https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Hydrology and Earth System Sciences
Discussions

Multi-variable process-based calibration of a behavioural hydrological model

Moritz M. Heuer[1], Hadysa Mohajerani[1], Markus C. Casper[1]

[1]Department of Physical Geography, University Trier, Trier, 54296, Germany

Correspondence to: Moritz M. Heuer (heuer@uni-trier.de)

Abstract. Behavioural hydrological modelling aims not only at predicting the discharge of an area within a model, but also at understanding and correctly depicting the underlying hydrological processes. Here, we present a new approach for the calibration and evaluation of water balance models, exemplarily applied to the Riveris catchment in Rhineland-Palatinate, Germany. For our approach, we used the behavioural model WaSiM. The first calibration step is the adjustment of the evapotranspiration (ETa) parameters based on MODIS evaporation data. This aims at providing correct evaporation behaviour of the model and at closing the water balance at the gauging station. In a second step, geometry and transmissivity of the aquifer are determined using the Characteristic Delay Curve (CDC). The portion of groundwater recharge was calibrated using the Delayed Flow Index (DFI). In a third step, inappropriate pedotransfer functions (PTFs) could be filtered out by comparing dominant runoff process patterns under a synthetic precipitation event with a soil hydrological reference map, Then, the discharge peaks were adjusted based on so-called signature indices. This ensured a correct depiction of high-flow volume in the model. Finally, the overall model performance was determined using signature indices and efficiency measures. The results show a very good model fit with values for the NSE of 0.88 and 0.9 for the KGE in the calibration period and an NSE of 0.81 and a KGE of 0.89 for the validation period. Simultaneously, our calibration approach ensured a correct depiction of the underlying processes (groundwater behaviour, runoff patterns). This means that our calibration approach allows selecting a behaviourally faithful one from many possible parameterisation variants.

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

1 Introduction

Traditionally, hydrological models are calibrated mainly on the basis of gauging data, with the aim of accurately predicting discharge. However, the underlying processes like groundwater behaviour or runoff generation processes are often neglected in this approach (Schaake et al., 1996; Xiong and Guo, 1999; Casper et al., 2019; Kheimi and Abdelaziz, 2022). Relying solely on statistical evaluations of overall runoff performance may not adequately capture model performance for high and low flow extremes (Westerberg et al., 2011; Althoff and Rodrigues, 2021). This means that although these models are then suitable for predicting runoff, they do not allow investigations of the underlying processes. This emphasises the necessity for physically-based models to be not just theoretically accurate but also empirically validated against the dynamics of natural hydrological systems (K. Beven, 2002).

Behavioural modelling addresses this issue by considering not only the discharge, but also the discharge-forming processes when calibrating the model. This means that methodological approaches must be incorporated in the calibration process that allow to align different simulated processes with the actual catchment responses (Vansteenkiste et al., 2014). For instance, Ferket et al. (2010), H. Zhang et al. (2011), and Meresa et al. (2023) implemented performance measures on the sub-surface flow (e.g., interflow and deep percolation to groundwater) components of runoff discharges. Casper et al. (2023) improved the reproduction of spatial and temporal evapotranspiration (ETa) patterns by applying a MODIS-based calibration approach to vegetation-related ETa parameters.

Groundwater's delayed response to precipitation and its role in baseflow during dry periods are critical for accurate water resource management (K. J. Beven and Alcock, 2012). The duration from groundwater recharge to baseflow discharge is influenced by topography, geology, vegetation, land use, and climate (Barthel, 2006; Götzinger et al., 2008). Baseflow fed streamflow is directly related to groundwater storage and its interaction with streams, which can vary heavily across catchments (Barkwith et al., 2015). This complexity necessitates incorporating groundwater flow into hydrological models to accurately simulate discharge under diverse hydrological conditions (Knisel Jr, 1963; Smakhtin, 2001; McNamara et al., 2011; Barkwith et al., 2015; Stoelzle et al., 2015). The behaviour of the groundwater component in water balance models must therefore be

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

considered when calibrating a model. This makes it necessary to implement a way of evaluating the model's ability to correctly represent groundwater behaviour and its temporal contribution to the overall discharge.

Pedotransfer functions (PTF) allow the estimation of soil hydraulic properties from widely available soil data like grain size, density, or depth. Simulation outcomes of different PTFs highly differ in runoff components (surface runoff, interflow and deep percolation) and evapotranspiration (ETa) rates in space and time (Refsgaard, 2001; Stisen et al., 2008; Koch et al., 2016, Koch et al., 2017; Casper et al., 2019; Mohajerani et al., 2021). Therefore, the correct choice of a PTF for soil parameterisation is crucial.

Liu et al. (2022) demonstrated that the incorporation of remote sensing data like ETa data or terrestrial water storage change (TWSC) for hydrologic model calibration can improve the depiction of those processes. It was also shown that combinations of different evaluation criteria increase the model accuracy regarding the underlying processes (Nesru et al., 2020; Nolte et al., 2021; Yáñez-Morroni et al., 2024). However, there have been no calibration approaches that include the evaluation of ETa, groundwater behaviour, runoff generation processes, and the overall discharge in one consecutive calibration process.

To address the above-mentioned challenges, our research introduces a new approach for the parameterisation and calibration of water balance models. This approach comprises the calibration of evapotranspiration patterns of different land uses based on remote sensing ETa data, ensuring correct ETa patterns and a closed water balance. In addition, the ground water behaviour is assessed by deriving the long term baseflow from the measured discharge of the catchment. This allows for calibration of the groundwater behaviour (storage, recession) as well as the groundwater recharge (deep percolation) within the model. Furthermore, the influence of the soil parameterisation on the spatial pattern of runoff generation is assessed. This ensures a correct depiction of runoff patterns over the catchment area. Lastly, high discharge volume is calibrated by deriving information about the catchment discharge characteristics from the flow duration curve. By incorporating the calibration and evaluation of these different model aspects, we aim at reaching a model calibration that correctly simulates the discharge as

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Hydrology and Earth System Sciences
Discussions

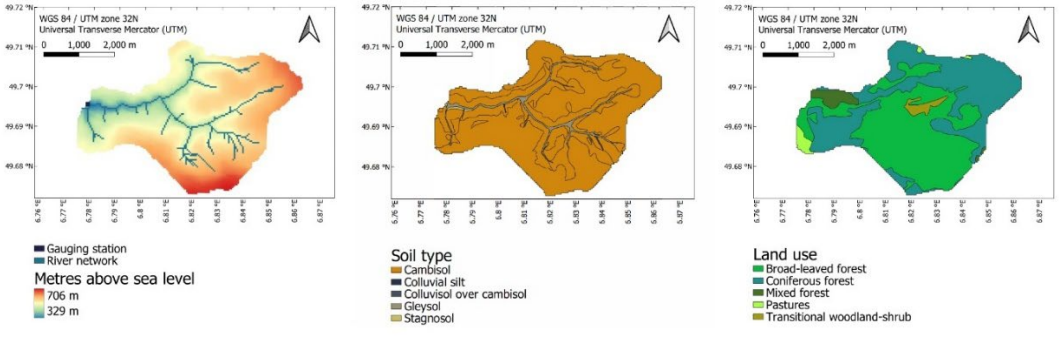
well as the underlying hydrological processes, leading to a behavioural model in the sense of Gupta et al. (2006), which simulates a correct hydrograph at the catchment outlet for the right reasons.

Our study is aimed at (i) systematically describing the calibration steps in connection to the structure of the hydrological model and its parameterisations; (ii) exploring the implications of process specific parameters on model behaviour; (iii) demonstrating how our novel approach of model calibration can lead to a more accurate simulation of hydrological processes in space and time, which we define as a behavioural model.

2 Methodology and Material

2.1 Study area

The Riverisbach catchment (Fig. 1) was selected as the study area for the demonstration of the parameterisation approach. This was due to the good availability of data on soil, land use, evaporation patterns and discharge, which is necessary for the evaluation of the model calibration. The catchment basin is located south-east of Trier in Rhineland-Palatinate, Germany. It covers an area of around 22 km² and ranges from 329 m above sea level in the north-west to 705 m above sea level in the south, resulting in a height amplitude of 376 metres and an average slope gradient of 4.49 %. The used gauging station 'Riveristalsperre' is located in the west of the catchment at 49° 41.771' N, 6° 46.741' E. The mean annual precipitation amounts to 918 mm per year.



(a) (b) (c)

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

90

Figure 1: Topography, soil types and land cover types within the Riverisbach catchment as it's used within our WaSiM based model.

The area is located above bedrock from the Drohntal strata, i.e. quartz sandstone and quartzitic sandstone with intercalations of claystone and siltstone. The soils are dominated by Cambisols, while Gleysols and Stagnosols can be found along the 95 watercourses in the floodplain area. The majority of the Riverisbach catchment area is covered by forest. Conifers are dominating in the north-east and west and deciduous trees in the centre and south. In the west there are also small areas of grassland and mixed woodland.

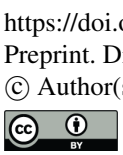
2.2 Data sources

Soil type information was taken from the 'Bodenflächendaten im Maßstab 1:50.000 (BFD50)' (Landesamt für Geologie und 100 Bergbau, 2021). The data for the landuse is derived from European Union's Copernicus Land Monitoring Service information (European Environment Agency, Copenhagen, 2018). INTERMET data (Gerlach, 2006) was used as time series for meteorological data. Wind data was taken from the Agrarmeteorologie Rheinland-Pfalz (2024). Values for the saturated hydraulic conductivity k_{sat} were taken from Ad-hoc-AG Boden (2006).

2.3 Model setup and parameterisation

105 The WaSiM model (Schulla, 1997) version 10.08.02 (Schulla, 2024a) was selected for the simulation and development of the parameterisation approach. It is a deterministic, hydrological catchment model that is suitable for the simulation of both small (< 1 km²) and very large areas (> 10000 km²). It also simulates the underlying processes that lead to discharge generation. This includes the ETa, groundwater flow, surface runoff and interflow, as well as groundwater recharge. It is therefore suitable for a behavioural modelling approach that includes the calibration of these processes. A schematic depiction of the WaSiM model 110 is shown in Fig. 2. The soil is represented in the model as a rectangular grid of 1-dimensional columns. Each of these columns is divided into soil horizons of different thicknesses, which in turn are subdivided into several layers. At the bottom, a section

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025



Hydrology and Earth System Sciences
Discussions

of aquifer layers is included. Surface runoff, interflow and groundwater-contributing deep percolation can be generated. Surface runoff and interflow of each subcatchment are delayed through a single linear reservoir (SLR) each.

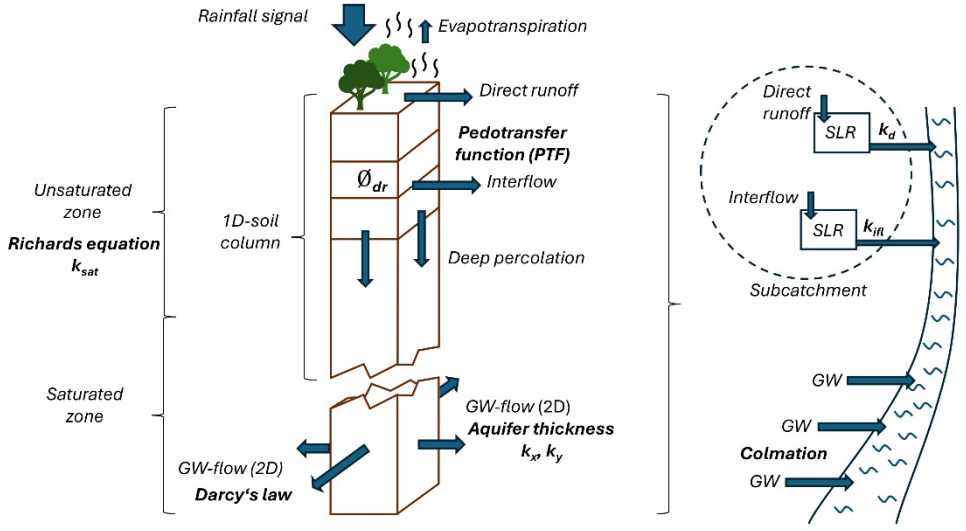


Figure 2: Conceptual diagram of the WaSiM model's structure. Bold text symbolises certain parameters or functions that are used to derive parameter values for the model parameterisation. Blue arrows indicate water fluxes within the model.

Spatially resolved data is differentiated within the model using grid structures. This also enables the model to interpolate climatic input data over the catchment area. The model uses the Richards equation (Richards, 1931) to calculate the water transport within the unsaturated soil zone. It is defined as:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[k(\Psi_m)\left(\frac{\partial \Psi_m}{\partial t}\right)\right] \quad (1)$$

where z is the depth, θ is the water content [vol.-%], t is the time [d], and Ψ_m is the hydraulic conductivity in dependence of the matrix potential $[cm \cdot d^{-1}]$. The van Genuchten parameters (Van Genuchten, 1980) are used to calculate the soil physical properties. The Penman-Monteith (Monteith, 1965) method is used to calculate evapotranspiration. A two-dimensional approach based on Darcy's law (Darcy, 1856) is used to calculate groundwater flow. It is defined as:

$$q = k \cdot \frac{\partial \Psi}{\partial z} \quad (2)$$

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

where q is the volume flow $[m^3 \cdot s^{-1}]$, k is the hydraulic conductivity $[m \cdot s^{-1}]$, and $[\frac{\partial \Psi}{\partial z}]$ is the hydraulic gradient [-].

For the model parameterisation, a spatial resolution of 40 m and a temporal resolution of 1 h were chosen. The 40 m spatial resolution showed to be the best trade-off between spatial resolution precision and model computation time. This also applies to the chosen temporal resolution of 1 h. INTERMET data (Gerlach, 2006) was used as input time series for meteorological data (temperature, precipitation, radiation, humidity). The data ranges from 01.01.2010 to 31.12.2020. Wind data was taken from the Agrarmeteorologie Rheinland-Pfalz (2024) for the stations Avelsbach [49.754° N, 6.693° E], Hermeskeil [49.655° N, 6.933° E] and Konz [49.687° N, 6.572° E]. Missing entries for periods of a few hours were manually resolved.

Following, the preprocessing tool of WaSiM, TANALYS (Schulla, 2024b), was used to calculate the required spatial information grids based on the digital elevation model. These spatial information grids include grids for the slope, exposition, subcatchments, river network, river width and depth, colmation, as well as lateral aquifer conductivities (k_x and k_y). A value of 50 was selected as the threshold for the river network. The threshold value describes from how many cells of runoff must be combined to form a water body cell in the model. Higher values for this threshold therefore result in a coarser river network, while lower values result in finer river networks. The resulting network, based on the threshold value of 50 cells, showed the best fit with the water body of the catchment. Based on the soil types and land use information, profiles of the individual soils were created. These profiles contained data on thickness, soil type, depth, bulk density, carbonate content, humus content and dry bulk density of the individual horizons.

Simulated soil hydraulic properties include hydraulic conductivity, soil water content at field capacity, and saturated water content. These are described using van Genuchten parameters and the saturated hydraulic conductivity k_{sat}. We used 12 different pedotransfer functions (PTFs) to calculate these parameter values. Pedotransfer functions can derive the required values for the van Genuchten parameters from measured soil data based on certain regression curves. Combinations of used pedotransfer functions are shown in Table 1. For the first seven PTF combinations, values for the saturated hydraulic conductivity k_{sat} were taken from the KA5 Ad-hoc-AG Boden (2006). For PTF combinations 8 to 12, the values were calculated by the respective PTF's equation for k_{sat}. The chosen PTFs mainly differ in their underlying data, soil sample size,

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

and considered soil parameters for the resulting predictive equations. A comprehensive analysis of the effects of PTFs 1 to 11 on hydrological soil properties has been provided by Mohajerani et al. (2021). Each soil was then initialised with 27 layers, including a groundwater layer, and their respective hydraulic properties derived by the PTFs.

Table 1: PTF combinations used to estimate the van Genuchten parameters and the saturated hydraulic conductivities.

PTF Combination	Van Genuchten Parameters	Soil Hydraulic Conductivity k_{sat}
1	Wösten et al. (1999)	Ad-hoc-AG Boden (2006) KA5
2	Renger et al. (2008)	Ad-hoc-AG Boden (2006) KA5
3	Weynants et al. (2009)	Ad-hoc-AG Boden (2006) KA5
4	Zacharias and Wessolek (2007)	Ad-hoc-AG Boden (2006) KA5
5	Teepe et al. (2003)	Ad-hoc-AG Boden (2006) KA5
6	Y. Zhang and Schaap (2017): Rosetta H2w	Ad-hoc-AG Boden (2006) KA5
7	Y. Zhang and Schaap (2017): Rosetta H3w	Ad-hoc-AG Boden (2006) KA5
8	Wösten et al. (1999)	Wösten et al. (1999)
9	Renger et al. (2008)	Renger et al. (2008)
10	Y. Zhang and Schaap (2017): Rosetta H2w	Y. Zhang and Schaap (2017): Rosetta H2w
11	Y. Zhang and Schaap (2017): Rosetta H3w	Y. Zhang and Schaap (2017): Rosetta H3w
12	Szabó et al. (2021): euptfv2	Szabó et al. (2021): euptfv2

2.4 Calibration scheme

The calibration approach and its individual steps are described and summarised in Table 2. In Fig. 3, the individual calibration steps are depicted schematically in connection to the corresponding hydrological processes conceptualised in the WaSiM model structure. In step 1, evapotranspiration parameters are calibrated using MODIS evaporation patterns. This step ensures a closed water balance as well as correct ETa patterns across different land uses. Step 2 adjusts the geometry and transmissivity of the groundwater model. In step 3, the rate of groundwater recharge via the amount of water entering the aquifer is calibrated. Both steps aim at correctly depicting the groundwater model behaviour with its contribution to total discharge. In step 4, the different PTFs are evaluated by comparing the patterns of dominant runoff processes under a synthetic heavy rainfall event.

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

This step allows for the identification and exclusion of unsuitable PTFs that generate inaccurate runoff patterns. In step 5, the peaks in the hydrograph, represented as the high flow volume on the flow duration curve, are then adjusted to calibrate the model parts that are directly influenced by precipitation. Finally, in step 6, the model is evaluated in terms of its ability to predict the overall discharge, based on hydrograph efficiency metrics in a split-sample test.

Table 2: Scheme for the calibration and evaluation approach applied in this study.

Step	Description	Aim	Scale	Behaviour
1	Adjustment of ETa (for each landuse)	Close the water balance, match spatial patterns with MODIS	Spatial and temporal pattern match	Mean long-term behaviour
2	Adjusting GW-model (transmissivity)	Calibrated baseflow within the DFI	Temporal match (DFI)	Mean long-term behaviour of GW-submodel
3	Adjusting GW-recharge	Partitioning GW / interflow	GW / interflow	Long-term GW-recharge
4	Checking runoff generation processes	Match runoff processes with reference map (BHK)	Spatial match	Model behaviour test for extreme precipitation event (100 mm)
5	Adjusting high flows	Adjusting signature indices	Match on flow duration curve	Rainfall-fed part of the hydrograph
6	Final model evaluation	Peak flow statistics, split-sample test	Flow duration curve, hydrograph	Consistency at catchment outlet

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Hydrology and Earth System Sciences
Discussions

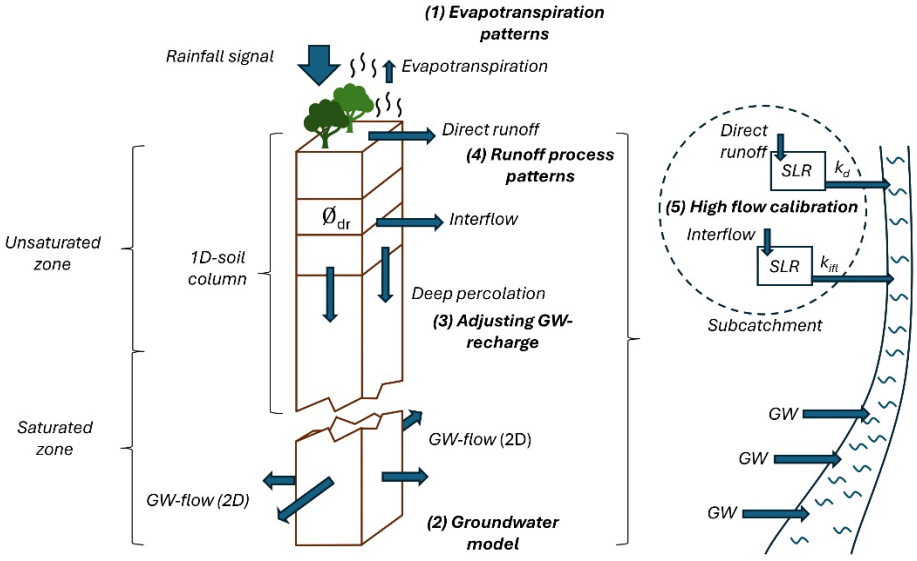


Figure 3: Conceptual diagram of the WaSiM model structure and the steps of the associated calibration approach. Evapotranspiration patterns are calibrated using MODIS evaporation data (1). The groundwater model flow is then calibrated using the transmissivity (2). Groundwater recharge, i.e. the amount of water, is adjusted by calibrating the amount of interflow with the scaling factor d_r (3). Dominant runoff process patterns derived from an extreme synthetic rainfall event are compared with the reference map to filter for matching patterns (4). Calibration of high discharge (peak flows) by adjusting the recession parameters of the direct runoff and interflow single linear reservoirs for each subcatchment (5). The last step, the evaluation of the hydrograph with efficiency metrics (6), is not shown in this concept figure.

2.5 Calibration of ETa patterns (Step 1)

The approach for calibrating the ETa patterns was originally described by Casper et al. (2023). According to this, the evaporation parameters were calibrated using land use-specific MODIS-derived data (MOD16A2) and validated against Landsat-derived ETa data. This calibration step enhances the representation of spatio-temporal ETa dynamics within the model and closes the water balance at the catchment outlet. All ETa related parameters are taken from Casper et al. (2023).

2.6 Calibration of transmissivity (Step 2)

Firstly, the model was calibrated in terms of its ability to reproduce the groundwater behaviour and the associated base flow. For this purpose, simulation runs were carried out with the initial parameterisations. A model run for the period from 1 January

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

2010 to 31 December 2014 served as a preliminary run for model spin-up, while the actual model run was then carried out for the period from 1 January 2010 to 31 December 2020 using the preliminary run as the initial model state.

We then examined the groundwater behaviour of the catchment and the model by applying the delayed flow index (DFI) method of Stoelzle et al. (2020) to the measured gauging data and the simulated hydrograph. For this, the series of discharge values of the hydrograph is divided into non-overlapping sections. These sections span a specific period of block-length n (days) with $1 \leq n \leq 180$. The minimum flow value of each interval is then compared with the ones from adjacent intervals. If a minimum value multiplied by a specific factor $f = 0.9$ is smaller than the adjacent minima, a turning point (TP) is defined at its position. These TPs are then connected and form a delayed-flow hydrograph, which results in a specific hydrograph for each block length n. From this, the delayed-flow index (DFI) is calculated for each block length as the ratio of the sum of the delayed-flow to the sum of the total flow. An example how the applied block lengths result in different hydrographs can be seen in Fig. 4.

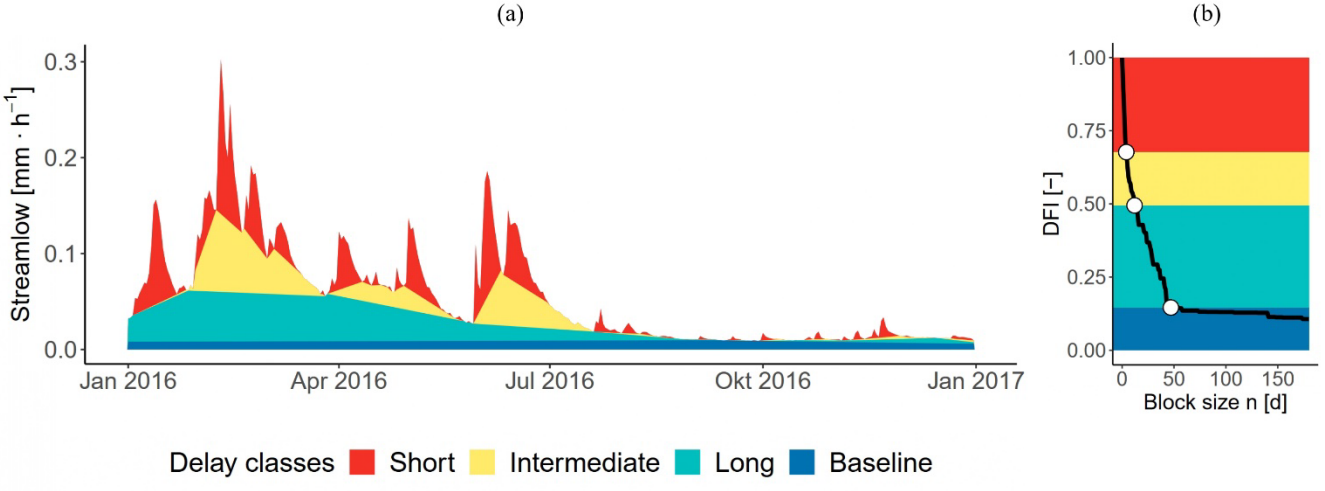


Figure 4: Application of the DFI approach. (a) is the hydrograph separation according to calculated break point values for block lengths. The corresponding characteristic delay curve (CDC) derived from the hydrograph separation over all block lengths of $1 \leq n \leq 180$ are shown in (b).

The DFI analysis was conducted using R (R Core Team, 2023) within RStudio (RStudio Team, 2020). The above-mentioned method was applied to the simulated hydrograph. DFI values for the individual block lengths n were calculated using the

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

function baseflow from the package *lfstat* (Gauster et al., 2022). The resulting DFI values for all block lengths n were then plotted in a diagram, creating a characteristic delay curve (CDC). The *find_bps* function from the R-package segmented (Muggeo et al., 2008) was then used to determine the breakpoints of the curve. Breakpoints are defined as those points of the curve at which a change in the discharge characteristic can be determined (sudden change in slope). For this, $n_{LS} = 4$ linear segments were fitted to the CDC by residual minimisation, resulting in a total of $n_{BP} = 3$ breakpoints along the curve. The area between the last breakpoint ($n = 48$) and $n = 180$ was then considered as the area of the CDC where the aquifer's baseflow is the dominant contribution. This was the area where our groundwater model calibration took place. This procedure was then done for each PTF, resulting in a CDC for each PTF parameterisation.

Calibration was done to fit the slope of the rear area of the CDC. As the slope is determined by the transmissivity of the aquifer, adjustments were made for the model parameters k_x, k_y, colmation, as well as the thickness of the aquifer. This was done until the slopes of the rear ends of the CDC for the simulations were identical with the slope of the CDC for the gauging station. A table with the calibrated model parameters can be found in the Appendix (Table B1).

2.7 Calibration of groundwater recharge (Step 3)

After the groundwater transmissivity was adjusted, the different PTFs showed varying proportions in their CDC curves' rear areas. This indicated that the different PTFs lead to different amounts of water that reached the aquifer. To fit the simulation's CDC curve height to the height of the curve for the measured discharge, the value for the model parameter drainage density (d_r) was adjusted for each PTF independently. This conceptual parameter describes how much of the infiltrating water in the soil passes into the interflow and thus does not reach the aquifer. It therefore controls the amount of water contributing to groundwater recharge. As per Schulla (1997), the parameter d_r is included in the formula for the interflow as

$$q_{ifl} = k_{s(\Theta_m)} \cdot \delta z \cdot d_r \cdot \tan\beta \quad (3)$$

with k_s being the saturated hydraulic conductivity [$m \cdot s^{-1}$], Θ_m being the water content in the actual layer m [-], d_r being the scaling parameter for the interflow to consider anisotropy of $k_{s,horizontal}$, compared to $k_{s,vertical}$, and β being the slope angle with a maximum of $\beta = 45°$.

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Hydrology and Earth System Sciences
Discussions

In this context, higher values of d_r represent soil with stronger lateral drainage capabilities. This usually leads to more interflow and therefore less water that can infiltrate into the aquifer and contribute to groundwater recharge. Regarding the groundwater recharge calibration, higher values for d_r lowered the curve, especially in the rear end. This brought the DFI values into the range of the reference curve (Fig. 5) for PTFs that initially showed higher CDCs in the rear area. For CDCs of PTFs that were lower than the reference CDC of the gauging station, the value for d_r had to be lowered. This reduced interflow and increased the groundwater recharge. A table with the values of d_r for the different PTFs can be found in the Appendix (Table B2).

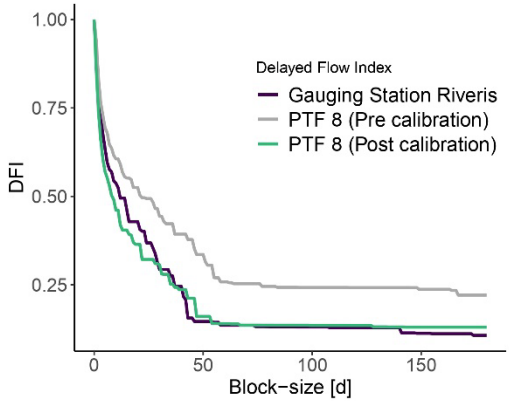


Figure 5: CDCs for the uncalibrated groundwater model and after groundwater model calibration, exemplarily for PTF 8.

2.8 Evaluation of dominant runoff process patterns (Step 4)

In the next step, the different PTFs were compared regarding their ability to accurately depict the surface runoff processes in the catchment area under a heavy precipitation event. This step served to filter out those PTFs that are not capable of simulating the correct runoff patterns. For this purpose, the approach developed by Mohajerani et al. (2023) for comparing the runoff processes was used and adapted for our calibration scheme.

The soil hydrological map (BHK) of Rhineland-Palatinate from Steinrücken and Behrens (2010) was used as a reference for our comparison. The BHK is a map that depicts which runoff type dominantly appears under a heavy precipitation event. It

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences Discussions

250 divides the runoff into saturated overland flow (SOF), subsurface flow (SSF) and deep percolation (DP). Two finer classifications for SOF and SSF are characterised by different delay times. However, the WaSiM model does not consider the delay but only the runoff type itself. Therefore, we only used the three main groups and not the subgroups for the comparison. We also refrained from subdividing the model processes according to the fractions, as suggested by Mohajerani et al. (2023). This was done because the soil hydrological map categorises the subclasses according to the delay and not to the proportions
255 of runoff processes. A division by fractions therefore wouldn't be fully comparable with a division by delay times (as in the BHK).

The BHK was adjusted to the Riverisbach catchment boundaries and rasterised to a resolution of 40 m x 40 m. This was done to facilitate a direct comparison between simulated runoff processes and the BHK as reference. For the comparison, the model
260 state at the end of 31 December 2014 was used as the initial state of this step's model run. This initial state was then used to carry out a 7-day run-up under controlled climatic conditions ($temperature = 10\,°C$, $radiation = 0\,W \cdot m^{-2}$, $wind\ speed = 0\,m \cdot s^{-1}$, $relative\ humidity = 100\,\%$ and $precipitation = 0\,mm$) for the entire duration. This was done to eliminate influences from melting snow on the runoff analysis during the following main run as well as bringing soil moisture to field capacity. The final state of this preliminary run then served as the initial state for another 7-day model run.
265 During this run, the catchment was irrigated with 100 mm of rain over the first seven hours ($14.286\,mm \cdot h^{-1}$). Over the simulation period of these seven days, the cumulative runoff fractions for each cell of the catchment grid were calculated. From the calculated fractions of runoff per grid cell, maps were created where each grid cell's dominant runoff process was attributed to. This resulted in a dominant runoff process map for each PTF.

270 The simulated runoff process patterns were then compared with the runoff process patterns of the BHK. For this purpose, the comparison approach using the spatial efficiency metric (SPAEF) (Stisen et al., 2017; Demirel et al., 2018), was adapted. The SPAEF is to be understood as a measure of spatial similarity. It is defined as:

$$SPAEF = 1 - \sqrt{(\alpha - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (4)$$

$$\alpha = \rho(A, B) \quad (5)$$

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

$$\beta = \left(\frac{\sigma_A}{\mu_A} \Big/ \frac{\sigma_B}{\mu_B}\right) \tag{6}$$

$$\gamma = \frac{\sum_{j=1}^{n} \min(K_j, L_j)}{\sum_{j=1}^{n} K_j} \tag{7}$$

with α being the Pearson correlation coefficient between the simulated grid (A) and the reference grid (B). β is the fraction of coefficient of variations as an indicator of spatial variability. γ is the percentage of histogram intersection (Demirel et al., 2018). The closer the SPAEF value is to 1, the higher the similarity between the compared patterns. During our analysis, however, we encountered a limitation with the standard SPAEF formula when applied to patterns consisting of only three groups. Specifically, the Pearson correlation coefficient, as a component of the SPAEF, tended to yield lower values if deviations occurred in marginal areas. This occurred even when there was substantial overall agreement. To address this issue, we adapted the SPAEF calculation by substituting the Pearson correlation component. Instead, we used a direct measurement of percentage agreement between the simulation and the reference map grids. This adjustment led to the development of a modified SPAEF formula:

$$SPAEF_{mod} = 1 - \sqrt{(\delta - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \tag{8}$$

$$\delta = \frac{\sum_{j=1}^{n_g} 1}{n_g} \; for \; A_j = B_j \tag{9}$$

where δ is the percentage match of all grid fields between simulated map (A) and reference map (B). It is calculated as the fraction of the amount of identical grid cell pairs between both maps to the number of grid cells in one map (n_g). β and γ remain unchanged. This new equation for $SPAEF_{mod}$ allowed us to correctly analyse the agreement between the simulated runoff patterns and the reference patterns of the hydrological map (BHK). A separate $SPAEF_{mod}$ value was then calculated based on the DRP map for each PTF.

2.9 Calibration of high flow discharge (Step 5)

The discharge peaks of the model were calibrated by adjusting the coefficients of the single linear reservoirs for the direct runoff (k_d) and the interflow (k_{ifl}). The metrics of the signature indices (Casper et al., 2012) were used to evaluate the

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

calibration of the individual linear reservoirs. These indices consider different sections and properties of the flow duration curves (FDCs) of simulated and measured discharge and compare them against each other. This yields a percentage bias for each signature indice parameter. The *BiasRR* describes the percent bias in the mean values. The *BiasFDCmidslope* describes the percent bias in slope of the mid-segment. The *BiasFHV* describes the percent bias in high-segment volumes (upper 2 %). The *BiasFLV* is the difference in the long-term baseflow. The *BiasFMM* depicts the percent bias in mid-range flow levels.

First, the coefficient for the direct runoff single linear reservoir, k_d, was calibrated. A low value of 2 seemed to fit best for all PTFs, as the proportion of direct runoff in the total runoff was low and did not need to be delayed any further. The value of *BiasFHV* was then minimised by adjusting the coefficient for the interflow runoff single linear reservoir, k_{ifl} . This was done to adjust the peaks of the simulated hydrograph to more closely resemble those of the measured hydrograph of the catchment. Higher values for k_{ifl} lead to a stronger delay of the interflow runoff. This results in lower peaks of the discharge.

2.10 Final model evaluation (Step 6)

2.10.1 Characteristic delay curve (CDC) comparison

The CDCs for the different PTFs were compared to determine how well the discharge is simulated in the interflow area. For this purpose, the Manhattan distance between the CDCs between $n = 1$ and $n = 43$ (last breakpoint of the measured data) was calculated according to the following formula:

$$d(A,B) = \sum_{i=1}^{n} |A_i - B_i| \tag{10}$$

where A represents the values of the CDC for the gauging station and B the values for the curve of the simulation.

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

2.10.2 High discharge histogram overlap (HDHO) analysis

In addition, a high discharge histogram overlap (HDHO) analysis was carried out based on the hydrographs. By comparing the histograms of the temporal peak discharge distribution for the simulated and measured hydrograph, the model's capability of simulating the strongest discharge events can be assessed. For this purpose, the maximum discharge value of each year was determined. This was done for each PTFs hydrograph and for the measured data. The data were plotted in a histogram. The histogram overlap between simulated and measured data was then calculated for each PTF according to following formula:

$$HDHO = \frac{\sum_{j=1}^{n} \min(K_j, L_j)}{\sum_{j=1}^{n} K_j} \quad (11)$$

where n is the number of bins, K_j the number of values within bin j for the reference (gauging station), and L_j the number of values in bin j for the simulation. This was done to determine a measure of the predictive accuracy of the discharge peaks. High histogram overlap values indicate a model's better predictive accuracy. Lower values represent poorer model capabilities of high discharge prediction.

2.10.3 Hydrograph efficiency metrics

The hydrographs of the final simulations were then compared with the measured hydrograph by applying a split sample test. This was done to evaluate the model's ability to correctly predict the overall discharge. For this purpose, three metrics were chosen. These include the Kling-Gupta efficiency (KGE) to evaluate the correspondence between observed and simulated hydrographs. It considers aspects like correlation, bias, and variability (Gupta and Kling, 2011). The Nash-Sutcliffe model efficiency coefficient (NSE) was used to evaluate how well simulated and measured values fit the 1:1 line. It puts a special focus on the prediction of correct volume (Nash and Sutcliffe, 1970). The third metric included was the coefficient of determination R^2. This metric is a measurement of the proportion of variance in the measured data that is predictable from the model data. The Kling-Gupta efficiency was calculated according to following formula:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (12)$$

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

where r is the Pearson correlation coefficient, α is a term representing the variability of prediction errors, and β is a bias term. The Nash-Sutcliffe model efficiency coefficient was calculated according to following formula:

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_O^t - Q_m^t)^2}{\sum_{t=1}^{T}\left(Q_O^t - \overline{Q_O}\right)^2} \quad (13)$$

where $\overline{Q_O}$ is the mean of observed discharges, Q_m is the simulated discharge, and Q_O^t is the observed discharge at time t. The coefficient of determination was calculated according to formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \hat{y})^2} \quad (14)$$

where y_i is the simulated discharge, $\hat{y}_i$ is the measured discharge, and $\overline{y}$ is the mean measured discharge. All three efficiency metric values were calculated for the calibrated model hydrographs for each PTF.

3 Results

3.1 ETa patterns (Step 1)

In step 1, we were able to use the already parameterised and calibrated values for the ETa-relevant plant properties from Casper et al. (2023). This made a separate evaluation of calibrated parameter values obsolete. The adequacy of the used values was also supported by the closed water balance in our model (see subsection 3.4), with deviations ranging from −10.99 % to 3 %.

3.2 Groundwater model parameterisation (Step 2 and 3)

The evaluation of the groundwater model adjustment (Fig. 6) shows that, in step 2 of our approach, we successfully matched the slope of the CDC to the observed data for all PTFs. This was achieved by using a single layer aquifer with a thickness of 1 m and lateral hydraulic conductivities of $3E-5\ m \cdot s^{-1}$. In step 3, the CDC height could also be adapted to the course of the gauging station curve for almost all PTFs. Only PTFs 9 and 10 could not be adjusted in height. The corresponding values for d_r range from 20 for PTF 11 up to 75 for PTF 2. The values for PTFs 9 and 10 were even higher but did not change the

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Hydrology and Earth System Sciences
Discussions

355 height of the CDC. In the front part of the curve, the simulations almost exclusively run below the reference curve of the gauging station. Only PTFs 9 and 10 run above the curve for the measured data.

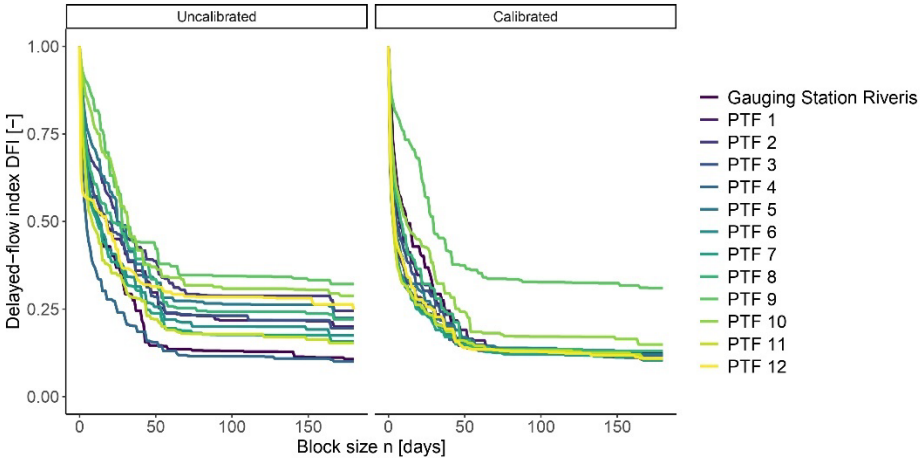
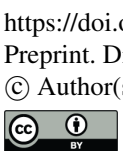


Figure 6: CDCs for the uncalibrated groundwater model and after groundwater model calibration for each PTF.

3.3 Dominant runoff process patterns (Step 4)

360 In step 4, the simulated dominant runoff processes for each PTF were compared to the reference map (BHK) to evaluate how well each PTF represents the spatial patterns of runoff (see Fig. 7). The overview of the simulated runoff processes shows that some PTFs deviate significantly from the reference map. Except for PTFs 4, 9 and 10, all show dominant interflow over most of the catchment area. PTFs 1, 3, 5, 8 and 12 show hardly any significant areas of deep percolation. However, in the reference map of the BHK, deep percolation can be found in the northern and southern edges of the catchment. Only PTFs 6, 7 and 11 365 show such areas with dominating deep percolation at the same positions as the BHK. PTF 4 shows almost exclusively dominant, extensive surface runoff. It also only shows interflow around the watercourse. This differs highly from the reference map. In comparison, PTF 9 and 10 show strongly dominating deep percolation over a large area. Also, only narrow areas with interflow can be found in the vicinity of the watercourse. The area with surface runoff in the west is also not depicted correctly in both PTFs. For all PTFs, the high correspondence between simulated and reference map for the direct runoff patterns results 370 from the fact that, by definition, surface runoff occurs in the model when a watercourse flows through a cell.

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025



Hydrology and Earth System Sciences
Discussions

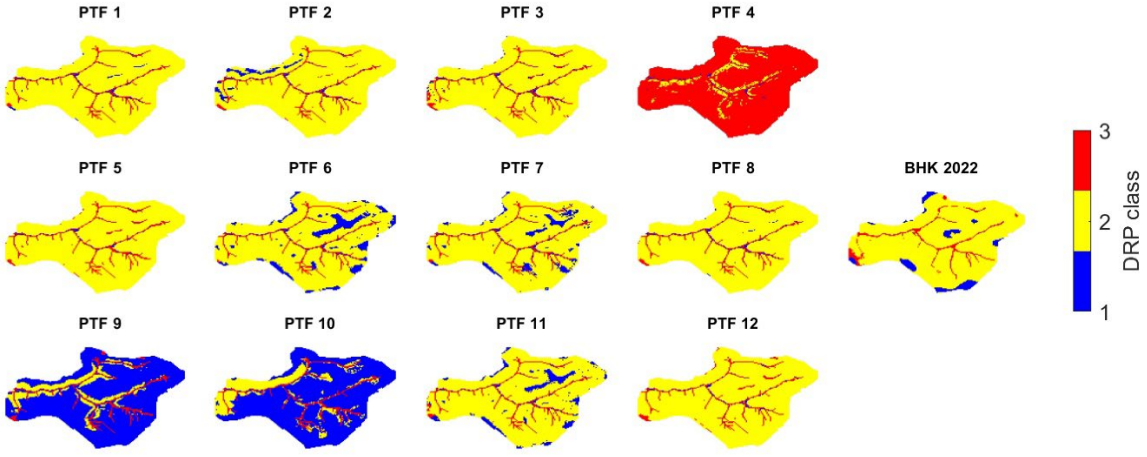


Figure 7: Spatial patterns for the simulated dominant runoff processes and the corresponding BHK reference map after a synthetic rainfall event.

The overall values as well as the individual metrics of the $SPAEF_{mod}$ metric are listed in Table 3. The $SPAEF_{mod}$ values summarise the values for the three individual parameters. PTFs 2, 7 and 11 achieve very high values of just over 0.8. Their simulated patterns for these PTFs therefore show high similarity to the patterns of the reference map. PTFs 1, 3, 4, 6, 8, and 12 show values in the mid-range. They show strong overall similarities between the patterns, while individual areas are not correctly depicted in the simulated patterns. PTFs 4, 9 and 10 have the lowest values of -0.34, -4.08 and -3.75.

Table 3: Metrics for the comparison of simulated dominant runoff processes and the BHK reference map.

PTF	% match	α	Histogram overlap	$SPAEF_{mod}$
1	0.88	0.77	0.96	0.74
2	0.86	0.90	0.98	0.82
3	0.88	0.80	0.96	0.76
4	0.13	0.46	0.14	-0.34
5	0.88	0.76	0.95	0.73
6	0.84	1.29	0.94	0.66
7	0.85	1.08	0.98	0.83
8	0.88	0.76	0.96	0.73
9	0.24	5.97	0.26	-4.08
10	0.25	5.63	0.29	-3.75
11	0.86	1.14	0.97	0.8

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

12	0.88	0.77	0.95	0.74

3.4 High flow calibration (Step 5)

The signature indices, including an evaluation of the high discharge (step 5), show a pronounced amplitude across the range of PTFs for some indices. For the *BiasRR*, which represents the mean deviation and thus the water balance, most PTFs show only small deviations of less than 5 %. Only PTFs 4 and 10 have higher deviations of over 10 %. It is striking that most PTFs underestimate the water balance, i.e. show negative deviations. Only PTFs 7 and 11 overestimate the water balance with positive deviations. The *biasFDCmidslope*, which describes the reactivity of the hydrograph, shows a large amplitude. PTFs such as 1, 2, 3, 8 and 10 show deviations of well below 10 %. PTF 7 shows an upward deviation of 24.57 %. PTF 9 shows a downward deviation of -33.65 %. Almost all PTFs show a *BiasFHV* close to 0. Only PTFs 9 and 10 show significant deviations of -44.26 % and -26.49 %. Most PTFs show a moderate underestimation of between -10 % and -15 % for the *BiasFLV*. Only PTFs 10 and 9 show a slight and a considerable upward deviation of 2.93 % and 44.65 % respectively. The deviation of the median (*BiasFMM*) shows a strong amplitude across the various PTFs. PTF 6 shows the largest negative deviation of -26.99 %. PTF 9 shows the largest positive deviation of 24.74 %. PTF 10 has the lowest deviation from zero at just 6.3 %.

Table 4: Signature indices of the calibrated model for different PTFs.

PTF	BiasRR	BiasFDC	BiasFHV	BiasFLV	BiasFMM
1	-4.2	3.5	0.65	-13.05	-8.79
2	-4.11	5.55	-0.34	-14.26	-6.96
3	-4.43	6	-0.38	-12.94	-4.05
4	-11.6	11.51	-1.32	-13.32	-27.1
5	-7.05	9.59	0.07	-8.97	-22.56
6	-1.51	21.2	0.32	-11.78	-26.99
7	3.15	24.57	-0.29	-10.33	-23.23
8	-4.89	6.34	0.27	-13.49	-17.82
9	-10.99	-33.65	-44.26	44.65	24.74
10	-5.6	-1.67	-26.49	2.93	6.3

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Hydrology and Earth System Sciences
Discussions

11	3	17.9	0.17	-7.48	-16.41
12	-4.33	16.73	-0.7	-21.85	-9.41

3.5 Final model evaluation (Step 6)

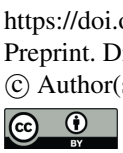
The Manhattan distances, calculated between the CDCs of simulated and observed data across the range of n values from $n = 1$ to $n = 43$, show considerable variabilities across all PTFs (Table 5). While PTF 8 has a distance value of only 1.8, the distance value of PTF 9 is several times higher with 8.3. PTFs 1 and 10 also show small distances, while the other PTFs are located in the middle range. For the high discharge histogram overlap (HDHO), PTF 4 shows the lowest value of 0.4. PTF 8 shows a high value of 0.9. Other PTFs are located in between.

Table 5: Efficiency metrics for the calibrated model for different PTFs.

PTF	MHd	HDHO	NSE_{cal}	KGE_{cal}	R^2_{cal}	NSE_{val}	KGE_{val}	R^2_{val}
1	2.22	0.8	0.834	0.896	0.842	0.758	0.872	0.803
2	4.41	0.7	0.659	0.820	0.691	0.591	0.795	0.674
3	3.13	0.7	0.721	0.848	0.740	0.638	0.822	0.705
4	5.05	0.4	0.644	0.733	0.694	0.346	0.612	0.561
5	4.34	0.7	0.719	0.819	0.747	0.551	0.754	0.666
6	5.91	0.7	0.746	0.867	0.771	0.649	0.827	0.729
7	5.95	0.7	0.597	0.808	0.673	0.477	0.758	0.640
8	1.5	0.9	0.881	0.901	0.888	0.807	0.889	0.843
9	8.3	0.5	0.542	0.557	0.589	0.573	0.545	0.648
10	1.62	0.5	0.808	0.814	0.821	0.780	0.789	0.782
11	5.61	0.7	0.553	0.788	0.634	0.417	0.737	0.594
12	4.31	0.7	0.667	0.821	0.700	0.585	0.795	0.671

The split-sample test carried out based on the simulated and measured hydrograph (Fig. 8) shows strong consistency with evaluation metrics of the model for the best parameterisation (PTF 8). The model shows high values for the efficiency measures for both the calibration and the validation period. Between calibration and validation, there is only a slight decrease in the NSE

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025



Hydrology and Earth System Sciences
Discussions

from 0.88 to 0.81, while the KGE decreases only minimally from 0.9 to 0.89. The R^2 also remains high at 0.89 to 0.84. Efficiency measures for the split-sample test of other PTFs (Table 5) show a large value range. For example, PTFs 1 and 10 also show relatively high values for the efficiency measures. However, PTFs 4, 9 and 11 show the lowest values. All other PTFs show values in between.

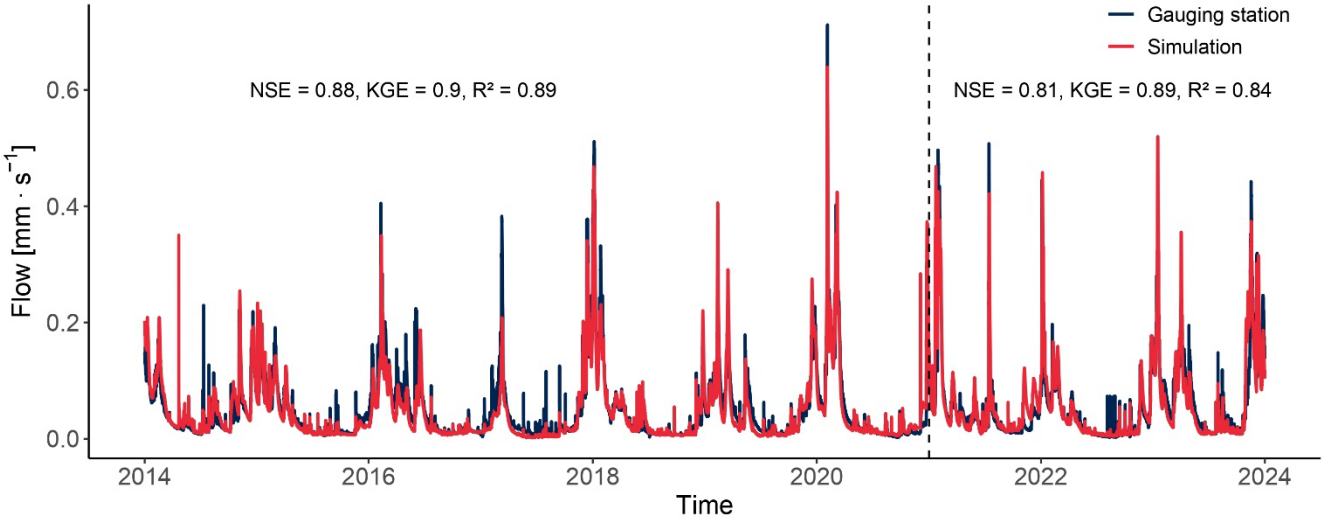


Figure 8: Measured (Gauging station) and simulated (PTF 8) hydrographs. Period before the dashed vertical line is the calibration period, while the one right of the dashed line marks the validation period. Efficiency metric values are shown for their respective period.

The hydrograph simulated by PTF 8 successfully replicates the measured hydrograph, with only slight underestimation of peak flows and a minor delay in response around December 2017. The model tends to smooth out finer fluctuations, resulting in a lower reactivity compared to observed data. Overall, however, PTF 8 closely mirrors the complex shape of the observed hydrograph. Hydrographs for other PTFs can be found in the appendix as Fig. A1 and Fig. A2.

The long-time discharge can also be depicted as a flow duration curve (Fig. 9). The flow duration curve for PTF 8 shows very good agreement in the high discharge volume. This corresponds to the discharge peaks of the hydrograph. In the middle part, the flow duration curve shows a kink. From there, it is no longer fully congruent with the curve for the measured discharge in areas for lower discharge volumes. The simulation slightly deviates from the measured flow duration curve in the area of very low discharges. However, it should be noted that the representation is logarithmic. The deviations occurring in the low

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Hydrology and Earth System Sciences
Discussions

discharge range therefore only account for a small proportion of the total discharge. PTF 8 therefore fits the flow duration curve of the reference the best. The other PTFs are deviating around the measured curve. Some overestimate the corresponding proportions and others underestimate the proportions. In the middle range, the results of the simulations are almost exclusively lower than the reference.

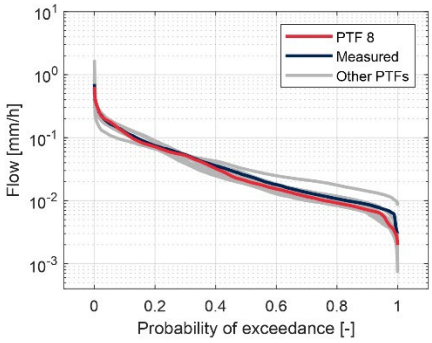


Figure 9: Flow duration curve for the gauging station, for the simulation with PTF 8 (red) and the other PTFs (grey).

4 Discussion

This study employed a multi-step calibration approach designed to incrementally improve the accuracy of hydrological simulations, by systematically targeting specific components of the water balance model. The following paragraphs discuss the results of each calibration step in detail.

4.1 Evapotranspiration/Water Balance (Step 1)

We used calibrated vegetation parameters from Casper et al. (2023). Because of the almost closed water balance (*BiasRR* in Table 4), an additional calibration step for evapotranspiration parameters was not necessary in our case. Only if the water balance can't be closed at the catchment outlet, it would have been necessary to adjust the evaporation parameters.

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

4.2 Groundwater model (Step 2 and 3)

Fitting to the Characteristic Delay Curve (CDC) is a perfect means for the calibration of the groundwater model in terms of its mean long-term behaviour (Fig. 6). The gradient of those segments of the CDCs which correspond to longer delay intervals (higher n-values) are highly sensitive to aquifer transmissivity parameters (k_x, k_y and thickness). On the other hand, the long-term groundwater recharge depends on the interflow intensity, which is adjusted by the parameter drainage density d_r. This approach effectively modified the height of the CDCs across most PTFs. However, two PTFs (PTFs 9 and 10) did not allow a good adjustment to the observed CDC height, due to lack of soil stratification in their parameterisation. These two PTFs estimate the hydraulic properties based on grain size, while key factors like depth or bulk density—typically considered in other PTFs or when using the KA5 standard for saturated hydraulic conductivity (k_{sat})—are not addressed. This means that, in the absence of stratification, there is little interflow and a large portion of water percolates into the aquifer (Ahuja et al., 1981). Without stratification, interflow cannot be controlled by the scaling factor d_r because there is too little interflow to begin with. The consistent underestimation of the initial segments of the CDCs suggests that the catchment is delaying certain parts of the water more than the model does (Yeh and Chen, 2022). This could theoretically be resolved by increasing the interflow delay through increasing values for k_{ifl}. However, as our catchment is mainly interflow dominated, the discharge peaks are almost exclusively interflow. Such an adjustment could reduce peak discharge significantly, which might compromise the hydrograph fit, as noted by Shrestha et al. (2013). Therefore, we assume that a two-layer aquifer model with distinct transmissivities would probably better represent the complex groundwater dynamics in our catchment.

4.3 Evaluation of dominant runoff processes (Step 4)

The evaluation of dominant runoff processes has shown that most PTFs can reproduce the pattern of the reference with reasonable accuracy (Fig. 7). However, PTFs 4, 9, and 10 showed significant deviations from the reference patterns, which indicate that these PTFs produce soil parameter estimates that differ substantially from actual field conditions. This results in either little interflow and too much surface runoff (PTF 4) or too much deep percolation and no interflow (PTFs 9 and 10). The high proportion of surface runoff and low fractions of interflow of PTF 4 are probably due to the low hydraulic

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

conductivities compared to other PTFs (Mohajerani et al., 2021). Therefore, the upper soil layers in the model quickly saturate during the synthetic rainfall event which results in a predominance of surface runoff. In contrast, PTFs 9 and 10 lead almost exclusively to dominant deep percolation. This is due to a lack of soil stratification, as only the grain size distribution is considered, but no other properties such as bulk density or depth (Renger et al., 2008; Y. Zhang and Schaap, 2017). Consequently, the model assumes uniform permeability, that allows most precipitation to infiltrate directly into the groundwater reservoir and bypass interflow pathways. However, the strong deviations in runoff pattern among these three PTFs can be systematically identified using the $SPAEF_{mod}$ metric. While the majority of PTFs achieved $SPAEF_{mod}$ values exceeding 0.65, which indicates good alignment with the reference map, PTFs 4, 9, and 10 showed significantly lower (in some cases, negative) values. This evaluation step serves as a reliable means to screen out PTFs that fail to capture dominant runoff processes accurately. This ensures that only soil parameterisations consistent with observed runoff fractions are considered in the final model selection process.

4.4 High flow calibration (Step 5)

The subsequent adjustment of the rainfall-fed part of the hydrograph, e.g. discharge fractions in the high volume based on the signature indices (Table 4), showed good applicability. For all PTFs except 9 and 10, the *biasFHV* could be brought close to zero. The water distribution could be shifted from peak discharge values towards mid-range discharge levels by adjusting k_d or k_{ifl}. PTFs 9 and 10 lack volume in the discharge peaks due to the large proportion of water that infiltrates very quickly into the aquifer. Therefore, hardly any direct runoff or interflow is present, which could contribute to high volume discharge (Seiler and Gat, 2007). This is also reflected in the patterns for the dominant runoff processes. In that case, the parameter k_{ifl} could not be used to shift more water from the peaks to the stronger delayed portions of discharge without losing a significant amount of water volume in the peaks. This is probably because our study area produces only little direct runoff, the contribution of which to the total runoff is delayed via k_d, mainly interflow contributes to the discharge. As a result, the hydrograph peaks in our model primarily reflect fast interflow rather than a balanced combination of direct runoff and interflow runoff. An independent adjustment via k_d and k_{ifl} would only be possible, if both runoff types are present to a certain extend. Adding a second aquifer layer with slightly higher conductivities than our current aquifer would enable us to represent a less delayed

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

groundwater discharge that currently is depicted through interflow. As a result, less interflow would be needed to represent parts of the slow components and therefore could be used to model part of the peak discharge. However, the necessity for this depends entirely on the catchment characteristics (Natkhin et al., 2012; Kraller et al., 2014) and can be derived from a repeated application of the Characteristic Delay Curve (Step 2 and 3), then with two aquifer layers.

4.5 Final model evaluation (Step 6)

The hydrograph of the best fitting model (based on PTF 8) shows that the model is capable of correctly mapping the discharge (Fig. 8). This is also supported by high values of efficiency measures such as NSE (0.81), KGE (0.89) and R^2 (0.84) for the validation period in the split-sample test. In addition, a high discharge histogram overlap (0.9) shows a good agreement in the peak discharge over time. However, the various PTFs show considerable deviations from each other. The choice of the pedotransfer function has a significant influence on the individual processes depicted by the model, and therefore the correct choice of the pedotransfer function is crucial to develop a behaviourally correct model parameterisation. This is also consistent with the findings of Mohajerani et al. (2021) and Paschalis et al. (2022). Our multi-criteria calibration framework, with its combination of parameterisation steps, proved effective both in evaluating PTFs and refining the calibration itself. Inconsistencies with both the CDCs and the patterns of dominant runoff processes proved the non-suitability of PTFs 9 and 10. Likewise, PTF 4 was found unsuitable due to deviations in runoff process patterns, despite its potential for further groundwater volume adjustments via drainage density d_r. This shows that a holistic view of the different processes is indeed necessary, as one PTF can be suited for a single process such as the groundwater flow but unsuited for other processes.

4.6 Transferability and Outlook

Our calibration approach is effectively transferable to other hydrological models and catchments, provided the necessary input parameters are available. For the first step, the calibration of ETa, remote-sensing ETa data is necessary. Here, readily available MODIS data can be used. Additionally, the application of the delayed flow index (DFI) requires only simulated and measured hydrographs, alongside a mechanism for adjusting groundwater recharge by percolating water. Models must support runoff

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences Discussions

partitioning into surface runoff, interflow, and deep percolation (groundwater recharge) to utilise the dominant runoff process comparison. For this, a spatial reference is necessary like the soil hydrological map used in our study. While certain methods necessitate only discharge data, we emphasize the benefits of incorporating multiple evaluation approaches. This comprehensive parameterisation captures the catchment behaviour across various hydrological processes more accurately. Consequently, our methodology demonstrates broad applicability for future parameterisations of hydrological water balance models, particularly those with a process representation similar to the WaSiM model.

Including tracer data as an additional evaluation criterion could enhance the robustness of our model parameterisation assessments (e.g., Wu et al., 2023). It offers valuable insights into discharge composition by distinguishing contributions from individual runoff components at the gauging station. For glacial and snow influenced catchments, the isotope approach of Penna et al. (2014) could be applied. For wetlands, Birkigt et al. (2018) and Schwerdtfeger et al. (2016) demonstrated approaches of tracer-based modelling. This could further improve the accuracy of selecting the correct model parameterisation by including this additional evaluation step.

5 Conclusions

Our study demonstrates that the multi-criteria calibration approach is highly effective not only in calibrating individual sub-processes within the model but also in providing a robust evaluation of the model's overall performance. By applying this approach, we were able to accurately identify specific parameterisations that resulted in incorrect representations of certain hydrological processes. This capability prevents the reliance on parameterisations that may yield satisfactory efficiency metrics (at catchment outlet) yet fail to adequately capture the underlying hydrological processes in the catchment area.

We consider this approach a significant advancement over traditional methods that prioritise hydrograph-based efficiency metrics alone when assessing model calibration and performance. Calibrating the ETa-relevant plant parameters ensured accurate spatio-temporal representation of ETa and a closed water balance. This step improved the model's ability to simulate plant-water interactions and maintain correct hydrological fluxes. Calibration of the groundwater model enhanced the

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

representation of groundwater behaviour, including baseflow contributions to discharge and groundwater recharge. This improved the accuracy of aquifer water storage and flow dynamics. Evaluating dominant runoff patterns enabled correct discharge partitioning and better spatial representation of runoff generation. Finally, applying signature indices and traditional efficiency metrics together allowed for both behavioural and quantitative model evaluation. Our multi-criteria framework adds depth to the calibration process by aligning the process representation with observed data in space and time. This enhances the model's reliability across varied hydrological conditions.

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Hydrology and Earth System Sciences
Discussions

Appendix

A Figures

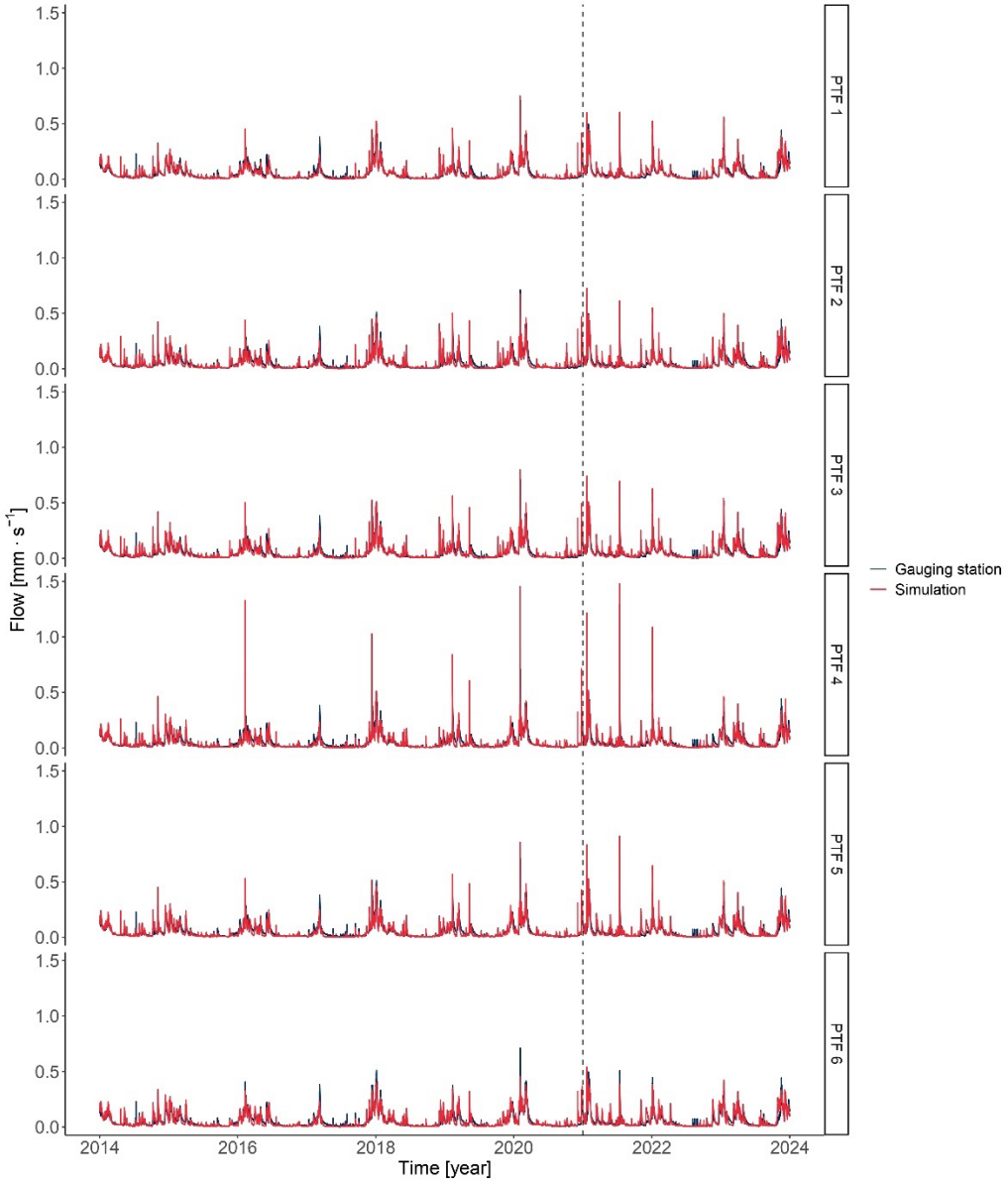
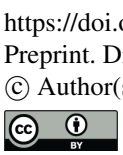


Figure A1: Full hydrographs for the gauging station and the simulation for PTFs 1 to 6. The hydrograph left of the dashed line was used as calibration period, while the part right of the dashed line served as calibration period.

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025



Hydrology and Earth System Sciences
Discussions

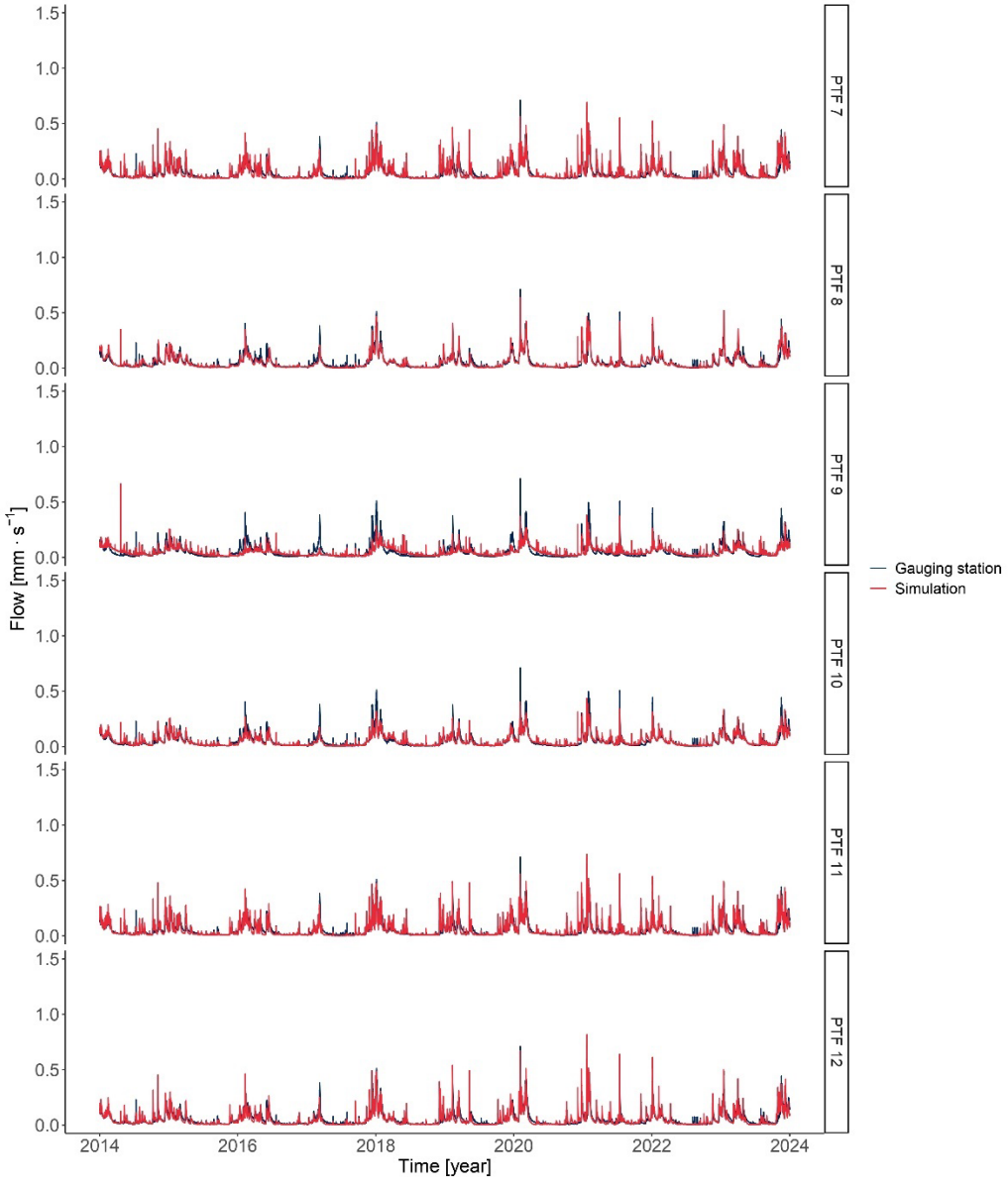


Figure A2: Full hydrographs for the gauging station and the simulation for PTFs 7 to 12. The hydrograph left of the dashed line was used as calibration period, while the part right of the dashed line served as calibration period.

555

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Hydrology and Earth System Sciences
Discussions

B Tables

Table B1: Parameters adjusted within our parameterisation and calibration approach.

Parameter	Unit	Values	Description
k_x	$[m \cdot s^{-1}]$	3E-5	Lateral conductivity of the aquifer in x-direction
k_y	$[m \cdot s^{-1}]$	3E-5	Lateral conductivity of the aquifer in y-direction
Colmation	$[m \cdot s^{-1}]$	3E-5	Hydraulic conductivity resistance between aquifer and waterbody
River network threshold	$[-]$	50	Threshold for the river network generation in TANALYS
d_r	$[-]$	10 to 75 (160)	Scaling factor for the interflow
k_d	$[h]$	2	Recession parameter for the direct runoff SLR
k_{ifl}	$[h]$	8 to 30	Recession parameter for the interflow SLR

Table B2: Calibrated parameters with values for different PTFs.

PTF	k_{ifl}	d_r	Comment
1	9	42	
2	21	75	
3	9	35	
4	27	10	
5	8	65	
6	30	30	
7	14	25	
8	26	65	
9	30	(160+)	Calibration of d_r not possible
10	30	(160+)	Calibration of d_r not possible
11	18	20	
12	18	50	

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025


Hydrology and Earth System Sciences
Discussions

Data and code availability

The calibrated model as well as the used input data can be found under https://doi.org/10.5281/zenodo.14185565.

Author contributions

M.C.C and M.M.H. conceptualised the study and methods. M.M.H. did the data curation, formal analysis, software development, and the original draft. M.C.C. did the funding acquisition, project administration and supervision. M.M.H., H.M., and M.C.C. die the review and editing.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We thank the Stadtwerke Trier (SWT) for providing gauging data for the catchment. We also thank the Landesamt für Umwelt (LfU) Mainz for providing high-resolution climate data.

Financial support

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — Project number 426111700 and Forstliche Forschungsförderung Nr. 5.2-04-2023 - Project „Klimawald2100 Modul Wald und Wasser“.

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

References

Ad-hoc-AG Boden (Ed.). (2006). Bodenkundliche Kartieranleitung (KA5) (5th ed.). Bundesanstalt für Geowissenschaften und Rohstoffe in Zusammenarbeit mit den Staatlichen Geologischen Diensten.

Agrarmeteorologie Rheinland-Pfalz. (2024). Retrieved February 5, 2024, from https://www.wetter.rlp.de /Agrarmeteorologie

Ahuja, L. R., Ross, J., & Lehman, O. (1981). A theoretical analysis of interflow of water through surface soil horizons with implications for movement of chemicals in field runoff. Water Resources Research, 17(1), 65–72.

Althoff, D., & Rodrigues, L. N. (2021). Goodness-of-fit criteria for hydrological models: Model calibration and performance assessment. Journal of Hydrology, 600, 126674.

Barkwith, A., Hurst, M. D., Jackson, C. R., Wang, L., Ellis, M. A., & Coulthard, T. J. (2015). Simulating the influences of groundwater on regional geomorphology using a distributed, dynamic, landscape evolution modelling platform. Environmental Modelling & Software, 74, 1–20.

Barthel, R. (2006). Common problematic aspects of coupling hydrological models with groundwater flow models on the river catchment scale. Advances in Geosciences, 9, 63–71.

Beven, K. (2002). Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. Hydrological processes, 16(2), 189–206.

Beven, K. J., & Alcock, R. E. (2012). Modelling everything everywhere: A new approach to decision-making for water management under uncertainty. Freshwater Biology, 57, 124–132.

Birkigt, J., Stumpp, C., Małoszewski, P., & Nijenhuis, I. (2018). Evaluation of the hydrological flow paths in a gravel bed filter modeling a horizontal subsurface flow wetland by using a multi-tracer experiment. Science of the total environment, 621, 265–272.

Casper, M. C., Grigoryan, G., Gronz, O., Gutjahr, O., Heinemann, G., Ley, R., & Rock, A. (2012). Analysis of projected hydrological behavior of catchments based on signature indices. Hydrology and Earth System Sciences, 16(2), 409–421.

Casper, M. C., Mohajerani, H., Hassler, S., Herdel, T., & Blume, T. (2019). Finding behavioral parameterization for a 1-D water balance model by multi-criteria evaluation. Journal of Hydrology and Hydromechanics, 67(3), 213–224.

Casper, M. C., Salm, Z., Gronz, O., Hutengs, C., Mohajerani, H., & Vohland, M. (2023). Calibration of Land-Use-Dependent Evaporation Parameters in Distributed Hydrological Models Using MODIS Evaporation Time Series Data. Hydrology, 10(12), 216.

Corine land cover [Data set]. (2018). http://data.europa.eu/88u/dataset/ispra_rm-meta_geo_cl001

Darcy, H. (1856). Les fontaines publiques de Dijon.

Demirel, M. C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., & Stisen, S. (2018). Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. Hydrology and Earth System Sciences, 22(2), 1299–1315.

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

European Environment Agency (EEA). (2020, September). Dominant Leaf Type 2018 (raster 10 m), Europe, 3-yearly, Sep. 2020. https://doi.org/10.2909/7b28d3c1-b363-4579-9141-bdd09d073fd8

Ferket, B. V., Samain, B., & Pauwels, V. R. (2010). Internal validation of conceptual rainfall–runoff models using baseflow separation. Journal of Hydrology, 381(1-2), 158–173.

Gauster, T., Laaha, G., & Koffler, D. (2022). lfstat: Calculation of Low Flow Statistics for Daily Stream Flow Data. https://doi.org/https://doi.org/10.32614/CRAN.package.lfstat

Gerlach, N. (2006). Niederschlags-Abfluss-modellierung zur Verlängerung des Vorhersagezeitraumes operationeller Wasserstands-Abflussvorhersagen. In B. für Gewässerkunde: Koblenz (Ed.), Gewässerkunde.

Götzinger, J., Barthel, R., Jagelke, J., Bardossy, A., et al. (2008). The role of groundwater recharge and baseflow in integrated models. Groundwater-surface water interaction: process understanding, conceptualization and modelling, 103–109.

Gupta, H. V., Beven, K. J., & Wagener, T. (2006). Model calibration and uncertainty estimation. Encyclopedia of hydrological sciences.

Gupta, H. V., & Kling, H. (2011). On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics. Water Resources Research, 47(10).

Kheimi, M., & Abdelaziz, S. M. (2022). A daily water balance model based on the distribution function unifying probability distributed model and the SCS curve number method. Water, 14(2), 143.

Knisel Jr, W. G. (1963). Baseflow recession analysis for comparison of drainage basins and geology. Journal of Geophysical Research, 68(12), 3649–3653.

Koch, J., Mendiguren, G., Mariethoz, G., & Stisen, S. (2017). Spatial sensitivity analysis of simulated land surface patterns in a catchment model using a set of innovative spatial performance metrics. Journal of Hydrometeorology, 18(4), 1121–1142.

Koch, J., Siemann, A., Stisen, S., & Sheffield, J. (2016). Spatial validation of large-scale land surface models against monthly land surface temperature patterns using innovative performance metrics. Journal of Geophysical Research: Atmospheres, 121(10), 5430–5452.

Kraller, G., Warscher, M., Strasser, U., Kunstmann, H., & Franz, H. (2014). Distributed hydrological modeling and model adaption in high alpine karst at regional scale (Berchtesgaden Alps, Germany). H2Karst Research in Limestone Hydrogeology, 115–126.

Landesamt für Geologie und Bergbau. (2021, June). Bodenflächendaten im Maßstab 1:50.000 (bfd50).

Liu, X., Yang, K., Ferreira, V. G., & Bai, P. (2022). Hydrologic model calibration with remote sensing data products in global large basins. Water Resources Research, 58(12), e2022WR032929.

McNamara, J. P., Tetzlaff, D., Bishop, K., Soulsby, C., Seyfried, M., Peters, N. E., Aulenbach, B. T., & Hooper, R. (2011). Storage as a metric of catchment comparison. Hydrological Processes, 25(21), 3364–3371.

Meresa, H., Zhang, Y., Tian, J., Ma, N., Zhang, X., Heidari, H., & Naeem, S. (2023). An integrated modelling framework in projections of hydrological extremes. Surveys in Geophysics, 44(2), 277–322.

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Mohajerani, H., Jackel, M., Salm, Z., Schütz, T., & Casper, M. C. (2023). Spatial Evaluation of a Hydrological Model on Dominant Runoff Generation Processes Using Soil Hydrologic Maps. Hydrology, 10(3), 55.

Mohajerani, H., Teschemacher, S., & Casper, M. C. (2021). A comparative investigation of various pedotransfer functions and their impact on hydrological simulations. Water, 13(10), 1401.

Monteith, J. L. (1965). Evaporation and environment. Symposia of the society for experimental biology, 19, 205–234.

Muggeo, V. M., et al. (2008). Segmented: an R package to fit regression models with broken-line relationships. R news, 8(1), 20–25.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. Journal of hydrology, 10(3), 282–290.

Natkhin, M., Steidl, J., Dietrich, O., Dannowski, R., & Lischeid, G. (2012). Differentiating between climate effects and forest growth dynamics effects on decreasing groundwater recharge in a lowland region in Northeast Germany. Journal of Hydrology, 448, 245–254.

Nesru, M., Shetty, A., & Nagaraj, M. (2020). Multi-variable calibration of hydrological model in the upper Omo-Gibe basin, Ethiopia. Acta Geophysica, 68(2), 537–551.

Nolte, A., Eley, M., Schöniger, M., Gwapedza, D., Tanner, J., Mantel, S. K., & Scheihing, K. (2021). Hydrological modelling for assessing spatio-temporal groundwater recharge variations in the water-stressed Amathole Water Supply System, Eastern Cape, South Africa: Spatially distributed groundwater recharge from hydrological model. Hydrological Processes, 35(6), e14264.

Paschalis, A., Bonetti, S., Guo, Y., & Fatichi, S. (2022). On the uncertainty induced by pedotransfer functions in terrestrial biosphere modeling. Water Resources Research, 58(9), e2021WR031871.

Penna, D., Engel, M., Mao, L., Dell'Agnese, A., Bertoldi, G., & Comiti, F. (2014). Tracer-based analysis of spatial and temporal variations of water sources in a glacierized catchment. Hydrology and Earth System Sciences, 18(12), 5271–5288.

R Core Team. (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Refsgaard, J. C. (2001). Towards a formal approach to calibration and validation of models using spatial data. Spatial patterns in catchment hydrology: observations and modelling, 329–354.

Renger, M., Bohne, K., Facklam, M., Harrach, T., Riek, W., Schäfer, W., Wessolek, G., & Zacharias, S. (2008). Ergebnisse und Vorschläge der DBG-Arbeitsgruppe „Kennwerte des Bodengefüges" zur Schätzung bodenphysikalischer Kennwerte. Wilhadi, 5, 21682.

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. physics, 1(5), 318–333.

RStudio Team. (2020). RStudio: Integrated Development Environment for R. RStudio, PBC. Boston, MA. http://www.rstudio.com/

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Schaake, J. C., Koren, V. I., Duan, Q.-Y., Mitchell, K., & Chen, F. (1996). Simple water balance model for estimating runoff at different spatial and temporal scales. Journal of Geophysical Research: Atmospheres, 101(D3), 7461–7475.

Schulla, J. (1997). Hydrologische Modellierung von Flussgebieten zur Abschätzung der Folgen von Klimaänderungen, Zürcher Geographische Schriften, Heft 69. Verlag Geographisches Institut ETH Zürich.

Schulla, J. (2024a). Model Description WaSiM (Water balance Simulation Model) - (version 10.08.00) [Accessed: 09-19-2024]. http://www.wasim.ch/downloads/doku/wasim/wasim_2024_en.pdf

Schulla, J. (2024b). TANALYS Topographisches Analyse-Tool [Accessed: 10-01-2024]. http://www.wasim.ch/de/products/tanalys.htm

Schwerdtfeger, J., Hartmann, A., & Weiler, M. (2016). A tracer-based simulation approach to quantify seasonal dynamics of surface-groundwater interactions in the Pantanal wetland. Hydrological Processes, 30(15), 2590–2602.

Seiler, K.-P., & Gat, J. R. (2007). Groundwater recharge from run-off, infiltration and percolation (Vol. 55). Springer Science & Business Media.

Shrestha, R. R., Osenbrück, K., & Rode, M. (2013). Assessment of catchment response and calibration of a hydrological model using high-frequency discharge–nitrate concentration data. Hydrology Research, 44(6), 995–1012.

Smakhtin, V. U. (2001). Estimating continuous monthly baseflow time series and their possible applications in the context of the ecological reserve. Water SA, 27(2), 213–218.

Steinrücken, U., & Behrens, T. Bodenhydrologische Karte. – LUWG-Bericht 6/2010. 2010.

Stisen, S., Demirel, C., & Koch, J. (2017). A novel spatial performance metric for robust pattern optimization of distributed hydrological models. AGU Fall Meeting Abstracts, 2017, H11D–1204.

Stisen, S., Jensen, K. H., Sandholt, I., & Grimes, D. I. (2008). A remote sensing driven distributed hydrological model of the Senegal River basin. Journal of Hydrology, 354(1-4), 131–148.

Stoelzle, M., Schuetz, T., Weiler, M., Stahl, K., & Tallaksen, L. M. (2020). Beyond binary baseflow separation: A delayed-flow index for multiple streamflow contributions. Hydrology and Earth System Sciences, 24(2), 849–867.

Stoelzle, M., Weiler, M., Stahl, K., Morhard, A., & Schuetz, T. (2015). Is there a superior conceptual groundwater model structure for baseflow simulation? Hydrological processes, 29(6), 1301–1313.

Szabó, B., Weynants, M., & Weber, T. K. D. (2021). Updated European hydraulic pedotransfer functions with communicated uncertainties in the predicted variables (euptfv2). Geoscientific Model Development, 14(1), 151–175. https://doi.org/10.5194/gmd-14-151-2021

Teepe, R., Dilling, H., & Beese, F. (2003). Estimating water retention curves of forest soils from soil texture and bulk density. Journal of Plant Nutrition and Soil Science, 166(1), 111–119.

Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil science society of America journal, 44(5), 892–898.

https://doi.org/10.5194/hess-2024-369
Preprint. Discussion started: 22 January 2025

Vansteenkiste, T., Tavakoli, M., Van Steenbergen, N., De Smedt, F., Batelaan, O., Pereira, F., & Willems, P. (2014). Intercomparison of five lumped and distributed models for catchment runoff and extreme flow simulation. Journal of Hydrology, 511, 335–349.

Westerberg, I., Guerrero, J.-L., Younger, P., Beven, K., Seibert, J., Halldin, S., Freer, J., & Xu, C.-Y. (2011). Calibration of hydrological models using flow-duration curves. Hydrology and Earth System Sciences, 15(7), 2205–2227.

Weynants, M., Vereecken, H., & Javaux, M. (2009). Revisiting Vereecken pedotransfer functions: Introducing a closed-form hydraulic model. Vadose Zone Journal, 8(1), 86–95.

Wösten, J., Lilly, A., Nemes, A., & Le Bas, C. (1999). Development and use of a database of hydraulic properties of European soils. Geoderma, 90(3-4), 169–185.

Wu, S., Tetzlaff, D., Yang, X., Smith, A., & Soulsby, C. (2023). Integrating Tracers and Soft Data Into Multi-Criteria Calibration: Implications From Distributed Modelling in a Riparian Wetland. Water Resources Research, 59(11), e2023WR035509.

Xiong, L., & Guo, S. (1999). A two-parameter monthly water balance model and its application. Journal of hydrology, 216(1-2), 111–123.

Yáñez-Morroni, G., Suárez, F., Muñoz, J. F., & Lagos, M. S. (2024). Hydrological modelling of the Silala River basin. 2. Validation of hydrological fluxes with contemporary data. Wiley Interdisciplinary Reviews: Water, 11(1), e1696.

Yeh, H.-F., & Chen, H.-Y. (2022). Assessing the long-term hydrologic responses of river catchments in Taiwan using a multiple-component hydrograph approach. Journal of Hydrology, 610, 127916.

Zacharias, S., & Wessolek, G. (2007). Excluding organic matter content from pedotransfer predictors of soil water retention. Soil Science Society of America Journal, 71(1), 43–50.

Zhang, H., Huang, G. H., Wang, D., & Zhang, X. (2011). Multi-period calibration of a semi-distributed hydrological model based on hydroclimatic clustering. Advances in Water Resources, 34(10), 1292–1303.

Zhang, Y., & Schaap, M. G. (2017). Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). Journal of Hydrology, 547, 39–53.