



Does the complexity in temporal precipitation disaggregation matter for a lumped hydrological model?

Hannes Müller-Thomy & Anna E. Sikorska-Senoner

To cite this article: Hannes Müller-Thomy & Anna E. Sikorska-Senoner (2019) Does the complexity in temporal precipitation disaggregation matter for a lumped hydrological model?, Hydrological Sciences Journal, 64:12, 1453-1471, DOI: [10.1080/02626667.2019.1638926](https://doi.org/10.1080/02626667.2019.1638926)

To link to this article: <https://doi.org/10.1080/02626667.2019.1638926>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Accepted author version posted online: 09 Jul 2019.
Published online: 08 Aug 2019.



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



Article views: 442



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Does the complexity in temporal precipitation disaggregation matter for a lumped hydrological model?

Hannes Müller-Thomy^{a,b,*} and Anna E. Sikorska-Senoner^{b,c}

^aInstitute of Hydraulic Engineering and Water Resources Management, Vienna University of Technology, Vienna, Austria; ^bInstitute of Hydrology and Water Resources Management, Leibniz Universität Hannover, Hanover, Germany; ^cDepartment of Geography, University of Zurich, Zürich, Switzerland

ABSTRACT

Flood peaks and volumes are essential design variables and can be simulated by precipitation–runoff (P–R) modelling. The high-resolution precipitation time series that are often required for this purpose can be generated by various temporal disaggregation methods. Here, we compare a simple method (M1, one parameter), focusing on the effective precipitation duration for flood simulations, with a multiplicative cascade model (M2, 32/36 parameters). While M2 aims at generating realistic characteristics of precipitation time series, M1 aims only at accurately reproducing flood variables by P–R modelling. Both disaggregation methods were tested on precipitation time series of nine Swiss mesoscale catchments. The generated high-resolution time series served as input for P–R modelling using a lumped HBV model. The results indicate that differences identified in precipitation characteristics of disaggregated time series vanish when introduced into the lumped hydrological model. Moreover, flood peaks were more sensitive than flood volumes to the choice of disaggregation method.

ARTICLE HISTORY

Received 1 October 2018
Accepted 21 May 2019

EDITOR

A. Castellarin

ASSOCIATE EDITOR

G. Thirel

KEYWORDS

precipitation disaggregation;
multiplicative cascade
model; effective daily
precipitation duration;
precipitation-runoff
modelling; floods

1 Introduction

High-resolution rainfall time series are needed in many fields in hydrology and water resources management, e.g. urban hydrology (Cross *et al.* 2018), erosion investigations (Jebari *et al.* 2012), preferential flow (Wiekenkamp *et al.* 2016) or precipitation–runoff modelling (Ding *et al.* 2016). For the latter, the temporal resolution of the precipitation time series has to match the temporal scale of the underlying hydrological processes. For example, Blöschl and Sivapalan (1995), in their review paper for urban drainage processes, recommend a sub-hourly resolution, while for irrigation purposes already a monthly resolution may be sufficient. For the simulation of highly dynamic processes such as floods, high-resolution time series are required to cope with (a) the high intermittency of precipitation (Ochoa-Rodriguez *et al.* 2015, Kim *et al.* 2018) and a movement of storm cells (Ran *et al.* 2012), especially for convective events, and (b) fast-responding catchments, defined depending on their attributes, such as size, drainage network, steepness and percentage of impervious area (Schilling 1991, Post and Jakeman 1996, Berne *et al.* 2004). Ficchi *et al.* (2016) found – for 240 catchments in France with a median catchment size of 356 km² – that the hydrological model performance did indeed improve with increasing the temporal resolution of the precipitation input time series (from daily to sub-daily time steps), especially for catchments with “shorter response times and flood durations, lower stream flow autocorrelation, and shorter and highly variable storm

events” (Ficchi *et al.* 2016). Melsen *et al.* (2016, and references therein) recommended hourly time series for the adequate calibration of hydrological models and for accurate runoff predictions in mesoscale catchments. This is agreement with Blöschl and Sivapalan (1995), who recommend an hourly resolution for flood warning purposes. In most cases, however, observed time series of this temporal resolution are too short to achieve a reliable calibration of the precipitation–runoff modelling, especially in mesoscale catchments with an area smaller than 1000 km² (Sikorska and Seibert 2018) and in which the network density is low in comparison to daily measurements.

Possible solutions are either (i) the generation of precipitation time series, e.g. by Poisson-cluster models (Onof *et al.* 2000, and references therein), alternating renewal models (Haberlandt *et al.* 2008, Callau Poduje and Haberlandt 2017, 2018), and combinations of rainfall generation models (Paschalis *et al.* 2014, Pohle *et al.* 2018), or (ii) the temporal disaggregation of observed precipitation time series of a coarser resolution, which are usually available for longer periods and higher network densities than high-resolution time series. Disaggregation has the advantage that it relies on real (measured) precipitation amounts and a correct representation of time series characteristics, e.g. autocorrelation, temporal asymmetry (Müller *et al.* 2017), and burstiness (Schleiss and Smith 2016). In addition, the disaggregation approach can be applied to increase a temporal resolution of satellite and

climate models that are usually available only at a daily scale. Well-known disaggregation methods for whole time series are multiplicative cascade models (Serinaldi 2010, Schertzer and Lovejoy 2011), method-of-fragments (Westra *et al.* 2012), or combinations of different methods (Koutsoyiannis and Onof 2001), and for events the application of mass curves (Blöschl and Sivapalan 1995 and references therein).

Disaggregation by a multiplicative cascade model is, in general, carried out on a station-by-station basis, without taking into account precipitation information from surrounding stations. Thus, the spatial patterns of the precipitation fields generated from surrounding stations are unrealistic, since the generation for each station is carried out independently from the others (Müller and Haberlandt 2015). To overcome this issue, Thober *et al.* (2014) suggested the application of cross-covariance for spatial consistence between raster fields during the disaggregation process. Recently, Müller and Haberlandt (2015) proposed the implementation of spatial consistence using a resampling algorithm as a subsequent step after the disaggregation. This should improve the coherence of disaggregated precipitation time series that are spatially associated. As an alternative, a disaggregation could be performed directly on areal precipitation totals to generate hourly areal precipitation time series. For example, Breinl (2016) generated univariate areal rainfall instead of multivariate station-based precipitation time series, but not with precipitation disaggregation methods as applied in this study.

However, to be used as an input for lumped hydrological models, these disaggregated station-based precipitation time series need to be spatially averaged to the mean areal catchment precipitation afterwards. Therefore, the added value of a spatial consistence of disaggregated precipitation time series, if captured with neighbouring stations, may be reduced or lost.

Sikorska *et al.* (2018) argued that the complexity in the temporal precipitation disaggregation approach may not matter that much for a lumped hydrological model, due to the assumptions of lumping precipitation amounts in space (the whole catchment receives the same areal precipitation) and, hence, in time (precipitation amounts at locations at different distances from the catchment outlet arrive at the outlet at the same time). This reduces the complex intermittency of the precipitation process and the dimension of spatial precipitation patterns and leads to the research question: “How complex does a precipitation disaggregation method (and thus the disaggregated time series) have to be for the application of the generated time series as input in a lumped hydrological model?” To the authors’ knowledge, there is a lack of investigations into precipitation time series generated by disaggregation methods of different complexities and, in particular, their impact on simulated runoff statistics when used as input for precipitation–runoff modelling. Instead, previous research has been more frequently focused on comparison of only the precipitation statistics resulting from different disaggregation methods (e.g. Pui *et al.* 2012, Paschalis *et al.* 2014), while the impact on precipitation–runoff simulation results remains unclear, or on the impact of the disaggregated time series, but only for similar disaggregation methods (e.g. Müller and Haberlandt 2018). Kandel *et al.* (2005) compared a scaling method based on a cumulative distribution function with a

simpler approach on daily mean intensities and investigated the impact of these two methods on the simulated infiltration excess runoff and the soil erosion response, showing a better performance for the more complex model. In another study, Hingray and Ben Haha (2005) tested two different disaggregation methods – a simple uniform splitting method and a micro-canonical cascade model – with different disaggregation time steps varying from 1 h to 10 min for continuous urban hydrological simulations. They found that the micro-canonical cascade model led to a better representation of the analysed precipitation and runoff characteristics, i.e. the maximum discharge and the maximum storage amount for a return period of 10 years. However, it is questionable whether these results for urban hydrology can be transferred to rural catchments with a pronounced catchment filtering role, as was found by Andres-Domenech *et al.* (2015) and Müller-Thomy *et al.* (2018).

Recently, Sikorska *et al.* (2018) proposed a simple one-parameter disaggregation approach that relies on uniform distribution of the daily precipitation totals into blocks of equal precipitation intensity. The only parameter of this approach is the effective daily precipitation duration and it determines the length of hourly blocks of equal precipitation sums. However, this method does not aim at reproducing observed precipitation characteristics; the disaggregated time series serve only as a direct input for a precipitation–runoff model and for flood analysis. Thus, the authors also did not investigate the effect of using more complex disaggregation approaches instead of the proposed one-parameter method.

Although it is clearly expected that a more complex disaggregation approach should provide a better representation of hourly precipitation time series, the greater value of such an approach as an input for a hydrological model is questionable. In particular, it is not clear whether the hydrological model can still benefit from precipitation time series disaggregated with a more complex approach than the simple one-parameter approach, or whether the added value of increasing complexity in the disaggregation approach is being lost after averaging the precipitation time series into the mean areal catchment values.

Therefore, the primary research question arises as: Does the complexity in precipitation disaggregation matter for a lumped hydrological model? To answer this question, we compare two precipitation disaggregation methods of a different complexity: the method of Sikorska *et al.* (2018), which we call M1, and the method of Müller and Haberlandt (2015), which is referred to as M2. The M1 method is considered as a backward approach, in which the optimal wet spell duration is estimated based on annual peaks simulated with a hydrological model (for details, see Section 3.1.1), while the M2 disaggregation method is a forward approach out of the family of micro-canonical cascade models and is based purely on precipitation data without considering any runoff-based simulations (for details, see Section 3.1.2). These two methods are applied to averaged areal catchment precipitation time series to derive disaggregated precipitation time series, which are then used directly as input for a lumped hydrological model – HBV (Seibert and Vis 2012) in nine Swiss mesoscale catchments. The method comparison is performed on two levels of reproducibility: of precipitation

characteristics and of flood characteristics assessed by (relative) peaks and volumes. Thus, the novelty of this study lies in comparing these two methods of the different disaggregation complexity on these two levels of analysis, as: M1 has not been evaluated in terms of precipitation characteristics until now, and, so far, M2 has been applied only to station data and not on the mean areal catchment precipitation series.

To answer the primary research question and to enable a comparison of the two disaggregation methods, the following secondary research questions must be addressed: Can the micro-canonical cascade model (M2) be applied for the disaggregation of the areal precipitation instead of station precipitation data?; What is the effect of the parameter uncertainty of a precipitation–runoff model on the runoff simulations for both methods, M1 and M2?; and, finally, What has a higher impact on the simulation results of the precipitation–runoff model in method M2: the parameter uncertainty of the precipitation–runoff model or the uncertainty resulting from the generated disaggregation realizations?

2 Study catchments and data

The comparison study was performed on nine mesoscale Swiss catchments (Fig. 1 and Table 1), with an areal range of 44–491 km², and a mean catchment elevation range of 511–2050 m a.s.l. Three of these catchments have glacier areas covering more than 5% of the total catchment area, which requires consideration in the modelling approach. The characteristics of the areal precipitation data are given in Table 1 (see Section 3.1.4 for a description of the precipitation characteristics).

Furthermore, the study area can be characterized climatologically according to Köppen-Geiger classification by two regions: a cold climate without dry season and with warm summers for catchments lying on the Swiss Plateau, Jura or in the pre-alpine region, and a polar climate (alpine Tundra) in the alpine catchments (Peel *et al.* 2007). For a more detailed description of the study catchments, the reader is referred to Sikorska *et al.* (2018).

The runoff data were available for the catchment outlets at an hourly resolution for the period 1980–2014. The mean

areal catchment precipitation and mean catchment temperature were estimated from ground stations nearby and within the catchment belonging to the automatic national gauging network that comprises about 75 Swiss meteorological ground stations (including tipping-bucket type gauges for measuring precipitation). The mean catchment estimates were calculated using the Thiessen polygon method for the same period as the runoff data. The average station coverage for the sample catchments was equal to 1.9 station per 100 km², which is rather high (Viviroli *et al.* 2011). The mean precipitation estimates were computed at an hourly resolution and then summed up to daily totals before being treated with both disaggregation methods. The hourly mean temperature data once computed were not changed in the further analysis.

3 Methods

3.1 Precipitation disaggregation

Two methods are investigated, namely M1, a backward approach, and M2, a micro-canonical cascade model (forward approach). For both methods, two versions are analysed, which are described below in detail. The evaluation of the disaggregated time series is explained in Section 3.1.3.

3.1.1 M1: backward approach

The disaggregation method M1 is a backward approach (see Sikorska *et al.* 2018), in which the optimal wet spell duration is estimated based on the annual peaks simulated with a hydrological model. For this purpose, daily precipitation totals are disaggregated into hourly sums on the basis of a uniform distribution on n evenly divided hourly intervals, where n is the only parameter and describes the effective precipitation length or an optimal wet spell duration within a day. To estimate n , different lengths of uniform blocks have to be tested in the range from 1 to 24 h, where for $n = 1$ the whole precipitation amount of one day falls in one hour. The first hour of the uniform block is always chosen randomly and independently for each simulation day ensuring that the entire daily total falls within the same day (see Fig. 2, M1, top panel). For example, for a 12-h distribution the first hour is selected randomly in the range between the

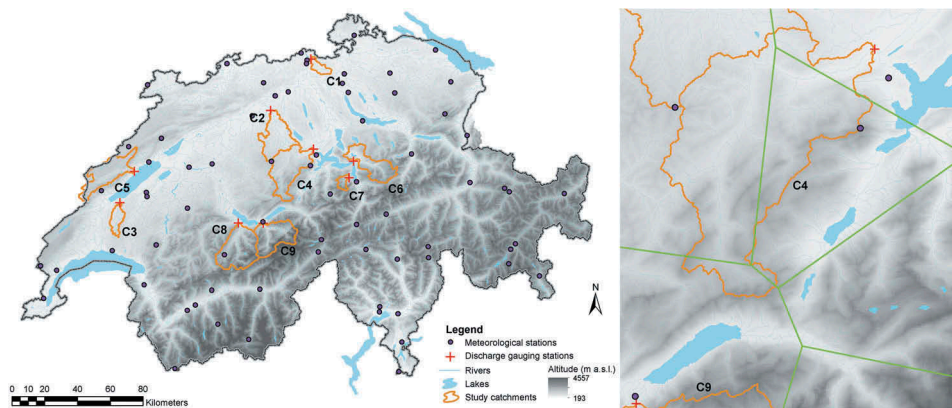


Figure 1. Overview of nine study catchments (left) and the Thiessen polygons (right) in the example catchment – Kleine Emme (C4) – chosen to demonstrate the results.

Table 1. Properties of the studied catchments.

Catchment	River	Discharge gauging station	Catchment attributes			Areal precipitation					
			Mean altitude (m a.s.l.)	Area (km ²)	Areal glacier ratio (%)	Number of rain gauges	Wet spell duration (h)	Wet spell amount (mm)	Dry spell duration (h)	Fraction of dry intervals (%)	Average intensity (mm/h)
C1	Surb	Döttingen	511	67	0	3	5.3	3.7	18.8	78	0.7
C2	Wigger	Zoffingen	660	368	0	4	6.0	4.0	17.5	75	0.7
C3	Mentue	Yvonand	679	105	0	3	5.3	3.6	18.8	78	0.7
C4	Kleine Emme	Littau	1050	478	0	4	6.0	4.1	17.2	74	0.7
C5	Areuse	Boundry	1060	377	0	3	5.2	3.3	20.2	80	0.6
C6	Muota	Ingenbohl	1360	316	<1	3	5.3	3.2	18.8	78	0.6
C7	Grosstalbach	Isenthal	1820	44	9.3	2	4.6	3.3	21.2	82	0.7
C8	Kander	Hondrich	1900	491	7.9	3	4.9	3.3	20.8	81	0.7
C9	Lütschine	Gsteig	2050	379	17.4	4	5.4	3.4	20.6	79	0.6

first and 13th hour of a day. By choosing the first wet hour randomly, we do not take into account typical seasonal precipitation occurrence during the day (e.g. summer storms in the afternoon or evening), to keep the disaggregation method as simple as possible.

Such generated hourly precipitation series with different disaggregation lengths are next used as input for a hydrological model to simulate runoff time series. An optimal wet spell duration within a day is chosen based on the simulation performance of annual peaks using these different disaggregated series. Annual peaks simulated with hourly observed precipitation data are used as a benchmark, which makes this method independent of runoff observations. Thus, the optimal duration is chosen as the duration being closest to the benchmark simulation. This duration was estimated in a previous study as being between 6 and 12 h for mesoscale mountainous catchments (Sikorska *et al.* 2018).

In addition to this uniform method, we introduce in this study a variation, which, instead of estimating the optimal wet spell duration from runoff simulations, directly assumes the optimal precipitation disaggregation to follow the triangle shape similar to the precipitation Euler Type II of DVWK (DVWK 1984). Within this disaggregation, the maximum hourly precipitation intensity occurs in the 8th hour of a day, with 48% of the daily precipitation amount. The slope of the ascending limb (1st–8th hour) is higher in comparison to the descending limb (8th–24th hour). Following this variant, daily totals are divided into 24 h, as presented in Fig. 2 (bottom panel). The advantage of this variant is that it provides a more realistic representation of the precipitation event than the uniform distribution of the M1 method. The disadvantage is that, in the case of precipitation events lasting more than one day, an artificial drop in the precipitation intensity may be generated at the end of each precipitation day.

Note that both method variants use the mean areal catchment precipitation totals and thus do not account for a spatial variability in the precipitation totals within the catchment.

3.1.2 M2: forward approach (micro-canonical cascade model)

The disaggregation method M2 is a multiplicative micro-canonical cascade model, which can be considered as a forward approach, i.e. the disaggregation of daily precipitation time series is performed based purely on precipitation data without considering any runoff-based simulations. This method has two variants (M2a and M2b), which are briefly explained below. The interested reader can find a more detailed description in Müller and Haberlandt (2015).

The general scheme of micro-canonical cascade model is shown in Fig. 3. A coarse time step is split into b finer time steps, where b is the branching number. For $b = 2$, the precipitation amount can either be forwarded to the first (1/0) or the second (0/1) of the two finer time steps (with the probabilities $P(1/0)$ and $P(0/1)$, respectively), or distributed on both finer time steps ($x/(1-x)$, with $P(x/(1-x))$). The numbers in brackets (e.g. 1/0) represent the fraction of precipitation forwarded to the first ($1 = 100\%$) and the second ($0 = 0\%$) finer time step, whereby x in $x/(1-x)$ represents a fraction of $0 < x < 100$. For the $x/(1-x)$ splitting, a distribution function $f(x)$ is required. Here, we use an

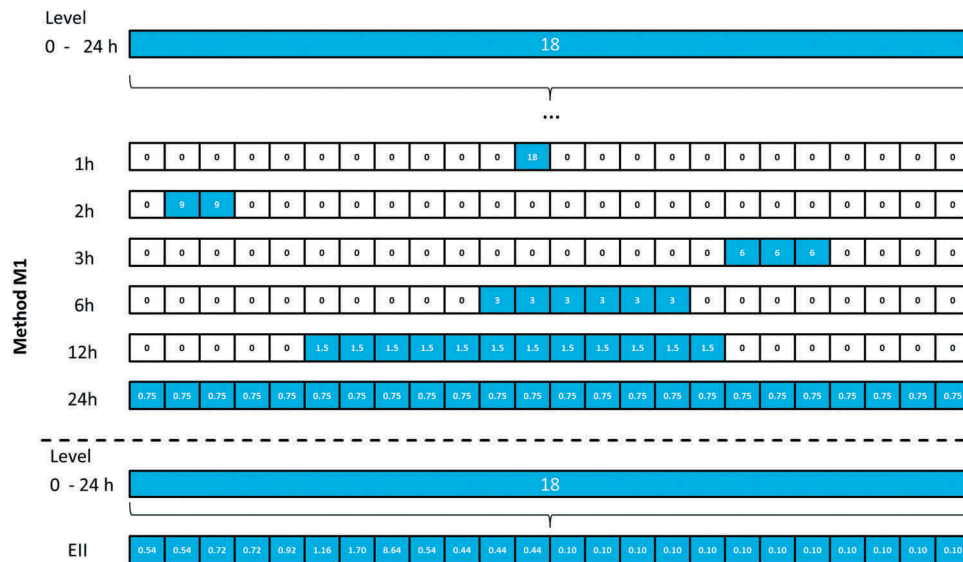


Figure 2. Scheme of the precipitation disaggregation method M1 with two variants: 1-h to 24-h (Uniform) and EII (Triangular). All precipitation amounts are given in mm/ Δt .

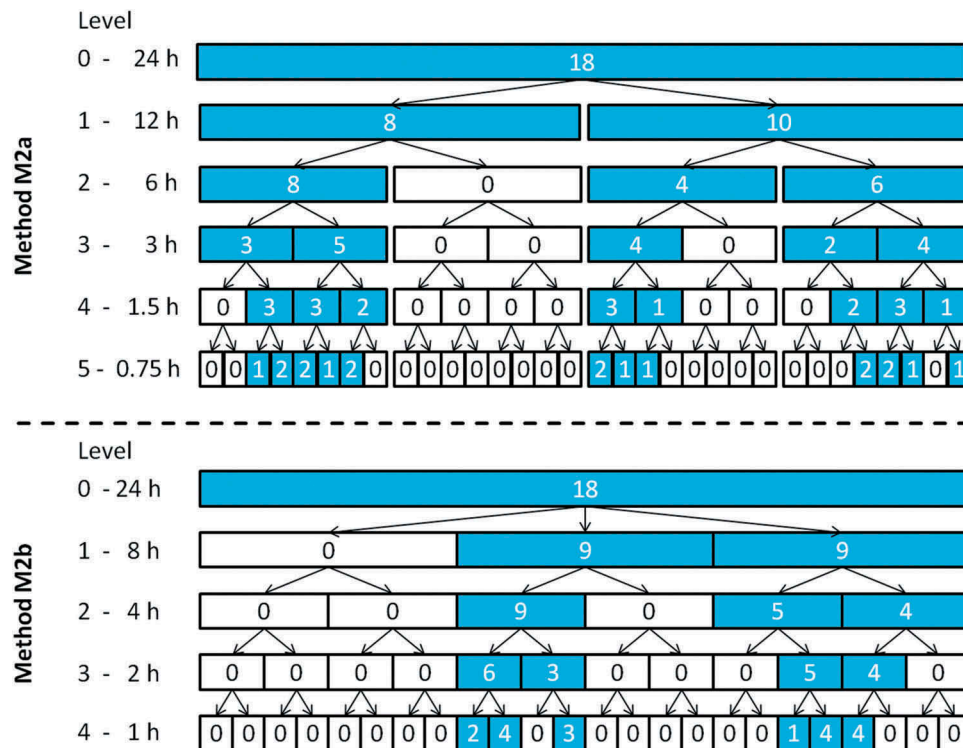


Figure 3. Scheme for cascade model M2a (Diversification) and M2b (Uniform Splitting). All precipitation amounts are given in mm/ Δt .

empirical distribution and so no additional parameters of a theoretical distribution function are required. However, the complexity of the disaggregation method is higher with $f(x)$ than without. Hence, we consider the empirical distribution function simplified as a fourth parameter in this investigation together with $P(1/0)$, $P(0/1)$, $P(x/(1-x))$ for the sake of comparability. Nevertheless, it remains an empirical distribution function and is not a single parameter value.

All four parameters can be estimated by running the cascade model backwards (Carsteanu and Foufoula-Georgiou 1996), i.e. by aggregating observed hourly time series, counting the three possible crossings and calculating the corresponding probabilities.

Olsson (1998) and Güntner *et al.* (2001) identified physical reasonable differences of the parameters if estimated in dependency (a) on the wetness-state of the time step before and after

the considered time step being disaggregated and (b) on its rainfall amount. To consider these dependencies, the disaggregation method parameters are estimated with a dependence of the position and the volume classes. Regarding (a), four different position classes are distinguished: starting (dry [time step before]-wet [considered time step being disaggregated]-wet [time step afterwards]); enclosed (wet-wet-wet); ending (wet-wet-dry); and isolated (dry-wet-dry). Regarding (b), for each position class, there is a lower and an upper volume class, with the mean of all precipitation intensities for this position and the temporal resolution of the disaggregation level being the threshold. All rainfall intensities below this threshold belong to the lower volume class, while all other rainfall intensities belong to the upper volume class.

For M2a, the so-called diversion approach, the same branching number ($b = 2$) is applied throughout the disaggregation steps (Fig. 3, top panel). When starting from daily values, a temporal resolution of 45 min is achieved after five disaggregation steps. Hourly time steps are achieved by a uniform distribution on 15-min time steps and, subsequently, aggregating four consecutive non-overlapping time steps.

For M2b (Fig. 3, bottom panel), the uniform splitting approach, $b = 3$ is applied in the first disaggregation step, resulting in 8-h time steps. No position classes are considered in this step, and only two volume classes are possible. The number of wet 8-h intervals is determined by $P(1/0/0)$, $P(1/2/1/2/0)$ and $P(1/3/1/3/1/3) = 1 - P(1/0/0) - P(1/2/1/2/0)$. For all further disaggregation steps, the branching number $b = 2$ is applied.

Hence, method M2a consists of 32 parameters, with 4 position classes \times 2 volume classes \times 4 splitting parameters, while method M2b has four additional parameters – $P(1/0/0)$ and $P(1/2/1/2/0)$ for both lower and upper volume classes – resulting in 36 parameters in total.

The disaggregation itself is a random process, which leads to different results, depending on the random number generator initialization. To cover this random behaviour, several disaggregation runs have to be performed for each daily precipitation time series. Müller and Haberlandt (2015) analysed the resulting precipitation characteristics (similar to the event-based and continuous characteristics in Section 3.1.4) of the disaggregated time series and found that the average values did not change significantly after 80 realizations by an increasing number of realizations. Accordingly, 80 realizations of disaggregation are carried out for each daily time series.

In this study, the variants M2a and M2b are applied directly to time series of the catchment mean areal precipitation. The disaggregation of areal precipitation rainfall time

series was carried out with M1 before (Sikorska *et al.* 2018, except M1-EII), but not with M2a or M2b to the authors' knowledge. Hence, it first needs to be validated whether the M2 variants are able to generate precipitation characteristics in the disaggregated areal precipitation time series comparable to the areal precipitation time series directly estimated from station values. However, the cascade model parameters are estimated for areal precipitation in the same way as for station-based precipitation by running the cascade model backwards and aggregating the hourly time series of estimated areal precipitation. A comparison of precipitation characteristics for individual stations *versus* areal precipitation was performed for the example catchment of Kleine Emme (C4) only.

3.1.3 Comparison of disaggregation methods and their limitations

The principle differences between the two methods M1 and M2 described above (and their two variants) are summarized in Table 2.

One of the major differences between these two methods is the number of parameters that need to be estimated. With the optimal wet spell duration being the only parameter for M1, it is also assumed that this parameter can be more easily regionalized for different regions or ungauged locations (without observed data) in comparison to the 32 parameters of M2a (or 36 for M2b). However, this hypothesis is not tested in the present study, which focuses only on the comparison between the disaggregation methods and their impact on the simulated flood characteristics. Indeed, the basis for a regionalization of M2 parameters was tested by Müller (2016), who showed that parameters in M2 demonstrate a small parameter variation for rain gauges located at similar altitudes in the same region. Optionally, the number of parameters for M2 could be reduced, because similarities between some parameters were found by Olsson (1998), Güntner *et al.* (2001) and Müller and Haberlandt (2018). Nevertheless, these simplifications were not applied in the present study (a) in order to enable comparisons with former authors' results and (b) due to the fact that the number of parameters for M2 after reduction would still be much higher in comparison to M1.

Both methods also differ regarding the parameter estimation method. For M1, the one parameter is estimated indirectly by precipitation–runoff modelling with the disaggregated time series and by analysing the impact of different disaggregation lengths on the flood peak efficiency (backward approach). For M2, all parameters are estimated from observed hourly precipitation time series without considering runoff simulations

Table 2. Properties of both disaggregation methods. EII: Euler type-II distribution.

Subject of comparison	M1		M2	
	1 h–24 h	EII	a	b
Number of parameters	1	–	32	36
Parameter estimation	Indirectly (backward estimation) from simulated peak flows	–	Directly (forward estimation) from observed hourly precipitation time series	
Shape of sub-daily precipitation	Block precipitation with uniform distribution	Block precipitation with triangular shape	Irregular precipitation intensities during the day	
Daily precipitation amount	Conserved exactly		Conserved exactly	

(forward approach). Thus, the choice of the disaggregation model may depend in practice on the available data (only precipitation or also runoff for calibration of the hydrological model). For catchments without runoff observations, the M1 parameter may be estimated with runoff peaks available for downstream gauges, or through a regionalization of the hydrological model prior to the M1 parameter estimation (Sikorska *et al.* 2018). For M2, satellite (Lütke-meier *et al.* 2018) or radar data can be a possible solution for catchments without observed hourly precipitation time series. The same applies for the availability of the hourly precipitation data for the method M1.

Also the sub-daily precipitation intensity distribution in the disaggregated time series differs between the two methods and their variants. While for M1 a block precipitation with uniformly distributed precipitation intensities is generated, for M1-EII, the precipitation intensities follow a triangular shape with a peak falling in the 8th hour of a day. For M2, precipitation intensities are distributed irregularly in different events during the day. A disadvantage of uniform M1 is the constant generation of a single block event during a day. Thus, for precipitation events lasting for several days, artificial dry spells may be generated in between. This issue is overcome with M1-EII, in which each event is extended over a day. However, hours with artificially low intensities may be generated at the end of each precipitation day. In contrast, with M2, it is possible to generate precipitation events starting and ending on different days, although dry spells can also be generated in between (Müller and Haberlandt 2018). Even so, it is unclear whether for long-lasting events the distribution of precipitation intensities during a day plays a significant role, or whether preserving the total precipitation amount is more important.

Despite the aforementioned differences, both methods M1 and M2 start with the disaggregation from daily totals and conserve the daily precipitation amounts exactly.

3.1.4 Evaluation of disaggregated precipitation time series

For the evaluation of the disaggregated precipitation continuous time series, event-based characteristics and extreme values are analysed. The continuous time series characteristics, i.e. the average intensity (mean intensity of all wet time steps) and the fraction of dry intervals (number of dry time steps, 0 mm/h) are calculated in relation to number of all non-missing time steps over the whole time series.

For the calculation of the event-based characteristics, i.e. wet spell duration, wet spell amount and dry spell duration, an event is defined by a minimum of one dry hour before and after one or more wet time steps. The event-based characteristics are next averaged over all events found in the precipitation time series.

For the analysis of precipitation extreme values, the event definition is slightly different to ensure the independence of the precipitation events, and it depends on the extreme event duration under investigation (1, 6 or 12 h). Here, we follow the suggestion of Schilling (1984), who ensured the independence of an extreme event by applying the same number of dry hours as the extreme event duration under investigation before and after the event, with a minimum of 4 h. For example, for precipitation extreme events of 1 h duration,

an event is considered as independent if there is a minimum of 4 dry hours before and after the extreme event, while for durations of 6 and 12 h, minimums of 6 and 12 h are required, respectively.

The aforementioned precipitation characteristics (PC) of the simulated time series (sim) are evaluated by a comparison with the observed time series (obs) using the relative error rE:

$$rE = \frac{PC_{sim} - PC_{obs}}{PC_{obs}} \quad (1)$$

3.2 Precipitation–runoff modelling

3.2.1 Lumped hydrological model

The HBV light model (Seibert and Vis 2012) is used in this study to simulate runoff at the catchment outlets. The model consists of four routines: (1) precipitation excess and snow-melt, (2) soil moisture, (3) response (groundwater) and (4) routing in the river. For catchments with an areal ratio of glacier areas $\geq 5\%$, an additional glacier routine is included. Here we implement the glacier routine for three catchments (C7–C9). The HBV model is run on an hourly time step in this investigation and was calibrated using the Kling-Gupta efficiency (KGE; Gupta *et al.* 2009) with the observed hourly areal precipitation and runoff time series. For the model calibration a Genetic Algorithm and Powell optimization (GAP) approach was applied, which relies on an evolutionary mechanism of selection and recombination of parameter sets (parameter population) randomly selected within parameter boundaries defined by the user. The evaluation of parameter sets is based on the selected objective function (here KGE) and only the sets maximizing the objective function are retained (see Seibert 2000 for details). The number of required parameter sets to be retained is specified by the user. In this study, this number was set to 100 for each catchment resulting in 100 optimal parameter sets. This enables us to represent uncertainty in the parameterization of the hydrological model.

The period of 1990–1999 was used for the calibration of the HBV model. The achieved median efficiency (KGE) over all catchments in this period was equal to 0.79 (range: 0.60–0.90), while the median KGE in the application period (1980–2014) was 0.75 (range: 0.50–0.88). No catchment-dependent relationships of the model performance could have been established.

For both precipitation disaggregation methods and both their variants, the same calibrated model parameters are used. As our previous study (Sikorska *et al.* 2018) showed, the parameter uncertainty of a hydrological model does not play any important role in the selection of the effective precipitation length. Thus, we perform the disaggregation comparison using only the best parameter set of the hydrological model. The effect of parameter uncertainty is investigated only in our example catchment (Kleine Emme, C4).

3.2.2 Evaluation of simulated runoff time series

The calibrated HBV model is next run for each catchment with the best parameter set and with disaggregated precipitation time series (1981–2014) derived from two different

Table 3. Rainfall characteristics of time series for four stations in the catchment Kleine Emme as well as the areal precipitation time series and relative error (rE) resulting from the disaggregation with M2a and M2b. Abs.: absolute value.

	Wet spell duration (h)			Wet spell amount (mm)			Dry spell duration (h)			Fraction of dry intervals			Average intensity (mm/h)		
	Abs.	rE (%)		Abs.	rE (%)		Abs.	rE (%)		Abs.	rE (%)		Abs.	rE (%)	
		M2a	M2b		M2a	M2b		M2a	M2b		M2a	M2b		M2a	M2b
Station 1	3.9	34	-21	3.7	2	-13	25.7	-2	-11	0.9	-5	1	1.0	-23	11
Station 2	3.6	34	-19	3.6	4	-12	23.5	-1	-11	0.9	-4	1	1.0	-23	9
Station 3	4.0	34	-20	4.2	6	-12	21.4	1	-11	0.8	-5	2	1.1	-21	10
Station 4	4.6	34	-22	5.2	9	-14	19.9	3	-12	0.8	-5	2	1.1	-19	11
Mean of stations 1–4	4.0	34	-21	4.2	5	-13	22.6	0	-11	0.8	-5	2	1.0	-21	10
Areal precipitation	6.0	34	-23	4.0	8	-15	17.2	-1	-13	0.7	-8	3	0.7	-19	10

methods, M1 and M2. The runoff simulation obtained with the observed hourly precipitation time series is used as a benchmark for comparing disaggregation methods and their two variants. Note that we do not use the observed runoff series for comparison purposes in order to focus entirely on the effect of disaggregating precipitation data on simulated flood events. In this way, we exclude possible side effects due to the impact of the model structure uncertainty (and the imperfect match between the calibration data and model simulations), as well as possible uncertainty of runoff data.

For the evaluation of simulated runoff time series (derived by different disaggregated precipitation time series) two metrics are used for annual maximum floods: relative peak and relative volume. Both metrics are essential for the validation of the disaggregated precipitation time series, since Brunner and Sikorska-Senoner (2019) have proven that flood volumes are not necessarily well reproduced if only flood peaks are considered in the calibration of the hydrological model.

The term “relative peak” was introduced by Sikorska *et al.* (2018) and it is computed from simulated annual maximum flood peaks by division by the benchmark peaks. In a similar way, we introduce here a “relative volume”, which is represented by simply dividing the volume estimated from simulated events by the event benchmark volumes. If the relative peak or the relative volume equals unity, the perfect agreement between the benchmark and the simulated peak (volume) is obtained. A value for a relative peak (volume) >1 implies that the peak (volume) is overestimated, whereas a value <1 indicates an underestimation of peak (volume). Thus, the disaggregation method that results in a relative peak (volume) lying closest to unity has the highest agreement with the benchmark (hourly data).

The volume of flood events is here defined based on the selected annual maxima using a fixed window of 72 h (i.e. 24 h prior to and 48 h after the flood event peak) and summing up the area under the hydrograph during this time window.

4 Results

4.1 Precipitation disaggregation

4.1.1 Comparison of station-based and areal precipitation disaggregation in M2

Daily values of areal precipitation were disaggregated by the aforementioned methods to hourly values. As an example, the precipitation characteristics for individual stations in the Kleine Emme (C4) catchment were compared with those of the catchment mean areal precipitation; these are shown in Table 3 along with the corresponding relative errors (rE) for M2a and M2b. It should be noted that precipitation characteristics of the areal precipitation time series differed from those of the station time series. As expected, the mean wet spell duration was longer for the areal precipitation (approx. +50%), presumably due to the movement of single precipitation events over the catchment occurring at different stations at different times. Hence, dry spell durations were shorter and the fraction of dry intervals was lower in comparison to station time series, leading to lower values for the average intensity of all wet time steps (approx. –30%).

However, the rE resulting from the disaggregation with M2a and M2b were comparable for both time series types. Wet spell duration was overestimated by 34% with M2a, independently of the time series type. For the wet spell amount, dry spell duration, average intensity and the fraction of dry intervals, estimated rE for areal precipitation was similar to the mean rE computed over stations 1–4 and lay within the range of the single rE values. Method M2b led to a better representation of the wet spell duration, the fraction of dry intervals and average intensity, while the wet spell amount and dry spell duration were better represented by M2a.

The empirical distribution functions of extreme precipitation values for the Kleine Emme catchment are presented in Fig. 4. Observed extreme values with durations of 1 and 12 h could be well reproduced by method M2a for the catchment mean areal values. For station precipitation amounts, extreme values were well represented for 12-hour duration, but underestimated for 1-h duration. Method M2b led to a general underestimation of extreme values for 12-h duration, for both station and areal precipitation, while for 1-h duration extreme values were overestimated, which means higher precipitation amounts were clustered in 1 h than in the observed extreme values. Extreme values were additionally analysed for a 6-h duration, and the obtained results were similar to those achieved for a 12-h duration.

In summary, the rE values for the precipitation characteristics of the continuous time series resulting from the disaggregation of areal precipitation were similar to those resulting from disaggregation of station time series for both M2a and M2b. Thus, our results show that, in general, method M2 can be applied for the disaggregation of areal precipitation time

series, whereby M2a leads to better representation of the extreme values, while M2b leads to better representation of the event and continuous time series characteristics.

4.1.2 Comparison of disaggregated areal precipitation time series

The disaggregated time series were analysed regarding their continuous and event-based characteristics, as well as their extreme precipitation values for all nine catchments. For both methods M1 and M2, the disaggregation is assumed to be a random process leading to hourly time series with different distributions of the precipitation over each day, also known as realizations. While for M2 the structure of the different time series differs regarding the sub-daily precipitation distribution with different intensities and dry spells in between, for M1 only the starting point of the precipitation event differs between days. Following Sikorska *et al.* (2018), for M1 one realization per catchment was generated, which is assumed to be representative because the starting hour is chosen randomly for each day independently, while for M2, 80 realizations were generated, as suggested by Müller and Haberlandt (2015). The characteristics of the time series generated by method M1 were then compared against the medians of the precipitation characteristics of 80 realizations generated by method M2 (see Figs. 5 and 6, and Table 4 for a quantitative comparison over all stations). It should be mentioned here again that method M1 does not aim at the replication of observed precipitation characteristics. Thus, the intention of the comparison is only to investigate whether any differences exist between time series generated by both methods.

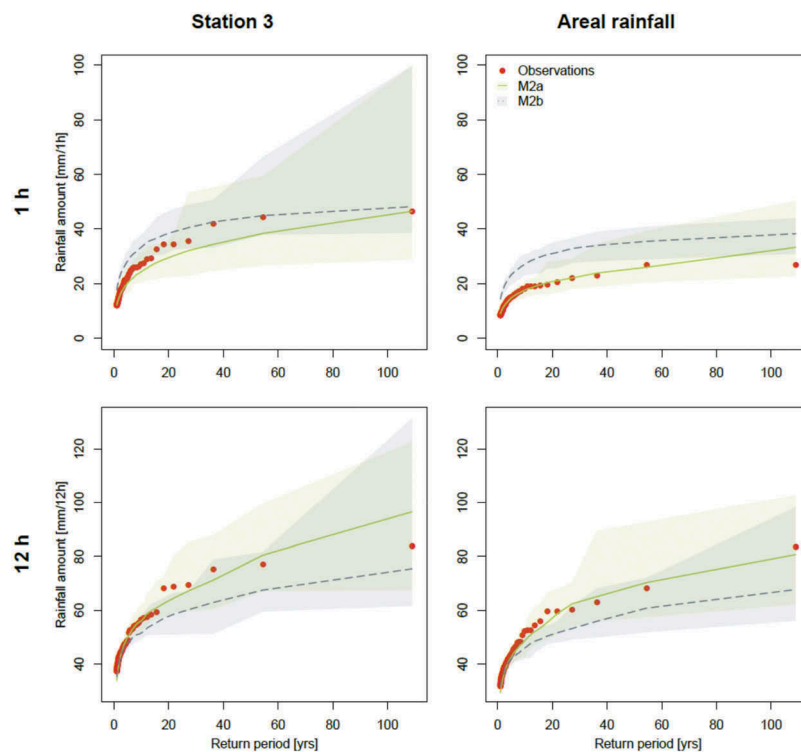


Figure 4. Extreme precipitation values for Station 3 (left column) and the catchment mean areal precipitation (right column) of the Kleine Emme catchment for durations of 1 h (upper panel) and 12 h (lower panel).

For M1, the dry spell duration decreased with increasing precipitation duration up to 12 h, while for a 24-h duration and the Euler type-II distribution (EII), a sudden increase could be detected. If the daily precipitation amount is distributed over 24 h, as in M1-24h and M1-EII, a disaggregated precipitation event could last for several days without any dry spells in between. Consequently, the fraction of dry time steps was much smaller in comparison to the other precipitation durations of M1 and was <50%. By contrast, dry spell durations achieved with M2 were comparable to the observations, but the fraction of dry intervals was underestimated by M2a ($rE = -7.1\%$) and slightly overestimated by M2b (2.6%). For the average precipitation intensity, overestimation was identified by M2b (10.3%) and underestimation by M2a (-20.4%). These results are in accordance with the findings of Müller and Haberlandt (2015) for stations in temperate oceanic and temperate continental climate (Peel *et al.* 2007) in Lower Saxony, Germany. For M1, the best results were obtained for the durations of 6 h ($rE = 50.1\%$) and 12 h (-25%). This finding is in agreement with Sikorska *et al.* (2018), who identified both these durations for the same study region as the most suitable for deriving a disaggregated input for precipitation-runoff simulations based only on flood peaks.

For the analysis of the precipitation extreme values of the disaggregated time series, the following empirical return periods (T_n) were analysed: T_n of 4.95, 9.9, 21.8, 54.5 and 109 years. Since the results were very similar for all return periods, only those for $T_n = 109$ years, which can be assumed as being close to a 100-year return period, are shown in Fig. 6.

For precipitation extreme events of 6- and 12-h duration, the results for M1-1h, M1-2h, M1-3h and M1-6h for 6-h duration, and M1-1h, M1-2h, M1-3h, M1-6h and M1-12h for 12-h duration, are identical. This is due to the fact that the applied methodology distributes the daily precipitation amount over a defined precipitation duration, which is shorter than the extreme value duration of the event under investigation. Due to the limited precipitation duration in a day, for extreme events with longer durations the generated event can only be extended by dry time steps, thus the absolute precipitation amount cannot increase. This results in a similarity for effective precipitation durations shorter than extreme events under investigation and is due to the assumptions of method M1.

For a 12-h duration extreme, all variants of method M1, apart from M1-24h (where an underestimation was observed), resulted in a good agreement between disaggregated data and observations. For a 6-h duration, M1-1h–M1-6h and M1-EII led to overestimation and M1-12h and M1-24h to underestimation. For a 1-h duration, M1-1h–M1-2h and M1-EII resulted in overestimation, M1-6h–M1-24h in underestimation, while M1-3h was close to the value estimated from observations.

With M2a and M2b, a slight underestimation was observed for 12-h duration. For a 6-h duration, an overestimation was found for M2a and underestimation for M2b, while for a 1-h duration a slight overestimation for M2b and a good agreement with the observations for M2a were identified. Thus, M2a represented precipitation extreme values slightly better for 1-h and 12-h duration than M2b.

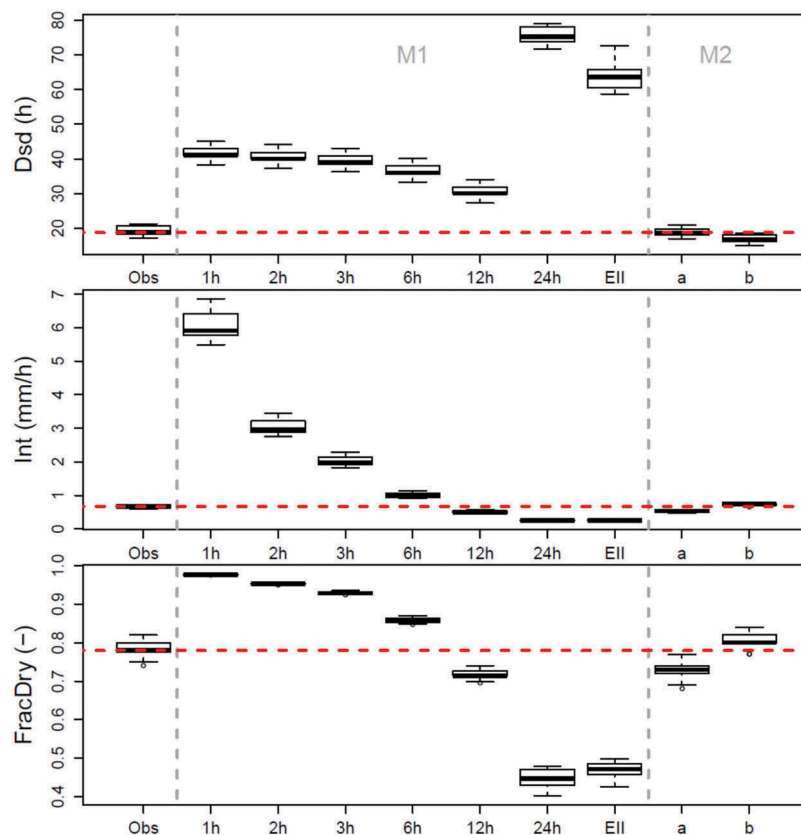


Figure 5. Comparison of dry spell duration (Dsd), average intensity (Int) and fraction of dry intervals (FracDry) of the disaggregated time series for all areal precipitation time series. The dashed (red) line indicates the median of the observed time series (Obs).

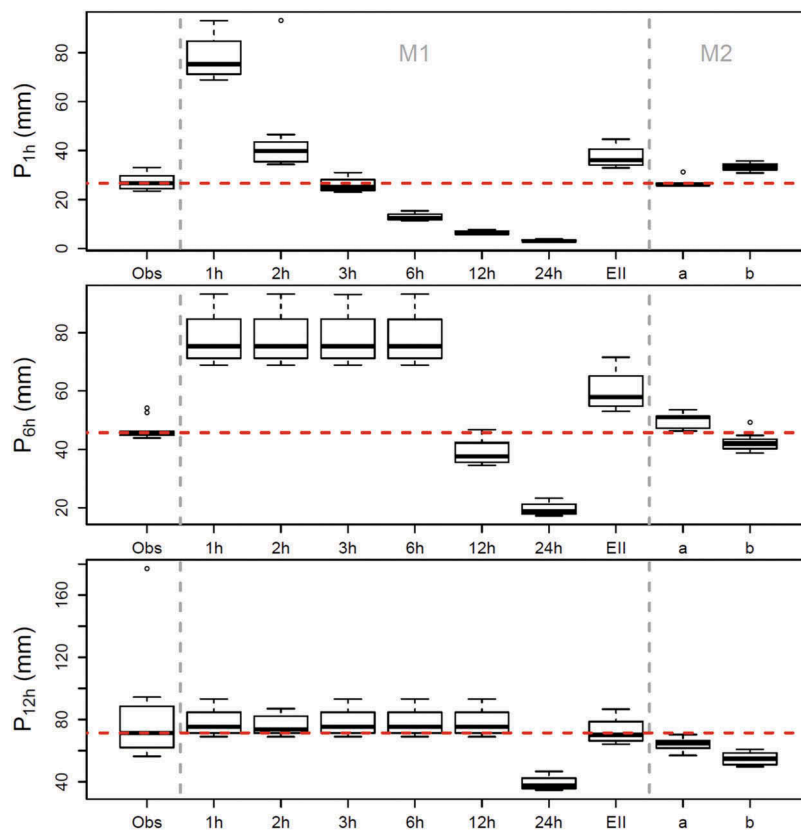


Figure 6. Comparison of precipitation extremes of the disaggregated time series for a return period of approx. 100 years for different durations for all areal precipitation time series. The horizontal dashed (red) line indicates the median of the observed time series (Obs).

Note, however, that the structure of the cascade model (M2) does not intend to reproduce extreme values primarily, but rather to generate continuous time series.

4.2 Precipitation–runoff modelling

4.2.1 Relative annual peaks and volumes for M1 (backward approach)

The precipitation time series disaggregated with method M1 and for different lengths were introduced into the precipitation–runoff model and the resulting runoff series were analysed for annual maximum floods (peak and volume). Next, relative values for both metrics were computed in accordance to the benchmark; these are visualized in Fig. 7 and summarized as relative errors (rE) in Table 5.

At first sight, it appears that longer disaggregation levels (between 6 and 12 h) yielded similarly good results, i.e. relative error close to unity for both variables (peak and volume), whereas the variant M1-EII led to a slight underestimation of peaks.

These observations were also confirmed with the computed rE values (Table 5). For most catchments, the effective precipitation length was 12 h and for catchments C3 and C5, 6 h, assessed by relative peak. Shorter disaggregation lengths resulted in overestimation, whereas longer lengths resulted in underestimation of annual maximum flood peaks. This is in agreement with the previous study (Sikorska *et al.* 2018), which used the same catchments but different type of

precipitation data (different source). The disaggregation M1-EII led to a higher agreement with the benchmark for four catchments and a lower agreement for five catchments.

Catchments C3 and C5, with an effective precipitation length of 6 h, are the most western catchments in this study and belong to the cold climate in the pre-alpine region (Peel *et al.* 2007), while all other catchments except C1 belong to the polar climate and the polar region. Also, for C1 the results for relative peaks (Fig. 7, y-axis) are very close for the effective precipitation length of 6 and 12 h. It could be hypothesized that the effective precipitation length in an alpine region is longer than in a pre-alpine region. However, this was not confirmed by the estimated mean wet spell duration of the areal precipitation (Table 1), which was quite similar for all catchments. Also, an investigation of the intra-annual distribution of precipitation extreme values for durations of 1, 6 and 12 h did not show any differences between these two catchments, i.e. C3 and C5, and all other catchments (not shown here).

Regarding the relative volume (Fig. 7, x-axis and Table 5), we found that all disaggregation lengths resulted in a similar agreement with the benchmark, and the best estimate was spread between 1 h and 24 h, depending on the catchment. Method M1-EII usually led to a slightly worse agreement than the best estimate from M1-1h to M1-24h.

Thus, the disaggregation length had a much higher impact on the efficiency in simulating the peak than the volume of annual maximum floods. This can also be seen by a larger spread of points along the y-axis than the x-axis on Fig. 7 for most of the catchments.

Table 4. Mean precipitation characteristics over observed and disaggregated areal precipitation time series and relative errors for all applied disaggregation methods and variants over all stations.

	Wet spell duration (h)		Wet spell amount (mm)		Dry spell duration (h)		Fraction of dry intervals (-)		Average intensity (mm/h)	
	Mean	rE (%)	Mean	rE (%)	Mean	rE (%)	Mean	rE (%)	Mean	rE (%)
Obs	5.3	-	3.5	-	19.3	-	0.8	-	0.7	-
M1-1	1.0	-83	6.1	50	41.7	142	0.98	32	6.1	800
M1-2	2.0	-67	6.1	50	40.7	137	0.95	28	3.0	350
M1-3	3.0	-50	6.1	50	39.7	131	0.93	25	2.0	200
M1-6	6.0	0	6.1	50	36.7	113	0.86	16	1.0	50
M1-12	12.0	100	6.1	50	30.7	78	0.72	-3	0.5	-25
M1-24	94.5	1479	24.5	505	75.6	340	0.45	-40	0.3	-62
M1-EII	73.2	1123	19.8	388	64.3	274	0.47	-37	0.3	-60
M2a	7.1	33	3.7	6	19.0	-2	0.73	-7	0.5	-20
M2b	4.1	-23	3.0	-15	16.9	-13	1.27	3	0.7	10

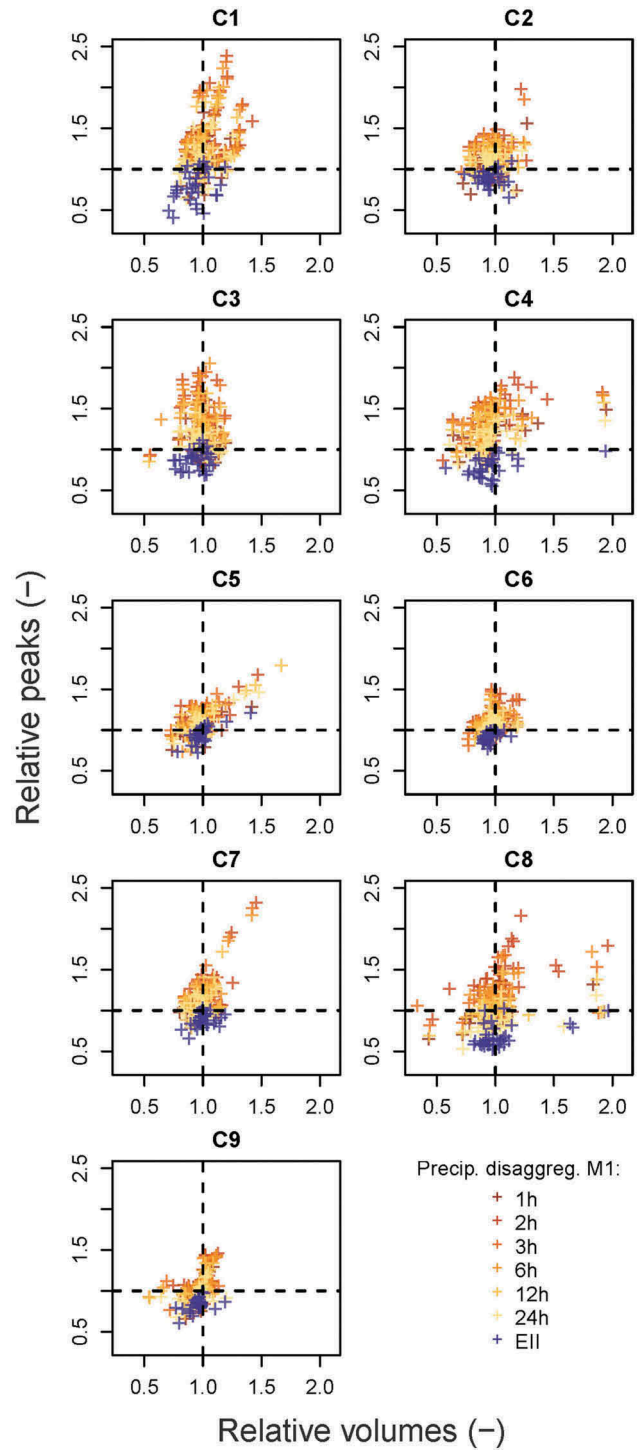


Figure 7. Relative peak vs volume on an annual basis (34 years) for nine study catchments using the disaggregation method M1 and different disaggregation variants.

4.2.2 Relative annual peaks and volumes for M2 (forward approach)

For method M2 (variants a and b), 80 different realizations of the disaggregated precipitation time series were simulated with the precipitation–runoff model. The resulting runoff time series were next analysed for annual maximum floods and the relative peaks and volumes were computed in accordance to the benchmark. For comparison purposes with the method M1, an ensemble mean was computed over all 80

Table 5. Summary of the comparison of runoff simulation results for methods M1 and M2 (mean of 34 annual relative values). The estimate closest to the benchmark (i.e. 1) is underlined, while the overall best variant is double underlined. Colours indicate deviations (%) from the benchmark results (green: $|rE| \leq 5\%$, yellow: $|rE| \leq 10\%$, red: $|rE| > 10\%$).

Catchment	M1					M2			
	1h	2h	3h	6h	12h	24h	EII	a	b
<i>Relative peak</i>									
C1	1.16	1.10	1.07	1.12	1.04	0.92	1.02	0.99	0.98
C2	1.23	1.19	1.18	1.10	0.99	0.86	0.99	0.96	0.97
C3	1.40	1.12	1.14	0.99	0.83	0.66	0.95	0.88	0.94
C4	1.37	1.32	1.32	1.22	1.05	0.76	1.15	0.96	0.98
C5	1.09	1.06	1.05	0.98	0.92	0.80	0.96	0.91	0.91
C6	1.38	1.38	1.34	1.23	1.07	0.85	1.07	1.01	1.04
C7	1.13	1.09	1.11	1.05	1.01	0.89	1.00	0.96	0.97
C8	1.44	1.40	1.38	1.29	1.09	0.77	1.10	0.99	1.00
C9	1.23	1.17	1.17	1.12	1.03	0.86	0.99	0.96	0.99
<i>Relative volume</i>									
C1	1.00	0.97	0.95	1.01	1.00	0.99	1.00	1.00	0.98
C2	1.04	1.03	1.02	0.98	0.97	1.00	0.97	0.99	1.00
C3	1.06	1.00	1.01	1.01	1.04	1.04	0.99	1.00	0.99
C4	1.00	0.95	0.96	0.95	0.95	0.97	0.97	0.96	0.95
C5	0.99	0.99	0.96	0.95	0.96	0.96	0.97	0.96	0.95
C6	0.98	0.99	0.98	0.99	0.99	0.94	1.00	0.97	0.96
C7	0.98	0.95	0.98	0.96	0.98	0.96	0.98	0.96	0.96
C8	1.08	1.05	1.05	1.04	1.00	0.93	0.99	0.98	0.97
C9	0.99	0.96	0.97	0.98	0.96	0.95	0.97	0.97	0.98

realizations for each catchment and these values are presented for relative peaks and volumes in Figs. 8 and 9 for variants M2a and M2b, respectively. The ensemble mean was computed as a mean over all simulated runoff hydrograph characteristics (peak and volume) based on 80 realizations.

Relative peaks of both method variants M2a and M2b were close to unity and in six catchments M2b gave a better agreement with the benchmark than M2a (see Table 5 for a quantitative comparison). Regarding relative volumes, both M2a and M2b slightly underestimated the flood volume, whereby M2a led to a better agreement for six catchments in comparison to M2b.

We also found that simulations resulting from different realizations of disaggregated data varied only slightly and for relative error in the range 0.5–1.5. Thus, this uncertainty had rather a small effect on the relative peaks or volumes.

4.2.3 Relative annual peaks and volumes – comparison of M1 and M2

The estimates for flood peaks and volumes resulting from all investigated variants of methods M1 and M2 are summarized in Table 5. The applied colour scheme of categories used for the deviations from the benchmark in Table 5 has no scientific background and is intended just to support visual interpretation of the results. In general, flood volume could be estimated more certainly (relative error for all estimations of $|rE| \leq 10\%$ and for most variants, $|rE| \leq 5\%$) compared to flood peak ($|rE| > 10\%$ for roughly 14% tested disaggregation set-ups over all method variants) independently of the disaggregation method applied.

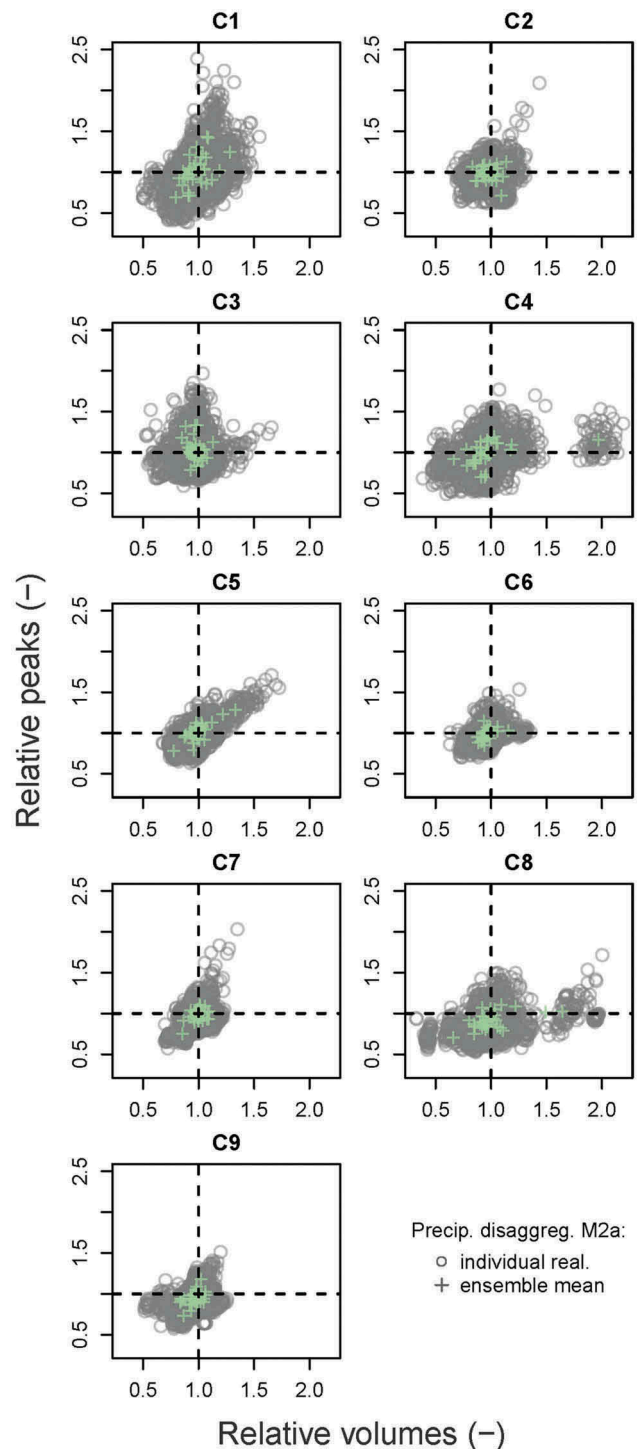


Figure 8. Relative peak vs volume on an annual basis (34 years) for nine study catchments using the disaggregation method M2a. The grey points represent different realizations from the disaggregation method, and the coloured points represent their ensemble mean.

For annual maximum flood volume, M1 led to the best results for all catchments, while M2 led to a comparably good estimation in only three catchments. However, it should be noted that, due to the very similar estimates, the best fit was achieved by six out of seven different variants of M1 (4 × M1-1h, 2 × M1-2h, 1 × M1-3h, 3 × M1-12h, 1 × M1-24h, 2 × M1-EII). Regarding annual maximum flood peak, M1 and M2 led to the best estimation for four out of nine analysed catchments each

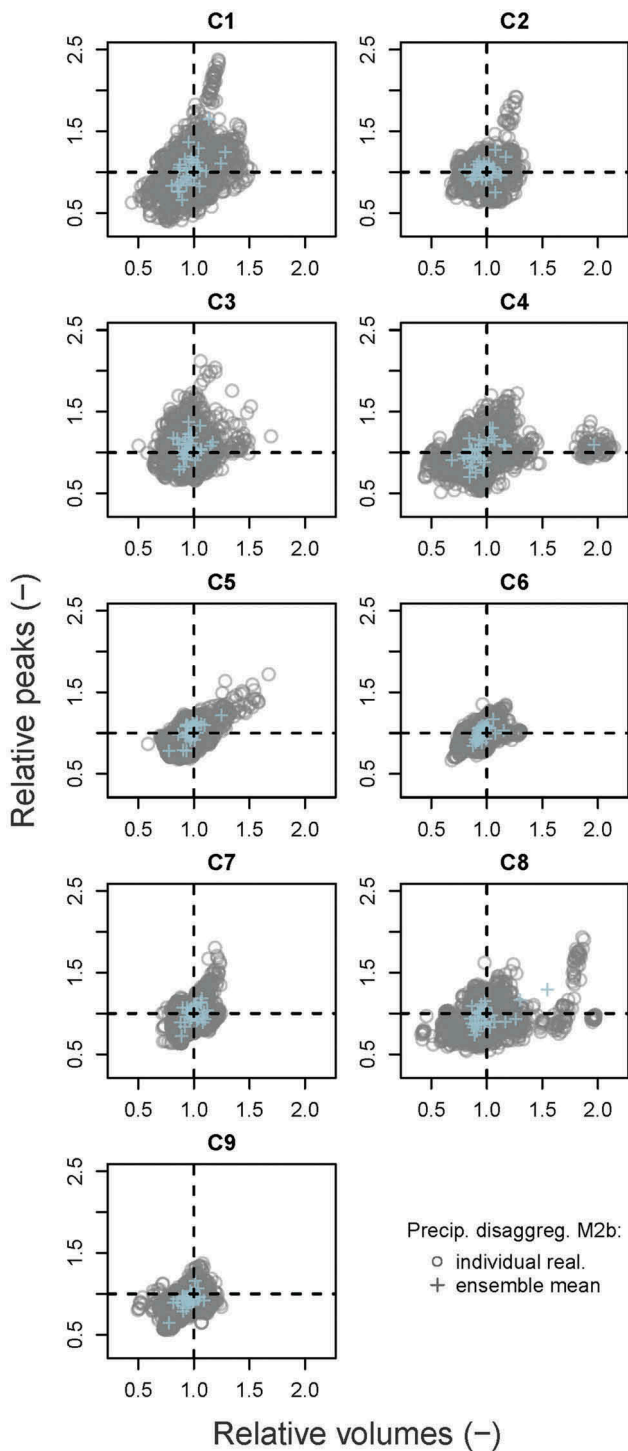


Figure 9. Relative peak vs volume on an annual basis (34 years) for nine study catchments using the disaggregation method M2b. The grey points represent different realizations from the disaggregation method, and the coloured points represent their ensemble mean.

(M1: $2 \times$ M1-6h, $1 \times$ M1-12h, $2 \times$ M1-EII; M2: $2 \times$ M2a, $2 \times$ M2b). For catchment C9, M1-EII and M2b gave similarly good estimations, with a relative peak value of 0.99. Thus, it may be concluded that both methods M1 and M2 have a similar value for simulating peaks of annual maximum floods with a lumped precipitation–runoff model (in this case, HBV), with a slight advantage of M1 for simulating the volume of annual maximum floods.

It has to be stressed here, however, that, on the one hand, method M1 aims at selecting the optimal distribution length among all tested variants and it should not be surprising that some variants resulted in a much worse fit than others. On the other hand, for method M2 only the medians of 80 realizations were investigated. Analysing individual realizations most likely would lead to a worse fit for some realizations but possibly to a better fit for other realizations in comparison to M1. This was not analysed here, since we consider these realizations as uncertainty of method M2, which was further investigated (see Section 4.3.2).

Summarizing the results, M1 (best variant) and M2 gave a slightly better agreement with the benchmark for relative peaks in four catchments each, while in catchment C9 the agreement was the same for both methods. Regarding relative volume, M1 gave a slightly better agreement for all catchments, but all variants of M1 and M2 gave results close to the best variant chosen. The effective precipitation duration in M1 was of greater importance for relative peak than for relative volume, and was estimated as lasting for 6 to 12 h. Smaller precipitation intensities caused longer but less peaky floods.

4.3 Sensitivity analysis of flood simulations

4.3.1 Station-based vs mean areal catchment precipitation in M2

The disaggregation within method M2 was applied to the mean areal catchment precipitation estimated for each catchment. Optionally, the disaggregation could have been applied to each station within each catchment and then averaged over all stations to estimate the mean areal precipitation without any subsequent steps to increase the spatial consistence (as, for example, analysed by Müller and Haberlandt 2018). We tested whether the order of averaging precipitation values matters for the precipitation–runoff modelling on the example catchment C4 (Kleine Emme). The results are presented in Fig. 10.

Comparing the relative peak and volume computed on the mean areal precipitation values *versus* station values showed that there is only a small difference between these two methods. While for relative peak better estimates were achieved with the mean areal catchment precipitation (relative peak values of 0.96 and 0.98 for M2a and M2b, respectively) for disaggregation of areal rainfall (cf. 0.90 and 0.96 for station values), for relative volume station values gave estimates that were slightly closer to unity (relative volume values of 0.96 and 0.95 for M2a and M2b, respectively, for areal rainfall, cf. 0.97 and 0.96 for station values). Hence, it appears that when disaggregated precipitation time series are used as input for a lumped precipitation–runoff model, it does not matter much at which step precipitation series are averaged (before or after disaggregation).

4.3.2 Effect of parameter uncertainty on the disaggregation results

Recently, Sikorska *et al.* (2018) showed that the parameter uncertainty did not have much effect on the choice of

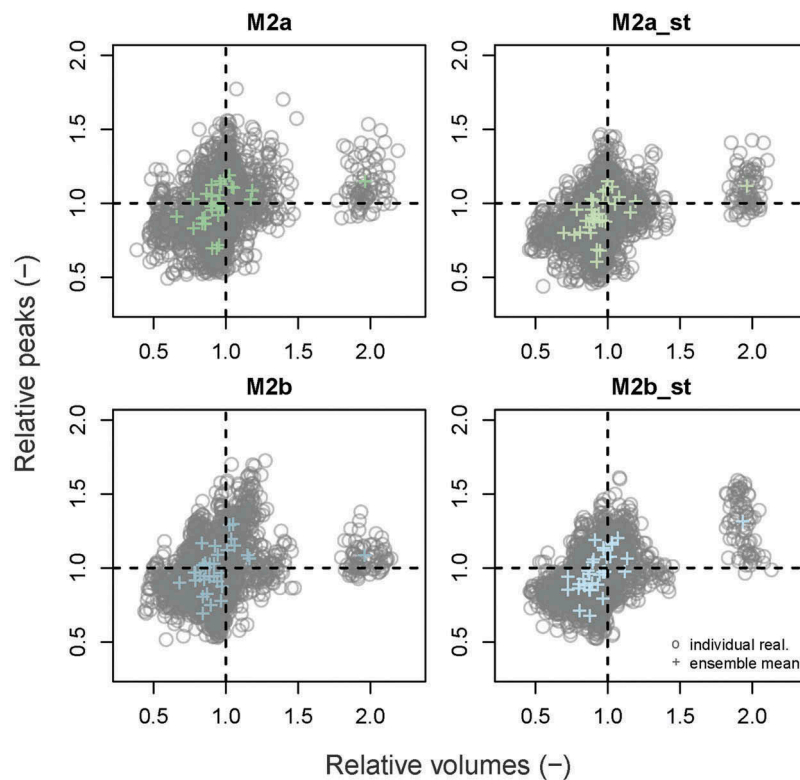


Figure 10. Relative peaks vs volumes on an annual basis (34 years) with the disaggregation M2 performed on the mean areal precipitation values and on the individual station records (_st) for the two variants M2a (upper panel) and M2b (lower panel) for the example catchment C4 (Kleine Emme), with the best HBV parameter set. The grey points represent 80 different realizations from the disaggregation methods, and the coloured points represent their ensemble mean.

disaggregation approach when using the method M1, and the disaggregation length was always between 6 and 12 h based on relative peak. Thus, in this study we assessed the effect of parameter uncertainty only for one example catchment – Kleine Emme (C4). As seen in Fig. 11, runoff simulations with 100 different parameter sets (grey points) follow generally the patterns of the best parameter set of the hydrological model for both relative peak and volume (coloured crosses), wherein the spread of uncertainty ranges (represented by the spread of grey points at x -axis and y -axis) is higher for short disaggregation lengths (1–6 h) and for EII, and smaller for longer lengths (12–24 h).

For method M2, 8000 simulations were computed for catchment C4 (Kleine Emme) that resulted from 100 parameter sets of the hydrological model and 80 different realizations of the disaggregated precipitation time series (seen as the method uncertainty). These results are presented in Fig. 12. As can be seen, the spread of grey points is much greater when both uncertainty sources are considered (parameter uncertainty of the hydrological model and M2 method uncertainty). It appears that the uncertainty due to parameters of the HBV model was greater than that due to different realizations of precipitation disaggregation time series (compare uncertainty spreads with Fig. 10, which presents only the uncertainty of method M2).

Note, however, that these two uncertainty sources cannot be separated here and thus a possible interaction between method uncertainty and uncertainty of the hydrological model cannot be excluded.

5 Discussion

5.1 Reproducibility of precipitation characteristics with M1 and M2

Our results show that method M2 could better represent the characteristics of precipitation time series than method M1, particularly for dry spell duration and fraction of dry intervals of the disaggregated time series.

Regarding the extreme precipitation values, the representation of the generated extremes when applying M1 depended on the investigated duration and the tested variant. While for shorter durations (1 h, 6 h) different M1 variants usually led to a worse representation of the observed characteristics than M2, for extreme events of 12-h duration, a similarly good representation could be achieved for M1 in comparison to M2. This was surprising because method M1, by definition, does not aim at reproducing precipitation characteristics. It seems that for the reproduction of a certain extreme precipitation characteristic a more complex disaggregation model is not necessarily better than a simpler model. Even though the structure of M2 is not intended to represent extreme precipitation values as well, the generated extreme precipitation values with 1-h, 6-h and 12-h durations were comparable to the observations.

By definition, M1 does not aim at reproducing precipitation characteristics exactly and was only developed to generate precipitation time series that are able to reproduce flood characteristics accurately. Hence, it was expected to achieve a better representation of precipitation characteristics with M2 than with M1. Moreover, although M1 preserves the daily

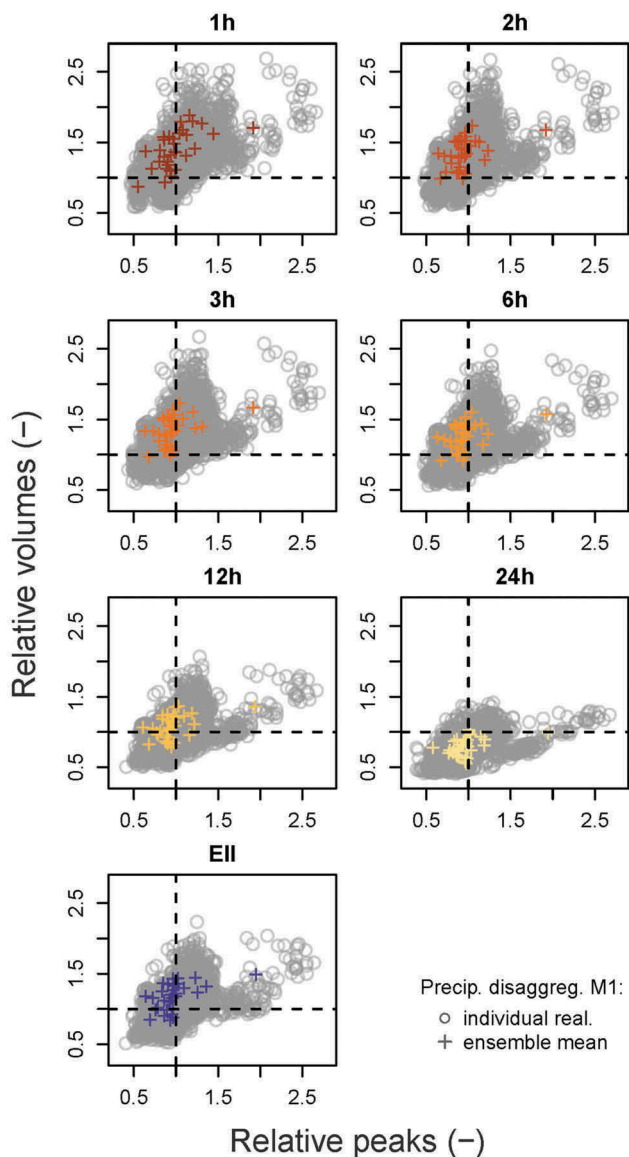


Figure 11. Effect of parameter uncertainty in the method M1 on the example of catchment C4 (Kleine Emme). The grey points correspond to simulations using 100 parameter sets of the HBV model, while the coloured points simulations with the best parameter set (relative peaks vs volumes on an annual basis, 34 years).

precipitation totals, it may generate artificial dry spells or cut out dry spells during long-lasting events. Not surprisingly, the properties of dry spells were thus better preserved with M2.

In addition, for M2 the comparison of the disaggregated time series of areal and station precipitation showed that the relative error (rE) was similar for several precipitation characteristics. This comparison on the level of precipitation characteristics revealed that method M2 is more suitable than M1 if the reproduction of precipitation characteristics is crucial, or the impact of precipitation characteristics on the achieved simulation results cannot be known.

5.2 Reproducibility of flood characteristics with M1 and M2

The comparison on the flood characteristics level showed that the differences seen between precipitation series disaggregated with

two different methods, M1 and M2, disappeared when these series were used as input for a lumped hydrological model. This is an interesting finding as it suggests that the complexity in the precipitation disaggregation method does not really matter for a lumped model. Thus, similarly good results could be obtained with a simpler 1-parameter approach (M1) as with a more complex 32-parameter method (M2). However, it is important to note that, although M1 contains seven variants that were analysed in this study, these variants must be seen as a single method and only the best variant selected should be considered as representative. The other variants are tested only for the purpose of selecting the best variant and should be rejected from the final application. Thus, it should not be a surprise that some individual variants of M1 led to a visible under- or overestimation of flood peak and volume. In contrast, M2 consists of two variants derived independently, which both provide an alternative way of driving precipitation time series. These two variants, M2a and M2b, provided flood estimates of a similar accuracy for different catchments.

This finding speaks for the suitability of both disaggregation methods: M1 (best variant), as well as M2a and M2b for generating hourly precipitation time series as input for a lumped hydrological model. This is also in agreement with previous research (Sikorska *et al.* 2018, Müller-Thomy *et al.* 2018), which suggested an existence of a dampening effect of the lumped hydrological model. Thus, a simpler 1-parameter disaggregation approach with a disaggregation length of 6–12 h remains attractive for deriving precipitation time series at an hourly resolution for precipitation–runoff modelling. This is particularly true for situations where 32 parameters of the more complex method M2 cannot be inferred.

We expect, however, that these differences between both methods will be larger for distributed hydrological models that rely on distributed input series per catchment unit. As we tested our methodology in mesoscale, natural mountainous catchments (area: 44–491 km²) only, these differences may also be higher in much smaller (area \ll 40 km²), fast reacting, or heavily urbanized catchments.

5.3 Limitations of the study

Some limitations of this study should be acknowledged, in terms of transferring our conclusions to other settings, catchments or precipitation–runoff models.

Firstly, the number of investigated catchments (9) is rather small and they share similar areal precipitation characteristics and physical properties. Thus, a generalization of our results for catchments in very different climate regions, or of different physical properties is not possible and the approach should be validated for the area of interest prior to application. Future research should include catchments with different hydro-climatic characteristics (for example different typology, land use, and different climate regions), to enable general conclusions to be drawn from the point of view of comparative hydrology (e.g. Gaál *et al.* 2012). In particular, the application of these two disaggregation methods in fast reacting or heavily urbanized catchments may yield different results.

Secondly, for the disaggregation of daily precipitation totals, only one variant of the simpler disaggregation method

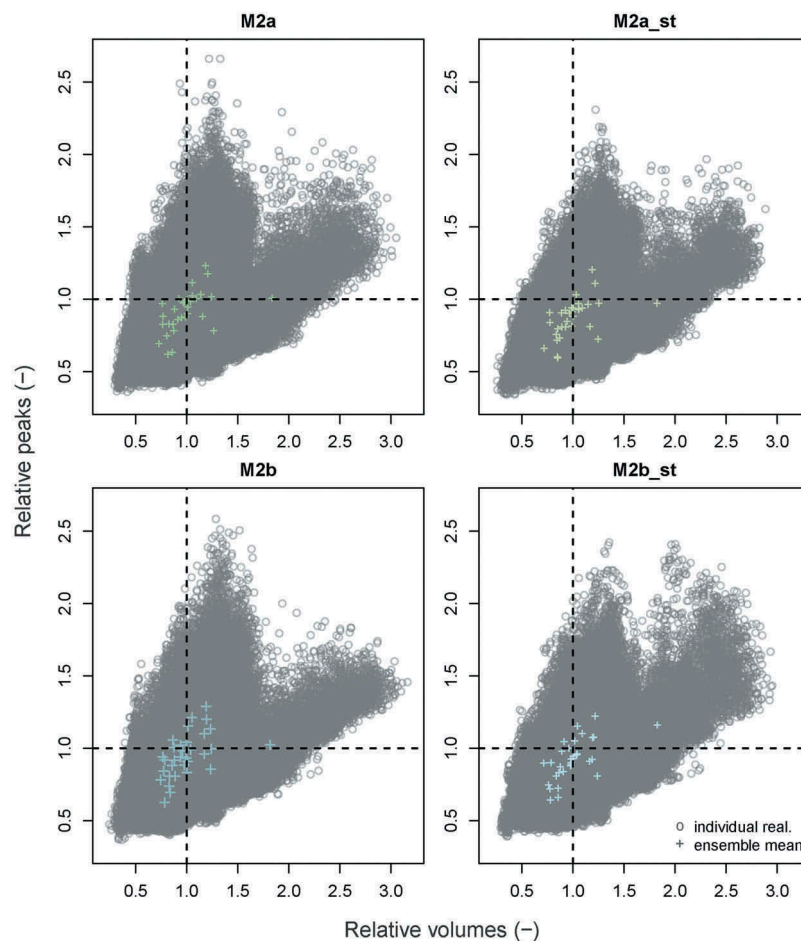


Figure 12. Effect of parameter uncertainty in the method M2 on the example of the catchment C4 (Kleine Emme) and the mean areal precipitation values and the individual station records (_st). The shaded (grey) points correspond to simulations using 100 parameter sets of the HBV model and 80 realizations of the method, and the coloured points represent simulations with the best parameter set and the ensemble mean of 80 realizations (relative peaks vs volumes on an annual basis, 34 years).

(M1) and a more complex (M2) disaggregation method have been applied. Our findings suggest that a simpler method provides similarly good results for runoff modelling compared to a more complex method (assessed by annual flood peak and volume). However, other disaggregation methods were not tested here. In addition, these results must be interpreted in terms of the hydrological model applied (i.e. a lumped model in this study) and, thus, cannot be generalized to other modelling approaches (see further below).

Third, for the precipitation–runoff modelling, only one hydrological model with one model structure was applied in this study. Although the applied HBV model demonstrated a good performance in this region before (Sikorska *et al.* 2018, Sikorska and Seibert 2018), testing other lumped hydrological models would lead to more robust and transferable conclusions. Also, for the calibration of the hydrological model, a single-objective calibration approach based on the Kling-Gupta efficiency was used. The choice of the model and its calibration strategy may have an impact on the simulation results. Yet, since the same hydrological model and the same parameters were used for both disaggregation methods, the differences observed between both methods are affected in the same way by these choices. Thus, the final conclusions should not depend on these choices.

We also found that the parameter uncertainty of the hydrological model was more important for the method M2 than the uncertainty of the method itself (in method M1 only the parameter uncertainty could have been assessed). However, different parameter sets led to a similar over- or underestimation of the relative peak and volume. This suggests that, although the parameter uncertainty of a hydrological model was substantial, the performance of the disaggregation methods does not depend on the choice of the parameter set, at least for the hydrological model considered here.

Note also that, in this study, hourly precipitation time series were available. In real-case applications, such high-resolution precipitation data are usually not available at the point of interest, or not with the required time series length. However, under the assumption that the parameters of the disaggregation method do not vary in space, they can be estimated at nearby stations, and if they do not vary in time, they can be estimated from short, high-resolution time series and then applied for the disaggregation of longer time series of daily records.

Finally, the aim of this study was to compare both disaggregation methods based on a simple lumped hydrological model. Thus, the obtained results cannot be generalized to more

complex (distributed) hydrological models. It is expected that the differences between the simple and complex disaggregation approaches could become more apparent with a distributed hydrological model. Such a distributed model could potentially benefit more from complex disaggregated data in terms of better representing spatial and temporal variability in precipitation fields at smaller sub-catchment units. Nevertheless, since lumped hydrological models are widely applied precipitation–runoff models, the authors consider that the findings of this investigation will be useful for the hydrological community, despite the aforementioned limitations.

6 Conclusions

Two precipitation disaggregation methods, M1 (Sikorska *et al.* 2018) and M2 (Müller and Haberlandt 2015), with different complexities were assessed regarding their ability to generate reliable hourly precipitation time series from daily areal precipitation totals as input for a lumped hydrological model. The study was carried out for nine mesoscale catchments in Switzerland and the results were analysed for precipitation characteristics: continuous and event-based, as well as for extreme precipitation values; and for flood characteristics: relative peak and relative volume of annual maximum floods. Based on our results, the following conclusions can be made:

- Both M2 method variants led to a better representation of continuous and event-based precipitation characteristics, as well as precipitation extreme values, than method M1.
- The differences between disaggregated precipitation time series with methods M1 and M2 identified for precipitation characteristics disappeared when these series were introduced into the lumped hydrological model. Similar efficiencies were found for both methods M1 (best variant) and M2 (variants a and b).
- As disaggregation of areal *versus* station precipitation time series resulted in similar results, a disaggregation of precipitation time series can be applied before or after the spatial averaging of records.
- The parameter uncertainty of the hydrological model was of small importance for selecting an optimal distribution in M1 and for assessing M2. Yet, the contribution of this parametric uncertainty in M2 was greater than the uncertainty resulting from the method M2 itself (80 disaggregation realizations).

Based on these findings, we conclude that the complexity in the precipitation disaggregation method is less important for a lumped hydrological model. However, independent of the identified differences, the choice between the disaggregation methods, i.e. M1 and M2, should depend on (a) the purpose of the intended application, and (b) the data availability. If there is no data limitation that precludes the application of one disaggregation method, both methods can be recommended for the generation of high-resolution precipitation time series to be used as input for precipitation–runoff modelling and derived flood frequency analyses. If precipitation

characteristics are a point of interest as well, the disaggregation methods M2a and M2b are recommended.

Acknowledgements

The authors thank the associate editor Guillaume Thirel, Claudia Brauer and one anonymous reviewer for their useful comments, which helped us improve this manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Funding was provided for Hannes Müller-Thomy as a Research Fellowship (MU 4257/1-1) by DFG e.V., Bonn, Germany. Anna E. Sikorska-Senoner acknowledges funding from the Swiss Federal Office for the Environment (FOEN) via the contract 15.0054.PJ/O503-1381. Runoff data were provided by the FOEN and the Canton of Argovia (Surb-Döttingen), and precipitation and temperature data were provided by MeteoSwiss. The authors acknowledge the TU Wien Library for financial support through its Open Access Funding Program.

ORCID

Hannes Müller-Thomy  <http://orcid.org/0000-0001-5214-8945>
Anna E. Sikorska-Senoner  <http://orcid.org/0000-0002-5273-1038>

References

- Andres-Domenech, I., *et al.*, 2015. Climate and hydrological variability: the catchment filtering role. *Hydrology and Earth System Sciences*, 19, 379–387. doi:10.5194/hess-19-379-2015
- Berne, A., *et al.*, 2004. Temporal and spatial resolution of rainfall measurements required for urban hydrology. *Journal of Hydrology*, 299 (3–4), 166–179. doi:10.1016/S0022-1694(04)00363-4
- Blöschl, G. and Sivapalan, M., 1995. Scale issues in hydrological modeling: a review. *Water Resources Research*, 9, 251–290.
- Breil, K., 2016. Driving a lumped hydrological model with precipitation output from weather generators of different complexity. *Hydrological Sciences Journal*, 61 (8), 1395–1414. doi:10.1080/02626667.2015.1036755
- Brunner, M.I. and Sikorska-Senoner, A.E., 2019. Dependence of flood peaks and volumes in modeled discharge time series: effect of different uncertainty sources. *Journal of Hydrology*, 572, 620–629. doi:10.1016/j.jhydrol.2019.03.024
- Callau Poduje, A.C. and Haberlandt, U., 2017. Short time step continuous rainfall modeling and simulation of extreme events. *Journal of Hydrology*, 552, 182–197. doi:10.1016/j.jhydrol.2017.06.036
- Callau Poduje, A.C. and Haberlandt, U., 2018. Spatio-temporal synthesis of continuous precipitation series using vine copulas. *Water*, 10, 862. doi:10.3390/w10070862
- Carsteanu, A. and Foufoula-Georgiou, E., 1996. Assessing the dependence among weights in a multiplicative cascade model of temporal rainfall. *Journal of Geophysical Research*, 101 (D21), 26363–26370. doi:10.1029/96JD01657
- Cross, D., *et al.*, 2018. Censored rainfall modeling for estimation of fine-scale extremes. *Hydrology and Earth System Sciences*, 22, 727–756. doi:10.5194/hess-22-727-2018
- Ding, J., *et al.*, 2016. Estimation of instantaneous peak flows from maximum mean daily flows using the HBV hydrological model. *Hydrological Processes*, 30, 1431–1448. doi:10.1002/hyp.v30.9
- DVWK (German Association for Water), 1984. Regeln: Arbeitsanleitung zur Anwendung von Niederschlags-Abfluss-Modellen in kleinen Einzugsgebieten. Teil II: Synthese. (*Regulations: instructions for the*

- use of rainfall-runoff models in small catchments, part II, in German). Technical Report. Germany: Verlag Paul Parey.
- Ficchi, A., Perrin, C., and Andréassian, V., 2016. Impact of temporal resolution on inputs on hydrological performance: an analysis based on 2400 flood events. *Journal of Hydrology*, 538, 454–470. doi:10.1016/j.jhydrol.2016.04.016
- Gaál, L., et al., 2012. Flood timescales: understanding the interplay of climate and catchment processes through comparative hydrology. *Water Resources Research*, 48, W04511. doi:10.1029/2011WR011509
- Güntner, A., et al., 2001. Cascade-based disaggregation of continuous rainfall time series: the influence of climate. *Hydrology and Earth System Sciences*, 5 (2), 145–164. doi:10.5194/hess-5-145-2001
- Gupta, H.V., et al., 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modeling. *Journal of Hydrology*, 377, 80–91. doi:10.1016/j.jhydrol.2009.08.003
- Haberlandt, U., Ebner von Eschenbach, A.-D., and Buchwald, I., 2008. A space-time hybrid hourly rainfall model for derived flood frequency analysis. *Hydrology and Earth System Sciences*, 12, 1353–1367. doi:10.5194/hess-12-1353-2008
- Hingray, B. and Ben Haha, M., 2005. Statistical performances of various deterministic and stochastic model for rainfall series disaggregation. *Atmospheric Research*, 77, 152–175. doi:10.1016/j.atmosres.2004.10.023
- Jebari, S., et al., 2012. Soil erosion estimation based on rainfall disaggregation. *Journal of Hydrology*, 436–437, 102–110. doi:10.1016/j.jhydrol.2012.03.001
- Kandel, D.D., Western, A.W., and Grayson, R.B., 2005. Scaling from process timescales to daily time steps: A distribution function approach. *Water Resources Research*, 41, W02003.
- Kim, D., Chun, J.A., and Aikins, M., 2018. An hourly-scale scenario-neutral flood risk assessment in a mesoscale catchment under climate change. *Hydrological Processes*, 32, 3416–3430. doi:10.1002/hyp.v32.22
- Koutsoyiannis, D. and Onof, C., 2001. Rainfall disaggregation and adjusting procedures on a Poisson cluster model. *Journal of Hydrology*, 246, 109–122. doi:10.1016/S0022-1694(01)00363-8
- Lütkemeier, R., et al., 2018. Uncertainty of rainfall products: impact on modelling household nutrition from rain-fed agriculture in Southern Africa. *Water*, 10, 499. doi:10.3390/w10040499
- Melsen, L.A., et al., 2016. Hydrology and earth system science opinions: the need for process-based evaluation of large-domain hyper-resolution models. *Hydrology and Earth System Sciences*, 20 (3), 1069–1079. doi:10.5194/hess-20-1069-2016
- Müller, H., 2016. Rainfall disaggregation for hydrological modeling. *Proceedings of the Institute of Water Resources Management, Hydrology and Agricultural Engineering*, 101. ISSN 0343-8090 (in German).
- Müller, H. and Haberlandt, U., 2015. Temporal rainfall disaggregation with a cascade model: from single-station disaggregation to spatial rainfall. *Journal of Hydrologic Engineering*, 20 (11), 04015026. doi:10.1061/(ASCE)HE.1943-5584.0001195
- Müller, H. and Haberlandt, U., 2018. Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology. *Journal of Hydrology*, 556, 847–864. doi:10.1016/j.jhydrol.2016.01.031
- Müller, T., Schütze, M., and Bárdossy, A., 2017. Temporal asymmetry in precipitation time series and its influence on flow simulations in combined sewer systems. *Advances in Water Resources*, 107, 56–64. doi:10.1016/j.advwatres.2017.06.010
- Müller-Thomy, H., Wallner, M., and Förster, K., 2018. Rainfall disaggregation for hydrological modeling: is there a need for spatial consistency? *Hydrology and Earth System Sciences*, 22, 5259–5280. doi:10.5194/hess-22-5259-2018
- Ochoa-Rodriguez, S., et al., 2015. Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: a multi-catchment investigation. *Journal of Hydrology*, 531 (Part 2), 389–407. doi:10.1016/j.jhydrol.2015.05.035
- Olsson, J., 1998. Evaluation of a scaling cascade model for temporal rainfall disaggregation. *Hydrology and Earth System Sciences*, 2 (1), 19–30. doi:10.5194/hess-2-19-1998
- Onof, C., Chandler, R.E., and Kakou, A., 2000. Rainfall modelling using Poisson-cluster processes: a review of developments. *Stochastic Environmental Research and Risk Assessment*, 6 (14), 384–411. doi:10.1007/s004770000043
- Paschalis, A., et al., 2014. On temporal stochastic modeling of precipitation, nesting models across scales. *Advances in Water Resources*, 63, 152–166. doi:10.1016/j.advwatres.2013.11.006
- Peel, M.C., Finlayson, B.L., and McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11 (5), 1633–1644. doi:10.5194/hess-11-1633-2007
- Pohle, I., et al., 2018. Coupling Poisson rectangular pulse and multiplicative micro-canonical random cascade models to generate sub-daily precipitation time series. *Journal of Hydrology*, 562, 50–70. doi:10.1016/j.jhydrol.2018.04.063
- Post, D.A. and Jakeman, A.J., 1996. Relationships between catchment attributes and hydrological response characteristics in small Australian mountain ash catchments. *Hydrological Processes*, 10, 877–892. doi:10.1002/(ISSN)1099-1085
- Pui, A., et al., 2012. A comparison of alternatives for daily to sub-daily rainfall disaggregation. *Journal of Hydrology*, 470–471, 138–157. doi:10.1016/j.jhydrol.2012.08.041
- Ran, Q., et al., 2012. Experimental study of the impact of rainfall characteristics on runoff generation and soil erosion. *Journal of Hydrology*, 424–425, 99–111. doi:10.1016/j.jhydrol.2011.12.035
- Schertzer, D. and Lovejoy, S., 2011. Multifractals, generalized scale invariance and complexity in geophysics. *International Journal of Bifurcation and Chaos*, 21 (12), 3417–3456. doi:10.1142/S0218127411030647
- Schilling, W., 1984. Univariate versus multivariate rainfall statistics – problems and potentials (a discussion). *Water Science Technology*, 16, 139–146. Copenhagen, Denmark. doi:10.2166/wst.1984.0185
- Schilling, W., 1991. Rainfall data for urban hydrology: what do we need? *Atmospheric Research*, 27 (1–3), 5–21. doi:10.1016/0169-8095(91)90003-F
- Schleiss, M. and Smith, J.A., 2016. Two simple metrics for quantifying rainfall intermittency: the burstiness and memory of interamount times. *Journal of Hydrometeorology*, 17, 421–436. doi:10.1175/JHM-D-15-0078.1
- Seibert, J., 2000. Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrology and Earth System Sciences*, 4, 215–224. doi:10.5194/hess-4-215-2000
- Seibert, J. and Vis, M.J.P., 2012. Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. *Hydrology and Earth System Sciences*, 16, 3315–3325. doi:10.5194/hess-16-3315-2012
- Serinaldi, F., 2010. Multifractality, imperfect scaling and hydrological properties of rainfall time series simulated by continuous universal multifractal and discrete random cascade models. *Nonlinear Processes in Geophysics*, 17 (6), 697–714. doi:10.5194/npg-17-697-2010
- Sikorska, A.E. and Seibert, J., 2018. Appropriate temporal resolution of precipitation data for discharge modelling in pre-alpine catchments. *Hydrological Sciences Journal*, 63 (1), 1–16. doi:10.1080/02626667.2017.1410279
- Sikorska, A.E., Viviroli, D., and Seibert, J., 2018. Effective precipitation duration for runoff peaks based on catchment modelling. *Journal of Hydrology*, 556, 510–522. doi:10.1016/j.jhydrol.2017.11.028
- Thober, S., et al., 2014. Stochastic temporal disaggregation of monthly precipitation for regional gridded data sets. *Water Resources Research*, 50, 8714–8735. doi:10.1002/2014WR015930
- Viviroli, D., et al., 2011. Climate change and mountain water resources: overview and recommendations for research, management and policy. *Hydrology and Earth System Sciences*, 15, 471–504. doi:10.5194/hess-15-471-2011
- Westra, S., et al., 2012. Continuous rainfall simulation: 1. A regionalized sub-daily disaggregation approach. *Water Resources Research*, 48, W01535. doi:10.1029/2011WR010489
- Wiekenkamp, I., et al., 2016. Spatial and temporal occurrence of preferential flow in a forested headwater catchment. *Journal of Hydrology*, 534, 139–149. doi:10.1016/j.jhydrol.2015.12.050