospheric Chemistry and Physics*, 25, 1725–1748, 2025.
- Zhan, C., Orth, R., Yang, H., Reichstein, M., Zaehle, S., De Kauwe, M. G., Rammig, A., and Winkler, A. J.: Estimating the CO<sub>2</sub> Fertilization Effect on Extratropical Forest Productivity From Flux-Tower Observations, *Journal of Geophysical Research: Biogeosciences*, 129, <https://doi.org/10.1029/2023jg007910>, 2024.
- 945 Zhang, L., Davis, K. J., Schuh, A. E., Jacobson, A. R., Pal, S., Cui, Y. Y., Baker, D., Crowell, S., Chevallier, F., Remaud, M., et al.: Multi-season evaluation of CO<sub>2</sub> weather in OCO-2 MIP models, *Journal of Geophysical Research: Atmospheres*, 127, e2021JD035457, 2022.
- Zhang, Q., Phillips, R. P., Manzoni, S., Scott, R. L., Oishi, A. C., Finzi, A., Daly, E., Vargas, R., and Novick, K. A.: Changes in photosynthesis and soil moisture drive the seasonal soil respiration-temperature hysteresis relationship, *Agricultural and Forest Meteorology*, 259, 184–195, 2018.
- 950 Zhang, Y., Yu, G., Yang, J., Wimberly, M. C., Zhang, X., Tao, J., Jiang, Y., and Zhu, J.: Climate-driven global changes in carbon use efficiency, *Global Ecology and Biogeography*, 23, 144–155, 2014.
- Zhou, T., Shi, P., Hui, D., and Luo, Y.: Global pattern of temperature sensitivity of soil heterotrophic respiration (Q<sub>10</sub>) and its implications for carbon-climate feedback, *Journal of Geophysical Research: Biogeosciences*, 114, <https://doi.org/10.1029/2008jg000850>, 2009.