https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


# Learning from model failure: insights from an AI/SHAP-based temporal error decomposition of tracer-aided ecohydrological modelling

Hyekyeng Jung[1,2], Chris Soulsby[3,1], Christian Birkel[1,4], Songjun Wu[1], Kristina Yordanova[5], Dörthe Tetzlaff[1,2,3]

[1]IGB-Leibniz Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany
[2]Department of Geography, Humboldt Universität zu Berlin, Germany
[3]School of Geosciences, University of Aberdeen, Aberdeen, Scotland
[4]Department of Geography, University of Costa Rica, San Jose, Costa Rica
[5]Institute of Data Science, University of Greifswald, Greifswald, Germany

*Correspondence to*: Hyekyeng Jung (hyekyeng.jung@igb-berlin.de)

**Abstract.** Process-based models (PBM) have served as testing grounds for hypotheses, being falsified and refined through model evaluation based on temporally complex PBM errors. Yet, conventional evaluation practices (e.g., performance metrics, visual inspection across time series, etc.) largely depend on the user's a priori knowledge to trace PBM errors to environmental forcings and are limited in resolving temporal error implications.

In this study, temporal characteristics of PBM errors were explored using a data-driven Artificial Intelligence (AI) model focusing on delayed and non-stationary signatures and linking them to the predictors (i.e., PBM input data). First, an ecohydrological isotope-enabled PBM, EcoHydroPlot, was calibrated against high-resolution datasets of water amounts and water stable isotopes ($\delta^2H$) in soil of two depth layers and tree xylem, monitored from June to October 2020 in a riparian willow plot in Berlin, Germany. Then, PBM errors were calculated for six calibration targets. Second, an ensemble of LSTMs (i.e., AI error analyser) was trained to reproduce the PBM error, which was then decomposed with SHapley Additive exPlanations (SHAP) across both the retrospective lag and the study period, yielding a two-dimensional (lag × time step) quantification of temporal attribution for each predictor.

The analyser reproduced 60-99% of the error variance, showing that these errors were not random noise but carried a systematic, learnable structure. The temporal characteristics of PBM errors depended on both predictors and targets, showing a general tendency that errors contributed by LAI, or targeting xylem $\delta^2H$, were attributed to earlier time lags, whereas those related to precipitation or sapflow were to recent lags. Over the study period, error propagations were observed in distinct patterns, i.e., Event-driven, Period-driven, and regime change signal, especially, near the transition period from growing into non-growing season.

The results showed that the AI error analyser could support evaluation of a PBM as a complementary tool, and as an example, its diagnoses were translated into hypotheses for a better representation of xylem $\delta^2H$ in the PBM.

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


## 1 Introduction

Under the concept of "Modeling as a learning process" (Dunn et al., 2008), process-based models (PBMs) are considered not as fixed representations of reality but as evolving hypotheses that are refined through an iterative learning cycle of simulation, error evaluation, hypotheses generation and hypothesis testing (Beven, 2018). This can also advance through additional data acquisition to inform the modelling and domain can be progressively accumulated (Dunn et al., 2008). In ecohydrological modelling, this domain knowledge refers to understanding of dynamics of water transport, partitioning and mixing in water storages across the Soil – Plant – Atmosphere Continuum (SPAC; e.g., Birkel et al., 2025). This cycle of learning through modeling is driven through using model error, which carries information about both aleatory (random, e.g., measurement noise) and epistemic (knowledge-related, e.g., weak representation of processes in model structure) uncertainties about a model's deficiencies (Beven, 2016). Conventionally, however, model errors often collapse to aggregated scalar metrics (e.g., KGE, NSE, RMSE) in model evaluation, together with visual comparison of simulated and observed time series and with statistical or sensitivity analyses, which can show non-linear parameter–response relations. While being useful for performance evaluation or assessing parameter sensitivity, these approaches provide little insight into the temporal structure of the error (i.e., when and through which time lags drivers act on the error) (Gupta et al., 2008).

One context in which such error-driven learning is particularly informative is tracer-aided ecohydrological modelling. Tracers, such as stable water isotopes, can add other auxiliary criteria (i.e., water velocity, the speed at which the water molecules moves through and mixes within the system) to complement simulations of water fluxes (i.e., water celerity, the speed at which storage or pressure responds to a perturbation) to help test whether models give the "right answer for the right reasons" to falsify the model) (Beven, 2019; Kirchner, 2006). This has been most widely used in conventional conceptual tracer-aided PBMs with mixing conceptualisations that can be used to test simulations of celerity, so isotope tracers can provide independent information to test structural assumptions a model (Jung et al., 2025).

A particularly challenging case in contemporary ecohydrological modeling is the simulation of isotopes in the SPAC where the isotopic composition of xylem water contains great epistemic uncertainties and discrepancies in relation to the potential sources of root water uptake (Barbeta et al., 2019; Orlowski et al., 2023). In such studies, it is usually assumed that xylem water is a simple average of root water uptake from different depths (i.e., sources) in the subsurface either with or without an explicit consideration of root distributions; while the conceptualization of water storage within plant has recently been incorporated in some modelling (e.g., Knighton et al., 2020; Seeger and Weiler, 2021; Smith et al., 2022). While instantaneous mixing of water weighted by different soil layers has been still used for practical reasons in such models (e.g., Gessler et al., 2022; Gillefalk et al., 2021, 2022; Landgraf et al., 2022), incorporating convolution approaches (Knighton et al., 2020; Seeger and Weiler, 2021) explicitly considering rooting distance (Smith et al., 2022) have shown significant improvement in simulation of isotope composition in xylem water.

Despite those advances, however, key limitations remain. Isotopic behaviour of water within xylem and related plant cells (where mixing may occur) is still not well understood under non-stationary and lagged processes: especially following rainfall

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


events (e.g., observed depletion and marked dynamics in xylem water isotopic signatures often contrast with enriched soil water) (Birkel et al., 2025; Seeger and Weiler, 2021) or in the springtime after the dormant season (when enriched isotopes may contrast with depleted snowmelt) (e.g., Sprenger et al., 2025). Such limitations might be attributable to the simplified mass balance equations that rule process-based models and can amplify errors under the conditions of large epistemic uncertainties in high uncertainties in input data or physical representations of the models (Beven, 2020).

In addition, there are still large knowledge gaps regarding processes such as how and when root water uptake occurs from different depths, the mixing dynamics with water already stored in the trees, and their elucidation through the use of water isotopes is at a relatively early stage in ecohydrology (Evaristo et al., 2026). The linked uncertainties are often compounded by simplistic representation and conceptualization of vegetation and evapotranspiration in ecohydrological models (Jung et al., 2025). For example, potential evaporative enrichment in trees during dry spells (Barbeta et al., 2020; Beyer et al., 2020), lag times in water transfer from roots to certain height of stems (Tetzlaff et al., 2021), species-specific compensatory effect of trees in selective water uptake from different depths in reaction to the weather condition (e.g., Ring et al., 2023; Sprenger et al., 2025) are all known to be possible. Yet, these are rarely conceptualised in models.

At the same time, new opportunities have opened to resolved such issues: in situ methods to measure the stable isotope composition of xylem water in trees have been recently developed, which have improved the temporal resolution (subdaily) and extent (over growing seasons or hydrological years) of data availability compared to traditional methods (using destructive sampling) usually at monthly time interval (Beyer et al., 2020). Such high temporal resolution data has been used in both process-based ecohydrological models and Bayesian mixing models to better resolve the sources and dynamics of root water uptake by vegetation in the critical zone (e.g., Birkel et al., 2025; Gessler et al., 2022; Landgraf et al., 2022; Seeger and Weiler, 2021; Smith et al., 2022; Sprenger et al., 2025).

Given the persistent epistemic uncertainties in tracer-aided ecohydrological PBMs and recent improvement of data quality in xylem water, data-driven approaches such as machine learning can provide complementary insights. Machine learning models effectively find generalized patterns in time and space based on the data without being strictly constrained by physical assumptions and have the potential to address site-specific information better than PBM (Jung et al., 2026, under review). This is particularly advantageous in ecohydrology, where responses are strongly heterogeneous, depending on different individual trees, species, and local soil conditions, etc. However, although machine learning is increasingly used in ecohydrology, it has been mostly applied to fluxes (e.g., transpiration, sap flux, etc.) to build data-driven emulators of ecohydrological fluxes (e.g., Kabala et al., 2025), to fill gaps or upscale observations (e.g., Wang and Renninger, 2025), or in some cases to embed data-driven components within a physical framework (e.g., Koppa et al., 2022). Using ML to interpret and diagnose the error of an ecohydrological PBM remains relatively underexplored, particularly for tracer-aided ecohydrological models.

In hydrological studies, eXplainable AI, most often SHAP (SHapley Additive exPlanations), has been applied to interpret PBM errors, which decomposes each prediction into additive contributions by each feature based on game theory (Lundberg and Lee, 2017) (e.g., Feigl et al., 2022; Husic et al., 2025). Husic et al. (2025) worked spatially, relating SHAP of spatial stream flow prediction to static catchment attributes to explain PBM KGE variance across catchments. Feigl et al. (2022)

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


clustered SHAP values to group errors by time steps showing SHAP over the time steps, but the short study period (ca. 1 month) limited insight into seasonal influences. Importantly, both studies also used tree-based models (i.e., Random Forest, XGBoost) that does not have temporal memory structure for previous conditions, so each time step was treated independently, and lagged or non-stationary responses cannot be learnt through the AI model.

In this study, to analyse the temporal dynamics in the errors of a PBM ecohydrological model, we use an ensemble of Long Short-Term Memory (LSTM) models that can capture delayed and time-varying dependencies (Hochreiter and Schmidhuber, 1997; Kratzert et al., 2018) with an eXplainable AI tool. The LSTM ensemble is trained to reproduce the PBM error, and then, is explained with SHAP across both the retrospective lag and the study period, yielding a two-dimensional (lag × time step) quantification of temporal attribution for each predictor. We anticipated that this would help us to detect the non-stationary (across time steps) and delayed structure (across lags) of the error and to diagnose structural deficiencies in a tracer-aided ecohydrological model. We apply this novel approach at the Müggelseedamm riparian willow plot (Berlin, Germany, (Ring et al., 2025). This site provides a rare, comprehensive dataset for a data-driven analysis, including high-resolution time series of xylem and soil water $\delta^2H$, sap flow, and soil moisture at a single plot for ca. 4 months including a transition from growing to non-growing season. This dataset has been studied previously in statistical analysis (Landgraf et al., 2022) and physically based ecohydrological PBM (EcH2O-iso) (Smith et al., 2022).

We address the following research questions by combining an ecohydrological PBM (EcoHydroPlot) with the ensemble of LSTMs and the SHAP-based temporal decomposition:

(1) Are the PBM errors systematic and learnable, carrying a reproducible structural signature rather than random noise?

(2) Which predictors, at which temporal lags, and through which seasonal propagation pathways do the errors arise in terms of non-stationarity, delayed responses?

(3) If possible, what hypotheses for improving the process representation of xylem water $\delta^2H$ in ecohydrological models can be generated from the diagnosed signatures?

## 2 Study Site and Monitoring Campaign

### 2.1 Location and Climate

The study site is located at the Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), south-eastern Berlin, Germany (52.45°N, 13.65°E, ca. 32 m a.s.l.), on the northern shore of Lake Müggelsee (**Fig. 1 (a)**). The site is a peri-urban green space of ca. 480 $m^2$, surrounded by Lake Müggelsee to the south (ca. 50 m away), a stream (ca. 30 m away) to the east which is connected to a pond, and buildings to the north (ca. 40 m away) and the west (ca. 70 m away) (**Fig. 1 (b)**). The Berlin water utility (*Berlin Wasserbetriebe*) operates a water abstraction facility approximately 300 m north of the IGB study site, where groundwater induced by bank filtration is extracted from 30–60 m depth (Berliner Wasserbetriebe, 2026). At the study site itself, the groundwater table lies approximately 2.2 m below the surface and shows low annual variability (< 0.1 m) (Smith

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


et al., 2022). As shown in previous studies (Landgraf et al., 2022; Smith et al., 2022), groundwater is not considered a relevant water source for the shallow soil profile of the study.

The climate is continental temperate with marine influence (Köppen-Geiger classification: Cfb) showing strong seasonality. Over the long term (1994-2025), mean temperature is 10.43 °C and annual precipitation is 566.3 mm $yr^{-1}$ (Berlin-Marzahn; DWD, 2024). Usually, the growing season (April to October) is warmer, and precipitation is storm-driven with short but strong rain events, resulting in an evapotranspiration-dominant environment. The year of this study (2020) was the last year of the European drought (2018-2020), and it was warmer (11.64 °C) and drier (481.2 mm $yr^{-1}$) with lower yearly mean humidity (69.73%; long term mean = 73.39%). Notably, the decrease of precipitation is focused in the growing season.

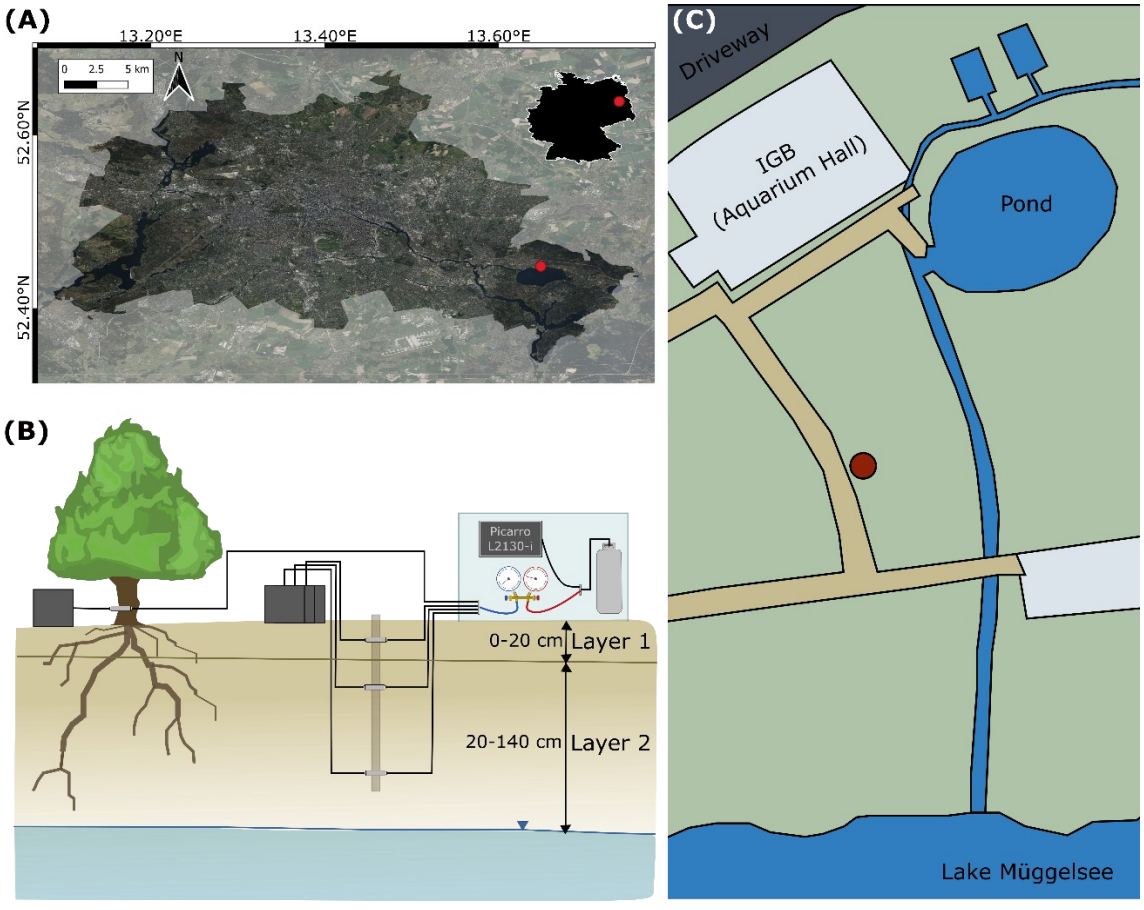


**Figure 1: Location of study site. (a) Site location (red) within Berlin, Germany. Base map: Bing Maps Aerial Imagery © Microsoft Corporation reprinted with permission. (b) Willow tree position on the IGB experimental site. (c) Data collection and monitoring design of in situ measurement of soil water and xylem water isotopes; While dry carrier gas was flowing from one end and desiccant (Drierite) on the other end, xylem water at a willow tree (2 heights) and soil water (3 depths) were collected and sent to a connected Picarro L2130-i analyzer.**

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


### 2.2 Vegetation and Soil

A willow tree (*Salix alba*) for this study stood over grassland (typical lowland riparian ecosystem) and was ca. 14 years old, with a height of ca. 8 m and a stem diameter of ca. 353 mm during the study period. *Salix alba* is a riparian species native to Europe adapted to relatively high water availability (Durrant et al., 2016) and can take water from mixed depths (ca. 70% from upper 60 cm depth) (Granda et al., 2025) in response to time-variant soil moisture conditions (Marttila et al., 2018). In the previous studies (Landgraf et al., 2022; Smith et al., 2022), root water uptake of the studied tree was shown to be largely sourced from shallow soil (ca. 90% of total absorbed water) and mostly derived from summer rains (56%) with xylem transport time of 6.2 – 8.1 days (Smith et al., 2022).

Local soils were emplaced as backfill during the construction of IGB, which was dominated by relatively homogeneous silty sands over depths (up to 100 cm), with organic matter content increasing toward the surface. The sand dominated texture promotes rapid infiltration and yields a low water holding capacity, so that surface soils respond quickly to precipitation inputs but offer limited subsurface water buffering Saxton and Rawls, 2006). This condition can be consistent with the shallow rooted vegetation of the study site (i.e., grassland and tree rooting denser in shallow depth; Visual inspection after completion of the monitoring campaign) especially during dry conditions.

### 2.3 Monitoring and Data Collection

A comprehensive monitoring campaign was conducted over the growing season (June to October) of 2020, and in this study, only the specific data used (i.e., in situ measurement in soil pits and southern willow, weather station) are described below. More details can be found in Landgraf et al. (2022).

#### 2.3.1 Soil

Soil moisture and stable water isotope ($\delta^2H$) in soil water were monitored at 10, 40 and 100 cm depth in two soil pits (Pit A and Pit B). Pit A was located beneath the canopy of the willow tree, and Pit B was at the open grass area, ca. 6 m to the south of the tree. In this study, measurements from Pit A and Pit B were averaged at each depth to obtain a site-representative time series of soil water content and $\delta^2H$, capturing conditions both beneath the willow canopy and in the surrounding open-grass area. Soil moisture was measured in situ using water content reflectometers (CS616, Campbell Scientific Inc. Logan, UT USA) at the three depths every 10 min, and stable water isotopes ($\delta^2H$) of soil water vapour were sampled in situ at the same three depths through buried polypropylene membranes (7 cm length, 0.2 µm pore size). Then, through the carrier gas (dried atmospheric air) offered from one end, the gas stream was transferred to a connected Picarro L2130-i cavity ring down spectrometer (CRDS) and analysed every two hours (**Fig. 1 (c)**).

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


#### 2.3.2 Willow tree and vegetation density

Sap flow rates were monitored at 15 minute intervals by the heat ratio method (Burgess et al., 2001) with a sap flow sensor (SFM1, ICT International, Australia), installed on the northern part of the trunk. Xylem water vapour was extracted from a cylindrical borehole (8-10 mm inner diameter) drilled at a height of 30 cm on the willow stem using the borehole equilibration approach with polypropylene membranes (Marshall et al., 2020). Samples were directed into the CRDS system at approximately two hour intervals. Wounding effect was observed in early days after installation showing abnormally enriched and then exponentially depleted signals, and the relevant data were filtered out in this study. As a general signal of vegetation density of the study area, Leaf Area Index (LAI) was obtained from MODIS Terra (MOD15A2) and linearly interpolated to daily resolution. Additionally, the canopy of the study site was photographed hourly between 6-23 May 2020 and 18 June - 8 July 2020, and in situ LAI was retrieved hourly from canopy photographs through gap fraction analysis based on a histogram derived unimodal threshold, implemented in the LAI R package (Martin, 2015). Due to the coarse temporal and spatial resolution of MODIS data (8 day temporal, 500 m spatial resolution), MODIS data were compared with daily mean of in situ LAI data.

#### 2.3.3 Climate data

Hydrometeorological data were measured by a mobile eddy covariance station (LI-COR Biosciences, Lincoln, NE, USA) deployed on site, at 10-20 Hz frequency. Measured variables included precipitation amount, air temperature, relative humidity, wind speed, and atmospheric pressure. Raw measurements were aggregated to 30 minute means and logged. Potential evapotranspiration (PET) was calculated using the FAO Penman Monteith method, driven by climate measurements from the on-site weather station (Allen et al., 1998). Additionally, for stable isotope analysis, rainfall samples were also collected at 4 hour intervals in sampling bottle with 0.5 cm layer of paraffin oil to prevent evaporation and analysed by Picarro L2130-i cavity ring down laser spectrometer (Picarro, Inc., Santa Clara, CA, USA).

## 3 Modeling

### 3.1 Base model

#### 3.1.1 Process-based Model: EcoHydroPlot

EcoHydroPlot is a plot scale, one dimensional (i.e., one profile over depths), conceptual ecohydrological model with a parsimonious group of input forcings (Stevenson et al., 2023) and subsequently extended with stable water isotope tracking function (Landgraf et al., 2023). In total, six forcing input data (atmospheric temperature, relative humidity, precipitation, $\delta^2H$ in precipitation, potential evapotranspiration (PET), Leaf Area Index (LAI)) are required as model forcing with fifteen parameters (ten for water tracking and five for isotope tracking) to be calibrated (**Table S1**). For detailed model equations, please refer to **Table S2** and Landgraf et al. (2023).

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


The model represents the partitioning of precipitation into the green and blue water fluxes through a cascade of interception, soil storages, and root water uptake by tree with isotope tracking implemented under an instantaneous mixing assumption in each storage compartment (**Fig. 2**). Canopy interception is represented as a storage bucket whose maximum depends on LAI and a calibrated interception parameter. Throughfall is generated when storage exceeds a calibrated threshold, and canopy storage is also limited by reductions in available storage when LAI decreases.

In this study, the soil layer structure of EcoHydroPlot was modified appropriate to the IGB experimental site. The soil compartment consists of two layers: a shallow (0-20 cm) and deep layer (20-140 cm), considering highly variable hydrological processes near the surface (e.g., weak rainfall and evaporation on the surface, water taken by roots concentrated in the upper 10 cm) and homogeneous soil texture over the soil pit profile (Smith et al., 2022). Transpiration from soil layers and evaporation from the tree canopy and upper soil layer are regulated with PET in total amount, and Surface cover fraction (SCF) calculated with LAI partitions PET into total possible evaporation ($E_P$) and transpiration ($T_P$). For each soil layer, the model first satisfies potential tree transpiration and soil evaporation proportional to available soil water in soil layer storage only if the remaining EP and TP are available. Percolation (restricted to the shallow soil layer) and groundwater flow (restricted to the deeper soil layer) are then calculated only when net precipitation is not zero. Preferential flow is permitted only during strong precipitation events (i.e., net precipitation > 4.2 mm $day^{-1}$).

Stable water isotopes ($\delta^2H$) are tracked through the canopy and soil layers via instantaneous weighted mixing of incoming and resident water at each time step, with option for passive mixing through parameters. Both temperature-dependent fractionation and kinetic fractionation-driven by diffusion differences are considered through the Craig-Gordon model (Craig et al., 1963). Additionally, transpired water from each soil layer is summed with complete and instantaneous mixing assumption at each time step, and the xylem $\delta^2H$ of the tree was calculated as weighted mean of transpired waters.

### 3.1.2 Model set-up and calibration

The model was driven at daily resolution by the following forcing variables: precipitation amount (P), precipitation $\delta^2H$ (P_D), potential evapotranspiration (PET), leaf area index (LAI), atmospheric temperature (AT), and relative humidity (RH). All forcing inputs were aggregated (Climate data) or interpolated (P_D and LAI) to daily resolution. Due to the short span of the monitored period and transient condition during the studied period, no spin up was performed. Instead, the initial soil water storages were prescribed in the model calculated from the observations (soil water storage for shallow layer = 15 mm, for deeper layer = 230 mm).

The model was calibrated against six targets: Shallow and deep soil water storages, shallow and deep soil water isotopes ($\delta^2H$), normalized transpiration, and the transpiration isotopic signal ($\delta^2H$). Observed soil water storages were derived from depth-specific soil moisture measurements representative of fixed layers. i.e., Soil moisture at 10 cm for 0-20 cm, soil moisture at 40 cm for 20-60 cm, and soil moisture at 100 cm for 60-140 cm sublayer. The deep soil layer storage was taken as the sum of the latter two. $\delta^2H$ at deep soil layer was calculated as a storage-weighted average of $\delta^2H$ signals at the 40 cm and 100 cm. Normalised transpiration was calibrated against normalised sap flow observations.

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


The DREAM (DiffeRential Evolution Adaptive Metropolis) algorithm was used for efficient calibration on all 15 parameters, which was an adaptive Markov chain Monte Carlo (MCMC)-based algorithm (Vrugt et al., 2009). Five chains were deployed over the global space, and the scale and direction of calibration of each chain was self-adapted in a subspace during the search. All chains converged after 10,000 iterations, through which a parameter group with higher likelihood function L = -800 x ln (1-KGE averaged for six targets) was more likely to be chosen, and 100 parameter groups with best performances were selected (hereafter, behavioural parameter sets). Model simulations were conducted using behavioral parameter sets, and the ensemble mean was reported as final product. Parameter boundaries and calibrated parameters can be seen in **Table S3**.

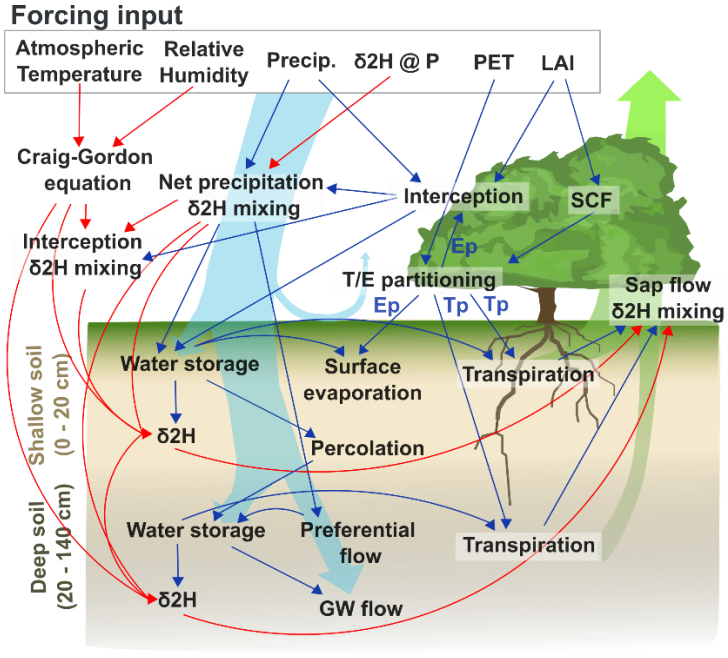


**Figure 2: Conceptual diagram of EcoHydroPlot (modified from Landgraf et al., 2023). Forcing input variables (top) drive water (blue) and δ²H (red) fluxes through shallow soil (0-20 cm), deep soil (20-140 cm), and the tree compartments.**

### 3.2 AI Error Analyser

The AI error analyser used is a post-hoc diagnostic framework (**Fig. 3**). A data-driven AI model (an ensemble of Long Short-Term Memory (LSTM) networks) is trained to simulate the residuals (Observation - Simulation) of each calibration target of the base PBM (**Sect. 3.1.2**), and the AI simulation based on AI understanding is then interpreted using SHAP (SHapley Additive exPlanation) (Lundberg and Lee, 2017).

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


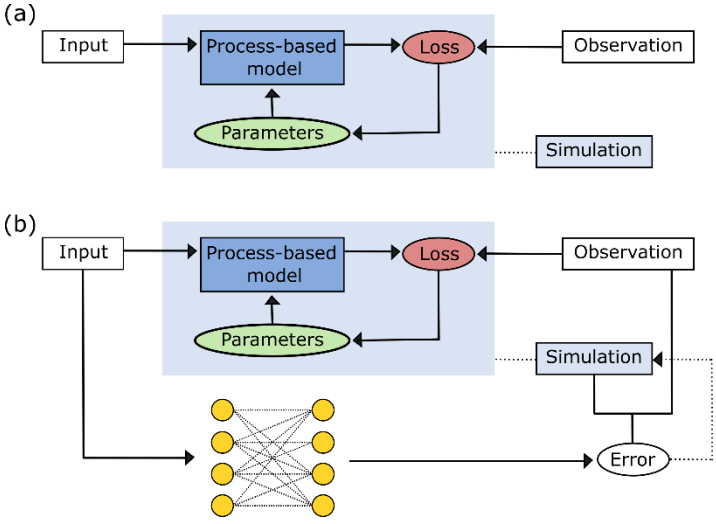


**Figure 3: (a) Calibration and simulation of the base PBM (EcoHydroPlot). (b) Post-hoc AI analyser is an ensembled LSTMs trained to simulate the error (residual) of the PBM.**

### 3.2.1 Ensemble Long Short-term Memory (LSTM)

An ensemble of 10 LSTM networks was trained with the same six forcing input data as predictors that drive the base model, one ensemble per target variable, to simulate the residual errors of the base model. The LSTM architecture was selected for its ability to learn temporal dependencies in sequential data. If the antecedent meteorological event is responsible for the present-day model error, the LSTM can capture that lagged relationship. A simple LSTM architecture suited to small-data hydrological applications was used: hidden size = 64, number of stacked LSTM layers = 2, dropout = 0.2, batch size = 32. Each ensemble member was trained for up to 100 epochs with the Adam optimiser (learning rate = 0.001), minimising the mean squared error (MSE) between predicted and observed residuals, with early stopping (patience = 15 epochs). For a retrospective (look-back) window (i.e., length of input sequence), 10 days was selected after a preliminary test with different input sequence lengths (10, 30, 60 days). Regardless of input sequence length, the model used the input data up to ca. 10 days ago. Each LSTM network was initiated with a different random seed, and ensembled inspired by deep ensemble (Lakshminarayanan et al., 2017), but due to the relatively small training data size, simplified to the median of simulations among the ensemble members and simulation range as uncertainty bounds. The entire study period was used both for model training and testing due to short span of monitoring campaign (132 days). In this study, high predictive accuracy and building a standalone model was not the primary purpose but rather to diagnose how the base PBM errors are temporally structured.

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


### 3.2.2 SHAP-based temporal error decomposition

SHAP is a post-hoc, model-agnostic method that attributes each single prediction to the input features using the Shapley value from cooperative game theory (Lundberg and Lee, 2017). Each predictor's contribution is its average marginal effect on the prediction (i.e., the difference between the model's expected output with and without that predictor, averaged over all combinations of the remaining predictors), so the prediction equals a base value (i.e., mean prediction in this study) plus the sum of all predictors' contributions (i.e., local accuracy).

SHAP was applied to each LSTM ensemble member and decomposed the simulated error at every time step into contributions from each predictor at each time lag within a 10 day retrospective window. The full raw SHAP tensor has the shape (10 x N x 10 x 6) for (ensemble members x time steps x time lags x predictors). Each SHAP value has information of how much and in which direction a predictor of certain days ago contributed the base model's error of a target at a time step in an ensemble member. Local accuracy of SHAP was checked for each target by summing the SHAP values over all predictors and retrospective lags and the base value (i.e., mean predicted error across the study period). The discrepancy of the predicted error of the ensemble LSTM and its SHAP-reconstructed value was within 0.7-1.6 % of the prediction range, confirming that the SHAP faithfully decomposes the model output from the AI error analyser.

To examine the temporal structure of errors, absolute SHAP values were averaged over the ensemble members (mean|SHAP|), and a distribution of mean|SHAP| was generated over lag times for each predictor-targeted model variable pair. Based on the distribution, two metrics were calculated showing the temporal characteristics of errors: *Lag centroid* is the lag-weighted mean of the mean|SHAP| distribution. It addresses the question: "on average, how many days back does the feature's contribution sit?" Higher values indicate that the feature influence is anchored in past information rather than in the current time step. *Lag concentration* measures how peaked or how flat the lag distribution is. It is computed as the normalised Shannon entropy ($H = -\Sigma\ p \cdot \log p$) of the mean|SHAP| distribution. Values near 1 indicate broadly distributed contributions across all lags (memory-anchored), while values close to 0 indicate contributions concentrated at a single lag (event-driven, spike-shaped).

Also, temporal SHAP attribution maps display how mean|SHAP| values of each feature for each target vary over the study period (y-axis = sample time steps) and the retrospective window (x-axis = lag days, 0 to 9). By aligning these axes, the maps reveal how a single forcing event propagates through the model's "memory", which cannot be obtained from the scalar summary metrics in the lag profile analysis. If the model depends on a single event, a diagonal pattern can be observed in the heatmaps.

## 4 Results

### 4.1 Performance of the PBM and its AI error analyser

The PBM (EcoHydroPlot) was run with the 100 behavioural parameter sets and evaluated against observations for all six calibration targets (i.e., celerity and velocity targets in shallow and deep soil, and in the tree) (PBM Simulation in Fig. S1),

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


and its errors were then reproduced by the ensemble LSTM error analyser and evaluated against the true error of the PBM (Fig. 4, Table 1). Generally, the PBM showed moderately good performance (KGE: 0.50-0.71) except for water storage in deep soil and $\delta^2$H in xylem vessel (KGE: 0.55 and 0.34 respectively) that were underestimated. The simulations across the behavioural parameter sets were relatively stable for all targets. In the model simulation, the tree absorbed more water from the deep soil (56.7 % of total root water uptake), which is lower than what the previous study reported (more than 80 % in the upper 10 cm in Smith et al., 2022). Deep soil water was more depleted than water in shallow soil, and thus xylem $\delta^2$H was underestimated. PBM errors were notably changed as the season transitioned from summer to winter in September. In particular, as the |PBM error| dropped in shallow and deep water storages, aforementioned PBM error in xylem $\delta^2$H was mitigated. In late July, $\delta^2$H in shallow soil was also underestimated by PBM, showing a similar strong diluting pattern observed in Pit A measured under tree canopy (Fig. S2).

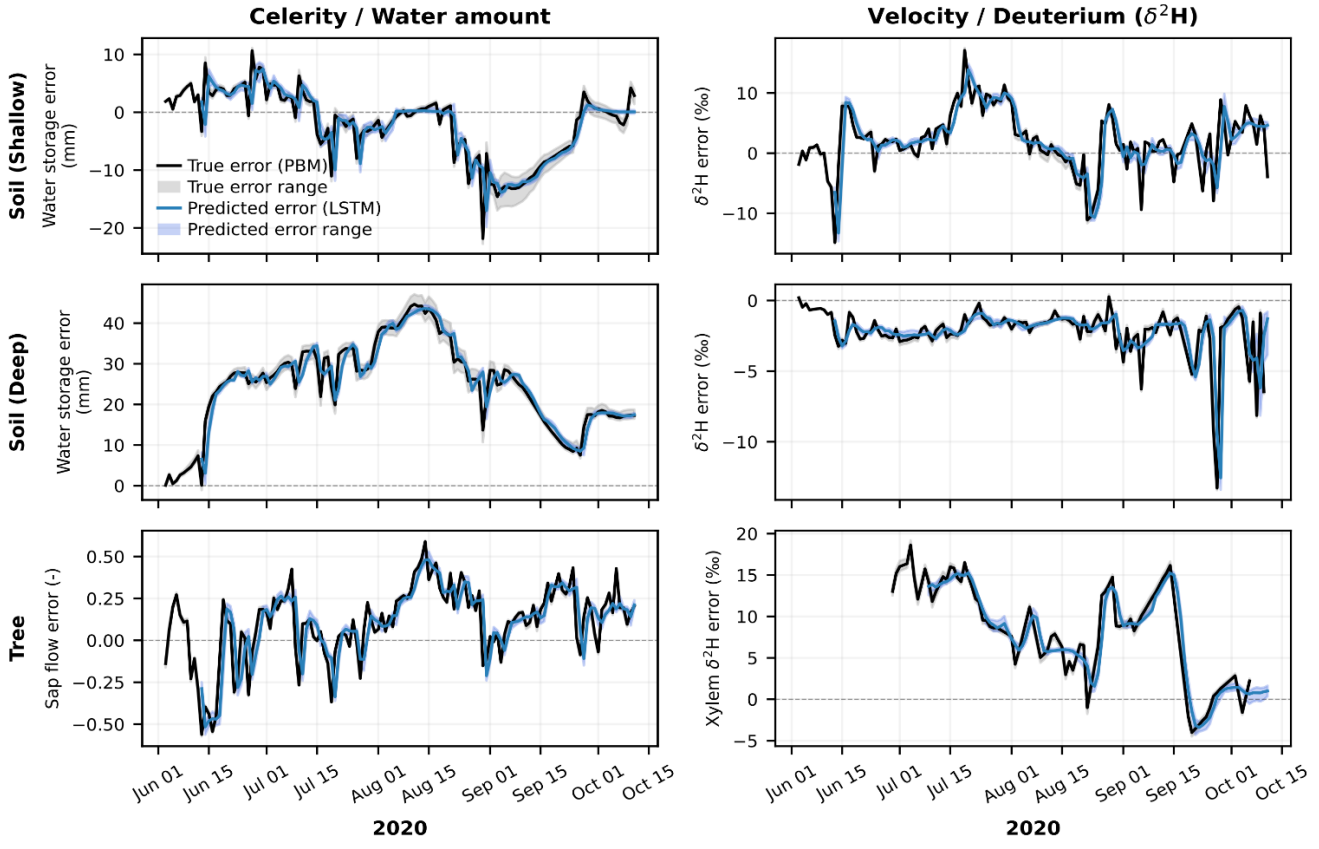


**Figure 4: Comparison of the true error (Observation - PBM simulation; Black line) and the predicted error by the ensemble LSTM error analyser (Blue line). True error range (Grey shade) denoted max - min across the predictions with 100 behavioural parameters, and predicted error range (Blue shade) showed max - min across 10 ensemble members.**

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


**Table 1: PBM and LSTM error analyser performance (RMSE, KGE, PBIAS) for each target. PBM simulation was evaluated against the observations (mean ± standard deviation across the simulations with 100 behavioural parameters), and LSTM simulated error was against the PBM errors (mean ± standard deviation across 10 ensemble members).**

| | | PBM simulation | | | LSTM simulated error | | |
|---|---|---|---|---|---|---|---|
| | | RMSE | KGE | Bias mean | RMSE | KGE | Bias mean |
| Celerity (Water flux) | Soil (Shallow) | 6.19±0.44 mm | 0.57±0.05 | -1.89±0.45 mm | 2.94±0.21 mm | 0.85±0.02 | -0.04±0.10 mm |
| | Soil (Deep) | 26.91±0.71 mm | 0.55±0.02 | 24.62±0.64 mm | 3.04±0.09 mm | 0.94±0.00 | 0.08±0.13 mm |
| | Tree | 0.24±0.00 | 0.50±0.00 | 0.11±0.00 | 0.13±0.00 | 0.79±0.01 | 0.00±0.00 |
| Velocity ($\delta^2H$) | Soil (Shallow) | 5.32±0.04 ‰ | 0.61±0.01 | 2.07±0.06‰ | 3.22±0.07 ‰ | 0.76±0.01 | -0.01±0.06 ‰ |
| | Soil (Deep) | 2.74±0.05 ‰ | 0.71±0.00 | -2.16±0.07 ‰ | 1.62±0.09 ‰ | 0.47±0.05 | -0.05±0.02 ‰ |
| | Tree | 10.10±0.20 ‰ | 0.34±0.01 | 8.38±0.24‰ | 2.94±0.16 ‰ | 0.81±0.02 | 0.60±0.13 ‰ |

The AI error analyser could reproduce the PBM errors with high accuracy for all targets apart from the deep soil $\delta^2H$ (KGE: 0.76-0.94), and the ten ensemble members provided stable estimates with a standard deviation among the ensemble members of 0.003 to 0.062. Deep soil $\delta^2H$ was the target whose error was least well reproduced (KGE: 0.47). This can be due to the lower information content of the PBM error which stayed relatively constant apart from a few rain event spikes (e.g., the largest on 27 Sep, after a 10.8 mm rain of depleted $\delta^2H$). The LSTM-based error analyser could learn most of PBM error variance (60-99%), which supported that PBM errors were not random but systematic and thus, largely learnable. It was also noteworthy that the AI error analyser could simulate model errors better for celerity-based targets (esp., water storage in shallow/deep soil) than velocity-based targets (i.e., $\delta^2H$ in shallow/deep soil).

### 4.2 Overall predictor contributions to the PBM errors for each target

Feature importance for simulating PBM errors of each target was quantified based on mean|SHAP| averaged over both the study period and retrospective window and compared between the targets by normalising the sum of mean|SHAP| by all predictors for a target to 1 (**Fig. 5**). Also, the directions of the SHAP values (i.e., contribution of a predictor) by changes of all input predictors could be checked in **Fig. S3** and for the whole time series, **Supplementary Videos**. The directions of contribution were interpreted only for the leading predictors of each target, since the sign of a non-dominant predictor could be incidental and the AI model was more likely to be trained opposite to physical meaning without being regulated by the loss function.

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


Overall, the PBM error of the celerity targets was explained by PET and LAI, whereas the PBM error of the velocity targets was explained by atmospheric temperature and precipitation $\delta^2H$. Importance of precipitation seemed more weakly coupled to the same-day predictors of rain events since PBM errors contributed by precipitation were lagged and spiked for a day.

PET ranked the highest in celerity targets such as water storage at shallow depth and sap flow. When PET is above average (standardized PET > 0), PET contributed PBM error (i.e., observation - simulation) positively in shallow water storage (Pearson R: +0.87) and negatively in sap flow (Pearson R: -0.86). This might imply the limitation in setting total potential transpiration and evaporation by partitioning PET, which can allow systematic overestimation of ET and led to underestimation of shallow water storage and overestimation sap flow when evaporative demand is high. This can be more critical to processes calculated earlier in the model structure (e.g., shallow depth soil).

LAI ranked the highest in deep soil water storage and xylem $\delta^2H$. In the LSTM model, both are largely regulated by LAI whose contribution decreased PBM error as LAI dropped (Pearson R: +0.93 and +0.92). In EcoHydroPlot, lower LAI determines lower ratios of total potential transpiration (Tp) and evaporation (Ep) that root water uptake from deep soil depth can be decreased.

Atmospheric temperature ranked higher in the targets related to velocity targets, (i.e., $\delta^2H$ in soil or xylem vessel) as well as sap flow. Lower atmospheric temperature contributed PBM errors negatively in shallow soil $\delta^2H$ (Pearson R: -0.61) and xylem $\delta^2H$ (Pearson R: -0.69), and positively in deep soil $\delta^2H$ (Pearson R: +0.50) and sap flow (Pearson R: +0.70). For velocity targets except deep soil $\delta^2H$, higher temperature enriches the simulated $\delta^2H$ and reduces the otherwise depleted underestimation in shallow soil and xylem water. However, the deep soil water $\delta^2H$ stood apart. Due to the low learnable information in PBM errors (Low AI model performance in 4.1.), its predictor relationships, including atmospheric temperature, largely reflect the conditions of that single event rather than a recurring mechanism. For sap flow, in contrast, higher temperature increased sap flow underestimation (e.g., in August). It might be related to the simple conceptualization of transpiration induced by temperature-dependent evaporation on tree leaves in EcoHydroPlot.

Precipitation $\delta^2H$ was important especially for the shallow soil $\delta^2H$. In shallow soil $\delta^2H$, more depleted precipitation (standardized Precipitation $\delta^2H < 0$) contributed positively to the error (Pearson R: -0.46). When incoming precipitation was more depleted than the resident soil water, the model overdepleted the shallow soil and underestimated $\delta^2H$. As mentioned in 4.1., EcoHydroPlot behaved closer to the exposed condition, more sensitive to precipitation.

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


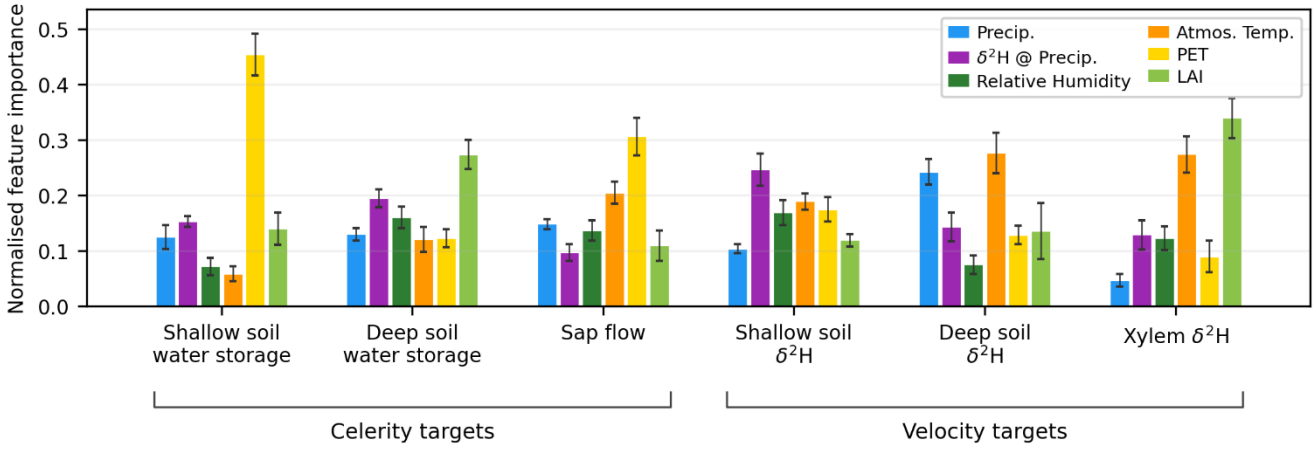


**Figure 5: Feature importance ranking by target. Feature importance was normalised such that the sum of feature importances of all predictors for each target is equal to 1.**

## 4.3 Predictor contributions to the PBM errors over time lags

For each predictor-target pair, the distribution of mean|SHAP| averaged over the study period was shown across the retrospective window (**Fig. 6**). Then, each distribution was summarised by its lag centroid (i.e., the mean lag at which the contribution is centered) and lag concentration (i.e., its spread), and the two are mapped against each other in **Fig. 7**.

Each type of data used as a predictor has characteristic temporal signatures, defined by their contribution is centered at previous or recent lags. Generally, precipitation and relative humidity showed event-driven characteristics with contributions concentrated at recent lags (mean lag centroid: 3.5 d and 3.9 d) and spike-shaped (mean lag concentration: 0.92 and 0.94). On the other hand, LAI, atmospheric temperature, Precipitation δ²H, and PET were memory-anchored, with contributions concentrated at previous lags (mean lag centroid: 5.0 d, 4.6 d, 4.4 d, 4.2 d) and uniformly distributed (mean lag concentration: 0.98, 0.98, 0.97, 0.96). However, the lag structure of each predictor could also be reshaped by the target. Precipitation became memory-anchored when explaining the xylem δ²H error (Lag centroid: 4.94d, Lag concentration: 0.99), and LAI shifted toward recent lags for sap flow (Lag centroid: 4.94d, Lag concentration: 0.99). The temporal structure of the error is therefore a property of the predictor-target pair, not of the predictor alone. A systematic gradient was observed within the targets. For the soil targets, depending on the depths, the lag centroid was slightly shorter for the shallow (Mean lag centroid: 4.45 d for shallow water storage and 3.77 d for its δ²H signal) than the deep layer (Mean lag centroid: 4.65 d for deep water storage and 4.27 d for its δ²H signal). The clearest contrast between celerity and velocity targets was in the tree. Sap flow error responded to recent forcing (Mean lag centroid: 3.53 d), whereas the xylem δ²H error depended on earlier forcing (Mean lag centroid: 5.01 d).

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


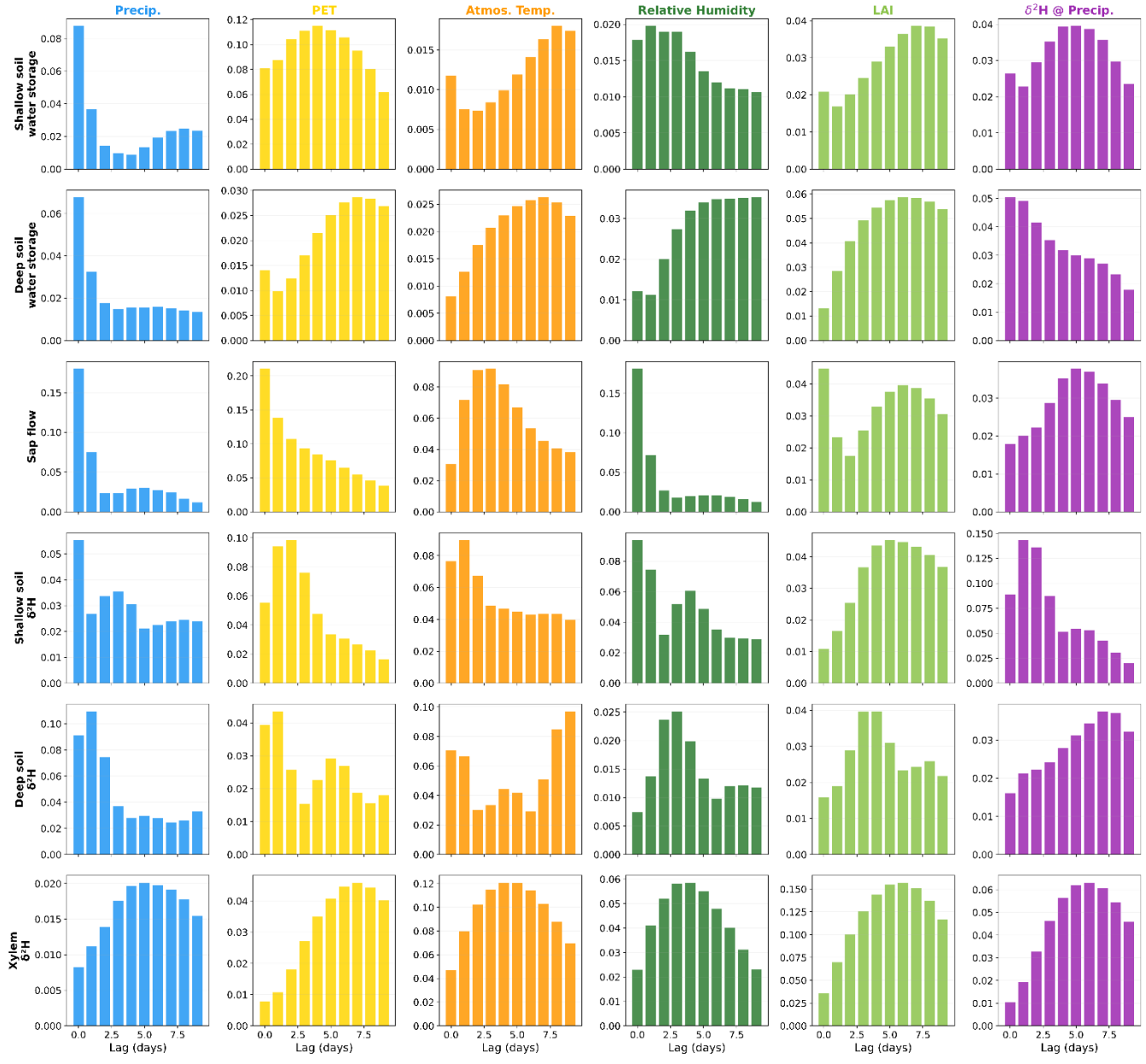


**Figure 6: Distribution of mean|SHAP| over the retrospective window by lag time (e.g., lag 0 = current time step, lag 1 = 1 day ago from the current)**

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


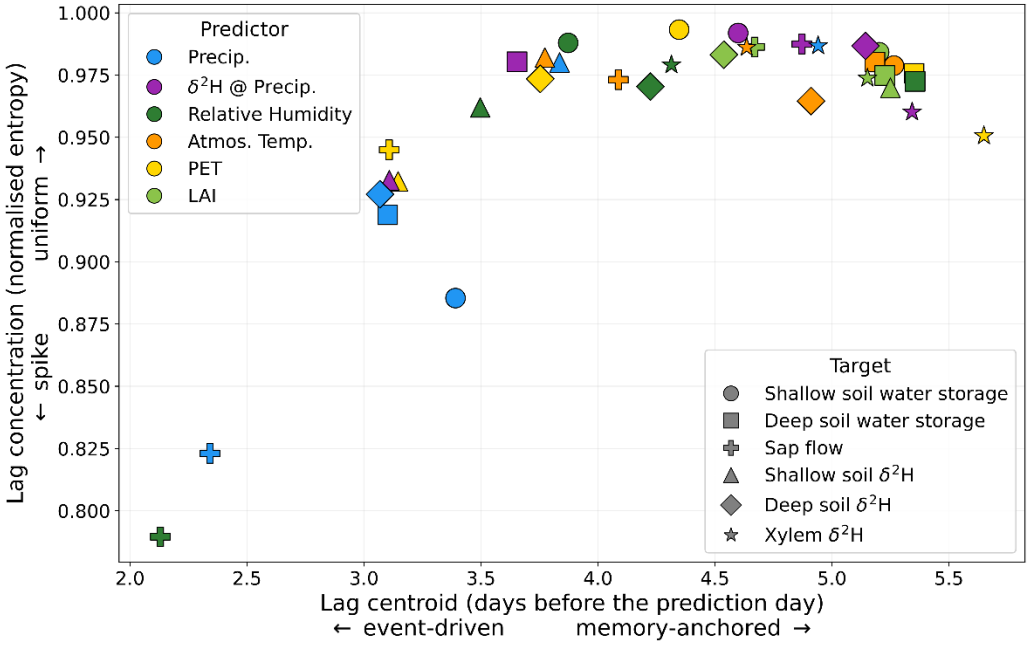


**Figure 7: Temporal characteristics of errors were summarised for every predictor-target pair using Lag centroid and Lag concentration.**

## 4.4 Predictor contributions to the PBM errors over the study period

All|SHAP| from each predictor-target pairs are shown in a matrix plot over both the retrospective window (horizontal axis) and the study period (vertical axis) (**Fig. 8**). Within these matrixes, a diagonal streak tracks a single day as it moves through the retrospective window over successive days.

Three distinct geometrical patterns could be identified visually with the following criteria, implying a distinct type of modes by which the errors draw information from the retrospective window and errors propagate.

- Type 1 (Event-driven) appears as an isolated diagonal sum spike that stands out sharply from neighbouring days, forming a *diagonal line* in the matrix.
- Type 2 (Period-driven) is defined by temporal continuity, with the influence of a single event or period spread across several consecutive days, forming an *oval* cluster in the matrix.
- Type 3 (Regime change signal) is showing the |SHAP| spread evenly across a partial or a whole retrospective window rather than following any single event, appearing as a *square* (horizontal band) in the matrix.

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


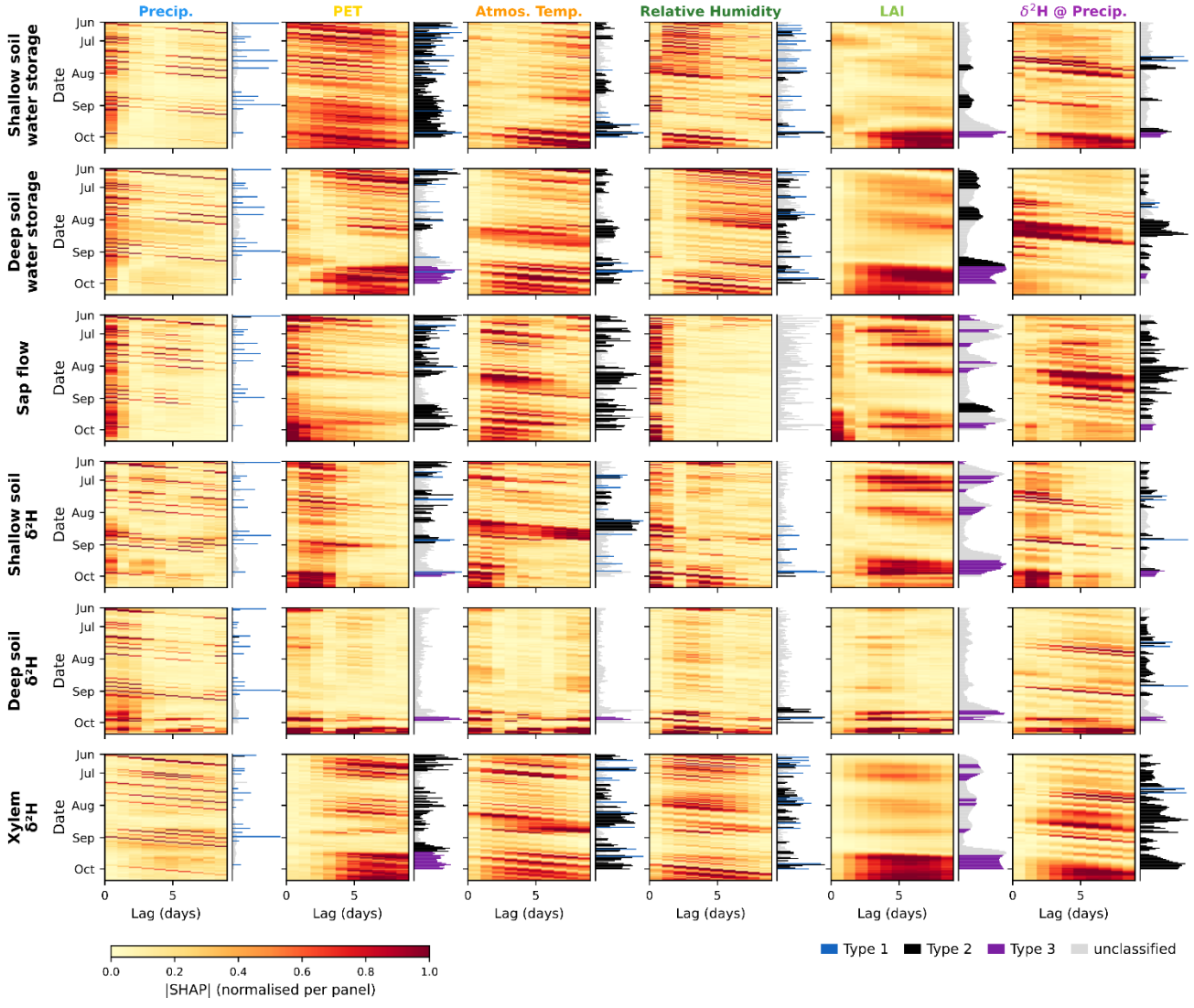


**Figure 8: Temporal changes of |SHAP| over the retrospective window (x axis) and the study period (y axis). In each panel the colour map shows normalised |SHAP|, and the adjacent bars give the diagonal sum of |SHAP| at each time step, coloured by the streak type assigned to it: Type 1: event-driven (blue), Type 2: period-driven (black) and Type 3: regime change (purple). Streaks were classified by visual inspection. When more than two patterns co-occur, overlapping labels were retained and shown with the priority of event-driven > regime change > period-driven. The diagonal sum of the final ten time steps was not considered because not all ten lags are available there.**

Generally, each type was associated with a distinct set of predictors: Type 1 was almost entirely tied to precipitation-related predictors (precipitation 42% and relative humidity 22.5% of all event-driven labels). Type 2 was dominated by smoothly changing predictors (PET 29%, precipitation $\delta^2H$ 28%, and air temperature 23%). Type 3 was overwhelmingly associated with LAI (63%) and were concentrated in a transition period (i.e., late September to October). However, these types can co-occur in a predictor-target matrix, and the dominant type shifted according to the event and period. For example, shallow soil $\delta^2H$

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


error illustrated this well, showing type changes and different predictors associated depending on periods. In late July and late August, it was associated with precipitation $\delta^2H$, PET, and relative humidity mostly in Type 1, implying mixing with depleted precipitation. In late August, the main predictors changed to atmospheric temperature and relative humidity in Type 2, implying a connection to surface evaporation. Finally, during the transition period, the attribution of almost all predictors except relative humidity reorganised into Type 3 bands drawn from the most recent lags, while relative humidity alone retained a Type 1 signature but from much earlier lags. These shifts showed that the PBM error was not tied to a fixed depth of retrospective window over the study period. Depending on which hydrological process dominated for certain time span, it connected to different parts of the retrospective window by different predictors. Additionally, the deep soil $\delta^2H$ showed only few Type 3 bands in the transition period, showing that its error structure was concentrated near the single late September rainfall event (27 September), consistent with the lower AI error analyser performance for this target (Sect. 4.1).

## 5 Discussion

### 5.1 How do AI-based and conventional evaluation of PBM help us understand non-stationarity and time lags between drivers and responses?

Ecohydrological processes are inherently complex, involving multiple process interactions with non-stationary and nonlinear responses. To the extent that PBMs do not fully capture this nonlinearity (e.g., due to the simplified representation of nonlinear responses) and non-stationarity (e.g., due to time-invariant parameters and structure), the resulting mismatch of model and data is imprinted as structure in their residuals (cf. Kirchner, 2024). In this study, contrasting nonlinear processes connected to various environmental drivers and different parts of the retrospective window dominated the PBM error of a target at alternative times during the growing season (Fig. 8 and S4). Through this, the AI error analyser showed strong potential in detecting not only non-stationary nonlinear hydrological signatures but also temporal delay of predictor signals on ecohydrological responses (e.g., transit time between model components). While previous AI approaches, particularly using SHAP, have shown dominant drivers over different time periods (e.g., Clark et al., 2025) and their nonlinear relationships (Jung et al., 2024) to their AI model prediction, the retrospective window introduced here as an additional analytical axis enables a more systematic characterisation of lagged responses in ecohydrological contexts.

Attributions from SHAP are known to become unreliable when predictors are correlated, as the credit assigned among correlated features can be arbitrary (Aas et al., 2021). Here, however, although the predictors are strongly correlated (e.g., Pearson R: +0.75 for PET-atmospheric temperature, +0.79 for PET-LAI, and +0.69 for atmospheric temperature-LAI), the ensemble predictions and the way information was drawn from each predictor remained consistent across ten ensemble members, and the analyser could learn the different processes at different periods over the time steps, separating the predictors through their non-shared variance. For example, the August heat could not be captured through PET, but through atmospheric temperature (Sect. 4.4), despite their high correlation. Such a separation is not possible with a period-averaged correlation matrix or linear regression, and the correlation can be weaker locally in time. This allowed the analyser to reveal the non-

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


stationarity of the error, although the SHAP result of strongly collinear predictors should still be interpreted with care. This is similar to statistical time series models such as Markov-switching autoregressive model (Birkel et al., 2012) that can learn parameters of previous time lags in linear regressions for each discrete, prespecified hydrological regime. The AI analyser generalised this concept by letting the relationships vary continuously and nonlinearly though time.

On the other hand, conventional PBM evaluation (Sect.4.1.), relies on aggregate metrics and visual inspection of the time series. Since the metrics carried little information averaged over many non-stationary periods of different hydrological signatures, it often requires hydrologists with local domain knowledge to visually explore patterns and relate them qualitatively to the predictors (Gupta et al., 2009). Beyond this routine practice, time-variant sensitivity analysis could offer more formal diagnostics, showing how each parameter's sensitivity changed quantitatively through time and was able to flag periods of low identifiability as possible structural deficits (e.g., Birkel et al., 2014; Herman et al., 2013; Reusser and Zehe, 2011). It captured non-stationarity in the parameter-output space and flagged directly which equation connected to parameter is responsible at certain time step (Pfannerstill et al., 2015). However, when the response in PBM output is delayed relative to a predictor's own time series, the two are hard to connect without a priori process knowledge at the study site, making the link between the hydrological signature and predictors difficult to detect. Moreover, the length of this delay can be itself non-stationary, which both qualitative inspection and parameter sensitivity analysis are less suited to capture, as the lag is not itself their target to analyse and can be noted only qualitatively or time-invariant quantity (cf. Razavi et al., 2021). Still, the AI error analyser captured non-stationarity in the environmental driver-PBM error space and detected which predictor is responsible for a hydrological signature, not only across time steps but also across lagged time window quantitively. Thus, it is complementary to time-variant sensitivity analysis. Also, coupling the analyser with a hybrid model might further link the diagnosed error signatures to the individual model equation responsible for them (See Sect. 5.3).

### 5.2 Interpreting PBM errors: composite information integrating error sources and structural signatures propagated through the model

The PBM error was not random but carried a systematic, structural signature, making it an informative target for diagnosing PBM structure. 60-99% of its variance could be reproduced by the AI-based analyser, with a small spread across the ten ensemble members (Sect. 4.1). Understanding what this error signal represents is therefore essential for interpreting the analyser's diagnoses correctly.

The structural signature that the AI error analyser targeted is observed in PBM errors when various error sources (e.g., input data, parameters, model structure, and measurement) pass through the model equations and are reshaped to conform to the model structure (Beven and Westerberg, 2011). Since the PBM error contains composite information, which is not separated into its component sources, the analyser's results are therefore specific to the particular input data, parameter sets, and model structure used in the PBM simulation. As a result of this, a structure that is in fact incorrect may leave no trace in the error (Kirchner, 2006; McMillan et al., 2018), as observed in deep soil $\delta^2H$ which contains almost no variance in the time series except a single rainfall event to inform the AI error analyser (Sect. 4.1 and 4.4). This is an inherent constraint of inference

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


from data of low information, applying to any model, whether data-driven or process-based, rather than a limitation specific to the AI analyser (Beven and Westerberg, 2011).

Furthermore, the structural signatures found in the final PBM errors of a variable can be associated to one or more equations through the model structure, but only dominant signatures can be detected distinctly by the AI error analyser. Inside the PBM, errors propagate like a cascade from the computationally earlier determined variable to the later variable. Components represented as mixtures, and those computed later in the model structure, carry errors that are integrals of several processes. In these cases, the individual contributions become diffuse in the later variable and hard to separate. Shared temporal signatures between PBM errors targeting xylem $\delta^2H$ and deep soil water storage (**Fig. 8**) may reflect this, as root uptake water from deeper in the soil was likely overestimated relative to shallow soil depths. This might be one of the reasons why the analyser tended to understand the celerity targets better than the velocity targets, and the shallow soil better than the deep soil, since the targets computed earlier in the PBM cascade accumulate less diffused signature from the earlier computed targets, so their errors remain more directly attributable.

### 5.3 Using PBM error signatures for improved modelling of time-variant plant xylem water sources

The analyser provided insight into delayed processes for each predictor-target pair, and the delay was clearest for xylem $\delta^2H$ (target) and LAI (predictor). This was consistent with earlier work in the same study period at the same site (Landgraf et al., 2022; Smith et al., 2022). Smith (2022) showed that applying a complex physically based PBM ($EcH_2O$-iso) with instantaneous mixing peaks from soil water sources cannot capture daily xylem variability because of its static transit time and absence of transient storage and/or interactions within the plant root stem system. This can be related to delayed response of xylem $\delta^2H$ and its variability according to dominant processes. Their estimated transit time from soil to xylem vessel was 6.1 to 8.1 days from the PBM, which was close to the lag days (Lag centroid: 4.3. – 5.7 days) found in this study. This was consistent with Landgraf et al. (2022) who showed that xylem water integrates inputs over weeks to months and does not respond immediately to single rain events. Also, xylem vessels in trees can rapidly respond to root water uptake during the rewetting transition from drought (i.e., dry condition) to recovery period (i.e., wetter condition) (Kühnhammer et al., 2023), matching the Type 3 temporal signature found at the period of season transition.

Considering PBMs as the foundation for domain knowledge (Nearing et al., 2021), this knowledge can continuously accumulate through a modelling cycle which repeats the routine of "PBM simulation → PBM error evaluation → Hypothesis generation → Test the hypothesis in the PBM" (i.e., "Modeling as a learning process"; Dunn et al., 2008). AI error analyser can contribute to the error evaluation process and building domain knowledge by helping generating the hypotheses for PBM improvement (cf. Jiang et al., 2024). Based on the aforementioned PBM error signatures identified by the analyser, three hypotheses can be generated: 1) introduction of non-instantaneous mixing and tree water storage could improve simulation, 2) flexible lag dynamic rather than a static one would be better, and 3) using a LAI-based phenology threshold would be more realistic for the change in tree behaviour during the season transition. Still, such hypothesis generation currently needs to be dependent on a model user with domain knowledge. However, a hybrid modeling approach such as Universal Differential

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


Equations (UDEs) (Bolibar et al., 2023; Höge et al., 2022; Huynh et al., 2026) in conjugation with AI error analysers has potential to rapidly identify the specific equation responsible for an error by replacing individual PBM equations with neural networks. Starting from the equation closest to the target variable and moving back through the linked computationally earlier equations, each equation can be replaced in turn, and the replacement that most reduces the error might narrow down its dominant sources. This error reduction, however, must be interpreted carefully in hybrid modelling contexts, as a structural deficiency in the model can also be compensated by high flexibility of neural networks rather than corrected (Acuña Espinoza et al., 2024; Höge et al., 2022).

### 5.4 Technical considerations

The results from the AI error analyser were based not directly on PBM errors but the model predictions of the LSTM. Thus, the influence of the technical limitation of the LSTM algorithm needs to be considered carefully while interpreting the results from the analyser. While training the AI model based on Recurrent Neural Network (RNN), the model gradients from earlier lags of the retrospective window tend to vanish. Although this issue has been alleviated in LSTM (Hochreiter and Schmidhuber, 1997), gradient-based attributions can still emphasise the most recent days of the retrospective window and underrepresent the earliest ones (Ismail et al., 2019).

This has the implication that the lag centroid should be read as a relative indicator of how far back the error draws on its drivers, not as an exact transit time between hydrological components. Also, the obvious importance of earlier lags for xylem $\delta^2H$ and LAI or Type3 error propagation can be strengthened since it emerged despite the bias towards recent lags. The length of the input sequence may limit how far back dependencies can be detected and furthermore may limit itself due to the vanishing gradients. Future applications of this method should test the sensitivity to sequence length and select an appropriate length for a given model purpose, studied site and period concerned. Beyond this, further research is needed to establish whether longer retrospective windows can be reliably resolved.

## 6 Conclusions

This study presents, as a proof of concept, a data-driven method for evaluating the errors of a process-based model (PBM). An ensemble of LSTM networks were trained to reproduce the PBM error, and SHAP then attributed that error to the model's input predictors, decomposing it along two dimensions (i.e., time lag and time step). We demonstrated the approach by coupling the analyser with the tracer-aided ecohydrological model EcoHydroPlot at a riparian willow plot in Berlin, Germany, across the 2020 growing season to dormant transition.

The ensemble reproduced most of the error variance, showing that the PBM error contains a learnable structure. This made it possible to analyse the non-stationarity and nonlinearity of the error, together with the delayed responses. Relative to conventional evaluation and more advanced diagnostics such as time-variant sensitivity analysis, its main potential lies in adding the time lag as a second axis of analysis, so that the lag itself can be resolved as a time-variant quantity. Also, the

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


analysis showed that temporal signatures of errors were dependent on both predictor-target pair and propagated along the time steps via three distinct modes (i.e., event-driven, period-driven and regime change signal).

From these signatures we generated three testable hypotheses for improving the representation of xylem $\delta^2H$. For example, non-instantaneous mixing with tree water storage, a flexible rather than static lag, and an LAI-based phenology threshold. Machine-learning-based error evaluation can therefore support the falsification of a model and the generation of alternative hypotheses, contributing to domain knowledge under the "modelling as a learning process" frame.

The PBM error remains composite information from various possible sources, so the diagnosis is specific to the input data, parameters and model structure used here; thus, applying the approach across other models, sites and periods is a next step. Furthermore, since the predictors do not need to be restricted to the model's own forcings in the AI error analyser, variables that the model structure cannot accommodate ("soft data") could also be attributed to its error. Together, these directions would allow the analyser to carry the error evaluation step of PBM and the learning process to be carried out more systematically

## Code and data availability

The process-based model (EcoHydroPlot) and the AI error analyser are archived in a Zenodo repository (https://doi.org/10.5281/zenodo.21457019). The input data are available in the open-access database FRED (https://fred.igb-berlin.de/data/package/582), where the isotope data are restricted and are available from the corresponding author upon request.

## Supplement link

Supplementary information and supplementary videos are available for this study.

## Author contributions

Jung and Soulsby conceptualised the modelling approach and methodology. Jung conducted model set-up, calibration, and validation, with contributions from Soulsby, Birkel, Wu, and the result of EcoHydroPlot was interpreted together. Jung visualized the results and prepared the manuscript which was discussed and contributed by all co-authors. Yordanova reviewed the machine-learning methodology and interpretation. The data were collected from the earlier work led by Tetzlaff and Soulsby. Tetzlaff acquired the main funding.

## Competing interests

The authors declare that they have no conflict of interest.

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


**Acknowledgements**

Jung and Tetzlaff were funded by the project Iso-Scale funded via the Leibniz Excellence fund (SAW K444/2022). Tetzlaff also acknowledges funding from the Wasserressourcenpreis 2024 of the Rüdiger Kurt Bode Stiftung. Tetzlaff, Yordanova, Wu and Soulsby are funded through the DFG project WETSCAPES2.0 (DFG TRR410/1 2025). Birkel and Tetzlaff also received funding from the DFG for "Ecohydrological feedbacks of tropical soil-plant-atmosphere Connectivity (ECCO, 534049325).

**Financial support**

The study was funded through the project ISO-SCALE funded via the Leibniz Collaborative Excellence (SAW K444/2022).

**References**

Aas, K., Jullum, M., and Løland, A.: Explaining individual predictions when features are dependent: More accurate approximations to Shapley values, Artif. Intell., 298, 103502, https://doi.org/10.1016/j.artint.2021.103502, 2021.

Acuña Espinoza, E., Loritz, R., Álvarez Chaves, M., Bäuerle, N., and Ehret, U.: To bucket or not to bucket? Analyzing the performance and interpretability of hybrid hydrological models with dynamic parameterization, Hydrol. Earth Syst. Sci., 28, 2705–2719, https://doi.org/10.5194/hess-28-2705-2024, 2024.

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration: guidelines for computing crop water requirements, FAO Irrigation and Drainage Paper 56, FAO, Rome, Italy, 1998.

Barbeta, A., Jones, S. P., Clavé, L., Wingate, L., Gimeno, T. E., Fréjaville, B., Wohl, S., and Ogée, J.: Unexplained hydrogen isotope offsets complicate the identification and quantification of tree water sources in a riparian forest, Hydrol. Earth Syst. Sci., 23, 2129–2146, https://doi.org/10.5194/hess-23-2129-2019, 2019.

Barbeta, A., Gimeno, T. E., Clavé, L., Fréjaville, B., Jones, S. P., Delvigne, C., Wingate, L., and Ogée, J.: An explanation for the isotopic offset between soil and stem water in a temperate tree species, New Phytol., 227, 766–779, https://doi.org/10.1111/nph.16564, 2020.

Berliner Wasserbetriebe: Wasserwerk Friedrichshagen, https://www.bwb.de/de/wasserwerk-friedrichshagen.php (last access: 20 July 2026), 2026.

Beven, K.: Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication, Hydrolog. Sci. J., 61, 1652–1665, https://doi.org/10.1080/02626667.2015.1031761, 2016.

Beven, K. J.: On hypothesis testing in hydrology: Why falsification of models is still a really good idea, WIREs Water, 5, https://doi.org/10.1002/wat2.1278, 2018.

Beven, K.: Towards a methodology for testing models as hypotheses in the inexact sciences, P. Roy. Soc. A-Math. Phy., 475, 20180862, https://doi.org/10.1098/rspa.2018.0862, 2019.

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


Beven, K.: Deep learning, hydrological processes and the uniqueness of place, Hydrol. Process., 34, 3608–3613, https://doi.org/10.1002/hyp.13805, 2020.

Beven, K. and Westerberg, I.: On red herrings and real herrings: disinformation and information in hydrological inference, Hydrol. Process., 25, 1676–1680, https://doi.org/10.1002/hyp.7963, 2011.

Beyer, M., Kühnhammer, K., and Dubbert, M.: In situ measurements of soil and plant water isotopes: A review of approaches, practical considerations and a vision for the future, Hydrol. Earth Syst. Sci., 24, 4413–4440, https://doi.org/10.5194/hess-24-4413-2020, 2020.

Birkel, C., Paroli, R., Spezia, L., Dunn, S. M., Tetzlaff, D., and Soulsby, C.: A new approach to simulating stream isotope dynamics using Markov switching autoregressive models, Adv. Water Resour., 46, 20–30, https://doi.org/10.1016/j.advwatres.2012.05.013, 2012.

Birkel, C., Soulsby, C., and Tetzlaff, D.: Developing a consistent process-based conceptualization of catchment functioning using measurements of internal state variables, Water Resour. Res., 50, 3481–3501, https://doi.org/10.1002/2013WR014925, 2014.

Birkel, C., Tetzlaff, D., Ring, A.-M., and Soulsby, C.: Does high resolution in situ xylem and atmospheric vapor isotope data help improve modeled estimates of ecohydrological partitioning? Agricultural and Forest Meteorology, 365, 110467, https://doi.org/10.1016/j.agrformet.2025.110467, 2025.

Bolibar, J., Sapienza, F., Maussion, F., Lguensat, R., Wouters, B., and Pérez, F.: Universal differential equations for glacier ice flow modelling, Geosci. Model Dev., 16, 6671–6687, https://doi.org/10.5194/gmd-16-6671-2023, 2023.

Burgess, S. S. O., Adams, M. A., Turner, N. C., Beverly, C. R., Ong, C. K., Khan, A. A. H., and Bleby, T. M.: An improved heat pulse method to measure low and reverse rates of sap flow in woody plants†, Tree Physiol., 21, 589–598, https://doi.org/10.1093/treephys/21.9.589, 2001.

Clark, S. R., Fu, G., and Janardhanan, S.: Explainable AI for Interpreting Spatiotemporal Groundwater Predictions, Water Resour. Res., 61, e2025WR041303, https://doi.org/10.1029/2025WR041303, 2025.

Craig, H., Gordon, L. I., and Horibe, Y.: Isotopic exchange effects in the evaporation of water: 1. Low-temperature experimental results, J. Geophys. Res., 68, 5079–5087, https://doi.org/10.1029/JZ068i017p05079, 1963.

Dunn, S. M., Freer, J., Weiler, M., Kirkby, M. J., Seibert, J., Quinn, P. F., Lischeid, G., Tetzlaff, D., and Soulsby, C.: Conceptualization in catchment modelling: simply learning? Hydrological Processes, 22(13), 2389–2393, https://doi.org/10.1002/hyp.7070, 2008.

Durrant, T., de Rigo, D., and Caudullo, G.: Salix alba in Europe: distribution, habitat, usage and threats, in: European Atlas of Forest Tree Species, edited by: San-Miguel-Ayanz, J., de Rigo, D., Caudullo, G., Houston Durrant, T., and Mauri, A., Publication Office of the European Union, Luxembourg, 2016.

DWD: Climate Data Center (CDC) of the German Weather Service, https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/kl/ (last access: 20 July 2026), 2024.

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


Evaristo, J., Wright, C., Bauser, H. H., Knighton, J., Johnson, D. M., and Kim, M.: Tracer Labelling and Transit Time Modelling in Soil–Plant Systems: Perspectives and a Call for Broader Dialogue in Ecohydrology, Ecohydrology, 19, e70182, https://doi.org/10.1002/eco.70182, 2026.

Feigl, M., Roesky, B., Herrnegger, M., Schulz, K., and Hayashi, M.: Learning from mistakes—Assessing the performance and uncertainty in process-based models, Hydrol. Process., 36, https://doi.org/10.1002/hyp.14515, 2022.

Gessler, A., Bächli, L., Rouholahnejad Freund, E., Treydte, K., Schaub, M., Haeni, M., Weiler, M., Seeger, S., Marshall, J., Hug, C., Zweifel, R., Hagedorn, F., Rigling, A., Saurer, M., and Meusburger, K.: Drought reduces water uptake in beech from the drying topsoil, but no compensatory uptake occurs from deeper soil layers, New Phytol., 233, 194–206, https://doi.org/10.1111/nph.17767, 2022.

Gillefalk, M., Tetzlaff, D., Hinkelmann, R., Kuhlemann, L.-M., Smith, A., Meier, F., Maneta, M. P., and Soulsby, C.: Quantifying the effects of urban green space on water partitioning and ages using an isotope-based ecohydrological model, Hydrol. Earth Syst. Sci., 25, 3635–3652, https://doi.org/10.5194/hess-25-3635-2021, 2021.

Gillefalk, M., Tetzlaff, D., Marx, C., Smith, A., Meier, F., Hinkelmann, R., and Soulsby, C.: Estimates of water partitioning in complex urban landscapes with isotope-aided ecohydrological modelling, Hydrol. Process., 36, e14532, https://doi.org/10.1002/hyp.14532, 2022.

Granda, E., Resco de Dios, V., and Castro-Díez, P.: Invasive tree species benefit from ecohydrological niche segregation and deeper soil water uptake in a Mediterranean riparian forest, Agr. Forest Meteorol., 363, 110433, https://doi.org/10.1016/j.agrformet.2025.110433, 2025.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling, J. Hydrol., 377, 80–91, https://doi.org/10.1016/j.jhydrol.2009.08.003, 2009.

Herman, J. D., Reed, P. M., and Wagener, T.: Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior, Water Resour. Res., 49, 1400–1414, https://doi.org/10.1002/wrcr.20124, 2013.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Comput., 9, 1735–1780, https://doi.org/10.1162/neco.1997.9.8.1735, 1997.

Husic, A., Hammond, J., Price, A. N., and Roundy, J. K.: Interrogating process deficiencies in large-scale hydrologic models with interpretable machine learning, Hydrol. Earth Syst. Sci., 29, 4457–4472, https://doi.org/10.5194/hess-29-4457-2025, 2025.

Huynh, N. N. T., Garambois, P.-A., Colleoni, F., and Monnier, J.: A hybrid physics–AI approach using universal differential equations with state-dependent neural networks for learnable, regionalizable, spatially distributed hydrological modeling, Geosci. Model Dev., 19, 1055–1074, https://doi.org/10.5194/gmd-19-1055-2026, 2026.

Höge, M., Scheidegger, A., Baity-Jesi, M., Albert, C., and Fenicia, F.: Improving hydrologic models for predictions and process understanding using neural ODEs, Hydrol. Earth Syst. Sci., 26, 5085–5102, https://doi.org/10.5194/hess-26-5085-2022, 2022.

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


Ismail, A. A., Gunady, M., Pessoa, L., Bravo, H. C., and Feizi, S.: Input-cell attention reduces vanishing saliency of recurrent neural networks, arXiv [preprint], arXiv:1910.12370, 27 Oct 2019.

Jiang, S., Sweet, L., Blougouras, G., Brenning, A., Li, W., Reichstein, M., Denzler, J., Shangguan, W., Yu, G., Huang, F., and Zscheischler, J.: How Interpretable Machine Learning Can Benefit Process Understanding in the Geosciences, Earth's Future, 12, e2024EF004540, https://doi.org/10.1029/2024EF004540, 2024.

Jung, H., Saynisch-Wagner, J., and Schulz, S.: Can eXplainable AI Offer a New Perspective for Groundwater Recharge Estimation?—Global-Scale Modeling Using Neural Network, Water Resour. Res., 60, e2023WR036360, https://doi.org/10.1029/2023WR036360, 2024.

Jung, H., Tetzlaff, D., Birkel, C., and Soulsby, C.: Recent Developments and Emerging Challenges in Tracer-Aided Modeling, WIREs Water, 12, e70015, https://doi.org/10.1002/wat2.70015, 2025.

Jung, H., Tetzlaff, D., Yordanova, K., and Soulsby, C.: Development of a new generic AI model for spatio-temporal prediction of soil moisture and soil water isotopes, Water Resour. Res., 62, https://doi.org/10.1029/2025WR041903, 2026.

Kabala, J. P., Massari, C., Niccoli, F., Natali, M., Avanzi, F., and Battipaglia, G.: Reconstruction of the dynamics of sap-flow timeseries of a beech forest using a machine learning approach, Agr. Forest Meteorol., 362, 110379, https://doi.org/10.1016/j.agrformet.2024.110379, 2025.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42, https://doi.org/10.1029/2005WR004362, 2006.

Kirchner, J. W.: Characterizing nonlinear, nonstationary, and heterogeneous hydrologic behavior using ensemble rainfall-runoff analysis (ERRA): Proof of concept, Hydrol. Earth Syst. Sci., 28, 4427–4454, https://doi.org/10.5194/hess-28-4427-2024, 2024.

Knighton, J., Kuppel, S., Smith, A., Soulsby, C., Sprenger, M., and Tetzlaff, D.: Using isotopes to incorporate tree water storage and mixing dynamics into a distributed ecohydrologic modelling framework, Ecohydrology, 13, e2201, https://doi.org/10.1002/eco.2201, 2020.

Koppa, A., Rains, D., Hulsman, P., Poyatos, R., and Miralles, D. G.: A deep learning-based hybrid model of global terrestrial evaporation, Nat. Commun., 13, 1912, https://doi.org/10.1038/s41467-022-29543-7, 2022.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22, 6005–6022, https://doi.org/10.5194/hess-22-6005-2018, 2018.

Kühnhammer, K., van Haren, J., Kübert, A., Bailey, K., Dubbert, M., Hu, J., Ladd, S. N., Meredith, L. K., Werner, C., and Beyer, M.: Deep roots mitigate drought impacts on tropical trees despite limited quantitative contribution to transpiration, Sci. Total Environ., 893, 164763, https://doi.org/10.1016/j.scitotenv.2023.164763, 2023.

Lakshminarayanan, B., Pritzel, A., and Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles, arXiv [preprint], arXiv:1612.01474v3, 4 Nov 2017.

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


Landgraf, J., Tetzlaff, D., Dubbert, M., Dubbert, D., Smith, A., and Soulsby, C.: Xylem water in riparian willow trees (Salix alba) reveals shallow sources of root water uptake by in situ monitoring of stable water isotopes, Hydrol. Earth Syst. Sci., 26, 2073–2092, https://doi.org/10.5194/hess-26-2073-2022, 2022.

Landgraf, J., Tetzlaff, D., Birkel, C., Stevenson, J. L., and Soulsby, C.: Assessing land use effects on ecohydrological partitioning in the critical zone through isotope-aided modelling, Earth Surf. Proc. Land., 48, 3199–3219, https://doi.org/10.1002/esp.5691, 2023.

Lundberg, S. M. and Lee, S.-I.: A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems 30, edited by: Guyon, I., Von Luxburg, U., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., Curran Associates, Inc., 4765–4774, 2017.

Marshall, J. D., Cuntz, M., Beyer, M., Dubbert, M., and Kuehnhammer, K.: Borehole Equilibration: Testing a New Method to Monitor the Isotopic Composition of Tree Xylem Water in situ, Front. Plant Sci., 11, 1–14, https://doi.org/10.3389/fpls.2020.00358, 2020.

Martin, C. A.: LAI: calculate indirect Leaf Area Index (LAI) from images, R package version 0.0.0.9004, Zenodo [code], https://doi.org/10.5281/zenodo.34690, 2015.

Marttila, H., Dudley, B. D., Graham, S., and Srinivasan, M. S.: Does transpiration from invasive stream side willows dominate low-flow conditions? An investigation using hydrometric and isotopic methods in a headwater catchment, Ecohydrology, 11, e1930, https://doi.org/10.1002/eco.1930, 2018.

McMillan, H. K., Westerberg, I. K., and Krueger, T.: Hydrological data uncertainty and its implications, WIREs Water, 5, e1319, https://doi.org/10.1002/wat2.1319, 2018.

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(3), e2020WR028091, https://doi.org/10.1029/2020WR028091, 2021.

Orlowski, N., Rinderer, M., Dubbert, M., Ceperley, N., Hrachowitz, M., Gessler, A., Rothfuss, Y., Sprenger, M., Heidbüchel, I., Kübert, A., Beyer, M., Zuecco, G., and McCarter, C.: Challenges in studying water fluxes within the soil-plant-atmosphere continuum: A tracer-based perspective on pathways to progress, Sci. Total Environ., 881, 163510, https://doi.org/10.1016/j.scitotenv.2023.163510, 2023.

Pfannerstill, M., Guse, B., Reusser, D., and Fohrer, N.: Process verification of a hydrological model using a temporal parameter sensitivity analysis, Hydrol. Earth Syst. Sci., 19, 4365–4376, https://doi.org/10.5194/hess-19-4365-2015, 2015.

Razavi, S., Jakeman, A., Saltelli, A., Prieur, C., Iooss, B., Borgonovo, E., Plischke, E., Lo Piano, S., Iwanaga, T., Becker, W., Tarantola, S., Guillaume, J. H. A., Jakeman, J., Gupta, H., Melillo, N., Rabitti, G., Chabridon, V., Duan, Q., Sun, X., et al. Maier, H. R.: The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support, Environ. Modell. Softw., 137, 104954, https://doi.org/10.1016/j.envsoft.2020.104954, 2021.

Reusser, D. E. and Zehe, E.: Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity, Water Resour. Res., 47, https://doi.org/10.1029/2010WR009946, 2011.

https://doi.org/10.5194/egusphere-2026-4371
Preprint. Discussion started: 4 September 2026


Ring, A.-M., Tetzlaff, D., Dubbert, M., Dubbert, D., and Soulsby, C.: High-resolution in situ stable isotope measurements reveal contrasting atmospheric vapour dynamics above different urban vegetation, Hydrol. Process., 37, e14989, https://doi.org/10.1002/hyp.14989, 2023.

Ring, A.-M., Tetzlaff, D., Birkel, C., and Soulsby, C.: Sub-daily stable water isotope dynamics of urban tree xylem water and ambient vapor, Hydrol. Earth Syst. Sci., 29, 6663–6683, https://doi.org/10.5194/hess-29-6663-2025, 2025.

Saxton, K. E. and Rawls, W. J.: Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions, Soil Sci. Soc. Am. J., 70, 1569–1578, https://doi.org/10.2136/sssaj2005.0117, 2006.

Seeger, S. and Weiler, M.: Temporal dynamics of tree xylem water isotopes: in situ monitoring and modeling, Biogeosciences, 18, 4603–4627, https://doi.org/10.5194/bg-18-4603-2021, 2021.

Smith, A., Tetzlaff, D., Landgraf, J., Dubbert, M., and Soulsby, C.: Modelling temporal variability of in situ soil water and vegetation isotopes reveals ecohydrological couplings in a riparian willow plot, Biogeosciences, 19, 2465–2485, https://doi.org/10.5194/bg-19-2465-2022, 2022.

Sprenger, M., Seeger, S., Berkelhammer, M., Bogie, N. A., Hess, R. J., Brown, W. S., Kuppel, S., and Knighton, J.: Opportunistic Short-Term Water Uptake Dynamics by Subalpine Trees Observed via In Situ Water Isotope Measurements, Water Resour. Res., 61, e2024WR039171, https://doi.org/10.1029/2024WR039171, 2025.

Stevenson, J. L., Birkel, C., Comte, J. C., Tetzlaff, D., Marx, C., Neill, A., Maneta, M., Boll, J., and Soulsby, C.: Quantifying heterogeneity in ecohydrological partitioning in urban green spaces through the integration of empirical and modelling approaches, Environ. Monit. Assess., 195, https://doi.org/10.1007/s10661-023-11055-6, 2023.

Tetzlaff, D., Buttle, J., Carey, S. K., Kohn, M. J., Laudon, H., McNamara, J. P., Smith, A., Sprenger, M., and Soulsby, C.: Stable isotopes of water reveal differences in plant – soil water relationships across northern environments, Hydrol. Process., 35, e14023, https://doi.org/10.1002/hyp.14023, 2021.

Vrugt, J. A., ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., and Higdon, D.: Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, Int. J. Nonlin. Sci. Num., 10, 273–290, https://doi.org/10.1515/IJNSNS.2009.10.3.273, 2009.

Wang, J. and Renninger, H. J.: SapFlower: an automated tool for sap flow data preprocessing, gap-filling, and analysis using deep learning, New Phytol., 246, 2324–2345, https://doi.org/10.1111/nph.70107, 2025.