Corrected Proof

Hydrology Research

Hydrology Research Vol 00 No 0, 1 doi: 10.2166/nh.2021.152

Daily vs. hourly simulation for estimating future flood peaks in mesoscale catchments

Marcus Beylich[a], Uwe Haberlandt[b,*] and Frido Reinstorf[a]

[a] Department of Water, Environment, Construction and Safety, Magdeburg-Stendal University of Applied Sciences, 39114 Magdeburg, Germany
[b] Institute of Hydrology and Water Resources Management, Leibniz University Hannover, 30167 Hannover, Germany
*Corresponding author. E-mail: haberlandt@iww.uni-hannover.de

ABSTRACT

Daily hydrological models are commonly used to study changes in flood peaks due to climate change. Although they often lead to an underestimation of absolute floods, it is assumed that future flood peaks in smaller mesoscale catchments are less underestimated when examining the relative change signal of floods. In this study, the applicability of this hypothesis is investigated by comparing the results of a daily hydrological model set, calibrated on runoff hydrographs, with an hourly model set calibrated on flood peak distributions. For analysis, a daily RCP8.5 climate model ensemble is disaggregated to hourly values and the runoff is simulated on a daily and hourly basis for six mesoscale catchments in Central Germany. Absolute floods and relative flood changes are compared between both model sets. The results show significant differences between the absolute floods of both model sets, in most cases caused by underestimations due to the daily modeling process. In contrast, the differences between the two model sets are not significant for the relative change signal of the floods, especially for higher return periods. To improve results in climate studies with coarse modeling time step, the use of relative change signal of floods instead of absolute values is recommended.

Key words: climate change, disaggregation, flood modeling, mesoscale, modeling time step

HIGHLIGHTS

- Flood peak changes in mesoscale catchments simulated using daily/hourly modeling time steps by disaggregating a daily RCP8.5 climate model ensemble.
- Future absolute flood values and relative flood change signals compared between daily/hourly model results with significance testing.
- Results: underestimation of absolute flood values (daily); no significant differences for relative flood change signals between daily/hourly models.

INTRODUCTION

In recent climate impact studies on future flood development, expected changes in large catchment areas are often examined by hydrological models on a daily basis. This is particularly the case in more recent scenario-neutral approaches for climate impact studies, which require many simulation runs to explore the response surface of a catchment (e.g., Prudhomme *et al.* 2010, 2013; Vormoor *et al.* 2017; Keller *et al.* 2019). In that field, studies are often based on weather generators, which are already associated with great computational efforts on a daily basis due to the complex parametrization as well as long calculation times (e.g., Steinschneider & Brown 2013; Guo *et al.* 2018). Studies with higher temporal resolution are extremely rare and often conducted for special cases and using simple calculation methods (e.g., Kim *et al.* 2018). Studies based on more common scenario-based approaches would require high temporal resolution input data from regional climate model ensembles, which usually provide daily time series.

Long-time high-resolution data with sub-daily time steps are important to calibrate hydrological models with sufficient accuracy. Therefore, another reason for using daily time steps in climate studies might be the lack of observed sub-daily data extending back to before the 1990s. Since then, most of the data are collected automatically in the digital form (Ficchì *et al.* 2016). Numerous studies revealed the relevance of the temporal distribution of rainfall for the correct generation of peak runoff and runoff volume, depending on the characteristics of the catchments (e.g., Gabellani *et al.* 2007; Viglione *et al.* 2010; Paschalis *et al.* 2014). By using models with a low temporal resolution, flood peaks are likely to be underestimated. To improve daily results, methods have been elaborated to estimate instantaneous peak flows from mean and maximum daily flows, often regarded to important catchment properties like catchment size, flow path or slope (e.g., Ding *et al.* 2015; Chen *et al.* 2017). Another way for providing high-resolution data is using disaggregation methods, in particular for rainfall (e.g., Olsson 1998; Müller & Haberlandt 2015), or stochastic models. However, the application of these methods is also demanding regarding computation time, in particular for scenario-neutral climate impact studies, which require many simulation experiments.

In most of the described daily scenario-neutral climate impact studies investigating the development of flood peaks and flood distribution, the objective is the relative change of floods from the present time to the future. More precisely, a climate change signal of floods is investigated instead of examining absolute runoff values, which is supposed to lower or negate a possible underestimation of the floods. Therefore, the aim of this study is to examine the applicability of commonly used daily models for investigating climate change signals of flood peak values and absolute flood peaks in comparison to hourly models for mesoscale study sites in Central Germany. Therefore, floods with different return periods are investigated, hereinafter referred to as RP floods.

This study structures as follows. In the beginning, measurement data, climate ensemble data consisting of 14 global-regional climate model chains and the investigated six catchments are further described. In the next section, the methods of this study are introduced. For simulating the past and future runoff, the applied rainfall-runoff model HBV and its components are shortly described. Because of a good daily database, but lacks in hourly data, the daily time series are disaggregated to hourly ones, using harmonic approaches for temperature and evapotranspiration and a multiplicative random cascade for precipitation, which is described afterwards. Then the calibration and validation of the two used model sets are described. One model set, based on daily data, is designed for reproducing the daily hydrograph, while the other model set is set up for reproducing the peak flow distribution, using hourly input time series. The statistical analysis of the simulated floods and its changes to be investigated, are also explained. Afterwards, the calibration results of both model sets are evaluated. The absolute RP floods and relative RP flood changes, simulated with both model sets, are compared and the differences are tested for significance. At the end, the results are summarized and concluded.

STUDY AREA AND DATA

Measurement data and study sites

Figure 1 shows the location of investigated catchments and the corresponding river gauges. Six catchments, located in the Harz Mountains and its foreland in Central Germany, are explored in this study: the river Rappbode at Trautenstein, the river Holtemme at Mahndorf and the river Großer Graben at Oschersleben. The investigation of the river Selke is explored by three subcatchments at three different gauges: Silberhütte, Meisdorf and Hausneindorf. The main land use of the plane areas Oschersleben, Hausneindorf and the lower parts of Holtemme is agriculture, while in the mountainous areas forest is dominating.

For important meteorological values, sufficiently long daily time series are available for simulation. The mean annual values for temperature and precipitation are shown in Table 1. The observed daily climate data for the catchments are taken from a local climate project in the Federal State Lower Saxony (NLWKN 2017). In the project, point station climate data, collected from the German Weather Service (DWD), are interpolated to space, selecting the best results from different geostatistical methods. Missing data, especially for the outlying Selke catchments, were complemented. For all gauges, sufficient long daily discharge data were made available by the governmental authority for water management of the Federal State of Saxony-Anhalt (Landesbetrieb für Hochwasserschutz, LHW). For model calibration, the selected time period covers 30 years from 1981 to 2010 (1987–2017 for Oschersleben catchment).

Higher-resolution data for precipitation with sufficient quality is available only for a limited time period, varying between the stations, as depicted in Figure 1. Hourly station data of precipitation are interpolated to the catchments using inverse distance weighting. These time series are only used for estimating the parameters of MRC. The maximum runoff peak values for every month are available for a time series as long as or longer than the daily runoff data.

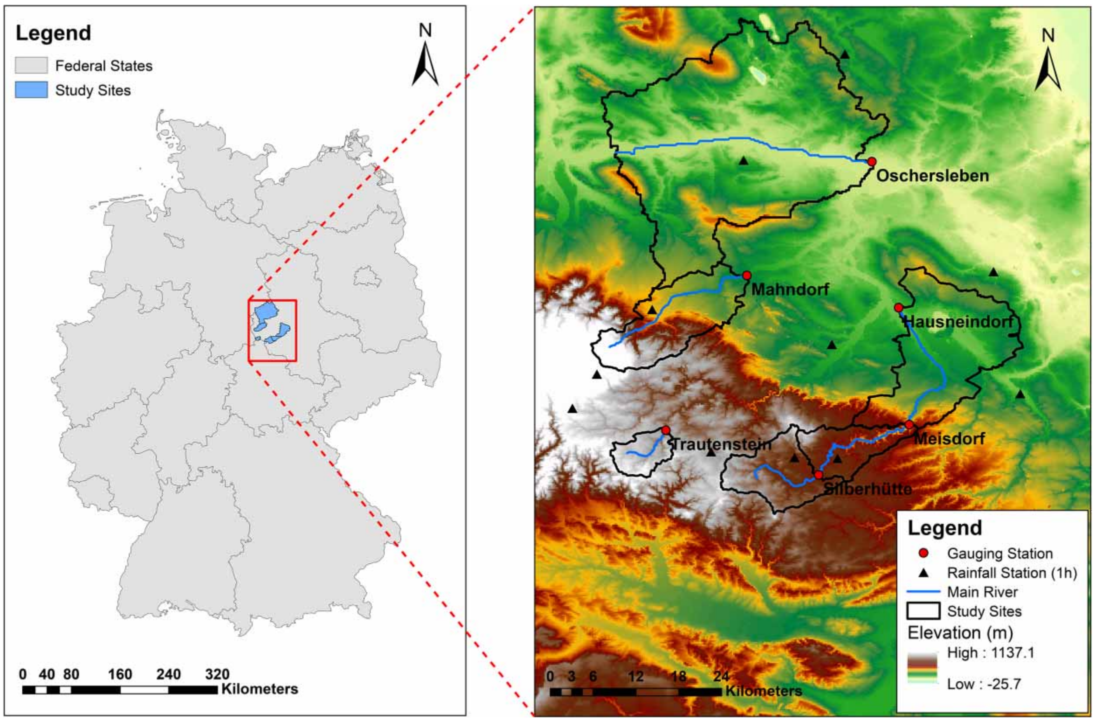

Figure 1 | Location of study sites and stations in Central Germany.

Table 1 | General characteristics of the investigated catchments with mean annual meteorological data (30 years mean) and mean annual runoff data (20 years mean)

		Percentage land use				
Catchment (gauge)	**Size (km^2)**	**Forest**	**Agriculture, grassland**	**Annual temperature mean (°C)**	**Annual precipitation sum (mm)**	**Annual runoff sum (mm)**
Trautenstein	39.1	74.7%	20.5%	6.6	1,024	658
Mahndorf	162.0	45.7%	40.2%	7.7	843	275
Oschersleben	823.5	7.5%	86.3%	9.1	625	83
Silberhütte	102.3	58.3%	35.7%	6.8	789	364
Meisdorf[a]	77.8	85.7%	9.8%	7.3	660	117
Hausneindorf[a]	271.8	12.8%	73.7%	8.9	652	30

[a] Values for the subcatchment area of the respective gauge.

Climate model ensemble

As climate models, a multi-model ensemble consisting of 14 global-regional model chains from EURO-CORDEX database is selected (see Table 2). The data were made available by the Lower Saxony Water Management, Coastal Defence and Nature Conservation Agency (Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz, NLWKN) in a grid of 10 × 10 km (NLWKN 2017). All climate models are dynamically based and rely on the concept of the representative concentration pathway (RCP) scenarios, according to the current 5 IPCC report (AR5; IPCC 2013). For this study, the RCP8.5 scenario was investigated, representing a base scenario without additional efforts to limit the high global greenhouse gas emissions, commonly known as 'continue as before' scenario.

The climate series are aggregated from the grid to the catchments using inverse distant interpolation. Aggregated time series are used due to the low spatial resolution of processes in the RCM, which are usually larger than the existing grid point spacing. In addition, the climate model time series show biases from the long-term measured values in the catchments. Bias correction is commonly necessary in many climate change studies due to inaccuracies in the recent climate model approaches. There are many approaches for correction, but

Table 2 | Global-regional climate model chains used in this study

Global Model (GCM)	Regional Model (RCM)	Acronym in this study	Data length (Years)	Assigned time slices in study (hydrol. year) Past	Near Future	Far Future
CNRM-CM5	CCLM-4-8-17	CNRM_CCLM	1970–2100			2071–2100
CNRM-CM5	RCA4	CNRM_RCA4	1970–2100			2071–2100
EC-EARTH	CCLM-4-8-17	ECE_CCLM	1970–2100			2071–2100
EC-EARTH	HIRHAM5	ECE_HIRHAM	1951–2100			2071–2100
EC-EARTH	RACMO22E	ECE_RACMO22E	1951–2100			2071–2100
EC-EARTH	RCA4	ECE_RCA4	1970–2100			2071–2100
HadGEM2-ES	RACMO22E	HadGEM2_RAMCO22E	1970–2099	1981–2010	2021–2050	2070–2099
HadGEM2-ES	RCA4	HadGEM2_RCA4	1970–2099			2070–2099
IPSL-CM5A-MR	RCA4	IPSL_RCA4	1970–2100			2071–2100
IPSL-CM5A-MR	WRF331F	IPSL_WRF331F	1971–2100			2071–2100
MPI-ESM-LR	CCLM-4-8-17	MPI_CCLM	1951–2100			2071–2100
MPI-ESM-LR	RCA4	MPI_RCA4	1970–2100			2071–2100
MPI-ESM-LR	REMO (1)[a]	MPI_REMO1	1951–2100			2071–2100
MPI-ESM-LR	REMO (2)[a]	MPI_REMO2	1951–2100			2071–2100

All model chains from RCP8.5-scenario. Daily data series originally available in ~ 12.5 × 12.5 km resolution.

[a] Different REMO runs.

they can also change the interaction between the various climate variables. In this study, only temperature and precipitation are bias corrected in a simple scaling approach, assuming that both variables are the most important ones for future changes in flood peak values. Bias correction is calculated, using the following scaling equations (NLWKN 2017):

$$P^*_{\mathrm{sim}}(d) = P_{\mathrm{sim}}(d) \cdot \left(\frac{\bar{P}_{\mathrm{obs,\,ref}}(m)}{\bar{P}_{\mathrm{sim,\,ref}}(m)} \right) \tag{1}$$

$$T^*_{\mathrm{sim}}(d) = T_{\mathrm{sim}}(d) + (\bar{T}_{\mathrm{obs,\,ref}}(m) - \bar{T}_{\mathrm{sim,\,ref}}(m)) \tag{2}$$

In the equations, $P_{\mathrm{sim}}(d)$ is the simulated daily precipitation sum and $T_{\mathrm{sim}}(d)$ is the simulated daily mean temperature. The bias-corrected values are $P^*_{\mathrm{sim}}(d)$ and $T^*_{\mathrm{sim}}(d)$. $\bar{P}_{\mathrm{obs,\,ref}}(m)$ and $\bar{T}_{\mathrm{obs,\,ref}}(m)$ are the observed and $\bar{P}_{\mathrm{sim,\,ref}}(m)$ and $\bar{T}_{\mathrm{sim,\,ref}}(m)$ are the monthly mean values of the climate models in the reference period (past). This scaling approach changes all values in a similar way, according to the deviations in the monthly means, following that frequencies as well as the amount and length of dry and wet days are remaining constant. However, no change of climate signal according to the mean of climate values is added and the applied means of temperature and precipitation fit well to the observed ones.

METHODS

Rainfall-runoff modeling

For rainfall-runoff modeling, a version of the often-used hydrological model Hydrologiska Byråns Vattenbalansavdelning (HBV) was applied. It was originally developed at the Swedish Meteorological and Hydrological Institute and modified by Wallner *et al.* (2013) at the Institute of Hydrology and Water Resource Management in Hannover, Germany (HBV-IWW). The model is a fast working conceptual model with different modules for calculating the relevant hydrological processes for runoff.

The model is capable of lumped and semi-distributed modeling and supports an hourly respectively daily time step. As input data precipitation *P*, temperature *T*, potential evapotranspiration *ETP*, usually as grass reference evapotranspiration ETP_0, and seasonal crop factors *C* are necessary. Snow melt is calculated with the Temperature-Index method, while flow components are computed using linear reservoir approaches. The runoff concentration is calculated with the unit hydrograph method for the sum of all flow components and subsequent flood routing is determined using the Muskingum method. In this study, 2 parameters of the Muskingum method and the 12 other model parameters are optimized on hourly and daily time steps as shown in the following sections.

Disaggregation methods

Due to a lack of hourly long-term data, disaggregation of daily time series to hourly time step is applied. In this way, preserving mass and mean values is ensured. All climate values are interpolated to the respective catchment or subcatchment before a disaggregation is performed.

For temperature T, a simple daily harmonic approach after Förster *et al.* (2016) was used according to the following equation:

$$T_{i,j} = T_{m,i} + \frac{T_{\max,i} - T_{\min,i}}{2} \cdot \cos\left(\frac{\pi \cdot (t_j - a)}{12}\right) \tag{3}$$

where temperature $T_{i,j}$ for every day i and every hour j is calculated; $T_{m,i}$, $T_{\max,i}$ and $T_{\min,i}$ represent mean, max and min temperature of day i and rely on measurements. Assuming that every day starts with hour 0 the shifting parameter a stands for the hour of the day where the maximum hourly value of temperature occurs. This approach implies that the minimum hourly value of temperature occurs with a shift of 12 h at night. Evaluations in the study areas on short time hourly measurements show slightly varying times for the maximum and minimum temperature according to seasons and days with differing shifting times. For this study, a fixed value for parameter a at 14 was set according to the best match with maximum hourly statistics. It is assumed that the occurrence of maximum temperature is more critical for melting and flood processes than the exact time of minimum temperature.

The grass reference evapotranspiration ETP_0 is first calculated from the observed daily meteorological time series following the FAO standard procedure (Allan *et al.* 1998). Subsequently, daily values are disaggregated similar to temperature values, using Equation (1) with the same parameter a. Due to missing maximum and minimum daily values, minimum values at night are set to zero while maximum values are set to the double daily mean. As an hourly mean value, the equally divided daily ETP_0 is used. The application of a harmonic approach for ETP_0 was assumed to be more realistic than simply using an equally divided hourly mean. Nevertheless, estimations comparing both disaggregation approaches only show very small differences in calculated flood peaks of 0–3%, concluding that the intra daily ETP_0 cycle has only a minor influence on simulated flood peaks.

For disaggregation of precipitation from daily to hourly time steps, the multiplicative random cascade (MRC) model after Müller & Haberlandt (2015) was applied. The principle of the model is to increase the temporal resolution gradually while preserving mass and general statistics in every step. For every disaggregation level, the model uses different statistics dependent on the volume of precipitation and the position of the step in the surrounding dry and wet values. An exception is the first split into three volumes, where the surrounding dry and wet values are not taken into account but two statistics are used for values above and below the 0.998 quantile of precipitation. The necessary parameters for the disaggregation statistics (multifractal parameters) are calculated from stepwise aggregating observed hourly series to daily ones.

In this study, the time series for every catchment are interpolated with an inverse distance method from surrounding long-term daily and short-term hourly precipitation measurement stations before disaggregating the daily ones. The model parameters for disaggregation are calculated from interpolated hourly catchment time series. This straightforward approach without considering spatial consistence of disaggregation results before interpolation has been proven to produce sufficiently good results for simple conceptual rainfall-runoff models (Müller-Thomy *et al.* 2018). The described precipitation disaggregation method is also applied for climate model outputs, assuming that the multifractal parameters of rainfall are stationary for the future. Due to the stochastic nature of MRC, several disaggregation-runoff runs are performed, using the median value of all results for the evaluation of the results. Test runs have shown that using 10 runs is sufficient for further evaluations in this study.

Calibration and validation strategies

Two HBV models per investigated catchment area with specific objectives for calibration are used for this study. The first set of models (set D) is used for daily rainfall-runoff modeling with daily climate time series as input. It is calibrated on daily runoff hydrograph as commonly used in hydrological studies. Therefore, set D is intended just to be capable of showing the general daily behavior of the runoff in every investigated catchment. The second set of models (set H) is used for hourly runoff modeling with hourly climate time series input. This set is calibrated on peak flow runoff statistics while simulating on an hourly time step. To preserve the time dependency between simulated and measured runoff values, the match between observed and simulated daily hydrographs is also considered in the calibration functions. By using a higher time resolution in modeling and due to calibrating directly to the peak flood statistics, modeling set H is expected to be capable of showing a better resolution runoff behavior of the catchments. For determining and comparing the recent and future return period floods (RP floods), it is assumed that set H is more accurate than set D.

The calibration and validation of the model set D are performed by splitting the observed daily climate and runoff time series for each catchment in the calibration and validation period as shown in Table 3 (split sampling). In order to avoid calibrating the initial conditions of HBV, a spin-up phase of at least 2 years and common initial values are used. For optimization of the other parameters during calibration, the multi-criteria optimization algorithm AMALGAM (Vrugt 2016), that is based on finding a well-distributed set of Pareto solutions for multi-criterial optimization problems, is applied. In this study, three criteria representing the quality of the fit between simulated and measured hydrograph are used for set D, namely the Nash–Sutcliffe efficiency (NSE), the corrected Nash–Sutcliffe efficiency (NSEcor) and the Kling–Gupta efficiency (KGE). The characteristics of each quality criterion are described in detail in Gupta *et al.* (2009). Following objective functions (OFs) are minimized during calibration of the model set D:

$$\mathrm{OF}(1) = 1 - \mathrm{NSE} \tag{4}$$

$$\mathrm{OF}(2) = 1 - \mathrm{NSEcor} \tag{5}$$

$$\mathrm{OF}(3) = 1 - \mathrm{KGE} \tag{6}$$

One parameter set with similar weighting of all OFs is selected from the Pareto front for this investigation. For validating the model set D, another time period of a similar range and a spin-up period is analyzed using the same performance criteria.

The calibration of the model set H is performed with hourly data, disaggregated from measured daily time series as described in the previous section. As well as for the model set D, a spin-up period of at least 2 years is used. For optimization of the parameters during calibration, the more suitable multi-objective Pareto Archived Dynamically Dimensioned Search Algorithm (PADDS; Asadzadeh & Tolson 2012), as part of the Optimization Software Toolkit for Research Involving Computational Heuristics (OSTRICH; Matott 2017), is applied. Three OFs are optimized during calibration, while one parameter set with equal weighting of all OFs is selected from the final Pareto front. As two OFs, the Weighted Sum of Squared Errors (WSSE) between observed and simulated RP floods, separately for winter and summer, are used. This ensures a good fit in both seasons. For simulated and observed statistics, a theoretical Gumbel distribution function instead of an empirical distribution

Table 3 | Investigated study sites with model types in HBV

		Model set D (1 d)		**Model set H (1 h)**	
Catchment (gauge)	**Model type**	**Calibration period**	**Validation period**	**Calibration and validation period**	**Extra validation period**
Trautenstein	Lumped	1981–1995	1996–2010	1981–2010	2008–2011
Mahndorf	Lumped	1981–1995	1996–2010	1981–2010	2010–2013
Oschersleben	Lumped	1987–1998	1999–2010	1988–2017	2008–2011
Silberhütte	Semi-distributed	1981–1995	1996–2010	1981–2010	2010–2013
Meisdorf	Semi-distributed	1981–1995	1996–2010	1981–2010	2010–2013
Hausneindorf	Semi-distributed	1981–1995	1996–2010	1981–2010	2010–2013
Calibration/validation objective		Hydrograph		Flood distribution, hydrograph	
(quality criteria)		(NSE, NSEcor, KGE)		($\mathrm{QRP_{Win}}$, $\mathrm{QRP_{Sum}}$, NSEcor)	

Periods and strategies for calibration and validation of the model sets D and H (periods in hydrological years). For further explanations, see Section 'Calibration and validation strategies'.

is used to lower the effects of single extreme flood peaks. For the statistics, the series of maximum values of every season is taken into account. As weighting of each error, the reversed value of the related return period is used in order to put the focus on larger return periods. The third OF is calculated from the NSEcor value between the observed daily hydrograph and the simulated hourly values, aggregated to daily hydrographs. Following OFs are minimized during calibration of the model set H:

$$\mathrm{OF}(1) = \sum_{i=1}^{6} \frac{1}{\mathrm{RP}_i} \cdot (\mathrm{QRP}_{i\mathrm{Win,sim}} - \mathrm{QRP}_{i\mathrm{Win,obs}})^2 \tag{7}$$

$$\mathrm{OF}(2) = \sum_{i=1}^{6} \frac{1}{\mathrm{RP}_i} \cdot (\mathrm{QRP}_{i\mathrm{Sum,sim}} - \mathrm{QRP}_{i\mathrm{Sum,obs}})^2 \tag{8}$$

$$\mathrm{OF}(3) = 1 - \mathrm{NSEcor} \tag{9}$$

where the RP floods for winter ($\mathrm{QRP}_{i,\mathrm{Win}}$) and summer ($\mathrm{QRP}_{i,\mathrm{Sum}}$) are calculated for six return periods ($\mathrm{RP}_1 = 2$, $\mathrm{RP}_2 = 5$, $\mathrm{RP}_3 = 10$, $\mathrm{RP}_4 = 20$, $\mathrm{RP}_5 = 25$, $\mathrm{RP}_6 = 50$) during calibration. Due to the stochastic nature of MRC, 10 disaggregation-runoff runs per optimizing step are performed, calculating the OFs with the median values of $\mathrm{QRP}_{i,\mathrm{Win}}$, $\mathrm{QRP}_{i,\mathrm{Sum}}$ and NSEcor for the 10 runs. For validation of the model set H, 10 other disaggregation-runoff runs are analyzed and compared with the statistics of the calibration runs. In addition, an hourly runoff hydrograph, simulated with the calibrated model set H and driven by observed climate values, is compared with the observed short-term hydrograph (extra validation period) for testing the robustness of the models.

Comparison of future RP flood changes and statistical analysis

The model sets D and H are driven by climate series coming from the climate model ensemble, described in more detail in the previous sections. For the model set D, the daily climate ensemble series are directly simulated, while the series for the model set H are disaggregated before simulation. Assuming that the fractal parameters remain constant for the future, the MRC parameters from observed precipitation are used. With both model sets, the three time slices ‘Past’, ‘Near Future’ and ‘Far Future’ are simulated for each climate model. Subsequently, for each time slice of every climate model series, the RP floods for year, summer and winter are calculated. The Gumbel distribution function was fitted using the highest runoff peaks of each year respectively season. The simulated RP floods for the time slices of Past, Near Future and Far Future are analyzed for the whole model ensemble. These results are called ‘absolute RP floods’ in this study. Furthermore, the changes of RP floods from Past to Near Future, and Past to Far Future, divided to the related past RP flood, can be calculated for the evaluation. These changes are also analyzed for the whole model ensemble and are called ‘relative RP flood changes’ in this study. The relative RP flood changes can be interpreted as climate change signals of RP floods. Absolute RP floods and relative RP flood changes are calculated from daily and hourly model results, simulated with the model sets D and H from the same original climate ensemble series.

The ensemble mean differences for absolute RP floods as well as relative RP flood changes (change signals) between the model sets D and H are tested for their significance by using a two-sided two-sample *t*-test. The null hypothesis is that the difference in means is equal to zero; hence, the alternative hypothesis indicates that the means are unequal to zero. To quantify the degree of similarity between both model set results, the *p*-values are calculated, where a low *p*-value indicates a strong difference between the results.

RESULTS AND DISCUSSION

Calibration and validation results

Results for the calibration and validation of the model set D (daily models) are summarized in Figure 2. To illustrate the matching of the hydrographs, parts of them are also shown exemplarily for the Selke–Meisdorf catchment.

The results for the model set D show a good match between daily measurements and calibrated models with NSE values of about 0.8 for calibration and 0.66 to 0.82 for validation. The Oschersleben catchment shows the smallest NSE values, probably due to its distinctly agricultural character with much drainage, irrigation and anthropogenic influence on its river Großer Graben. The hydrographs of the Meisdorf catchment imply slight lacks in reproducing the daily flood values, while overall a good match is evident.

Catchment (gauge)		NSE	NSEcor	KGE
Trautenstein	Cal	0.82	0.82	0.87
	Val	0.80	0.79	0.78
Mahndorf	Cal	0.78	0.79	0.88
	Val	0.68	0.70	0.80
Oschersleben	Cal	0.76	0.77	0.87
	Val	0.66	0.70	0.68
Silberhütte	Cal	0.82	0.82	0.89
	Val	0.82	0.81	0.84
Meisdorf	Cal	0.83	0.84	0.91
	Val	0.79	0.78	0.82
Hausneindorf	Cal	0.86	0.86	0.91
	Val	0.75	0.75	0.66

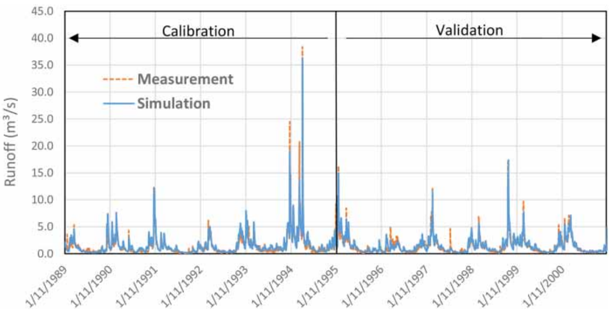

Figure 2 | Calibration and validation results (15 years each) for the model set D. Measured and simulated daily hydrographs for the catchment Selke–Meisdorf (cutout).

Results for the calibration and validation of the model set H are exemplary shown in Figure 3 for the Trautenstein catchment and summarized for all catchments in Table 4. The summarized results include the match between a selection of observed and simulated return period floods (RP floods), both smoothened by a fitted Gumbel distribution. For comparison, the corresponding RP floods from daily observations are also listed. To aggregate the results, only flood values calculated from annual maximum flood peaks are shown instead of the calibrated winter and summer floods.

Figure 3 shows a good fit between calibrated/validated RP floods and the observed distribution function, calculated from flood peak values, in the Trautenstein catchment. The band, spanned by 10 different disaggregated precipitation runs, is rather small, indicating small uncertainty due to the disaggregation of precipitation. The match is closer at higher return periods, probably due to higher weighting of less frequent floods. The large difference between the Gumbel distributions, calculated from peak values and daily values, reveals that in particular for this catchment a daily flood base underestimates the observed flood peaks. Table 4 shows similar results for some other catchments, where the bigger catchments Oschersleben and Hausneindorf in flatter regions show few or no differences between peak and daily base. NSEcor values of up to 0.8 imply a good match between the aggregated hydrograph of simulation and the observed daily one. Despite no measured hourly hydrograph being used for calibrating the model set H, NSEcor values >0.6 for the additional validation show a sufficient match for most of the catchments. Only the Oschersleben catchments have a lower value, probably due to the previously described anthropogenic influence.

The model set D is capable of simulating the general hydrograph from daily values, while the model set H is capable of simulating sufficiently accurate flood peaks from hourly climate series, disaggregated from daily ones. In the next sections, both model sets are applied to compare absolute RP floods and relative RP flood changes using the described climate model ensemble.

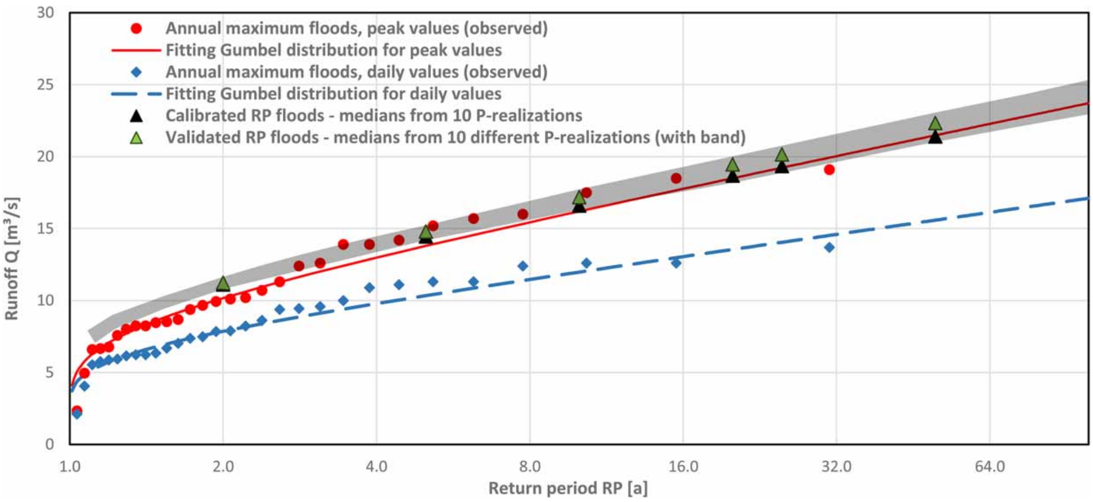

Figure 3 | Calibration and validation results for the Trautenstein catchment of the model set H.

Table 4 | Match of annual RP5, RP20 and RP50 floods between observation (Obs) and median values of calibration (Cal)/validation (Val) of the model set H

	RP5 (annual) (m^3/s)				RP20 (annual) (m^3/s)				RP50 (annual) (m^3/s)					
Catchment (gauge)	**Obs (p)**	**Cal**	**Val**	**Obs (d)**	**Obs (p)**	**Cal**	**Val**	**Obs (d)**	**Obs (p)**	**Cal**	**Val**	**Obs (d)**	**NSEcor (1 d, aggr.) 30 years period**	**NSEcor (1 h, val) 4 years period**
Trautenstein	13.8	14.4	14.8	10.3	18.5	18.7	19.4	13.5	21.5	21.4	22.3	15.6	0.75	0.75
Mahndorf	25.4	25.6	25.3	15.8	39.2	39.9	39.2	23.2	48.0	49.0	48.0	27.8	0.65	0.70
Oschersleben	13.2	12.2	12.3	12.7	18.9	19.2	19.5	18.4	22.5	23.7	24.1	22.0	0.80	0.56
Silberhütte	15.9	15.6	15.6	13.0	23.0	23.4	23.6	18.3	27.5	28.4	28.6	21.6	0.80	0.62
Meisdorf	19.7	19.5	19.4	16.4	28.4	27.1	26.5	23.2	33.9	31.9	31.1	27.5	0.75	0.63
Hausneindorf	21.1	21.6	21.4	18.8	31.3	30.1	29.3	27.5	37.7	35.5	34.3	33.1	0.75	0.69

Obs (p): RP flood calculated from observed peak flow values.

Obs (d): RP flood calculated from observed daily flow values.

NSEcor between daily observed hydrograph and simulated hourly hydrograph, aggregated to daily values (1 d, aggr.). NSEcor between hourly observed and hourly simulated hydrograph, as extra validation period (1 h, val).

Comparison of daily and hourly ensemble simulation results

Figure 4 shows the simulated results of absolute RP20 annual floods of both model sets for all study sites. Most catchments show considerable differences for both seasons in the absolute RP20 flood between the model set D (daily hydrograph simulation) and the model set H (hourly peak flow simulation). For example, in the Holtemme–Mahndorf catchment the mean RP20 flood for the Far Future from hourly simulation is two times higher than the daily one. In contrast, the Großer Graben–Oschersleben catchment has lower differences, probably due to prolonged flood events in this large lowland area.

This catchment even shows somewhat smaller floods for hourly than for daily simulation, indicating that the temporal resolution of simulation in this large catchment has a minor influence compared with the

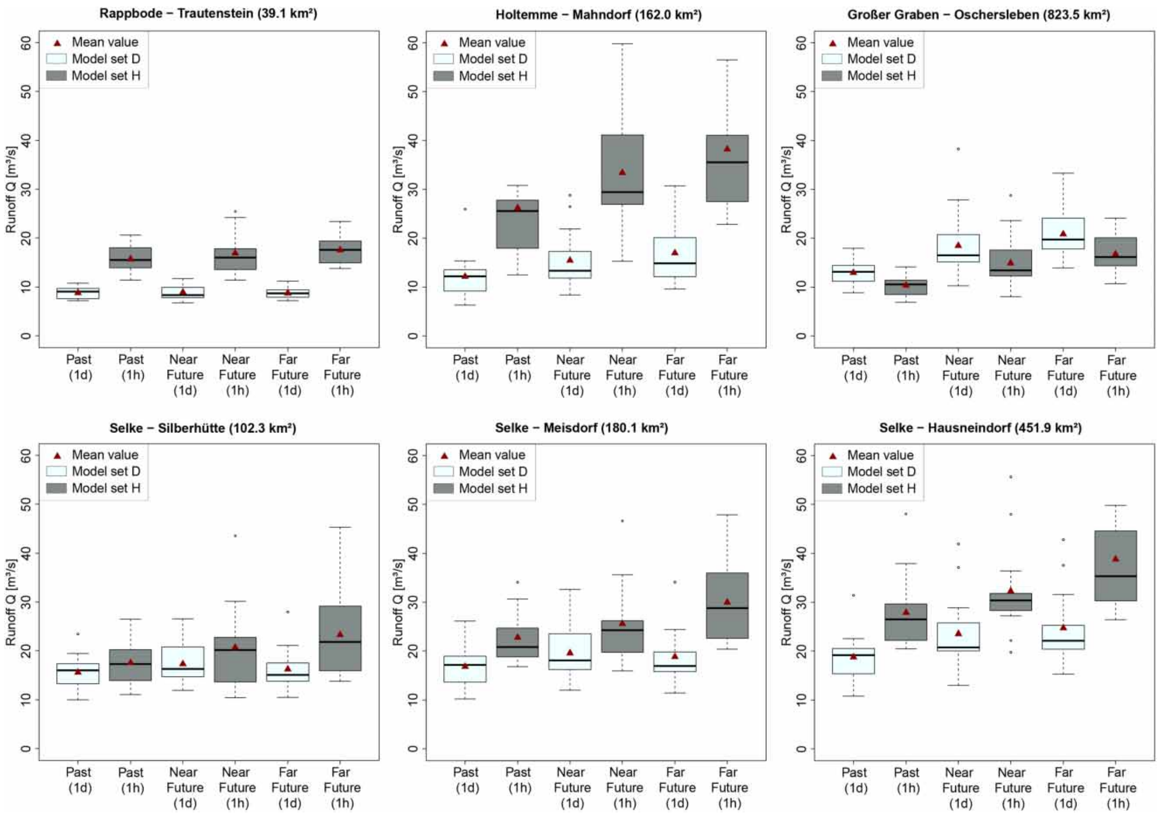

Figure 4 | Simulated absolute RP20 annual floods for Past, Near Future and Far Future of the model set D (1 d) and the model set H (1 h).

parameterization of the model. Furthermore, catchments with higher floods also appear to generate a wider range of results within the ensemble. Overall, absolute floods apparently differ between both model sets for most of the catchments.

Figure 5 shows the simulated results of the relative changes of the RP20 annual floods from Past to Near Future and Far Future for both model sets and for all study sites. The relative changes are interpreted as climate change signals of RP floods. The relative RP flood changes between both model sets are more similar than the absolute RP floods, in particular for changes to the Near Future. In some catchments, somewhat greater differences for Far Future become apparent, e.g., see floods in Selke–Silberhütte. Overall, relative RP flood changes from Past to Near Future and Far Future are quite similar between both model sets for most of the catchments.

Testing differences between daily and hourly ensemble simulation

To quantify the differences for the simulated absolute RP floods and relative RP flood changes between the model sets D and H, the *p*-values are calculated from a *t*-test of the ensemble mean values, shown in Figure 6 for the Near Future and in Figure 7 for the Far Future. For comparison, results for six different floods ranging from the return periods 2 to 50 are depicted. To aggregate the results, only the *p*-values, calculated from annual maximum flood peaks, are shown. The higher the *p*-values, the more similar are the results between daily and hourly models.

Overall, the *p*-values, calculated between the means of daily and peak flow models, are greater for relative flood changes than for differences of absolute floods, in particular for the Near Future. Here, almost all catchments show larger *p*-values for relative flood changes than for absolute floods. This means that there are no significant differences between daily and hourly simulations, when relative flood changes are considered. Furthermore, all catchments are exceeding the significance threshold of $p = 0.01$ for all return periods. For differences in absolute floods, the results are diverse, depending on the catchment and return period. In particular, the catchments Rappbode–Trautenstein and Holtemme–Mahndorf show *p*-values <0.01, indicating significant differences between absolute daily and peak flow model results. Both catchments are mountainous, while Rappbode–Trautenstein

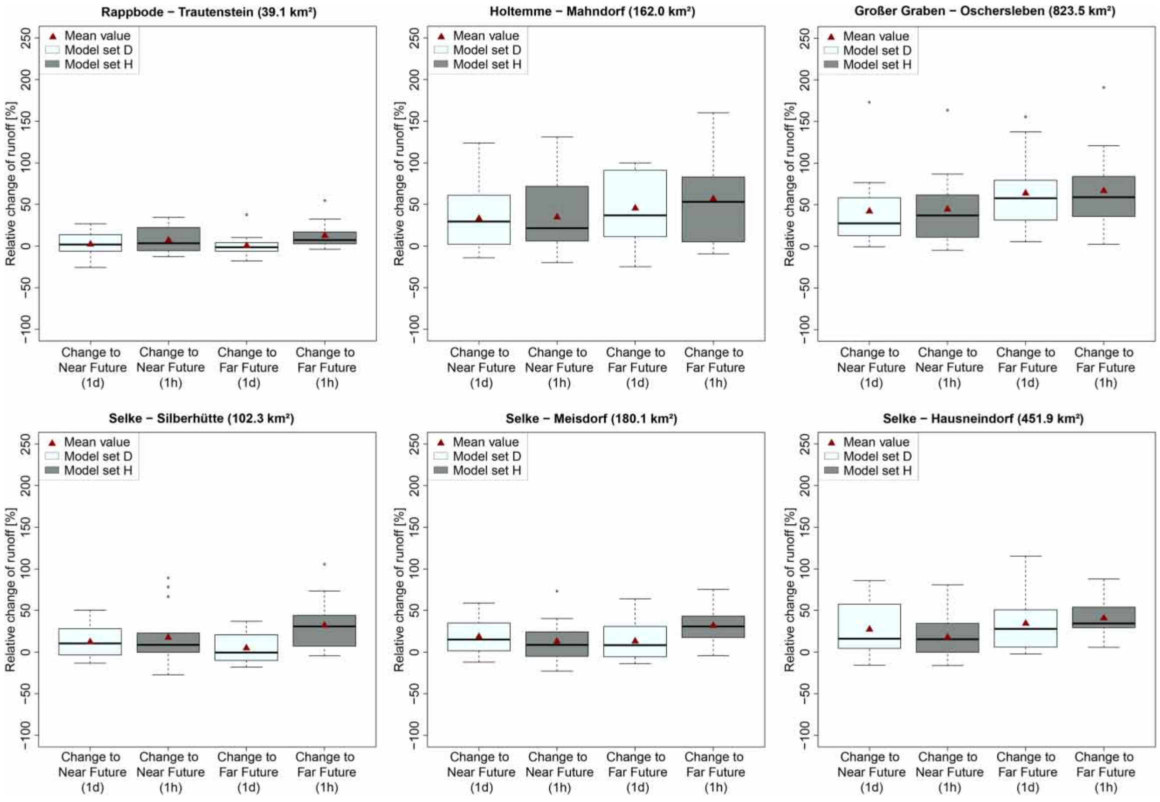

Figure 5 | Simulated relative changes of RP20 annual floods from Past to Near Future and Far Future of the model set D (1 d) and the model set H (1 h).

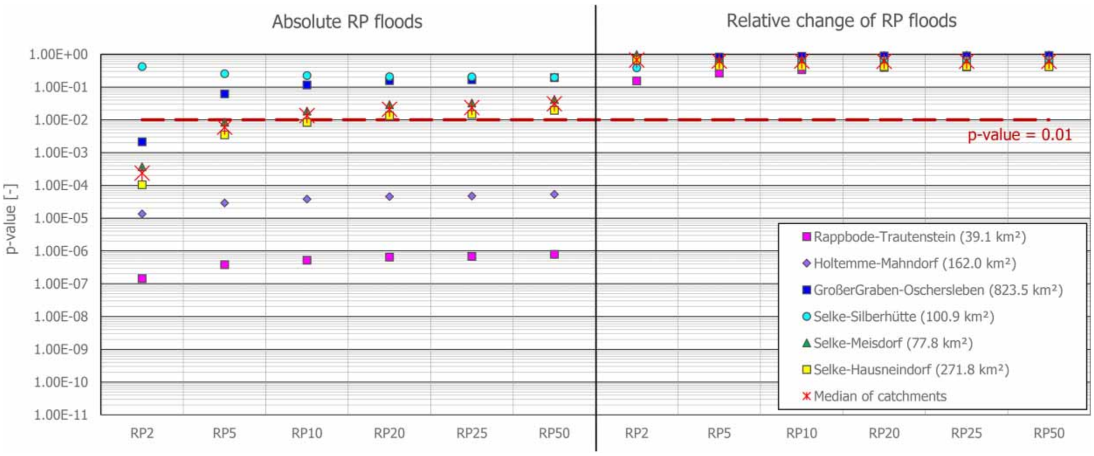

Figure 6 | Near Future – *p*-values of annual RP floods from student *t*-test between climate ensemble means, simulated with the model set D and the model set H. Depicted are *p*-values for means of absolute RP floods and relative RP flood changes for all investigated catchments and return periods.

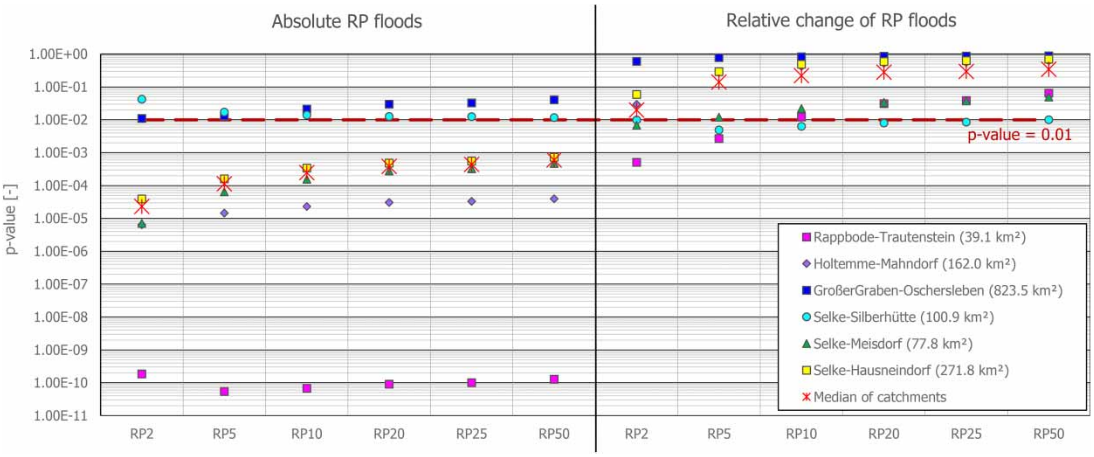

Figure 7 | Far Future – *p*-values of annual RP floods from student *t*-test between climate ensemble means, simulated with the model set D and the model set H. Depicted are *p*-values for means of absolute RP floods and relative RP flood changes for all investigated catchments and return periods.

is very small and Holtemme–Mahndorf has higher floods compared with the other catchments. In contrast to this, for relative flood changes the results for both catchments are equal, no matter if daily or hourly models are applied.

In the Far Future, somewhat lower *p*-values were calculated overall, indicating less similarity between both model sets with increasing flood changes. Rappbode–Trautenstein and Holtemme–Mahndorf show again the largest differences for absolute RP floods between daily and hourly models. However, the similarity of relative RP flood changes is high between both model sets for most of the catchments, while the absolute RP floods are not significant for most of them. In general, the *p*-values increase with increasing RP floods, indicating more similar results between the model sets for larger return periods. For the investigated catchments and climate ensemble, the differences between daily and hourly modeling negligible from RP20 floods and above considering relative flood changes. This is also the case for the seasonal evaluation, while relative changes of summer RP floods somewhat better match between both model sets (not shown).

CONCLUSIONS

The aim of this study was to examine the capability of daily hydrological models to simulate the future development of extreme flood events in mesoscale catchments. Calibration and validation results show that the daily model set D is able to simulate proper daily hydrographs for all investigated catchments. By using proper disaggregation methods to increase the temporal resolution of input data to hourly ones and by calibrating on the distribution function of RP floods, the resulting model set H is capable of generating flood distribution functions matching the observed ones. The comparisons between both model sets using climate model ensembles show the following results:

1. For most of the catchments, the calculated absolute RP floods between the model sets are significantly different for any time slice. An exception is the larger lowland catchments. Furthermore, the differences between both model outputs increase up to the end of the century. Increasing return periods result in increasing similarities between both models in most of the cases. The daily model set underestimates the absolute flood magnitudes in most of the cases, so it can be assumed that the peak flow model set is more capable of representing the expected future flood magnitudes.
2. Different results emerge from analyzing the relative RP flood changes from Past to both future time periods (flood change signals). Here, the ensemble mean differences between both model sets are not significant for all catchments in most of the cases. Increasing return periods lead to increasing similarities between both model sets. For all investigated catchments, differences are not significant for the Near Future and for all return periods, while for the Far Future differences in return periods from RP20 and above are not significant. Considering relative RP flood changes, both model sets can be regarded as being equivalent in most of the cases.

The results for this study indicate that an estimation of the relative RP flood changes is possible by using common daily models calibrated on daily hydrographs. In this way, a robust climate change signal for the future development of RP floods is calculated. To obtain absolute values for RP floods, the change signal should be applied on observed RP floods. Because the study examines catchments with the main land uses of agriculture, grassland and forest, it can be expected that the sample is probably more sensitive to the total rainfall amount than to the temporal distribution. For future generalization of the study results for other regions, other climatic conditions and for linking catchment or climate properties to the results, a larger sample size has to be investigated.

ACKNOWLEDGEMENTS

The authors thank the Lower Saxony Water Management, Coastal Defence and Nature Conservation Agency (Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz, NLWKN) for providing and preparing the ensemble data within the KliBiW project. The authors also thank an anonymous reviewer for the helpful comments and Hannes Müller-Thomy for providing the MRC model. The publication of this article was funded by the Open Access Fund of the Leibniz Universität Hannover.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

REFERENCES

Allan, R. G., Pereira, L. S., Raes, D. & Smith, M. 1998 *Crop Evapotranspiration – Guidelines for Computing Crop Water Requirements – FAO Irrigation and Drainage Paper 56*. FAO – Food and Agriculture Organization of the United Nations, Rome, Italy. ISBN 92-5-104219-5.

Asadzadeh, M. & Tolson, B. 2012 Hybrid Pareto archived dynamically dimensioned search for multi-objective combinatorial optimization: application to water distribution network design. *Journal of Hydroinformatics* **14** (1), 192–205. https://doi.org/10.2166/hydro.2011.098.

Chen, B., Krajewski, W. F., Liu, F., Fang, W. & Xu, Z. 2017 Estimating instantaneous peak flow from mean daily flow. *Hydrology Research* **48** (6), 1474–1488. https://doi.org/10.2166/nh.2017.200.

Ding, J., Haberlandt, U. & Dietrich, J. 2015 Estimation of the instantaneous peak flow from maximum daily flow: a comparison of three methods. *Hydrology Research* **46** (5), 671–688. https://doi.org/10.2166/nh.2014.085.

Ficchì, A., Perrin, C. & Andréassian, V. 2016 Impact of temporal resolution of inputs on hydrological model performance: an analysis based on 2400 flood events. *Journal of Hydrology* **538**, 454–470. http://dx.doi.org/10.1016/j.jhydrol.2016.04.016.

Förster, K., Hanzer, F., Winter, B., Marke, T. & Strasser, U. 2016 An open-source MEteoroLOgical observation time series DISaggregation Tool (MELODIST v0.1.1). *Geoscientific Model Development* **9**, 2315–2333. https://doi.org/10.5194/gmd-9-2315-2016.

Gabellani, S., Boni, G., Ferraris, L., von Hardenberg, J. & Provenzale, A. 2007 Propagation of uncertainty from rainfall to runoff: a case study with a stochastic rainfall generator. *Advances in Water Resources* **30** (10), 2061–2071. https://doi.org/10.1016/j.advwatres.2006.11.015.

Guo, D., Westra, S. & Maier, H. R. 2018 An inverse approach to perturb historical rainfall data for scenario-neutral climate impact studies. *Journal of Hydrology* **556**, 877–890. https://doi.org/10.1016/j.jhydrol.2016.03.025.

Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. 2009 Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. *Journal of Hydrology* **377** (1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

IPCC 2013 Climate change 2013: the physical science basis. In: *Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. & Midgley, P. M., eds). Cambridge University Press, Cambridge, UK and New York, NY, USA, p. 1535.

Keller, L., Rössler, O., Martius, O., Bronstert, A. & Weingartner, R. 2019 Comparison of scenario-neutral approaches for estimation of climate change impacts on flood characteristics. *Hydrological Processes* **33** (4), 535–550. https://doi.org/10.1002/hyp.13341.

Kim, D., Chun, J. A. & Aikins, C. M. 2018 An hourly-scale scenario-neutral flood risk assessment in a mesoscale catchment under climate change. *Hydrological Processes* **32** (22), 3416–3430. https://doi.org/10.1002/hyp.13273.

Matott, L. S. 2017 *OSTRICH: An Optimization Software Tool, Documentation and User's Guide, Version 17.12.19*. University at Buffalo Center for Computational Research, p. 79. Available from: www.eng.buffalo.edu/~lsmatott/Ostrich/OstrichMain.html (accessed 4 June 2020).

Müller, H. & Haberlandt, U. 2015 Temporal rainfall disaggregation with a cascade model: from single-station disaggregation to spatial rainfall. *Journal of Hydrologic Engineering* **20** (11), 04015026. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001195.

Müller-Thomy, H., Wallner, M. & Förster, K. 2018 Rainfall disaggregation for hydrological modeling: is there a need for spatial consistence? *Hydrology and Earth System Sciences* **22**, 5259–5280. https://doi.org/10.5194/hess-22-5259-2018.

NLWKN 2017 *Globaler Klimawandel – Wasserwirtschaftliche Folgenabschätzung für das Binnenland (Global Climate Change – Inland Water Management Impact Assessment)*. Oberirdische Gewässer Band 41, NLWKN – Lower Saxony Water Management, Coastal Defence and Nature Conservation Agency, Norden, Germany.

Olsson, J. 1998 Evaluation of a scaling cascade model for temporal rainfall disaggregation. *Hydrology and Earth System Sciences* **2** (1), 19–30. https://doi.org/10.5194/hess-2-19-1998.

Paschalis, A., Fatichi, S., Molnar, P., Rimkus, S. & Burlando, P. 2014 On the effects of small scale space-time variability of rainfall on basin flood response. *Journal of Hydrology* **514**, 313–327. https://doi.org/10.1016/j.jhydrol.2014.04.014.

Prudhomme, C., Wilby, R. L., Crooks, S., Kay, A. L. & Reynard, N. S. 2010 Scenario-neutral approach to climate change impact studies: application to flood risk. *Journal of Hydrology* **390**, 198–209. https://doi.org/10.1016/j.jhydrol.2010.06.043.

Prudhomme, C., Crooks, S., Kay, A. L. & Reynard, N. 2013 Climate change and river flooding: part 1 classifying the sensitivity of British catchments. *Climatic Change* **119**, 933–948. https://doi.org/10.1007/s10584-013-0748-x.

Steinschneider, S. & Brown, C. 2013 A semiparametric multivariate, multisite weather generator with low-frequency variability for use in climate risk assessments. *Water Resources Research* **49**, 7205–7220. https://doi.org/10.1002/wrcr.20528.

Viglione, A., Chirico, G. B., Komma, J., Woods, R., Borga, M. & Blöschl, G. 2010 Quantifying space-time dynamics of flood event types. *Journal of Hydrology* **394** (1–2), 213–229. https://doi.org/10.1016/j.jhydrol.2010.05.041.

Vormoor, K., Rössler, O., Bürger, G., Bronstert, A. & Weingartner, R. 2017 When timing matters-considering changing temporal structures in runoff response surfaces. *Climatic Change* **409** (24), 5403. https://doi.org/10.1007/s10584-017-1940-1.

Vrugt, J. 2016 *MANUAL Multi-Criteria Optimization Using the AMALGAM Software Package: Theory, Concepts, and MATLAB Implementation*. Available from: https://www.researchgate.net/publication/299458251 (accessed 4 June 2020).

Wallner, M., Haberlandt, U. & Dietrich, J. 2013 A one-step similarity approach for the regionalization of hydrological model parameters based on Self-Organizing Maps. *Journal of Hydrology* **494**, 59–71. http://dx.doi.org/10.1016/j.jhydrol.2013.04.022.

Received 29 July 2020; accepted in revised form 13 May 2021. Available online 1 June 2021