





Effect of data length, spin-up period and spatial model resolution on fully distributed hydrological model calibration in the Moselle basin

Ömer Ekmekcioğlu, Mehmet Cüneyd Demirel & Martijn J. Booij


To cite this article: Ömer Ekmekcioğlu, Mehmet Cüneyd Demirel & Martijn J. Booij (2022) Effect of data length, spin-up period and spatial model resolution on fully distributed hydrological model calibration in the Moselle basin, Hydrological Sciences Journal, 67:5, 759-772, DOI: [10.1080/02626667.2022.2046754](https://doi.org/10.1080/02626667.2022.2046754)

To link to this article: <https://doi.org/10.1080/02626667.2022.2046754>

 View supplementary material [↗](#)

 Published online: 24 Mar 2022.

 Submit your article to this journal [↗](#)

 Article views: 222

 View related articles [↗](#)

 View Crossmark data [↗](#)

Effect of data length, spin-up period and spatial model resolution on fully distributed hydrological model calibration in the Moselle basin

Ömer Ekmekcioğlu ^{a,b}, Mehmet Cüneyd Demirel ^{a,b} and Martijn J. Booij ^c

^aDepartment of Civil Engineering, Istanbul Technical University, 34469 Maslak, Istanbul, Turkey; ^bIstanbul Technical University, Hydraulics and Marine Sciences Research Center, Maslak, Istanbul, Turkey; ^cWater Engineering and Management, Faculty of Engineering Technology, University of Twente, Enschede 7500 AE, The Netherlands

ABSTRACT

Subjective decisions in hydrologic model calibration can have drastic impacts on our understanding of basin processes and simulated fluxes. Here, we present a multicase calibration approach to determine three pillars of an appropriate hydrological model configuration, i.e. calibration data length, spin-up period, and spatial resolution, using a spatially distributed meso-scale hydrological model (mHM) together with a dynamically dimensioned search (DDS) algorithm and Nash-Sutcliffe efficiency (NSE) for the Moselle basin. The results show that a 10-year calibration data length, 2-year spin-up period, and 4-km model resolution are appropriate for the Moselle basin to reduce the computational burden while simulating streamflow with a decent performance. Although the calibration data length and spatial resolution are related to the extent and quality of the data, and the spin-up period is basin dependent, analysing the combined effects further allowed us to understand the interactions of these three usually overlooked pillars in the mHM configuration.

ARTICLE HISTORY

Received 12 August 2021
Accepted 25 January 2022

EDITOR

A. Castellarin

ASSOCIATE EDITOR

M. Newcomer

KEYWORDS

mHM; mesoscale hydrologic model; Moselle River; model calibration; spatial resolution; spin-up period

1 Introduction

Hydrological models are crucial tools to evaluate physical processes and quantify water balance components in a catchment (Maina *et al.* 2020, Koc *et al.* 2021). They can be classified according to the amount of physics incorporated as empirical (or data-driven), conceptual and physically based models. The focus in this study is physically based regarding the amount of physics and fully distributed regarding the spatial resolution of the models. Obviously, the choice of the model type together with the availability of data such as the spatial resolution of inputs, the length of the spin-up period and the parameter calibration strategy all affect the model performance (Blöschl and Sivapalan 1995). The determination of all these aspects in a calibration framework is related to appropriate modelling in hydrology and should be based on the modelling objective, the data availability and a systematic analysis of the model–catchment interaction (Booij 2005). We focus on user-defined options in hydrological modelling as we are interested in identifying the appropriate calibration data length, spin-up period and spatial model resolution in the Moselle River basin.

The calibration process, which is of the utmost importance in minimizing the parameter uncertainty (Sreedevi and Eldho 2019, Westerberg *et al.* 2020), is described as the optimization of uncertain parameter values in the model to obtain sufficient accuracy in model outcomes (Simunek *et al.* 2012). Since calibration can be performed by trial and error for different conditions, i.e. manual calibration (Gelleszun *et al.* 2017), and also with mathematical algorithms, i.e. automatic calibration (Madsen 2003), time-efficiency is a major challenge. Some

scholars have recommended forgoing the validation set and evaluating the model performance according to the selection of the entire dataset as calibration (Arsenault *et al.* 2018), while others have highlighted the validation period by also considering different calibration period selection strategies to increase the reliability of the hydrological model (Gharari *et al.* 2013, Liu *et al.* 2018). In addition, one can reach competitive model performance using a surrogate sub-period for model calibration instead of using an entire dataset, in which the main point in that case is the selection of an appropriately representative sub-period based on expert knowledge (Razavi and Tolson 2013). The main constraint in determining the calibration period is the availability of data, i.e. long time series of runoff or other model output or state variables (Sorooshian *et al.* 1997). In general, using 20 years' worth of data for the calibration period is assumed to be sufficient for large basins to account for climatological and hydrological variability (Epstein *et al.* 1998). Although data records for large basins might be available for more than 30 years, keeping the calibration period as long as possible is computationally inefficient and not always meaningful, in particular when climatic or other trends are presented in the time series and the model should only be calibrated on the most representative (i.e. most recent) time period (Daggupati *et al.* 2015). For instance, Perrin *et al.* (2007) found that a much smaller number of random days (~300 days) is sufficient for the calibration of models with a small number of parameters.

Another factor affecting the calibration performance of hydrological models is the length of the spin-up period, which provides the required initial model state (Yang *et al.*

1995). The required spin-up period highly depends on the input data of the catchment and the hydrological response (Rodell *et al.* 2005). In addition, determining the optimum spin-up period is essential, since both shorter and longer spin-up periods may have negative effects on the calibration performance. A shorter spin-up period inevitably leads to a low (even wrong) performance evaluation, whereas a longer spin-up period can lead to a waste of data and misinterpretation of the results (Ajami *et al.* 2014a). Practitioners generally consider the first two or three years acceptable as a spin-up period, depending on the model structure. There have been studies using a spin-up period of only one year for lumped models (Rahman *et al.* 2016), semi-distributed models (Abdo *et al.* 2009, Xu *et al.* 2013) and distributed models (Lohmann *et al.* 1998, Cuo *et al.* 2006, Revilla-Romero *et al.* 2016).

Although it is taken as common sense that the spin-up period should vary from one year to several years, up to 10 years (Shi *et al.* 2008), no consensus has been reached in this regard (Kim *et al.* 2018). Sood *et al.* (2013) performed simulations with a monthly time step, since they had monthly streamflow observations, and the first two years of a 13-year data period were used as a spin-up period, while the remaining 11 years were utilized for model calibration. Ashraf (2013) performed simulations on a monthly basis as well and divided the entire dataset into two periods with six years as the spin-up period and 10 years as the calibration period. With a few exceptions, studies conducted to identify the optimum spin-up period surprisingly have not attracted the research community's attention, particularly for physically based distributed hydrological models.

Furthermore, heterogeneous land surface conditions require a sufficiently long spin-up period (Shrestha and Houser 2010). Ajami *et al.* (2014a) emphasize the importance of a multicriteria approach, which includes the groundwater storage, unsaturated zone storage, depth to water table, root zone storage, discharge, snow water equivalent and energy fluxes, in determining the spin-up period of integrated hydrological models. The length of spin-up periods also depends on the initial soil moisture content, soil depth, soil and vegetation type and groundwater storage at the start of the simulations, in addition to the temperature and rainfall forcings (Cosgrove *et al.* 2003). With a method based on relative changes in monthly groundwater storages, Ajami *et al.* (2014a) presented a hybrid approach on the basis of integration of ParFlow, which is an integrated hydrological model, and the empirical depth-to-water-table function, to satisfy equilibrium state conditions. They reduced the spin-up period by approximately 50% (from 20 years to 10–12 years) compared to the conventional continuous recursive simulation approach, which is widely employed for the determination of spin-up periods in land surface models.

Regardless of the model complexity, another issue that has a significant impact on hydrological model performance is the spatial model resolution (Koren *et al.* 1999). The size of the catchment (Wallace *et al.* 2018), heterogeneity of rainfall (Nicolina *et al.* 2008) and karstic geomorphology can greatly affect the spatio-temporal variations of hydrological processes (Zhang *et al.* 2020). The spatial resolution to be used in

a model is related not only to the availability of meteorological input data but also to the computational resources (Sood and Smakhtin 2015). Accordingly, simulation performance may either increase or decrease depending on the spatial resolution (Booij 2002, 2005, Bucchignani *et al.* 2016, Pang *et al.* 2020). However, in some cases, a considerable change is not observed, indicating that the model structure is suitable for all resolutions (Merz *et al.* 2009). In line with this, Nkiaka *et al.* (2017) evaluated the performance of reanalysis precipitation estimates on streamflow simulations using the SWAT model in a large river basin and concluded that the spatial resolution of the reanalysis datasets has little effect on streamflow simulation.

In addition, the spatial variability of storm events has an influence on the appropriate spatial resolution of the model. Lumped models may perform accurately with a spatially uniform input distribution, while they may need a higher spatial resolution (e.g. sub-basins) in the case of a non-uniform spatial input distribution (Tian *et al.* 2020). Pang *et al.* (2020) evaluated the precipitation model input, both temporally and spatially, based on the differences in various open-access precipitation products. In semi-distributed conceptual models, the spatial resolution is determined based on the sub-basin distribution. Distributed models provide distributed outputs since spatial heterogeneity is taken into account (Dehotin and Braud 2008). Etchevers *et al.* (2001) performed simulations for spatial resolutions of 1, 8 and 46 km using the soil-vegetation-atmosphere transfer (SVAT) model. They obtained mediocre simulation results for the 46 km resolution, whereas flash-flood events were better captured in the model with an 8 km resolution. Chen *et al.* (2017) employed the Liuxihe model, i.e. a physically based distributed hydrological model, to investigate flood events in the Liujiang River basin, China, which covers an area of about 60 000 km². They calibrated the model using particle swarm optimization (PSO) for a total of 29 flood events. Considering five spatial model resolutions, i.e. 200, 400, 500, 600 and 1000 m, they concluded that the results for the 1 km grid were not meaningful. The peak values were captured when applying resolutions of 500 m or smaller. Although slightly better results were obtained for 400 m, they chose 500 m grids as the appropriate spatial resolution considering the computational burden. Fully distributed models are more sensitive to the resolution of the rainfall input as compared to semi-distributed models (Gires *et al.* 2015). Most current studies have investigated the effects of either the model input resolution or the spin-up period on the model results. No study is known to the authors that explicitly assesses the effects of the spatial resolution of the model together with the length of the spin-up period and the calibration period on the model performance.

We aim to comprehensively investigate the impact of these three major but overlooked pillars – (1) calibration period, (2) spin-up period and (3) spatial model resolution – on the calibration and validation performance of a physically based distributed hydrological model for the Moselle River basin in France and Germany. The study area and data are introduced in section 2. The model and calibration framework are

presented in section 3. The calibration and validation results are presented and discussed in sections 4 and 5, respectively. Finally, the key conclusions are drawn in section 6.

2 Study area and data

2.1 Study area

The focus of this study is the Moselle River basin (Fig. 1), i.e. the largest sub-basin of the Rhine River. The main channel of the Moselle River has a length of about 545 km (Demirel *et al.* 2013). The Moselle River basin, covering parts of the three countries France, Germany and Luxembourg, has a surface area of approximately 27 262 km². The three longest tributaries of the Moselle River are the Saar, Sauer and Meurthe. The basin has varying lithological and topographic characteristics, while it has a rain-dominated regime (Brenot *et al.* 2007). The minimum, mean and maximum discharge values observed for the Moselle (at Cochem station) are 14 (dry summer), 130 (long-term average until 2009) and 4000 m³/s (winter), respectively (Demirel *et al.* 2013). The mean altitude of the basin is around 340 m and the land use is dominated by agriculture (54%) with arable areas, pastures and natural grasslands (Uehlinger *et al.* 2009), and forests (37%) in the mountains and on hillslopes (Demirel *et al.* 2019).

2.2 Data

Distributed hydrological models not only need hydrometeorological and geographical data as input but also require parameters relevant for different hydrological processes such as interception and infiltration. At this point, the data availability and the spatio-temporal resolution of the input data play a vital role in the accuracy of a model. In this study, the model uses spatially distributed precipitation, temperature and potential evapotranspiration data as input (Table 1). Meteorological data are from the ENSEMBLES daily gridded observational dataset (E-OBS) (Haylock *et al.* 2008), and the discharge data at Cochem station was obtained from the Global Runoff Data Center (GRDC) in Koblenz (Germany) (see the Supplementary material for the peak flow of each year utilized for model calibration based on observations in Fig. S1).

The digital elevation model (DEM) is based on the Shuttle Radar Topography Mission (SRTM) of National Aeronautics and Space Administration (NASA) (Ballabio *et al.* 2016). The soil classes are derived from the Harmonized World Soil Database (Fischer *et al.* 2012), while land cover data is provided from the Coordination of Information on the Environment (CORINE) dataset (Girard *et al.* 2019). Table 1 provides a brief summary of the data used in this study.

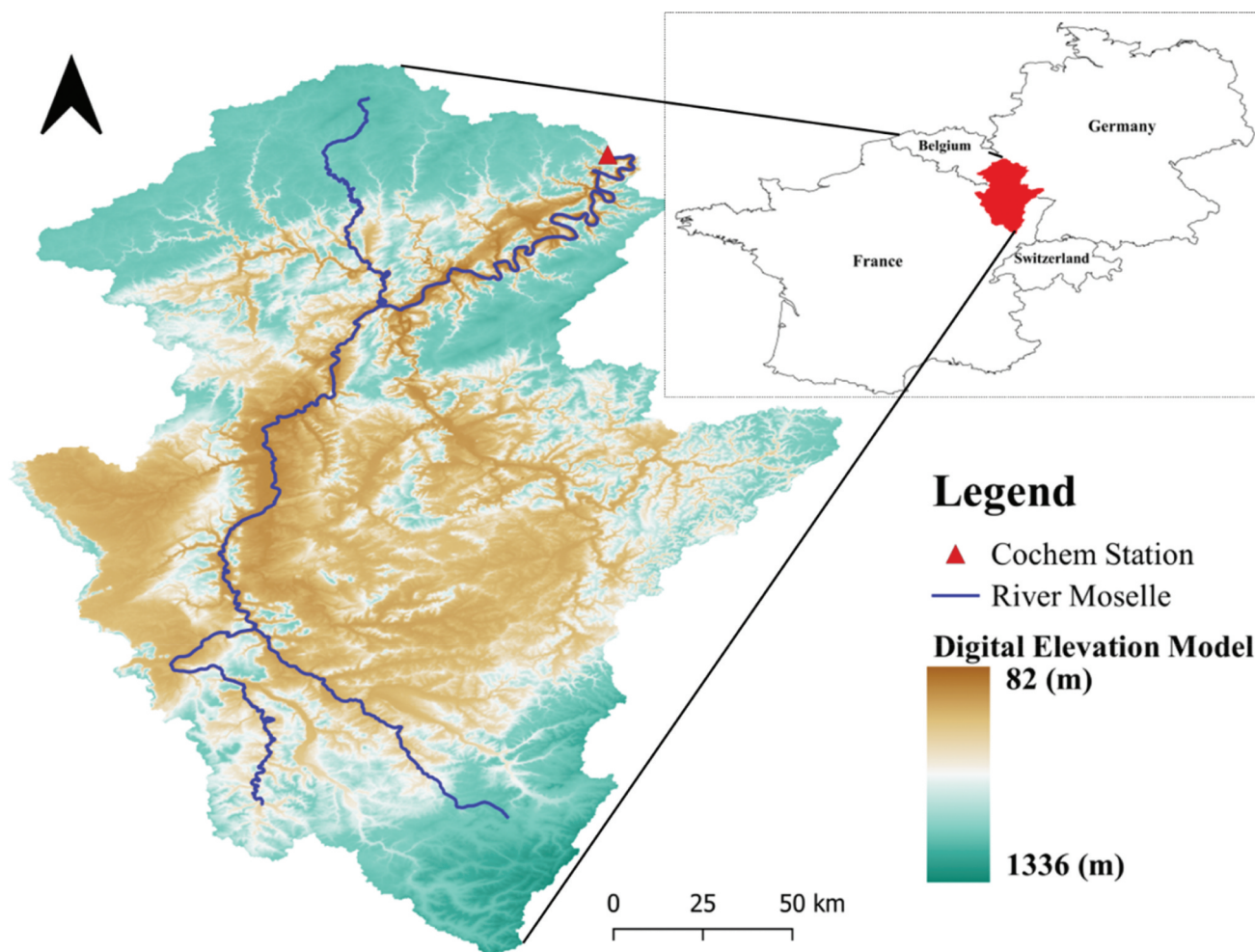


Figure 1. Moselle River network, basin boundary and elevation map.

Table 1. Summary of geographical and meteorological data used as input for meso-scale hydrological model (mHM).

| Variable | Description | Spatial resolution | Temporal resolution | Source |
|---------------------------|--|--------------------|------------------------|---------------------------------|
| Q (daily) | Streamflow (Cochem station, #6 336 050) | Point | Daily | GRDC |
| P (daily) | Precipitation | 24 km | Daily | E-OBS 20.0e, MODIS |
| ET _{ref} (daily) | Reference evapotranspiration | 24 km | Daily | E-OBS 20.0 ^e , MODIS |
| T _{avg} (daily) | Average air temperature | 24 km | Daily | E-OBS 20.0 ^e , MODIS |
| Land cover | Pervious, impervious and forest | 250 m | 1 map for whole period | CORINE |
| DEM data | Slope, aspect, flow accumulation and direction | 250 m | 1 map for whole period | SRTM |
| Geology class | Two main geological formations | 250 m | 1 map for whole period | EUROPEAN SOIL DATABASE |
| Soil class | Soil texture data | 250 m | 1 map for whole period | HARMONIZED WORLD SOIL DATABASE |

CORINE: Coordination of Information on the Environment; GRDC: Global Runoff Data Center; SRTM: Shuttle Radar Topography Mission.

3 Methods

3.1 Meso-scale hydrological model

The grid-based meso-scale hydrological model (mHM) is a fully distributed model in which for each grid cell incoming and outgoing fluxes for different storage compartments are calculated and the water balance of each compartment is updated after each time step (Samaniego *et al.* 2010, Kumar *et al.* 2013, Dembélé *et al.* 2020). In mHM, runoff is transferred to the downstream cells along the basin and river using three routing methods, i.e. Muskingum, adaptive time step with constant celerity and adaptive time step with varying celerity (Thober *et al.* 2019). In this study, we used adaptive time step with constant celerity method as it only requires one parameter, i.e. streamflow celerity. In the last decade, mHM has been applied to basins in many countries in Europe (Marx *et al.* 2017, Samaniego *et al.* 2018), including Germany (Höllering *et al.* 2018, Baroni *et al.* 2019, Jing *et al.* 2019) and Denmark (Demirel *et al.* 2018b), as well as to various large basins worldwide (Eisner *et al.* 2017, Huang *et al.* 2018).

mHM is an open-source software program written in the Fortran 2003 language and accessible from www.ufz.de/mhm; the model is compatible with many platforms, such as Linux, Mac and Windows (Nijssen *et al.* 2001, Brenner *et al.* 2021). One of the most appealing features of the model code is the transferability between different input resolutions (Fig. 2) for the desired computational resolutions (mesh). The model handles different resolutions of soil-related data and meteorological data (Fig. 2) by automatic upscaling and downscaling of high-resolution geographical data (L0) and coarse meteorological data (L2) to reach the user-defined hydrological output resolution (L1). Also, the model provides flexibility to select a routing resolution (L11) different than the hydrological resolution (L1), so that the user can benefit from high-resolution geographical input (soil, geology, aspect, LAI, elevation, etc.) and does not lose time with preprocessing of meteorological data to fit the

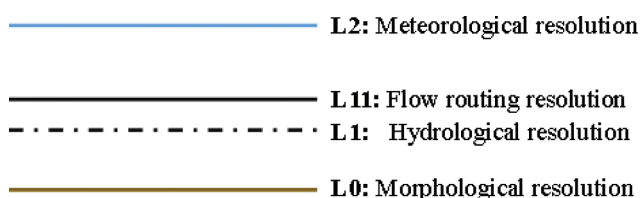


Figure 2. Model input and output scale configuration in meso-scale hydrological model (mHM).

resolutions for model runs. Transferring data to a coarser resolution is done based on harmonic averaging instead of arithmetic averaging. In addition, different temporal resolutions for the model outputs can be used, e.g. daily, monthly, or annual model results. For details of the process formulations, readers and potential users may refer to the model papers (Samaniego *et al.* 2010, Kumar *et al.* 2013).

3.2 Parameter sensitivity analysis

Sensitivity analysis (SA) is an important step before the calibration and validation of complex hydrological models to reduce the dimension of the search space. This will increase the effectiveness of the calibration process by reducing the run time. mHM includes around 55 global parameters used in physically based equations representing the different hydrological processes. Although there are comprehensive SA methods considering parameter interactions, such as Sobol's and sequential screening methods (Nossent *et al.* 2011, Cuntz *et al.* 2015), we used a local sensitivity analysis method available in the Parameter ESTimation (PEST) tool to identify the most important parameters (Doherty 2010). The parameters are perturbed one at a time with a particular percentage (i.e. 5%) and the change in the performance metric is observed. PEST allows one-sided (only increase) or two-sided (increase and decrease) sensitivity analysis. We analysed the sensitivity of 55 parameters using $2n + 1$ model runs (n is the number of parameters). Readers are referred to Demirel *et al.* (2018b) for the details of the SA method with smart sampling and the Jacobian matrix.

3.3 Model calibration and validation

Since we are interested in capturing peak flows, we selected the Nash-Sutcliffe efficiency (NSE), i.e. the most commonly used metric in flood hydrology (Knoben *et al.* 2019), to present our calibration results. In this study, mHM version 5.10 was set up for the Moselle River basin, and the effects of the three factors (pillars) on the model performance were examined. Accordingly, we tested all possible combinations of three factors, i.e. a total of 105 cases comprising three calibration data lengths, seven spin-up periods and five spatial model resolutions, to design an appropriate calibration framework for the Moselle River basin. Here, we tested spatial model resolutions varying from 1 to 12 km. The mHM model internally upscales and downscales the input data to match the input scale to the hydrological model

scale. Since we identified very small effects of the routing scale on the model performance, the routing scale was fixed to 6 km to save a substantial amount of run time using the workstation configuration of the AMD Ryzen Threadripper 1900X 8-Core Processor (Win-10, 4.10 GHz and 64 GB RAM). Further, we used three calibration periods of 1991–2005, 1996–2005 and 2001–2005, corresponding to data lengths of 15, 10 and 5 years, respectively. The four-year period between 2006 and 2009 was selected as the validation period for each model since we had data from 1991 to 2009

(Fig. 3) (see Supplementary material for monthly statistics of the validation data in Table S1). We tested seven spin-up periods of 0, 1, 2, 3, 4, 5 and 10 years independent from the calibration period and five spatial model resolutions of 1, 2, 4, 8 and 12 km. It should be noted that we diligently chose the spin-up periods independent from the calibration period to provide a fair comparison as the length of a spin-up period directly related to the basin’s initial conditions. Figure 4 illustrates the monthly average discharges along with the annual average discharge values. One can see from

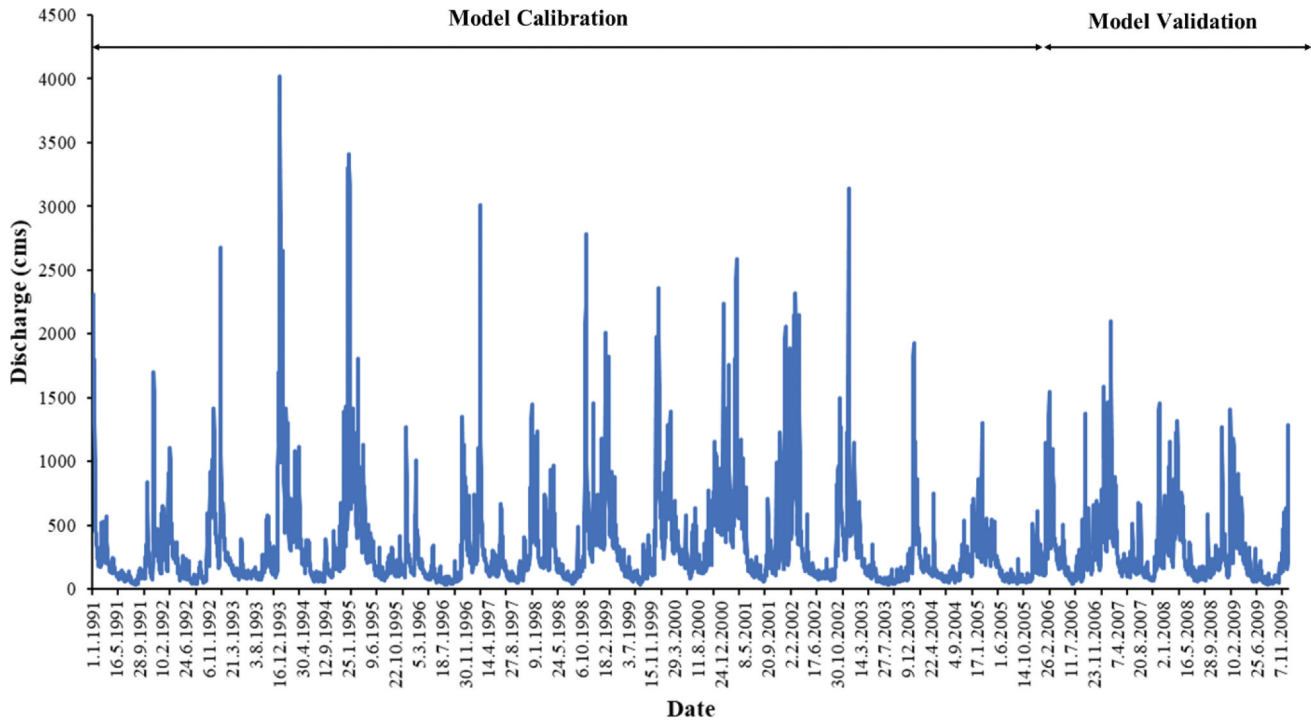


Figure 3. Time series of the utilized discharge data.

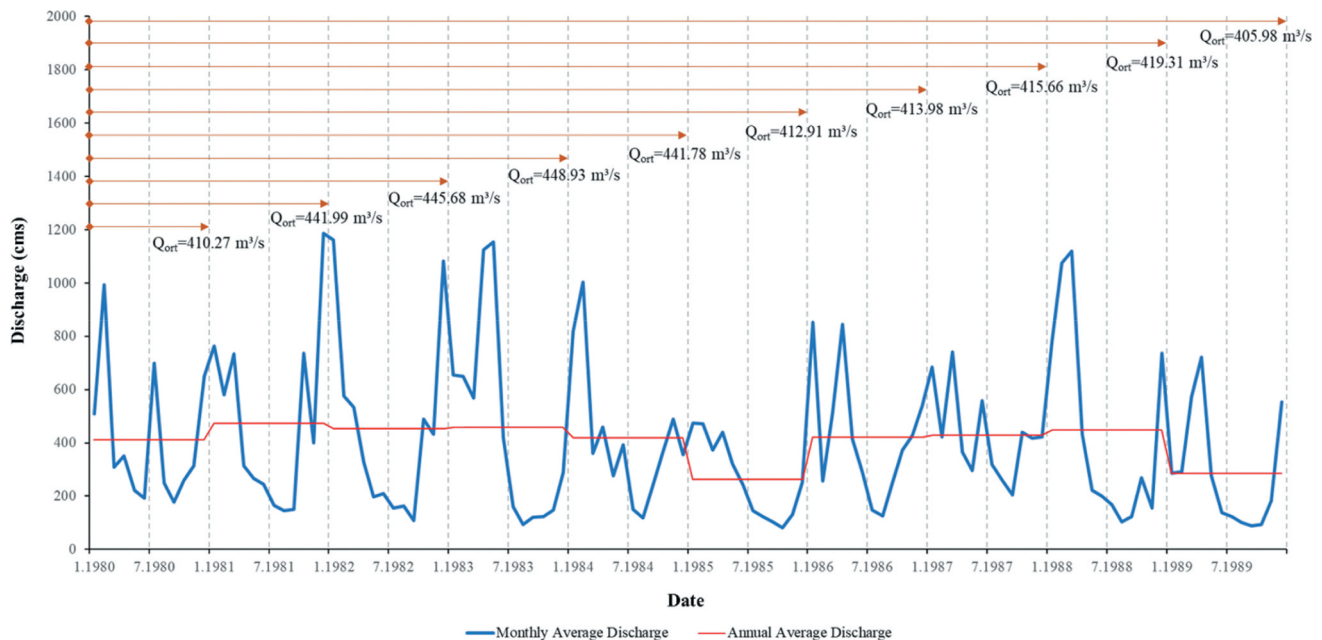


Figure 4. Variations in recorded discharge values for selected spin-up periods.

the figure that there are no major changes in terms of recorded discharges within the seven spin-up periods selected. In addition, the geographical and geomorphological data of the mHM model is at a 250 m resolution, and meteorological inputs, i.e., precipitation (P), reference evapotranspiration (ET_{ref}), and average temperature (T_{avg}) are at a 24 km resolution. The discharge data at Cochem station was used in both the calibration and the validation.

In addition, the mHM internal auto-calibration tool provides four search algorithms. In this study, the dynamically dimensioned search (DDS) algorithm (Tolson and Shoemaker 2007) is used to calibrate the model parameters, since DDS is a fast-converging method compared to local gradient-based methods such as the steepest descent algorithm (Huot *et al.* 2019). Tolson and Shoemaker (Tolson and Shoemaker 2007) also proved that DDS outperforms the shuffled complex evolutionary algorithm (Duan *et al.* 1992). For a comprehensive analysis of the search space, we set the maximum number of iterations to 3000 model runs. The framework of the study, consisting of 105 scenarios, is presented in Fig. 5.

4 Results

4.1 Parameter sensitivity analysis and optimized parameters

Table 2 shows the most important 11 parameters using the NSE metric and sorted based on the normalized sensitivities. The normalized values are used to take into account both initial parameter values and raw sensitivity indicators from the Jacobian matrix. This is a more objective way compared to using raw sensitivities directly, since a small change in some very small-valued parameters may have a huge impact on the results, whereas high-valued geo-parameters may have a small raw sensitivity. In this approach, initial parameter values and raw sensitivities are multiplied (4th column) and then normalized by the maximum of this column. The normalized sensitivity value of the most sensitive parameter is 1 in this approach. Around four or five of the 55 parameters were not influential on the streamflow dynamics, and similar parameters were found to be sensitive in other mHM studies in different basins by Demirel *et al.* (2018b). In addition, the model ended up with a different set of parameters

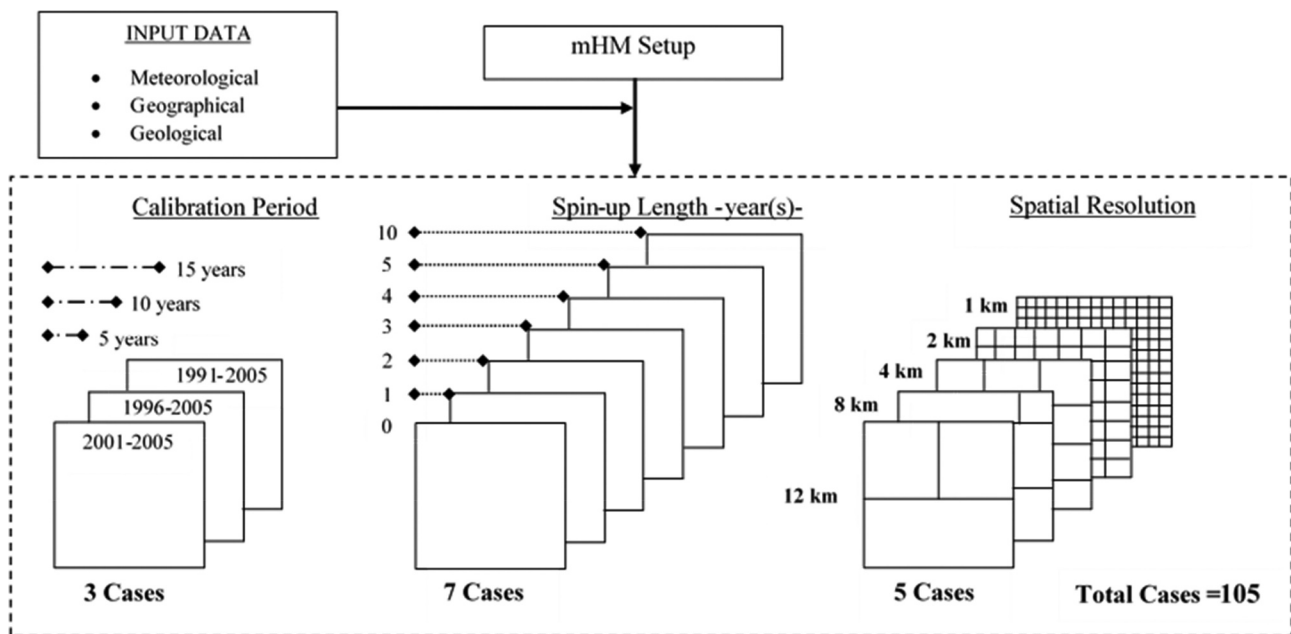


Figure 5. The framework of the study.

Table 2. Most sensitive parameters of meso-scale hydrological model (mHM) based on Nash-Sutcliffe Efficiency (NSE) performance.

| Parameter | Lower bound | Upper bound | Initial value (-) | Raw sensitivity (-) | Abs (init. value * Raw sensitivity) (-) | Normalized sensitivity (-) |
|------------------------------------|-------------|-------------|-------------------|---------------------|---|----------------------------|
| rootFractionCoefficient_forest | 0.9000 | 0.9990 | 0.9878 | 3.0199 | 2.9831 | 1.0000 |
| PTF_Ks_constant | -1.2000 | -0.2850 | -1.3251 | 0.4033 | 0.5344 | 0.1790 |
| rootFractionCoefficient_impervious | 0.9000 | 0.9500 | 0.9352 | 0.4676 | 0.4374 | 0.1470 |
| PTF_lower66_5_constant | 0.6462 | 0.9506 | 0.7518 | 0.3340 | 0.2511 | 0.0840 |
| rechargeCoefficient | 0.0000 | 50.0000 | 6.4266 | 0.0260 | 0.1674 | 0.0560 |
| PTF_Ks_sand | 0.0060 | 0.0260 | 0.0094 | 16.2841 | 0.1527 | 0.0510 |
| PTF_lower66_5_Db | -0.3727 | -0.1871 | -0.3323 | 0.4565 | 0.1517 | 0.0510 |
| exponentSlowInterflow | 0.0500 | 0.3000 | 0.0568 | 2.4514 | 0.1391 | 0.0470 |
| slowInterflowRecession_Ks | 1.0000 | 30.0000 | 13.3225 | 0.0077 | 0.1027 | 0.0340 |
| PTF_Ks_clay | 0.0030 | 0.0130 | 0.0035 | 11.2824 | 0.0399 | 0.0130 |
| GeoParam(4,) | 1.000 | 1000.0000 | 215.6520 | 0.0002 | 0.0335 | 0.0110 |

in each calibration. Table 3 provides the statistical details regarding the optimized parameter values based on the whole set of 105 scenarios.

4.2 Effect of calibration data length on model performance

Figure 6 shows the model performance results in the calibration (left column) and validation (right column) periods as a function of the calibration data length for different spin-up periods and spatial resolutions. Besides the calibration results, we also present validation results as an independent test to evaluate the effects of the 105 cases.

For a 1 km resolution, Fig. 6a shows that the model calibration performance varies depending on the spin-up period when the calibration data length increases from 5 to 15 years. Furthermore, Fig. 6b indicates that the model validation performance increases

with increasing calibration data length independently from the spin-up period. The results obtained for the 4 km resolution show that the model calibration performance decreased when the calibration data length increased from 5 to 15 years, except with a 1-year spin-up period (Fig. 6c). However, the model validation performance increased when the calibration data length increased from 5 to 10 years and did not show a significant change between 10 and 15 years for 4 km resolution (Fig. 6d). Figure 6e shows that the increase in calibration data length from 10 years to 15 years did not lead to significant changes in model calibration performance for an 8 km resolution except with a 0-year spin-up period. In addition, Fig. 6f illustrates that the increase in calibration data length from 10 to 15 years deteriorates the model validation performance for spin-up periods of 4, 5 and 10 years. Overall, a calibration data length of 10 years is sufficient for resolutions of 4 and 8 km, whereas setting the calibration data length to 15 years is required when the spatial resolution of the model is 1 km.

Table 3. Descriptive statistics for the optimized parameter values through 105 scenarios.

| Parameter name | Interval | Minimum | Median | Maximum | Standard Deviation |
|------------------------------------|-------------------|----------|----------|-----------|--------------------|
| canopyInterceptionFactor | [0.15, 0.4] | 0.1518 | 0.3656 | 0.3656 | 0.0809 |
| snowTreshholdTemperature | [-2, 2] | -1.9105 | 1.0450 | 1.9690 | 0.6396 |
| degreeDayFactor_forest | [0.0001, 4] | 0.1002 | 3.5610 | 3.5610 | 1.3649 |
| degreeDayFactor_impervious | [0, 1] | 0.0033 | 0.4840 | 0.9603 | 0.1681 |
| degreeDayFactor_pervious | [0, 2] | 0.1348 | 0.9360 | 1.5761 | 0.3245 |
| increaseDegreeDayFactorByPrecip | [0.1, 0.9] | 0.1455 | 0.5360 | 0.8598 | 0.1490 |
| maxDegreeDayFactor_forest | [0, 8] | 0.0322 | 7.0430 | 7.7957 | 2.7212 |
| maxDegreeDayFactor_impervious | [0, 8] | 4.5037 | 5.0990 | 7.7209 | 1.1559 |
| maxDegreeDayFactor_pervious | [0, 8] | 0.0374 | 6.1900 | 6.5210 | 2.4352 |
| orgMatterContent_forest | [0, 20] | 0.2550 | 14.0399 | 19.4790 | 5.2802 |
| orgMatterContent_impervious | [0, 1] | 0.0111 | 0.6900 | 0.6993 | 0.2684 |
| orgMatterContent_pervious | [0, 4] | 0.4612 | 1.9889 | 3.9425 | 0.7849 |
| PTF_lower66_5_constant | [0.64, 0.95] | 0.6613 | 0.9492 | 0.9492 | 0.0826 |
| PTF_lower66_5_clay | [0.0001, 0.0029] | 0.0001 | 0.0029 | 0.0029 | 0.0010 |
| PTF_lower66_5_Db | [-0.37, -0.18] | -0.3630 | -0.1873 | -0.1873 | 0.0521 |
| PTF_higher66_5_constant | [0.53, 1.12] | 0.6314 | 0.6314 | 1.1153 | 0.1871 |
| PTF_higher66_5_clay | [-0.0055, 0.0049] | -0.0049 | 0.0044 | 0.0044 | 0.0029 |
| PTF_higher66_5_Db | [-0.55, -0.0913] | -0.5033 | -0.1059 | -0.1058 | 0.1032 |
| PTF_Ks_constant | [-1.2, -0.28] | -1.2000 | -1.2000 | -0.2977 | 0.2689 |
| PTF_Ks_sand | [0.006, 0.026] | 0.0063 | 0.0100 | 0.0195 | 0.0027 |
| PTF_Ks_clay | [0.003, 0.013] | 0.0032 | 0.0041 | 0.0085 | 0.0014 |
| PTF_Ks_curveSlope | [0, 100] | 55.9795 | 55.9795 | 60.9600 | 2.4516 |
| rootFractionCoefficient_forest | [0.9, 1] | 0.9501 | 0.9501 | 0.9985 | 0.0197 |
| rootFractionCoefficient_impervious | [0.9, 0.95] | 0.9144 | 0.9333 | 0.9479 | 0.0064 |
| rootFractionCoefficient_pervious | [0.001, 0.09] | 0.0036 | 0.0834 | 0.0834 | 0.4502 |
| infiltrationShapeFactor | [1, 4] | 1.5415 | 1.7766 | 3.6055 | 0.6458 |
| imperviousStorageCapacity | [0, 5] | 0.3801 | 2.0570 | 3.1708 | 0.7701 |
| minCorrectionFactorPET | [0.7, 1.3] | 0.9438 | 1.0620 | 1.2842 | 0.0982 |
| maxCorrectionFactorPET | [0, 0.2] | 0.0554 | 0.1502 | 0.1856 | 0.0328 |
| aspectTreshholdPET | [160, 200] | 166.5593 | 173.0489 | 199.9828 | 6.2721 |
| interflowStorageCapacityFactor | [75, 200] | 75.6743 | 75.6743 | 199.0370 | 54.1785 |
| interflowRecession_slope | [0, 10] | 0.1108 | 3.7160 | 9.2953 | 2.1195 |
| fastInterflowRecession_forest | [1, 3] | 1.1370 | 1.1370 | 2.8340 | 0.5038 |
| slowInterflowRecession_Ks | [1, 30] | 1.5934 | 29.9435 | 29.9435 | 12.4273 |
| exponentSlowInterflow | [0.05, 3] | 0.0502 | 0.0502 | 0.2921 | 0.0905 |
| rechargeCoefficient | [0, 50] | 24.6774 | 28.0913 | 49.9744 | 8.9782 |
| rechargeFactor_karstic | [-5, 5] | -2.2244 | -1.9360 | 4.9481 | 2.7856 |
| gain_loss_GWreservoir_karstic | [1, 1] | 1.0000 | 1.0000 | 1.0000 | 0.0000 |
| slope_factor | [0.1, 100] | 30.0000 | 30.0000 | 30.0000 | 0.0000 |
| GeoParam(1,;) | [1, 1000] | 5.0244 | 5.0244 | 1000.0000 | 388.3279 |
| GeoParam(2,;) | [1, 1000] | 5.6564 | 161.0998 | 1000.0000 | 340.1956 |
| GeoParam(3,;) | [1, 1000] | 71.8306 | 232.4547 | 1000.0000 | 321.0455 |
| GeoParam(4,;) | [1, 1000] | 4.3755 | 4.3755 | 1000.0000 | 386.4119 |
| GeoParam(5,;) | [1, 1000] | 33.1643 | 175.7752 | 1000.0000 | 329.6667 |
| GeoParam(6,;) | [1, 1000] | 100.0000 | 100.0000 | 1000.0000 | 355.1041 |
| GeoParam(7,;) | [1, 1000] | 2.7255 | 11.4506 | 1000.0000 | 391.0604 |
| GeoParam(8,;) | [1, 1000] | 100.0000 | 100.0000 | 1000.0000 | 355.1041 |
| GeoParam(9,;) | [1, 1000] | 10.8886 | 24.3286 | 1000.0000 | 381.9377 |
| GeoParam(10,;) | [1, 1000] | 39.8612 | 672.9531 | 1000.0000 | 263.3028 |

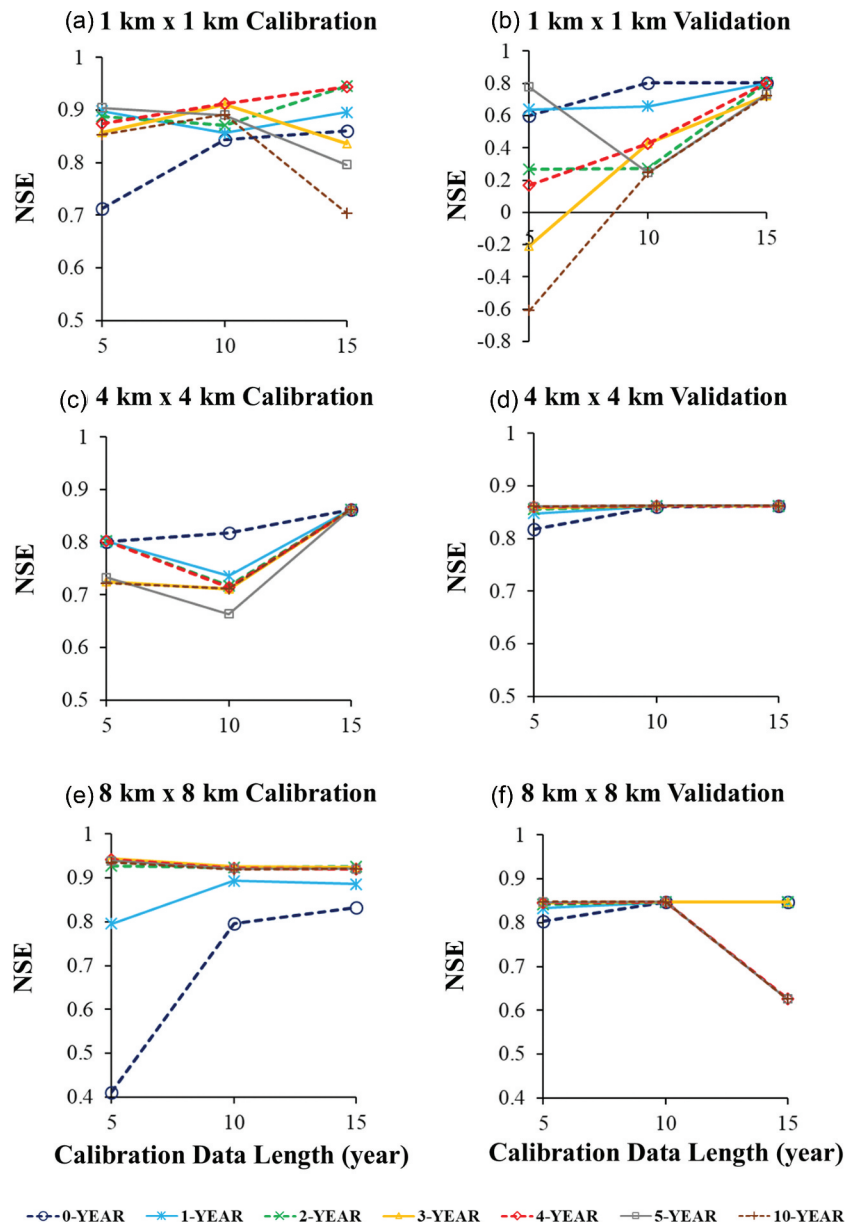


Figure 6. Model results (Nash-Sutcliffe efficiency) as a function of calibration period (length) for different spin-up periods (0 to 10 years) and different spatial resolutions.

4.3 Effect of spin-up period on model performance

Figure 7 highlights the impacts of the different spin-up periods on model performance by means of the NSE for different spatial resolutions and calibration data lengths. It is apparent from Fig. 7 that an increase in spin-up period results in a higher model calibration performance (except in the case with a calibration data length of 5 years and a 1 km resolution) as the model better adapts to the basin states. However, one can observe a decreasing trend in the validation performance when the spin-up period was set between 0 and 5 years, particularly at a spatial resolution of 1 and 2 km, while for a calibration data length of 15 years (Fig. 7f), we see a similar behaviour for almost every spatial resolution (except for a 4 km resolution). Interestingly, for a calibration length of 15 years, from meso to coarse spatial model resolution (from 4 to 12 km), the model calibration performance jumps from an NSE value of 0.4 to 0.9 as the spin-up period increases

from 0 to 2 years (Fig. 7e). With a few exceptions, model calibration and validation results show less sensitivity to changing spin-up periods after 2 years. On the other hand, the model calibration performance with 1 and 2 km resolutions shows high sensitivity to the spin-up period. This is a clear indication of the importance of selecting an appropriate spin-up period for a chosen spatial resolution in a systematic model calibration framework. In summary, considering a calibration data length of 10 years, a spin-up period of 2 years is found to be adequate for the application of mHM to the Moselle River basin.

4.4 Effect of spatial resolution on model performance

Figure 8 shows the variation of NSE in model calibration and validation as a function of spatial resolution. Coloured lines represent different spin-up periods. Two adjacent sub-plots in

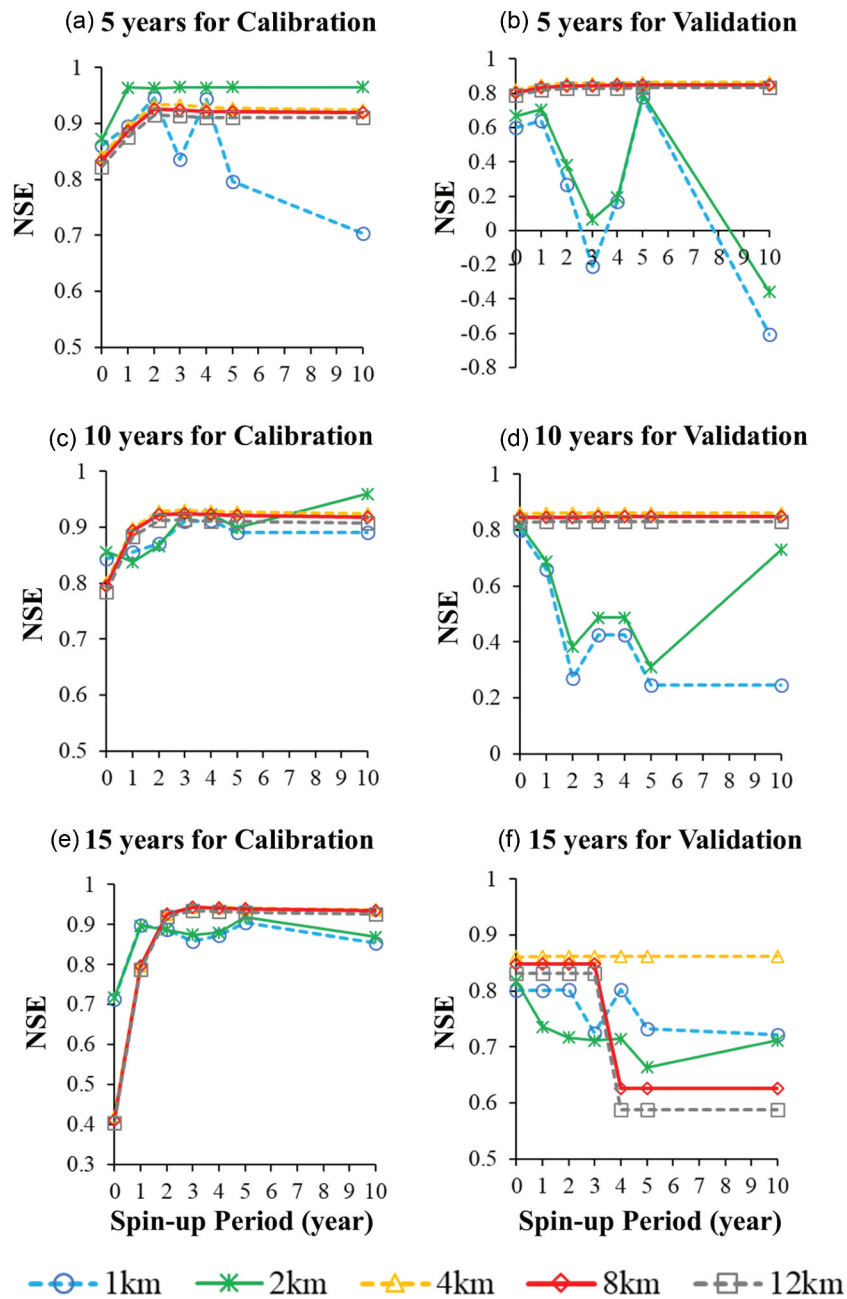


Figure 7. Model results (Nash-Sutcliffe efficiency) as a function of the spin-up period for different spatial resolutions (1, 2, 4, 8 and 12 km) and calibration periods.

each of the three rows illustrate 5, 10 and 15 years of calibration data lengths, respectively. The results obtained for both model calibration and validation demonstrate that the model performance increased as the model resolution increases from 1 to 4 km (except for the validation performance of 15 years' calibration data length). Even though this is contrary to the expectations considering the physical point of view, it may result from the fact that different uncertainties in the input data are less influential (reduced) after averaging data to coarser scales (upscaling). In addition, Fig. 8a depicts that a 2 km spatial resolution gave satisfactory results in model calibration, while the model shows the best validation performance when the spatial resolution is set to 4 km (Fig. 8b). Also, for a calibration data length of 10 years, a 4 km resolution seems the best option for both calibration and validation (Fig. 8 c and d). However, some inconsistencies may exist for shorter spin-up periods (such as a 0-year spin-up period).

What is striking about the cases with a 15-year calibration period is that there is no improvement in model performance beyond a spatial resolution of 4 km (Fig. 8 e and f), as the NSE values tend to decrease towards 8 and 12 km resolutions.

5 Discussion

5.1 Model calibration

Model calibration is usually executed with the available data and computational resources. More data and higher model resolutions are assumed to provide a more realistic simulation requiring less need for model calibration than those with coarser data. In this study, we analysed 105 different model calibrations to identify an appropriate configuration of three pillars, i.e. calibration data length, spin-up period and spatial

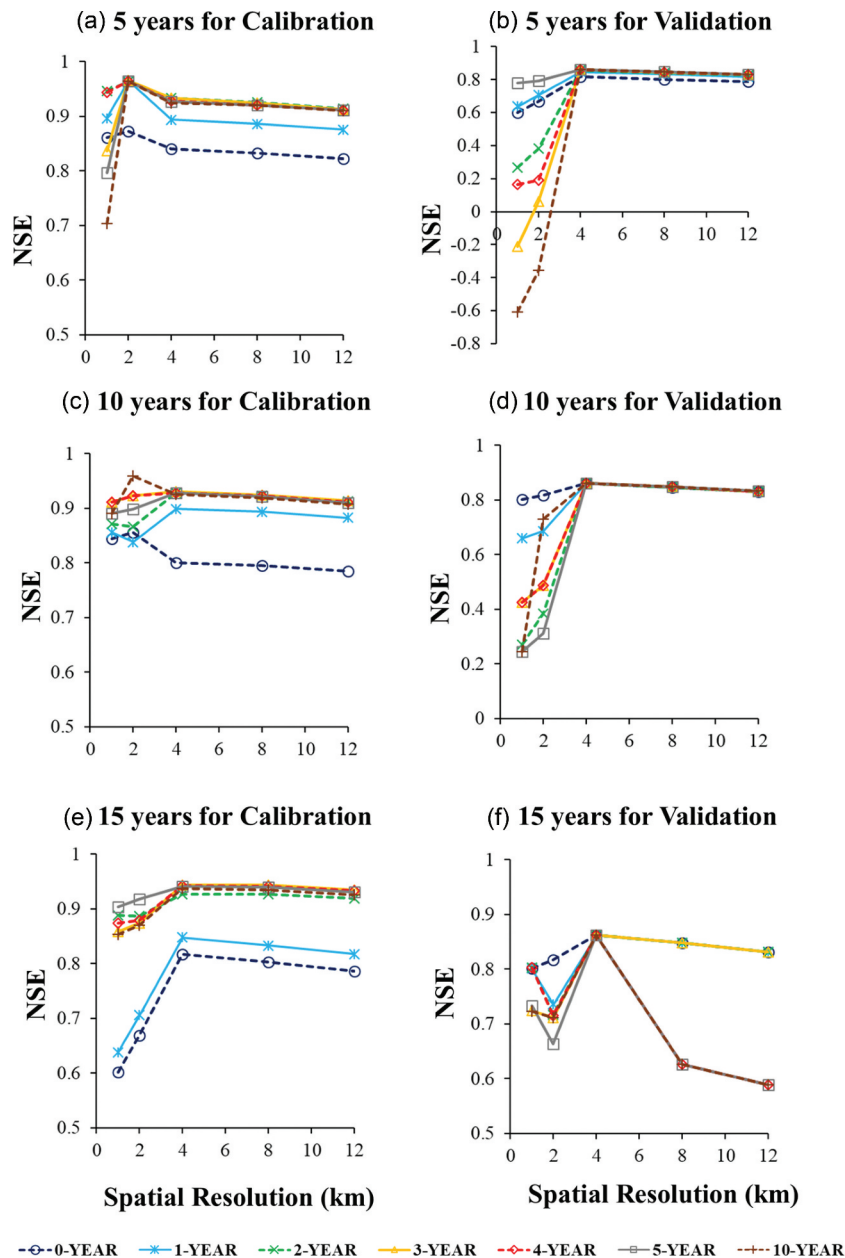


Figure 8. Model results (Nash-Sutcliffe efficiency) as a function of spatial resolution for different spin-up periods (0 to 10 years) and calibration periods.

resolution, for the Moselle basin. We followed a smart sampling approach for the choice of experimental details. For instance, instead of testing all spin-up periods of 1–10 years, we only focused on 0–5 years with 1-year intervals and added an experiment with a 10-year spin-up period as the last case. Similarly, we included only some of the most commonly used spatial model resolutions, i.e. 1, 2 and 4 km. Although we could include more spatial resolutions between 250 m (L0 geographical data resolution) and 24 km (L2 meteorological data resolution) such as 3, 6 and 24 km, we only considered two additional resolutions (8 and 12 km). Testing 11 spin-up periods (i.e. 0–10 years) together with 10 spatial resolutions (i.e. 250 and 500 m, and 1, 2, 3, 4, 6, 8, 12 and 24 km) would enormously increase the number of cases, directly affecting the total duration of the calibration experiments. This would also raise the question of redundancy due to the testing of minor changes in the resolutions and spin-up periods. Furthermore,

the model is capable of upscaling and downscaling of model inputs only for the integer multiplications. For example, the model cannot handle 5×5 km spatial resolution as meteorological data at 24×24 km grid size cannot be divided into 5×5 km pieces.

We used the DDS method, which is available in the model tool, to calibrate our model. To develop a full picture of hydrological model behaviour, additional studies will be needed that consider multi-objective calibrations using Pareto archived DDS (Asadzadeh and Tolson 2009) with additional metrics such as the Kling-Gupta efficiency (Gupta *et al.* 2009) and spatial efficiency (Demirel *et al.* 2018b). We chose a sufficiently large number of iterations (3000 runs) and reached reasonable performance results. Here, our motivation was to scan a wide spectrum of the parameter domain instead of a short calibration with several hundreds of iterations. Also, we only focused on single-gauge temporal calibration with

NSE. Further research should investigate the effect of multi-gauge and spatial model calibrations using spatial efficiency (SPAEF) as the objective function to assess the model performance (Demirel *et al.* 2018b). In addition, illustrating the capability of the proposed framework against the hydrological and meteorological extremities is of significance in terms of robustness and stability of the acquired results (see the Supplementary material for the performance of the mHM in dry and wet conditions that were provided by Hänsel (2020), in Fig. S2). Further attempts should also cover a detailed examination of the impact of extreme hydrological conditions on model performance.

5.2 Effect of three pillars on model performance

Based on the trade-off between available data and computational resources, the modeller has to choose an appropriate combination of the three pillars. In this study, we assessed the effect of each pillar on the model performance. It is somewhat surprising that higher spatial model resolutions (1 and 2 km) lead to a higher sensitivity to the length of the calibration period. For spin-up periods longer than 2 years, the model performance is relatively less sensitive. This indicates that using a longer spin-up period in hydrological simulations does not always have a positive effect on the model performance (see Supplementary material for the variations of model performances based on model calibration and validation on a three-dimensional graph in Figs S3 and S4, respectively). From a physical point of view, the spin-up period should be basin dependent and influenced by factors such as geographical heterogeneity, land cover and use and flow regime. For instance, in rainfed catchments, the performance of hydrological models is relatively higher than that in snowmelt-dominated regions, which can reduce the dependency of the model for longer data length and spin-up period (Demirel *et al.* 2019, Busari *et al.* 2021). However, capturing rainfall heterogeneity at higher spatial resolutions is necessary for better performance. Larger grid-size (coarser spatial resolution) can be used in larger basins, whereas especially for the latter cases (rainfall heterogeneity and complex geology) the need for better quality data and longer time series increases significantly.

Spatial model resolution directly affects the number of cells and the pattern of the hydrological variable, e.g. actual evapotranspiration (AET), over the model domain (Etchevers *et al.* 2001, Booij 2002, Cosgrove *et al.* 2003, Chen *et al.* 2017, Zheng *et al.* 2018). For instance, a single cell with a spatial resolution of 24 km does not provide any pattern of AET depending on the vegetation and soil type. To have a decent histogram of the spatial patterns, resolutions that result in around 1000–2000 cells (pixels) are required to calculate spatial performance, as shown in other basins (Demirel *et al.* 2018b).

5.3 Uncertainties and data

Assessing uncertainties arising from model structure, inputs and parameters is important for assessing the reliability of the results. Model structure uncertainty can be analysed by using multiple models (Demirel *et al.* 2013). Here, we only focused on one distributed model (i.e. mHM) and the E-OBS meteorological dataset. Parameter uncertainty is assumed to be reduced

during the model calibration. There are still many unanswered questions about the model input uncertainty. To compare the effect of input uncertainty on the results, the European Centre for Medium-Range Weather Forecasts Reanalysis v5 (ERA5) meteorological dataset (Hersbach *et al.* 2020) can be used in the model in addition to the E-OBS dataset (Cornes *et al.* 2018). Further, we chose the aspect-based potential ET correction in the model as leaf area index (LAI)-based potential ET correction will be a topic of our future study. It is assumed that the LAI-based potential ET correction would yield better AET estimates and therefore better discharge performance as compared to those with aspect data (Demirel *et al.* 2018b).

Data quality and length can be big issues for modellers from developing countries. Even if the modeller has a long time series with unlimited computational resources, a 10-year portion of the new dataset with a spin-up period of two or three years is sufficient for the model calibration. Then, the remaining – i.e. not wasted – data can be used for model validation (Royer-Gaspard *et al.* 2021).

6 Conclusions

This study was designed to comprehensively investigate the effects of three user-defined model configurations that are usually determined based on local expert knowledge and data availability. We focused on identifying the appropriate length of the calibration period, the length of the spin-up period and the appropriate spatial model resolution for the Moselle River basin. To do so, we used a fully distributed hydrological model (mHM) and performed 105 calibrations with the DDS optimization algorithm and NSE objective function. The 105 cases are combinations of three calibration periods, seven spin-up periods and five spatial model resolutions.

The main conclusions from this work can be summarized as follows:

- Based on the results of the comparison of three calibration data lengths, 10 years is found to be an appropriate length for the Moselle River basin. The interaction between the calibration period and 1–2 km spatial resolution has the strongest effect on the results.
- Based on the results of the comparison of three spin-up periods, two years of spin-up period in addition to the 10 years of calibration data is found to be sufficient for the model to adapt to the initial conditions in the Moselle River basin. Longer spin-up periods than two years did not significantly improve the model calibration and validation performances.
- Based on the results of the comparison of five spatial resolutions, 4 km is found to be the most appropriate model resolution for the Moselle River basin, as the performance slightly deteriorated at coarser resolutions (i.e. 8 and 12 km).

Overall, the three factors analysed in our study are usually overlooked in hydrological modelling. However, the results showed that we should carefully analyse the different combinations of calibration data length, spin-up period and spatial resolution instead of selecting them arbitrarily. It is also worth

noting that although these pillars of a fully distributed physically based hydrological model configuration are related to the extent and quality of the data, as well as basin dependencies, outcomes of the analysis regarding the combined effect of those factors convey a significant insight to prospective practitioners who intended to set different types of the hydrological model up for different regions.

Acknowledgements

We acknowledge the financial support for the SPACE project by the Villum Foundation (<http://villumfonden.dk/>) through their Young Investigator Programme (grant VKR023443). The second author is supported by the National Center for High Performance Computing of Turkey (UHeM) [under Grant number 1007292019] and Ir. Cornelis Lely Stichting [under Grant number 20957310].

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Dr. Ir. Cornelis Lely Stichting [Grant number 20957310]; National Center for High Performance Computing of Turkey (UHeM) [Grant number 1007292019]; Villum Fonden [Grant number VKR023443].

ORCID

Ömer Ekmeçcioğlu  <http://orcid.org/0000-0002-7144-2338>
Mehmet Cüneyd Demirel  <http://orcid.org/0000-0003-4402-906X>
Martijn J. Booij  <http://orcid.org/0000-0001-6208-9045>

References

- Abdo, K.S., *et al.*, 2009. Assessment of climate change impacts on the hydrology of Gilgel Abay catchment in Lake Tana Basin, Ethiopia. *Hydrological Processes*, 2274 (November 2008), n/a–n/a. doi:10.1002/hyp.7363
- Ajami, H., *et al.*, 2014. Technical Note: reducing the spin-up time of integrated surface water–groundwater models. *Hydrology and Earth System Sciences*, 18 (12), 5169–5179. doi:10.5194/hess-18-5169-2014
- Arsenault, R., Brissette, F., and Martel, J.-L., 2018. The hazards of split-sample validation in hydrological model calibration. *Journal of Hydrology*, 566, 346–362. doi:10.1016/j.jhydrol.2018.09.027
- Asadzadeh, M. and Tolson, B.A. (2009) A new multi-objective algorithm, pareto archived DDS. *Proceedings of the 11th annual conference companion on Genetic and evolutionary computation conference - GECCO '09*, 1963. New York, New York, USA: ACM Press. 10.1145/1570256.1570259
- Ashraf, A., 2013. Changing hydrology of the himalayan watershed. In: P. M. Bradley ed. *Current perspectives in contaminant hydrology and water resources sustainability*. London: IntechOpen. doi:10.5772/54492
- Ballabio, C., Panagos, P., and Monatanarella, L., 2016. Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma*, 261, 110–123. doi:10.1016/j.geoderma.2015.07.006
- Baroni, G., *et al.*, 2019. A comprehensive distributed hydrological modeling intercomparison to support process representation and data collection strategies. *Water Resources Research*, 55 (2), 990–1010. doi:10.1029/2018WR023941
- Blöschl, G. and Sivapalan, M., 1995. Scale issues in hydrological modeling: a review. *Hydrological Processes*, 9 (3–4), 251–290.
- Booij, M.J., 2002. Modelling the effects of spatial and temporal resolution of rainfall and basin model on extreme river discharge. *Hydrological Sciences Journal*, 47 (2), 307–320. doi:10.1080/02626660209492932
- Booij, M.J., 2005. Impact of climate change on river flooding assessed with different spatial model resolutions. *Journal of Hydrology*, 303 (1–4), 176–198. doi:10.1016/j.jhydrol.2004.07.013
- Brenot, A., *et al.*, 2007. Geological and land use control on $\delta^{34}\text{S}$ and $\delta^{18}\text{O}$ of river dissolved sulfate: the Moselle river basin, France. *Chemical Geology*, 244 (1–2), 25–41. doi:10.1016/j.chemgeo.2007.06.003
- Bucchignani, E., *et al.*, 2016. Analysis of ERA-Interim-driven COSMO-CLM simulations over Middle East - North Africa domain at different spatial resolutions. *International Journal of Climatology*, 36 (9), 3346–3369. doi:10.1002/joc.4559
- Busari, I.O., Demirel, M.C., and Newton, A., 2021. Effect of using multi-year land use land cover and monthly LAI inputs on the calibration of a distributed hydrologic model. *WATER*, 13 (11), 1538. doi:10.3390/w13111538
- Chen, Y., *et al.*, 2017. Large-watershed flood forecasting with high-resolution distributed hydrological model. *Hydrology and Earth System Sciences*, 21 (2), 735–749. doi:10.5194/hess-21-735-2017
- Cornes, R.C., *et al.*, 2018. An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123 (17), 9391–9409. doi:10.1029/2017JD028200
- Cosgrove, B.A., *et al.*, 2003. Land surface model spin-up behavior in the North American Land Data Assimilation System (NLDAS). *Journal of Geophysical Research D: Atmospheres*, 108 (22). doi:10.1029/2002jd003316
- Cuntz, M., *et al.*, 2015. Computationally inexpensive identification of noninformative model parameters by sequential screening. *Water Resources Research*, 51 (8), 6417–6441. doi:10.1002/2015WR016907
- Cuo, L., *et al.*, 2006. Use of the distributed hydrology soil vegetation model to study road effects on hydrological processes in Pang Khum Experimental Watershed, northern Thailand. *Forest Ecology and Management*, 224 (1–2), 81–94. doi:10.1016/j.foreco.2005.12.009
- Daggupati, P., *et al.*, 2015. A recommended calibration and validation strategy for hydrologic and water quality models. *Transactions of the ASABE*, 58 (6), 1705–1719. doi:10.13031/trans.58.10712
- Dehotin, J. and Braud, I., 2008. Which spatial discretization for distributed hydrological models? Proposition of a methodology and illustration for medium to large-scale catchments. *Hydrology and Earth System Sciences*, 12 (3), 769–796. doi:10.5194/hess-12-769-2008
- Dembélé, M., *et al.*, 2020. Improving the predictive skill of a distributed hydrological model by calibration on spatial patterns with multiple satellite data sets. *Water Resources Research*, 56 (1). doi:10.1029/2019WR026085
- Demirel, M.C., *et al.*, 2018b. 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, 10.5194/hess-22-1299-2018, Copernicus Publications
- Demirel, M.C., Booij, M.J., and Hoekstra, A.Y., 2013. Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models. *Water Resources Research*, 49 (7), 4035–4053. doi:10.1002/wrcr.20294
- Doherty, J., 2010. *PEST, Model-Independent Parameter Estimation—User Manual*. 5th Edition, with Slight Additions, Watermark Numerical Computing, Brisbane.
- Duan, Q.Q.-Y.-Y., Sorooshian, S., and Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28 (4), 1015–1031. doi:10.1029/91WR02985
- Eisner, S., *et al.*, 2017. An ensemble analysis of climate change impacts on streamflow seasonality across 11 large river basins. *Climatic Change*, 141 (3), 401–417. Climatic Change: 10.1007/s10584-016-1844-5.
- Epstein, D.J., Welles, E., and Day, G.N. (1998) Probabilistic Hydrologic Forecasting Methods and Tools. *Proceedings of the First Federal Interagency Hydrologic Modeling Conference*, 6(17–24). Las Vegas, Nevada: Subcommittee on Hydrology of the Interagency Advisory Committee on Water Data.
- Etchevers, P., *et al.*, 2001. Impact of spatial resolution on the hydrological simulation of the Durance high-Alpine catchment, France. *Annals of Glaciology*, 32, 87–92. doi:10.3189/172756401781819337
- Fischer, G., *et al.*, 2012. *Global Agro-Ecological Zones (GAEZ): Model Documentation 1–179*.

- Gelleszun, M., Kreye, P., and Meon, G., 2017. Representative parameter estimation for hydrological models using a lexicographic calibration strategy The Authors. *Journal of Hydrology*, 553, 722–734. doi:10.1016/j.jhydrol.2017.08.015
- Gharari, S., et al., 2013. An approach to identify time consistent model parameters: sub-period calibration. *Hydrology and Earth System Sciences*, 17 (1), 149–161. doi:10.5194/hess-17-149-2013
- Girard, M.-C., et al., 2019. Corine Land Cover. *Processing of Remote Sensing Data*. doi:10.1201/9780203741917-19
- Gires, A., et al., 2015. Impacts of small scale rainfall variability in urban areas: a case study with 1D and 1D/2D hydrological models in a multifractal framework. *Urban Water Journal*, 12(8), 607–617, 10.1080/1573062X.2014.923917, Taylor & Francis
- Gupta, H.V., et al., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology*, 377(1–2), 80–91, 10.1016/j.jhydrol.2009.08.003, Elsevier
- Hänsel, S., 2020. Changes in the characteristics of dry and wet periods in Europe (1851–2015). *Atmosphere*, 11 (10), 1080. doi:10.3390/atmos11101080
- Haylock, M.R., et al., 2008. A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research Atmospheres*, 113 (D20). doi:10.1029/2008JD010201
- Hersbach, H., et al., 2020. The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146 (730), 1999–2049. doi:10.1002/qj.3803
- Höllering, S., et al., 2018. Regional analysis of parameter sensitivity for simulation of streamflow and hydrological fingerprints. *Hydrology and Earth System Sciences*, 22 (1), 203–220. doi:10.5194/hess-22-203-2018
- Huang, S., et al., 2018. Multimodel assessment of flood characteristics in four large river basins at global warming of 1.5, 2.0 and 3.0 K above the pre-industrial level. *Environmental Research Letters*, 13 (12), 10.1088/1748-9326/aae94b. IOP Publishing doi:10.1088/1748-9326/aae94b
- Huot, P.-L., et al., 2019. A hybrid optimization approach for efficient calibration of computationally intensive hydrological models. *Hydrological Sciences Journal*, 64(10), 1204–1222, 10.1080/02626667.2019.1624922, Taylor & Francis
- Jing, M., et al., 2019. Influence of input and parameter uncertainty on the prediction of catchment-scale groundwater travel time distributions. *Hydrology and Earth System Sciences*, 23 (1), 171–190. doi:10.5194/hess-23-171-2019
- Kim, K.B., Kwon, H.H., and Han, D., 2018. Exploration of warm-up period in conceptual hydrological modelling. *Journal of Hydrology*, 556, 194–210. doi:10.1016/j.jhydrol.2017.11.015
- Knoben, W.J.M., Freer, J.E., and Woods, R.A., 2019. Technical note: inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23 (10), 4323–4331. doi:10.5194/hess-23-4323-2019
- Koc, K., Ekmekcioğlu, Ö., and Özger, M., 2021. An integrated framework for the comprehensive evaluation of low impact development strategies. *Journal of Environmental Management*, 294, 113023. doi:10.1016/j.jenvman.2021.113023
- Koren, V.I., et al., 1999. Scale dependencies of hydrologic models to spatial variability of precipitation. *Journal of Hydrology*, 217 (3–4), 285–302. doi:10.1016/S0022-1694(98)00231-5
- Kumar, R., Samaniego, L., and Attinger, S., 2013. Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resources Research*, 49 (1), 360–379. doi:10.1029/2012WR012195
- Liu, D., et al., 2018. Statistics for sample splitting for the calibration and validation of hydrological models. *Stochastic Environmental Research and Risk Assessment*, 32 (11), 3099–3116. doi:10.1007/s00477-018-1539-8
- Lohmann, D., et al., 1998. Hydrologie à l'échelle régionale: II. Application du modèle VIC-2L sur la rivière Weser, Allemagne. *Hydrological Sciences Journal*, 43 (1), 143–158. doi:10.1080/02626669809492108
- Madsen, H., 2003. Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Advances in Water Resources*, 26 (2), 205–216. doi:10.1016/S0309-1708(02)00092-1
- Maina, F.Z., et al., 2020. Determining the impact of a severe dry to wet transition on watershed hydrodynamics in California, USA with an integrated hydrologic model. *Journal of Hydrology*, 580, 124358. doi:10.1016/j.jhydrol.2019.124358
- Marx, A., et al., 2017. Climate change alters low flows in Europe under a 1.5, 2, and 3 degree global warming. *Hydrology and Earth System Sciences Discussions*, 1–24. doi:10.5194/hess-2017-485
- Merz, R., Parajka, J., and Blöschl, G., 2009. Scale effects in conceptual hydrological modeling. *Water Resources Research*, 45 (9), W09405. AGU. doi:10.1029/2009wr007872
- Nicótina, L., et al., 2008. On the impact of rainfall patterns on the hydrologic response. *Water Resources Research*, 44 (12). doi:10.1029/2007WR006654
- Nijssen, B., et al., 2001. Predicting the discharge of global rivers. *Journal of Climate*, 14 (15), 3307–3323. doi:10.1175/1520-0442(2001)014<3307:PTDOGR>2.0.CO;2
- Nkiaka, E., Nawaz, N., and Lovett, J., 2017. Evaluating Global Reanalysis Datasets as Input for Hydrological Modelling in the Sudano-Sahel Region. *Hydrology*, 4 (1), 13. doi:10.3390/hydrology4010013
- Nossent, J., Elsen, P., and Bauwens, W., 2011. Sobol' sensitivity analysis of a complex environmental model. *Environmental Modelling and Software*, 26(12), 1515–1525, 10.1016/j.envsoft.2011.08.010, Elsevier Ltd
- Pang, J., et al., 2020. Hydrological evaluation of open-access precipitation data using SWAT at multiple temporal and spatial scales. *Hydrology and Earth System Sciences*, 24 (7), 3603–3626. doi:10.5194/hess-24-3603-2020
- Perrin, C., et al., 2007. Impact of limited streamflow data on the efficiency and the parameters of rainfall–runoff models. *Hydrological Sciences Journal*, 52(1), 131–151, 10.1623/hysj.52.1.131, Taylor & Francis
- Rahman, M.M., Lu, M., and Kyi, K.H., 2016. Seasonality of hydrological model spin-up time: a case study using the Xinanjiang model. *Hydrology and Earth System Sciences Discussions*, 1–22. doi:10.5194/hess-2016-316
- Rakovec, O., et al., 2016. Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins. *Journal of Hydrometeorology*, 17 (1), 287–307. doi:10.1175/JHM-D-15-0054.1
- Razavi, S. and Tolson, B.A., 2013. An efficient framework for hydrologic model calibration on long data periods. *Water Resources Research*, 49 (12), 8418–8431. doi:10.1002/2012WR013442
- Revilla-Romero, B., et al., 2016. Integrating remotely sensed surface water extent into continental scale hydrology The Authors. *Journal of Hydrology*, 543, 659–670. doi:10.1016/j.jhydrol.2016.10.041
- Rodell, M., et al., 2005. Evaluation of 10 methods for initializing a land surface model. *Journal of Hydrometeorology*, 6 (2), 146–155. doi:10.1175/JHM414.1
- Royer-Gaspard, P., Andréassian, V., and Guillaume, T., 2021. Technical note: PMR – a proxy metric to assess hydrological model robustness in a changing climate. *Hydrol. Earth Syst. Sci. Discuss., Preprint*. doi:10.5194/hess-2021-58
- Samaniego, Brenner, J., et al., 2021. Mesoscale Hydrologic Model - mHM v5.11.1. *Leipzig*, doi:10.5281/ZENODO.4462822
- Samaniego, L., et al., 2018. Anthropogenic warming exacerbates European soil moisture droughts. *Nature Climate Change*, 8 (5), 421–426. Springer US: 10.1038/s41558-018-0138-5.
- Samaniego, L., Kumar, R., and Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, 46 (5), W05523. doi:10.1029/2008WR007327
- Shi, X., Wood, A.W., and Lettenmaier, D.P., 2008. How Essential is Hydrologic Model Calibration to Seasonal Streamflow Forecasting? *Journal of Hydrometeorology*, 9 (6), 1350–1363. doi:10.1175/2008JHM1001.1
- Shrestha, R. and Houser, P., 2010. A heterogeneous land surface model initialization study. *Journal of Geophysical Research Atmospheres*, 115 (D19), 1–10. doi:10.1029/2009JD013252
- Simunek, J.J., Van Genuchten, M.T., and Šejna, M., 2012. HYDRUS: model Use, Calibration, and Validation. *Transactions of the ASABE*, 55 (4), 1263–1276. doi:10.13031/2013.42239
- Sood, A., Muthuwatta, L., and McCartney, M., 2013. A SWAT evaluation of the effect of climate change on the hydrology of the Volta River basin. *Water International*, 38 (3), 297–311. doi:10.1080/02508060.2013.792404

- Sood, A. and Smakhtin, V., 2015. Revue des modèles hydrologiques globaux. *Hydrological Sciences Journal*, 60(4), 549–565. doi:10.1080/02626667.2014.950580, Taylor & Francis
- Sorooshian, S., Gupta, H.V., and Rodda, J.C., 1997. *Land Surface Processes in Hydrology: Trials and Tribulations of Modeling and Measuring*. S. Sorooshian, H.V. Gupta, and J.C. Rodda, Eds. Berlin, Heidelberg: Springer 10.1007/978-3-642-60567-3
- Sreedevi, S. and Eldho, T.I., 2019. A two-stage sensitivity analysis for parameter identification and calibration of a physically-based distributed model in a river basin. *Hydrological Sciences Journal*, 64(6), 701–719. doi:10.1080/02626667.2019.1602730, Taylor & Francis
- Tayşi, Demirel, Ö., et al., 2019. Additional value of using satellite-based soil moisture and two sources of groundwater data for hydrological model calibration. *Water*, 11 (10), 2083. doi:10.3390/w11102083
- Thober, S., et al., 2019. The multiscale routing model mRM v1.0: simple river routing at resolutions from 1 to 50 km. *Geoscientific Model Development*, 12 (6), 2501–2521. doi:10.5194/gmd-12-2501-2019
- Tian, J., et al., 2020. A coupled atmospheric-hydrologic modeling system with variable grid sizes for rainfall-runoff simulation in semi-humid and semi-arid watersheds: how does the coupling scale affects the results? *Hydrology and Earth System Sciences Discussions*, 1–36. doi:10.5194/hess-2019-587
- Tolson, B.A. and Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, 43 (1). doi:10.1029/2005WR004723
- Uehlinger, U., et al., 2009. The Rhine River Basin. In: *Rivers of Europe*. Elsevier, 199–245.
- Wallace, C., Flanagan, D., and Engel, B., 2018. Evaluating the Effects of Watershed Size on SWAT Calibration. *Water*, 10 (7), 898. doi:10.3390/w10070898
- Westerberg, I.K., et al., 2020. Hydrological model calibration with uncertain discharge data. *Hydrological Sciences Journal*. Taylor & Francis, 1–16. doi:10.1080/02626667.2020.1735638
- Xu, Y.-P., et al., 2013. Impact of climate change on hydrology of upper reaches of Qiantang River Basin, East China. *Journal of Hydrology*, 483, 51–60. doi:10.1016/j.jhydrol.2013.01.004
- Yang, Z.-L., et al., 1995. Preliminary study of spin-up processes in land surface models with the first stage data of project for inter-comparison of land surface parameterization schemes phase 1(a). *Journal of Geophysical Research*, 100 (D8), 16553. doi:10.1029/95jd01076
- Zhang, Z., et al., 2020. Coupled hydrological and biogeochemical modelling of nitrogen transport in the karst critical zone. *Science of the Total Environment*, 732, 138902. doi:10.1016/j.scitotenv.2020.138902
- Zheng, F., et al., 2018. On lack of robustness in hydrological model development due to absence of guidelines for selecting calibration and evaluation data: Demonstration for data-driven models. *Water Resources Research*, 54 (2), 1013–1030. doi:10.1002/2017WR021470