Adaptive clustering: Reducing the computational costs of distributed (hydrological) modeling by exploiting time-variable similarity of model elements

Uwe Ehret[1], Rik van Pruijssen[1], Marina Bortoli[1], Ralf Loritz[1], Elnaz Azmi[2], Erwin Zehe[1]

[1]Institute of Water Resources and River Basin Management, Karlsruhe Institute of Technology - KIT, Karlsruhe, Germany
[2]Steinbuch Centre for Computing, Karlsruhe Institute of Technology - KIT, Karlsruhe, Germany

Correspondence to: Uwe Ehret (uwe.ehret@kit.edu)

Abstract.

In this paper we propose adaptive clustering as a new method to reduce computational efforts of distributed modelling. It consists of identifying similar acting model elements during runtime, clustering them, running the model for just a few representatives per cluster, and mapping their results to the remaining model elements in the cluster. Key requirements for the application of adaptive clustering are the existence of i) many model elements with ii) comparable structural and functional properties and iii) only weak interaction (e.g. hill slopes, sub catchments or surface grid elements in hydrological and land surface models). The clustering of model elements must not only consider their time-invariant structural and functional properties, but also their current state and forcing, as all these aspects influence their current functioning. Joining model elements into clusters is therefore a continuous task during model execution rather than a one-time exercise that can be done beforehand. Adaptive clustering takes this into account by continuously checking the clustering, and re-clustering when necessary.

We explain the steps of adaptive clustering and provide a proof-of-concept at the example of a distributed, conceptual hydrological model fit to the Attert basin in Luxembourg. The clustering is done based on normalized and binned transformations of model element states and fluxes. Analysing a 5-year time series of these transformed states and fluxes revealed that many model elements act very similarly, and the degree of similarity strongly varies with time, indicating the potential for adaptive clustering to save computation time. Compared to a standard, full-resolution model run used as a virtual reality 'truth', adaptive clustering indeed reduced computation time by 75 percent, while model quality, expressed as Nash-Sutcliffe efficiency of sub catchment runoff, declined from 1 to 0.84. Based on this proof-of-concept application, we believe that adaptive clustering is a promising tool for reducing the computation time of distributed models. Being adaptive, it integrates and enhances existing methods of static grouping of model elements, such as lumping or grouped response units (GRU). It is compatible with existing dynamical methods such as adaptive time-stepping or adaptive gridding, and unlike the latter, does not require adjacency of the model elements to be joined.

As a welcome side effect, adaptive clustering can be used for system analysis: In our case, analysing the space-time patterns of clustered model elements confirmed that the hydrological functioning of the Attert catchment is mainly controlled by the spatial patterns of geology and precipitation.

1 Introduction

Hydrological systems are often characterized by considerable spatial heterogeneity of relevant properties such as topography or soils (Schulz et al., 2006), and considerable temporal variability due to time changing boundary conditions such as precipitation or radiation, and time changing system properties such as vegetation cover (Zehe and Sivapalan, 2009). If we are mainly interested in aggregated characteristics and dynamics of such systems, such as mean wetness, mean travel times,

or discharge at a catchment outlet, a spatially lumped representation in a model will suffice. Models designed for such coarse spatial resolutions, such as Topmodel (Beven and Kirkby, 1979) or HBV (Bergström, 1976) are easy to set up and computationally highly efficient, but necessarily conceptualize process patterns and redistribution processes and the underlying controls by means of effective dynamical laws, effective states, effective parameters and effective fluxes.

5   Often, however, we want to analyse and predict hydrological systems in higher spatial detail, which requires spatially distributed models. In distributed models, spatial variability of hydrological systems is captured by dividing the model domain into sub domains, which are assumed to be internally homogeneous with respect to their main structural and functional properties. Such model elements have been referred to as hydrological response units HRU (Flügel, 1995; Kouwen et al., 1993), representative elementary areas REA (Wood et al., 1988), or representative elementary watersheds

10   REW (Reggiani et al., 1998). Beyond incorporating the spatial variability of the hydrological system, distributed models offer additional advantages: They incorporate distributed forcing, they can be parameterized and validated by distributed observations, and they permit more fundamental process representations (Kouwen et al., 1993). Distributed, physically based models such as MIKE SHE (Abbott et al., 1986), HYDRUS (Šimunek et al., 1999) or CATFLOW (Zehe et al., 2001) therefore have the desirable property of providing physically meaningful, distributed answers based on distributed internal

15   dynamics.

The major drawback of distributed models is their large demand of high-resolution data for model setup and operation, and a CPU demand that rapidly grows with system resolution. The question about the optimal balance of spatial resolution and computational burden has therefore been a long-standing issue (not only) in the hydrological sciences (Melsen et al., 2016; Liu et al., 2016; Dehotin and Braud, 2008; Booij, 2003; Gharari et al., 2020). In this context, a range of methods have been

20   proposed to address the computational problem: It can be crushed by massive parallel computing (Kollet, 2010), or reduced by avoiding redundant computations. Redundancy occurs if several model elements act similar, such that knowing the behaviour of one is a good proxy for the behaviour of the others. In this context, and throughout the remainder of this text, we define similarity of two model elements as follows: "Two model elements act similarly if they share similar structural and functional properties, are in a similar state and exposed to similar forcing, such that they produce similar responses

25   based on similar internal fluxes and state changes" (see also Zehe et al., 2014). Similarity - and its counterpart, redundancy - among model elements can be considered as a time-invariant (static) or time-variant (dynamic) phenomenon, and methods for redundancy reduction have been proposed on the basis of either of these views. Grouped response units GRU (Kouwen et al., 1993) for example rely on the static similarity paradigm. GRUs are groups of HRUs close enough to be subject to uniform forcing and negligible differences in routing. All HRUs in a GRU are then treated as a single computational unit.

30   This considerably reduces computational effort, but it comes at the cost of losing spatial detail and spatial positioning. Time-variant (or adaptive) methods do not rely on a single, time-invariant grouping. Instead, groups of model elements are dynamically established and adjusted during model run time by identifying and exploiting patterns of similarity in either time or space. Adaptive time stepping (Minkoff and Kridler, 2006) exploits patterns of similarity in time, adaptive gridding (Pettway et al., 2010; Berger and Oliger, 1984) exploits patterns of similarity in space; combinations of both approaches are

35   possible (Miller et al., 2006). Due to their generality, adaptive methods have been used to improve distributed modelling of a large variety of systems such as the universe (Teyssier, 2002), the atmosphere (Bacon et al., 2000; Aydogdu et al., 2019), oceans (Pain et al., 2005), and groundwater systems (Miller et al., 2006). While adaptive methods are highly useful, they all require direct adjacency - in either time or space - of the model elements to be joined. However, similarity, in both nature and models, is not necessarily restricted to contiguous regions. For example, there may be many non-contiguous south-

40   facing forested hillslopes with shallow soils in a watershed, in an intermediate wetness state, which will act very similarly at a particular sunny day.

In this context, we suggest a new adaptive method for clustering of model elements, which is not limited to contiguous regions. It is motivated by the suggestions of Melsen et al. (2016) "to further investigate and substantially improve the

representation of spatial and temporal variability in large-domain hydrological models" and contributes to solving the computational challenges of hydrological modelling as formulated by Clark et al. (2017). It comprises several steps: Clustering of model elements, choice of cluster representatives, mapping of results from representatives to recipients, and continuous evaluation of the clustering to decide when re-clustering is needed. We demonstrate adaptive clustering at the example of a distributed, conceptual hydrological model of the Attert basin in Luxembourg. Besides evaluating adaptive clustering in terms of computational gains and related losses of modelling quality, we also discuss how the normalized and binned representations of model states and fluxes that we used for clustering contribute to hydrological system analysis by revealing space-time patterns of similarity in the catchment.

The remainder of the manuscript is structured as follows: In Sect. 2, we first describe the general, application-independent steps of adaptive clustering. This is the key methodological contribution of the paper. We then introduce the SHM hydrological model and its set-up for the Attert basin in Luxembourg. The model constitutes the test environment for the proof-of-concept of adaptive clustering. We then describe how adaptive clustering is implemented in the SHM Attert model, and finally describe our approach and metrics used for evaluating adaptive clustering and measuring hydrological similarity. In Sect. 3, we present and discuss results from distributed modelling with and without adaptive clustering, and compare them to a range of benchmark models in terms of computational efficiency and model quality. In the same section, we show results of the hydrological similarity analysis. These results are relevant in relation to adaptive clustering, as the time-varying degree of similarity among model elements directly controls adaptive clustering. In addition, they are also useful for hydrological system analysis. In Sect. 4, we summarize the results, draw conclusions, discuss limitations of adaptive clustering and suggest further research.

2 Data and methods

2.1 Adaptive clustering

As explained in the introduction, the main goal of adaptive clustering is to reduce computational efforts of distributed and high-resolution modelling. The main idea is to avoid redundant computations by clustering similar acting model elements, and then to infer the dynamics of all elements in a cluster from just a few representatives. Key requirements for the successful application of adaptive clustering are the existence of i) many model elements with ii) comparable structural and functional properties and iii) only weak interaction. If there are only few model elements, there will be nothing to cluster, if they are not structurally and functionally similar, it will be impossible to assign results from representatives to the remaining cluster members (recipients), and if there is strong interaction, ignoring it - which is inevitable in adaptive clustering - will cause large modelling error. It is important to keep in mind that even if two model elements are identical with respect to all time invariant (structural) properties, they can still act differently when starting from different initial conditions, or when exposed to different boundary conditions. Therefore, while similarity among model elements can have a strong time invariant component, and static clustering can be beneficial, the full potential of clustering will be exploited if it is treated as time-variant (Loritz et al., 2018). This is the core idea of adaptive clustering. Its main steps are illustrated in Fig. 1, and we will explain the method along steps 'a' to 'j' in the plot.

Step 'a': Start the model from a fully distributed (non-clustered) initial state. Each model element - depicted by a circle - is in a particular initial state – indicated by the value in the circle. Model elements typically possess several state variables and hence an array of state values, but for simplicity only a single one is shown in the plot.

Step 'b': Based on the similarity of their states, combine the model elements in clusters. In the plot, the clusters are depicted by bold circles and labelled 'A', 'B' and 'C'. The clustering involves two important choices: Choice of a suitable clustering algorithm and values of its hyper parameters, and choice of a state variable by which the clustering is done. In the following, we will refer to this variable as 'clustering control variable'. The clusters, determined with states at the current time step, will

3

until further decision (see step 'h') be used for all further modelling time steps. For these time steps, we refer to clusters determined in the past as 'inherited clusters'.

Step 'c': Select from each cluster a subset of model elements. These serve as cluster representatives. In the plot, the representatives are indicated by blue circles. The number of representatives per cluster controls the performance of adaptive clustering: A large number will guarantee high modelling quality, but small computational gains, and vice versa.

Step 'd': Execute the model for the next time step, but only for the representatives. From running the model, the representatives obtain updated values for each of their state variables. In the plot, the updated states are indicated by red colour.

Step 'e': The representatives 'donate' their updated states (and fluxes) to all recipients in their cluster by using a suitable mapping technique. In the plot, this is indicated by arrows. Note that due to the mapping, conservations laws are potentially violated. This is a drawback of adaptive clustering and requires further attention.

Step 'f': Based on the updated states of the clustering control variable, combine the representatives into a new set of clusters. In the plot, the new clusters are depicted by bold circles labelled 'I' and 'II' in red colour. These clusters may differ from the inherited clusters: When the model is executed in step 'd', each representative is driven by its particular forcing, which potentially leads - even within a cluster - to a divergence of states. Clusters may therefore break apart, unite, or exchange elements as the states of the model elements evolve over time. Inherited clusters may therefore at some point become invalid and must be replaced by an up-to-date version.

Step 'g': Compare the new clusters to the inherited clusters. Please note that the new clustering in step 'f', and the cluster comparison is done only for the representatives. This is much more efficient than considering all model elements. Comparing clusters involves identifying matching clusters, and then measuring their degree of agreement. In the plot, this is illustrated by a table: Each column represents one of the inherited clusters, each row one of the new clusters. Matching clusters are indicated by cells with blue background colour ('A' and 'I', 'B' and 'II', 'C' has no match). The larger the number of representative model elements in matching clusters, the higher the agreement of the new and the inherited clusters.

Step 'h': Decide if the agreement of the new and the inherited clusters is sufficiently high. If the answer is 'Yes', the inherited clusters are still valid, and steps 'd' to 'h' are repeated for the next time step. If the answer is 'No', the inherited clusters are replaced. Obviously, the new clusters can be used as a replacement, but they only contain the representatives. Recipients can be assigned to the new clusters based on their current state. The problem is, however, that their states were transferred from the representatives, and depending on the mapping method, they may be more or less averaged, smoothed states. If these values are used for clustering, there is a risk that recipients are always clustered in the same manner, limiting the model's ability to adapt to changing conditions and to represent heterogeneous situations. This risk can be reduced by operating the model in full resolution for some time, as explained in steps 'i' and 'j', allowing the recipient model elements to evolve towards their particular state.

Step 'i': From the current time step, jump back in time. In the plot, this is indicated by a curved arrow extending back over two time steps.

Step 'j': Set the model to a fully distributed (non-clustered) mode. In the plot, this is indicated by all model elements arranged in a row, without surrounding clusters. Starting from the state of the model elements at the jumped-to time, execute the model in full resolution until the current time step is reached again. Generally, the length of the jump back is a trade-off between enabling the recipient model elements to evolve towards their particular states free from cluster constraints, and additional computational expenses. Based on the new states of all model elements, continue at step 'b'.
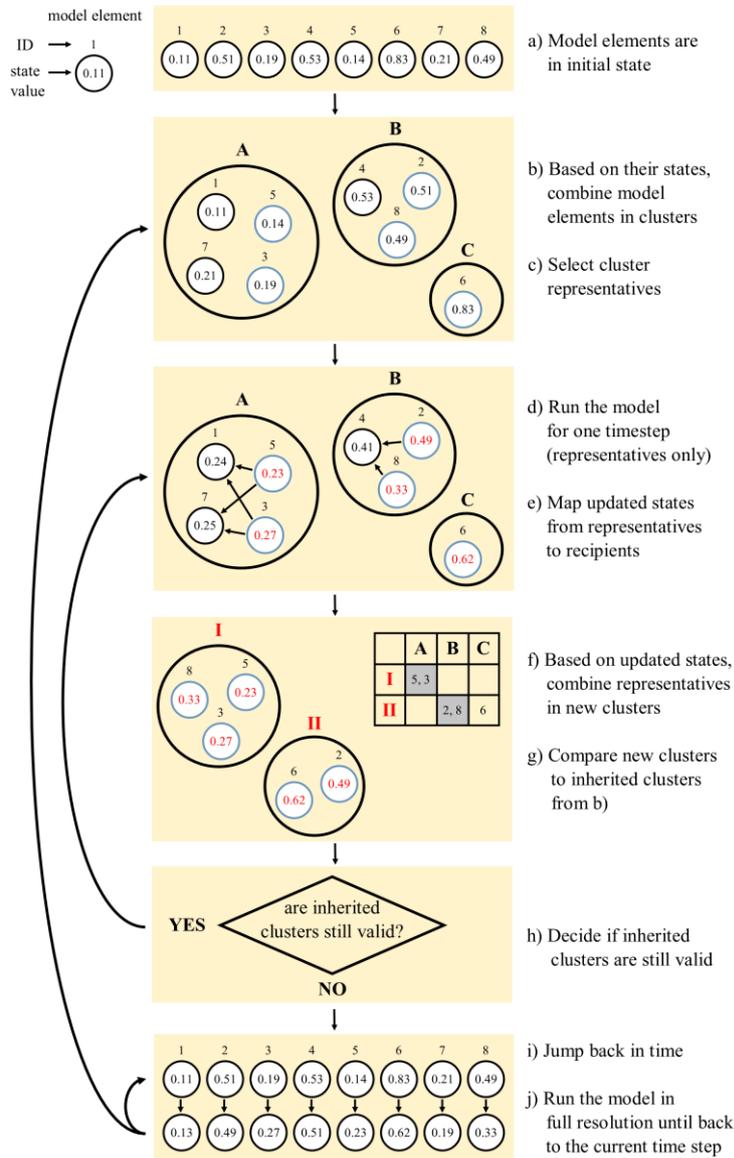
Figure 1. Main steps of adaptive clustering. A detailed explanation is given in the text.

2.2 Study area and hydrological model

The Attert basin, our test site, is located in the central western part of the Grand Duchy of Luxembourg and partially in Eastern Belgium with a total catchment area of 288 km² up to gauge Useldange (Fig. 2). The landscape shows topographical, geological and pedological diversity, with a small area underlain by sandstones in the South and Northeast, a wide area of sandy marls in the centre part, and an elevated region underlain by schist in the North, which is part of the Ardennes massif. The schist region reaches elevations up to 539 m a.s.l. and contains deeply incised river valleys. The Attert basin is situated in the temperate oceanic climate zone, and snow-related processes play a negligible role. Precipitation is mainly associated with westerly synoptic flow regimes and reaches annual amounts of about 850 mm (Pfister et al., 2005; Pfister et al., 2000).

We selected the Attert basin for several reasons: A large body of existing hydrological knowledge (Pfister et al., 2009; Juilleret et al., 2012) including modelling studies (Fenicia et al., 2014; Fenicia et al., 2016), access to a comprehensive data set compiled in the CAOS (Catchments as Organized Systems) project (Zehe et al., 2014), and own prior modelling studies

5

in the Colpach, a sub basin of the Attert, that revealed pronounced and time-variable similarity of model element behaviour (Loritz et al., 2018).

Instead of using one of the existing hydrological models for the Attert basin, we decided to set up a new one. This was mainly to ensure full code control, which greatly facilitates prototyping and testing of adaptive clustering. We chose a simple, conceptual, yet distributed model architecture tailored to the structure and hydrological function of the Attert basin. It is closely related to established hydrological models such as HBV (Bergström, 1976), and due to its simplicity, we named it 'SHM' (Simple Hydrological Model). Its general structure and process inventory is explained in detail in Appendix A.1. The setup of SHM for the Attert basin is described in Appendix A.2, and its multi-criteria calibration and validation - based on five years of hourly data - in Appendix A.3. Overall, the model achieves acceptable performance (Nash-Sutcliffe efficiency of 0.73 for validation), making it a suitable test bed for exploring adaptive clustering.

Figure 2. Map of the Attert basin up to gauge Useldange. Black labels are SHM sub catchment ID's, blue labels are SHM river element IDs. Yellow and blue labels are rain and river gauge IDs, respectively. The yellow lines indicate the area of influence of each rain gauge as determined by the Nearest Neighbor method. Further information about the gauges and the SHM model are given in Appendices A.1-A.3.

2.3 Implementation of adaptive clustering in SHM Attert

In this section, we explain how adaptive clustering is implemented in the SHM Attert model. We do so along the lines of the general steps 'a' to 'j' of adaptive clustering as described in Sect. 2.1 and Fig. 1.

For clustering of model elements - SHM sub catchments in our study - (step 'b'), we apply a straightforward yet effective approach based on binning: First, all model states of all sub catchments are normalized to [0,1] values. The state variable-

and sub catchment-specific minima and maxima required for normalization were obtained from running the model – in full resolution – for the entire five years of available data. The [0,1] value range is then subdivided into 64 bins of uniform width. Choosing the number of bins was guided by the objective to balance resolution (many bins) and sufficiently populated bins (few bins). All sub catchments with normalized values of the clustering control variable falling into the same bin are assigned to the same cluster. Each non-empty bin therefore defines a cluster. The possible number of clusters is limited to a minimum of one and a maximum of 64, and the number of clusters at a given point in time expresses the degree of similarity among the sub catchments at that time. We selected sub catchment runoff ($q_{cat,out}$, see Fig. A1 and Table A1) as a single clustering control variable for three reasons: Firstly, for catchment hydrologists runoff is the main variable of interest, secondly, sub catchment runoff is influenced by all sub catchment states and fluxes, hence similarity of two sub catchments with respect to their runoff is a reasonable single-value indicator of overall similarity, thirdly we used only a single control variable to keep things simple.

For each cluster, representatives are selected (step 'c') by random picking controlled by three parameters (see Table 1): *Perc_reps* defines the total number of representatives, expressed as percentage of the total number of sub catchments in the model. Applied to each cluster, it provides a first estimate about how many representatives should be picked from it. We found that besides controlling the total number of representatives, it is also useful to set a limit to the minimum and maximum number of representatives per cluster. This is controlled by parameters *min_reps_per_clus* and *max_reps_per_clus*.

Mapping states and fluxes from cluster representatives to recipients (step 'e') applies the normalized values already used for clustering: Recipients are forced to assume the representative's normalized state (or flux), and these normalized states (or fluxes) are then re-converted by each recipient's min-max range to dimensionful values. If there is more than one representative in a cluster, a single best one is selected as the representative closest to the median value of the clustering control variable of all representatives. The results of that single best representative are then mapped to all recipients. This method clearly leaves room for improvement, but we considered it good enough for a first proof-of-concept.

Comparing inherited clusters with current clusters (step 'g') involves two steps: Identifying matching clusters, and then measuring the degree of agreement between them. Note that with respect to the first step, it is not possible to simply define matching clusters as those with the same bin number. For example, assume that the inherited clusters are determined, and afterwards uniform rainfall falls in the catchment, uniformly shifting all sub catchment states to a higher, 'wetter' bin. Clustering the sub catchments with the new states will yield clusters identical to the inherited ones, but the cluster labels will have changed. We used the well-known Hungarian method (Kuhn, 1955; Munkres, 1957), which matches clusters by maximizing the agreement of their content instead of comparing their labels. When the cluster matches are established (compare the table in Fig. 1), the degree of similarity between the clusterings is measured by the number of elements in matching clusters divided by the total number of elements (in Fig. 1, this is the number of elements in the blue cells divided by the total number of elements in the table). Clustering similarity is hence expressed by a number between zero (clusterings are incomparable) and one (clusterings are identical).

The inherited clusters are replaced (step 'h') if clustering similarity falls below an acceptance limit set by *sim_crit* (Table 1). The jump back in time (step 'i') is controlled by parameter *sim_uncrit*, which like *sim_crit* is a similarity threshold: The jump goes back to the last time at which this threshold was still exceeded. Depending on the prevailing hydro-meteorological situation, the jump can be shorter or longer, but it will never extend beyond the time at which the inherited clusters were established.

All parameters controlling adaptive clustering are summarized in Table 1. For the SHM Attert application, we determined their values by manual, iterative trial-and-error with the objective of maximizing computational savings while minimizing quality loss.

Table 1. Parameters of adaptive clustering and their values for SHM Attert

Name	Description	Value chosen
num_bins	The value range of a variable is divided into *num_bins* bins of uniform width	64[a]
perc_reps	Ratio of sub catchments used as representatives and total number of sub catchments	10
min_reps_per_clus	Minimum number of representatives per cluster	1
max_reps_per_clus	Maximum number of representatives per cluster	8
sim_crit	Replace inherited clustering if its similarity to the current clustering is < *sim_crit.*	55
sim_uncrit	If re-clustering is required, go back in time until the similarity between the inherited and the current clustering is ≥ *sim_uncrit.*	75

[a] for normalized variables with value range [0,1], this means bin edges [0, 0.0156, 0.0313, … , 0.9688, 0.9844, 1]

2.4 Experimental design and evaluation criteria

The existence of time-variant similarity among model elements is a precondition for a useful application of adaptive clustering (see the related discussion in Sect. 2.1). We therefore precede the analysis of adaptive clustering performance with an analysis of space-time patterns of similarity among SHM Attert sub catchments. The approach and related metrics are explained in Sect. 2.4.1. In Sect. 2.4.2, we describe SHM Attert model variants used as benchmarks for adaptive clustering, and we introduce the evaluation criteria for measuring both computational effort and simulation quality of the competing models.

2.4.1 Entropy as a measure of hydrological similarity

How to measure the similarity among sub catchments, and its variation with time? Sub catchments differ in size, and many of their states and fluxes are size-dependent. Therefore, instead of directly comparing their values, we use the (0,1)-normalized and binned time series of states and fluxes of all sub catchments as described in Sect. 2.3, step 'b'. At each point in time, the occupations of the 64 bins together form a histogram, which can be normalized to a discrete probability distribution by dividing the bin populations with the total number of sub catchments. The overall degree of similarity among sub catchments can then be measured - in the same manner for any state or flux of interest - by Shannon information entropy in the universal unit of 'bit' (Eq. 1). We adopted this approach from Loritz et al. (2018); a more detailed introduction to concepts, measures and applications of information theory is given in Neuper and Ehret (2019), Singh (2013), and Cover and Thomas (1991).

$$H(X) = -\sum_{x \in X} p(x) \, log_2 \, p(x) \qquad (1)$$

As a basis for the similarity analysis, we operated the SHM Attert model in full resolution (no clustering) for the entire 5-year period of available data and then converted all sub catchment states and fluxes to normalized and binned values. Spatial maps of these states are then used to analyze spatial patterns of similarity, and time series of entropy to analyze temporal patterns. For each state and flux, we also calculated the time-averaged entropy as a measure of overall, mean variability.

It is an interesting property of Shannon entropy that for a discrete distribution with a given number of bins, there exists an upper and a lower bound: If all elements fall into a single bin, the entropy of the distribution will take its minimum value, zero. If the elements are uniformly distributed over the bins, entropy will take its maximum value of $H = log_2(n)$, where n is the number of bins. As we use the same 64 bins for all variables, the same lower bound of zero and the same upper bound of six bit applies to all of them, which facilitates comparison. In terms of similarity, entropies close to zero indicate a high degree of similarity among model elements, entropies close to six indicate low similarity.

2.4.2 Evaluation criteria and benchmark models for adaptive clustering

With respect to adaptive clustering, two aspects are important: Computational savings, and related losses of modeling quality. The savings we measure by overall model run time, as it measures the entire modeling effort, i.e. the effort for operating the actual hydrological model and the effort for the adaptive clustering overhead. In order to make run times comparable, we performed all model runs on the same machine and with no additional processes active. We verified the reproducibility of the results by repeating the runs many times: The observed spread was less than one per cent of total run time and therefore considered negligible. Modeling quality we measure by Nash-Sutcliffe efficiency (NSE) of sub catchment runoff $q_{cat,out}$, as sub catchment runoff is a comprehensive single-valued indicator of overall sub catchment state, and NSE is the best-known quality measure in hydrology. We calculate NSE in a distributed manner, i.e. separately for each sub catchment runoff, and then take its mean, weighted by the area of each sub catchment. Unlike directly calculating NSE of discharge at the basin outlet, this avoids potential compensation of under- and overestimations in particular sub catchments when aggregating their discharge in the river network.

As for similarity analysis, we use the results of a fully distributed model run for the entire 5-year time series of available data as a virtual reality benchmark. This 'reference' run was created by operating SHM Attert in a standard mode, i.e. without any adaptive clustering functionality implemented. In addition to the 'reference' run, we established further benchmark cases: For the 'static' benchmark, we implemented the adaptive clustering functionality into SHM Attert, but set its parameters such that throughout the entire model run, each sub catchment was treated separately, and any clustering was suppressed. This means that adaptive clustering was in action, causing its computational overhead, but nevertheless the model was operated in the same fully distributed manner as the 'reference' run. We also established a 'static optimal' benchmark based on an offline, prior similarity analysis of the sub catchments: All sub catchments with identical structural properties - except size - and identical forcing were joined into a set of time-invariant clusters (see Table 2). As we set up the SHM Attert model in a straightforward manner, with sub catchment parameters varying only between different geology and land use classes, and sub catchment forcings varying only between rain gauges, the 173 sub catchments could be grouped into only 24 time-invariant, yet optimal clusters. 'Optimal' here means that there is no within-cluster variability - except size, which vanishes due to the (0,1)-normalization - and any single sub catchment picked from a cluster is a perfect representative of all others. This is of course a simplified and idealized case due to the simplified set up of the model. Adding further structural properties and forcing, and in higher resolution, will result in more clusters. Nevertheless we used the 'static optimal' benchmark to evaluate the merits of advance knowledge about time-invariant sub catchment similarity. Its model setup and operation was equal to the 'static' case, but this time the 24 static optimal clusters were used, instead of treating each sub catchment as a single cluster.

Table 2. Sub catchments of SHM Attert, grouped into 24 time-invariant clusters by agreement in attributes geology, land use and meteorological forcing. The clusters are used in the 'static optimal' benchmark model. Sub catchment locations are shown in Fig. 2. The possible number of unique combinations of geology, land use, and rain gauge is 3 * 5 * 3 = 45. The cluster sizes range from 1 to 28, the average number of elements in a cluster is 7.2.

Geology class[a]	Land use class[b]	Rain gauge[c]	# of cluster members	Sub catchment IDs
1	1	ROD	13	6, 7, 8, 10, 14, 15, 40, 127, 131, 133, 134, 135, 137
1	1	RCL	1	124
1	2	ROD	7	21, 36, 91, 142, 144, 152, 153
1	2	RCL	1	123
1	3	ROD	13	9, 27, 30, 33, 41, 52, 126, 128, 129, 136, 143, 154, 155
1	3	RCL	1	122
1	4	ROD	5	24, 28, 31, 130, 132
2	1	ROD	22	19, 22, 26, 47, 71, 74, 82, 88, 94, 95, 96, 97, 98, 106, 108, 110, 116, 118, 157, 161, 168, 169
2	1	RCL	28	29, 34, 38, 44, 48, 51, 53, 62, 64, 67, 70, 79, 83, 84, 85, 90, 92, 99, 111, 117, 119, 138, 139, 146, 147, 150, 158, 164
2	1	USL	15	32, 72, 73, 76, 86, 102, 103, 104, 105, 112, 113, 115, 121, 140, 151
2	2	ROD	19	35, 42, 45, 54, 56, 58, 59, 65, 68, 69, 75, 78, 80, 81, 89, 93, 101, 145, 148
2	2	RCL	6	16, 43, 46, 60, 61, 100
2	2	USL	6	39, 49, 50, 57, 87, 141
2	3	ROD	2	149, 167
2	4	ROD	7	37, 77, 107, 109, 156, 160, 162
2	4	RCL	5	25, 55, 63, 120, 159
2	5	ROD	2	66, 163
3	1	RCL	6	2, 4, 11, 12, 23, 125
3	2	ROD	1	114
3	2	RCL	1	20
3	3	ROD	1	166
3	3	RCL	5	1, 3, 5, 13, 18
3	4	RCL	4	17, 165, 170, 171
3	4	USL	2	172, 173

[a] 1: Schist, 2: Marl, 3: Sandstone

[b] 1: Meadow, 2: Agriculture, 3: Coniferous forest, 4: Broad-leaf forest, 5: Sealed area

[c] ROD = Roodt, RCL = Reichlange, USL = Useldange (compare Table A2)

3 Results and discussion

We first show results from analysing sub catchment similarity, as it is a precondition for adaptive clustering, and then results from comparing models runs with adaptive clustering to benchmark cases. We also discuss if, and to which degree, space-time patterns of similarity apparent in the fully distributed 'reference' model run are preserved by adaptive clustering.

3.1 Results for hydrological similarity

We discuss hydrological similarity by three aspects: Time-variant behavior, time-averaged values, and spatial patterns. For the first, time series of Shannon entropy for selected variables of the fully distributed 'reference' model run are shown in Fig. 3, panel 'a'. For better visibility, the plot is restricted to a single year. High entropies indicate little similarity among sub catchments, low entropies indicate high similarity. It is apparent that for all variables, entropies remain well below the benchmark maximum entropy (shown as red line), and often they are close to zero. This indicates that many of the sub catchments act similarly, leading to redundancies in modeling which can potentially be reduced by adaptive clustering. Also, entropies vary with time. While the variability differs among variables (e.g. high for interflow storage s_i, and low for base flow storage s_b), it is present for all of them, which emphasizes that clustering should be done in a time-variant manner. The entropy of the clustering control variable, $q_{cat,out}$, shows a high correlation with discharge magnitude as shown in panel 'b': In times of rising and high discharge, entropies are high, which is likely due to i) the interplay of spatially distributed precipitation and catchment states, and ii) the onset of fast runoff components, which may differ among sub catchments. As through times of recession, precipitation is zero for all sub catchments, and fast runoff components are dormant, low flow situations are accompanied by low entropies, indicating a high degree of similarity among sub catchments. All of these observations agree with the findings of Loritz et al. (2018).
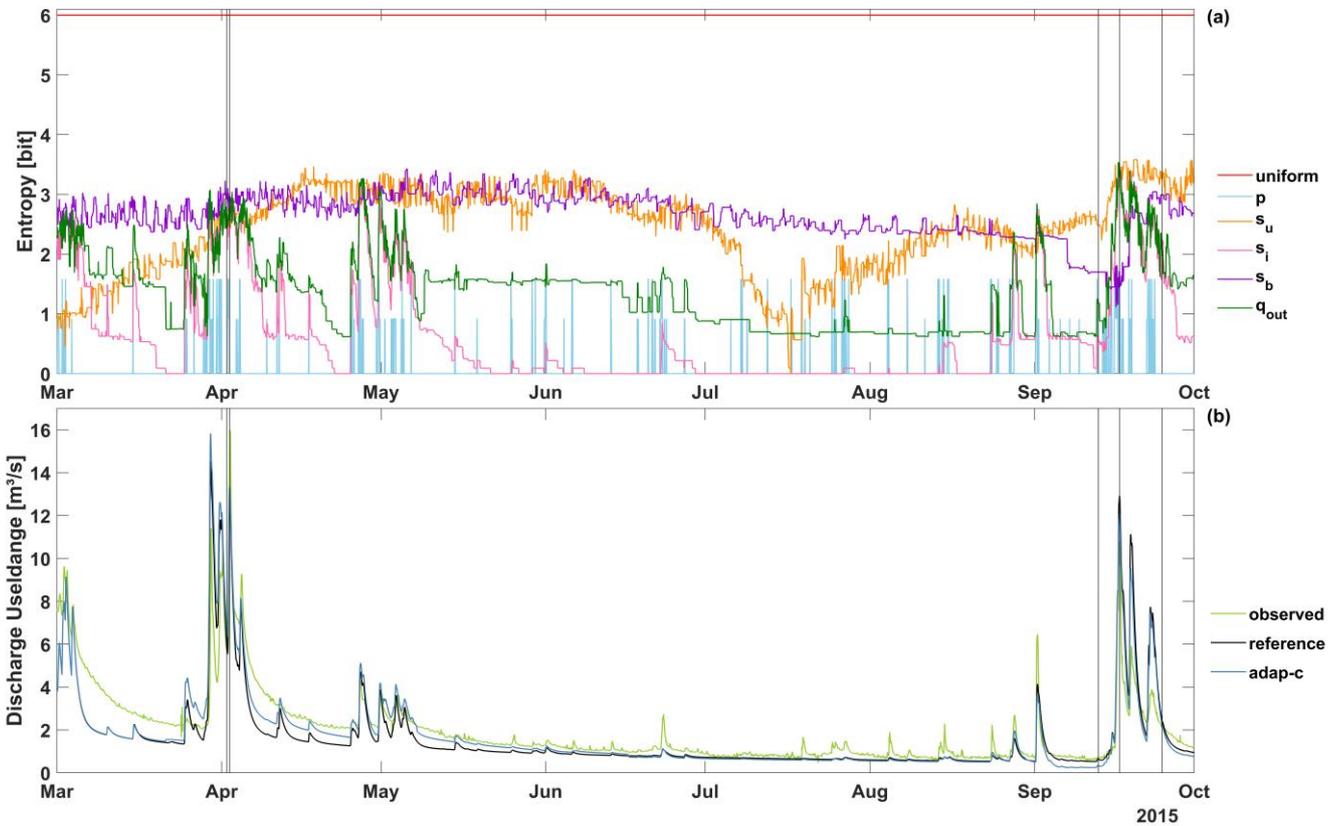
Figure 3. Panel a: Time series of Shannon entropy of distributions of normalized and binned sub catchment states and fluxes. Distributions are based on the fully distributed 'reference' model run. 'P' is precipitation, 's_u' is unsaturated zone storage, 's_i' and 's_b' are interflow and base flow storage, 'q_{out}' is sub catchment runoff (see Fig. A1 and Table A1). 'Uniform' indicates the benchmark maximum entropy of 6 bit for a

11

64-bin distribution. Black vertical lines indicate times for which spatial maps are shown in Fig. 4. Panel b: Discharge time series at catchment outlet gauge Useldange. 'Observed' are observations, 'reference' are results from the fully distributed 'reference' model run, 'adap-c' are from an adaptive clustering run with optimized parameters as shown in Table 1.

5   Time-averaged entropies for all SHM state and flux variables are shown in Table 3. Like in Fig. 3, the values differ quite substantially among the variables, with precipitation 'p' showing the lowest entropy of only 0.1 bit, and base flow related variables s_b and $q_{b,out}$ showing the highest entropy of 2.88 bit. The low value of precipitation entropy can be explained by two effects: Firstly, during the frequent times of no rain, precipitation entropy is also zero as all stations show the same value, and secondly, even if it rains, at most three different bins of the distribution can be occupied as precipitation is measured by

10 only three stations. This limits precipitation entropy to a possible maximum of $\log_2(3) = 1.58$ bit. The high values for base flow can be explained by the pronounced, geology-induced differences of the base flow behavior across the catchment (compare the values of k_b in Table A3), and the fact that due to the slow-changing nature of base flow, these differences prevail for a long time, keeping entropies high throughout the year (see Fig. 3, panel 'a'). Interestingly, the entropies of several variables are identical ($q_{i,in}$, $q_{b,in}$, and $q_{u,out}$; s_i and $q_{i,out}$; s_b and $q_{b,out}$). This is not a coincidence, but a consequence of

15 how they are related: $q_{i,in}$, and $q_{b,in}$ are percentages of $q_{u,out}$; runoff from both the interflow and the base flow reservoir are linear functions of the respective storages (see Fig. A1 and Table A1). All of these relations are entropy-preserving transformations, i.e. the entropies of all variables involved are necessarily equal.

Table 3. Mean entropy of all normalized and binned SHM Attert states and fluxes of the 'reference' run, for period 2011/11/01 00:00 –
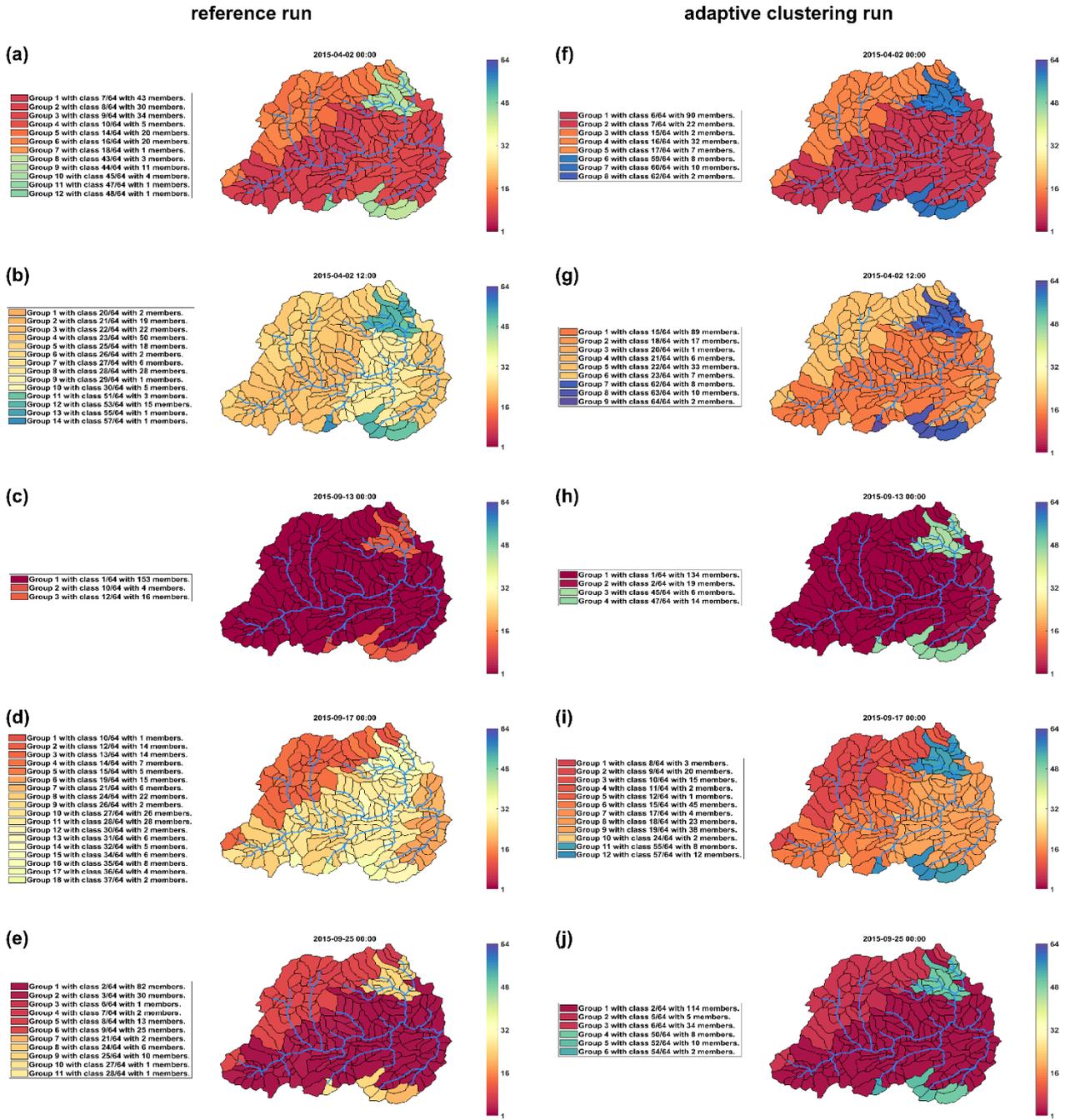20 2016/10/31 23:00. The states and fluxes are explained in Table A1.

State or flux	Entropy H [bit]
uniform	6^a
p	0.10
et	0.76
s_u	2.38
$q_{u,out}$	0.18
$q_{i,in}$	0.18
s_i	1.23
$q_{i,out}$	1.23
$q_{b,in}$	0.18
s_b	2.88
$q_{b,out}$	2.88
$q_{cat,out}$	1.84

[a] Entropy of the benchmark uniform distribution

Fig. 4 shows spatial patterns of normalized and binned values of the clustering control variable $q_{cat,out}$ for selected points in time. Plots in the left column ('a'-'e') are based on the 'reference' run, and we will focus on these in the following. We
25 selected the times such as to cover a wide range of different hydrological situations (compare the black vertical lines in Fig. 3). Plots 'a' and 'b' are both in spring and related to the same rainfall-runoff event, the last in a sequence of three, with plot 'a' showing the values just before the onset of precipitation, and plot 'b' at the time of peak runoff. Comparing the plots it is obvious that the general magnitude of runoff has increased, indicated by the main colours shifting from red (low values) to

yellow (intermediate values). Additionally we see that the spatial pattern of similarity also shifted from a geology-dominated pattern, reflecting the geology-based parametrization of sub catchments, to a pattern reflecting the joint influence of both geology and the spatial distribution of rainfall (see geological map and rain gauge areas of influence in Fig. 2). Interestingly, while the grouping of the sub catchments into clusters obviously changed, the overall number of clusters did only increase by two, from 12 to 14, indicating that the overall degree of similarity among sub catchment remained largely constant. The next plot, 'c', shows a very different situation at the end of a long summer drought: Most sub catchments show very low runoff, only the sand stone areas, where groundwater flow dominates, maintain runoff above their absolute minimum. Overall, the entire catchment is in a very homogeneous state and sub catchments group into only three clusters. This state of high similarity comes to a sudden end with the onset of precipitation (plot 'd'), increasing the diversity of sub catchment runoff to 18 clusters and the development of a spatial pattern which is mainly influenced by rainfall spatial distribution and only to a lesser degree by geology: the sand stone area is not as clearly separated from the other geologies as usual. Finally, after a period of extended rainfall (plot 'e'), a spatial pattern similar to the initial one in plot 'a' has re-established, but overall runoff magnitudes are still lower; a heritage of the long dry summer.

Altogether, the analysis of time-variant behavior, time-averages, and spatial patterns of similarity reveals several important points: i) for most variables, there is pronounced hydrological similarity among sub catchments, ii) similarity is time-variant, and iii) the spatial patterns of similarity and their variation with time are in accordance with hydrological reasoning, which increases our confidence that expressing similarity among sub catchments by entropies is reasonable. In the next section we discuss to which degree, and at which price, adaptive clustering can capitalize on this.

reference run

adaptive clustering run

(a) 2015-04-02 00:00
Group 1 with class 7/64 with 43 members.
Group 2 with class 8/64 with 30 members.
Group 3 with class 9/64 with 34 members.
Group 4 with class 10/64 with 5 members.
Group 5 with class 14/64 with 20 members.
Group 6 with class 16/64 with 20 members.
Group 7 with class 18/64 with 1 members.
Group 8 with class 43/64 with 3 members.
Group 9 with class 44/64 with 1 members.
Group 10 with class 45/64 with 4 members.
Group 11 with class 47/64 with 1 members.
Group 12 with class 48/64 with 1 members.

(f) 2015-04-02 00:00
Group 1 with class 6/64 with 90 members.
Group 2 with class 7/64 with 22 members.
Group 3 with class 15/64 with 2 members.
Group 4 with class 16/64 with 32 members.
Group 5 with class 17/64 with 7 members.
Group 6 with class 59/64 with 8 members.
Group 7 with class 60/64 with 10 members.
Group 8 with class 62/64 with 2 members.

(b) 2015-04-02 12:00
Group 1 with class 20/64 with 2 members.
Group 2 with class 21/64 with 19 members.
Group 3 with class 22/64 with 22 members.
Group 4 with class 23/64 with 50 members.
Group 5 with class 25/64 with 18 members.
Group 6 with class 26/64 with 2 members.
Group 7 with class 27/64 with 6 members.
Group 8 with class 28/64 with 28 members.
Group 9 with class 29/64 with 1 members.
Group 10 with class 30/64 with 5 members.
Group 11 with class 51/64 with 3 members.
Group 12 with class 53/64 with 15 members.
Group 13 with class 55/64 with 1 members.
Group 14 with class 57/64 with 1 members.

(g) 2015-04-02 12:00
Group 1 with class 15/64 with 89 members.
Group 2 with class 18/64 with 17 members.
Group 3 with class 20/64 with 1 members.
Group 4 with class 21/64 with 6 members.
Group 5 with class 23/64 with 33 members.
Group 6 with class 23/64 with 7 members.
Group 7 with class 62/64 with 8 members.
Group 8 with class 63/64 with 10 members.
Group 9 with class 64/64 with 2 members.

(c) 2015-09-13 00:00
Group 1 with class 1/64 with 153 members.
Group 2 with class 10/64 with 4 members.
Group 3 with class 12/64 with 16 members.

(h) 2015-09-13 00:00
Group 1 with class 1/64 with 134 members.
Group 2 with class 2/64 with 19 members.
Group 3 with class 45/64 with 6 members.
Group 4 with class 47/64 with 14 members.

(d) 2015-09-17 00:00
Group 1 with class 10/64 with 1 members.
Group 2 with class 12/64 with 14 members.
Group 3 with class 13/64 with 14 members.
Group 4 with class 14/64 with 7 members.
Group 5 with class 15/64 with 5 members.
Group 6 with class 19/64 with 15 members.
Group 7 with class 21/64 with 6 members.
Group 8 with class 24/64 with 22 members.
Group 9 with class 26/64 with 2 members.
Group 10 with class 27/64 with 26 members.
Group 11 with class 28/64 with 28 members.
Group 12 with class 30/64 with 2 members.
Group 13 with class 31/64 with 6 members.
Group 14 with class 32/64 with 1 members.
Group 15 with class 34/64 with 6 members.
Group 16 with class 35/64 with 8 members.
Group 17 with class 36/64 with 4 members.
Group 18 with class 37/64 with 2 members.

(i) 2015-09-17 00:00
Group 1 with class 8/64 with 3 members.
Group 2 with class 9/64 with 20 members.
Group 3 with class 10/64 with 15 members.
Group 4 with class 11/64 with 2 members.
Group 5 with class 12/64 with 1 members.
Group 6 with class 15/64 with 45 members.
Group 7 with class 17/64 with 4 members.
Group 8 with class 18/64 with 23 members.
Group 9 with class 19/64 with 38 members.
Group 10 with class 24/64 with 2 members.
Group 11 with class 55/64 with 8 members.
Group 12 with class 57/64 with 12 members.

(e) 2015-09-25 00:00
Group 1 with class 2/64 with 82 members.
Group 2 with class 3/64 with 30 members.
Group 3 with class 6/64 with 1 members.
Group 4 with class 7/64 with 2 members.
Group 5 with class 8/64 with 13 members.
Group 6 with class 9/64 with 25 members.
Group 7 with class 21/64 with 2 members.
Group 8 with class 24/64 with 6 members.
Group 9 with class 25/64 with 10 members.
Group 10 with class 27/64 with 1 members.
Group 11 with class 28/64 with 1 members.

(j) 2015-09-25 00:00
Group 1 with class 2/64 with 114 members.
Group 2 with class 5/64 with 5 members.
Group 3 with class 6/64 with 34 members.
Group 4 with class 50/64 with 8 members.
Group 5 with class 52/64 with 10 members.
Group 6 with class 54/64 with 2 members.

Figure 4. Spatial maps of normalized and binned values of the clustering control variable, $q_{cat,out}$, for all sub catchments and for selected points in time as marked by the black vertical lines in Fig. 3. Colors indicate the values, which correspond to the bin numbers ranging from 1 (lowest normalized state, red) to 64 (highest normalized state, blue). Values in the left column are from the fully distributed 'reference' run, values in the right column are from an adaptive clustering run with optimized parameters as shown in Table 1.

5

3.2 Results for adaptive clustering

As explained in Sect. 2.4.2, we evaluate adaptive clustering for both computational effort and associated quality losses against several benchmarks. Fig. 5 shows the results as a two-dimensional plot. The black square indicates the 'reference' run, which corresponds to the standard case of running a model run in full resolution and without any adaptive clustering functionality. It took 816 seconds, and as the 'reference' run is our virtual reality 'truth', the model shows perfect simulation quality, indicated by an NSE of 1. When adaptive clustering functionality is integrated into the model, but by choice of its parameters a fully distributed run is enforced (the 'static' benchmark case), the model still shows perfect simulation quality,

14

but the overhead of adaptive clustering increases computation times by 707 seconds to a total of 1523 seconds (black triangle in Fig. 5). This is almost a doubling compared to the 'reference' case, a computational extra cost which clustering needs to over-compensate in order to be worth the effort. This is indeed the case, even for the simple 'static optimal' benchmark case (red triangle in Fig. 5): Representing 173 sub catchments by 24 representatives (one per cluster, see Table 2) reduced

5   computation time to 233 seconds, despite the overhead, at no loss of quality. How does time-variable, adaptive clustering compare to that? The blue dots in Fig. 5 depict results for selected parameter choices of adaptive clustering (we tested many but only show the pareto-optimal results), and reveal a general pattern of trade-off between effort and quality: The higher the computational effort, the higher the modelling quality, and vice versa. The red dot indicates the - in our eyes - optimal tradeoff based on the optimized parameter set shown in Table 1. The related computation time is 207 seconds, NSE is 0.84.

10   This means that compared to the 'reference' case, computation time is reduced by 75% at the price of worsening NSE by 0.16. The effect of adaptive clustering on the quality of discharge simulations at the catchment outlet is shown in Fig. 3, panel 'b': Differences between the 'reference' and the 'adap-c' optimal adaptive clustering run are visible, but they are generally much lower than the differences between the 'reference' simulation and the observed discharge. This is encouraging. But when comparing the 'adap-c optimal' run to the 'static optimal' benchmark, we may ask whether the small

15   reductions in computation time, at the cost of a decrease in model quality, makes adaptive clustering worth the effort. For the given model, our answer will likely be 'No'. However, as discussed in Sect. 2.4.2, SHM Attert is extremely well suited for static clustering due to its simple set up, and for most distributed models it will not be possible to group model elements into an equally small number of time-invariant yet optimal clusters. In such cases, the relative gains of adaptive clustering will be potentially more pronounced.

20
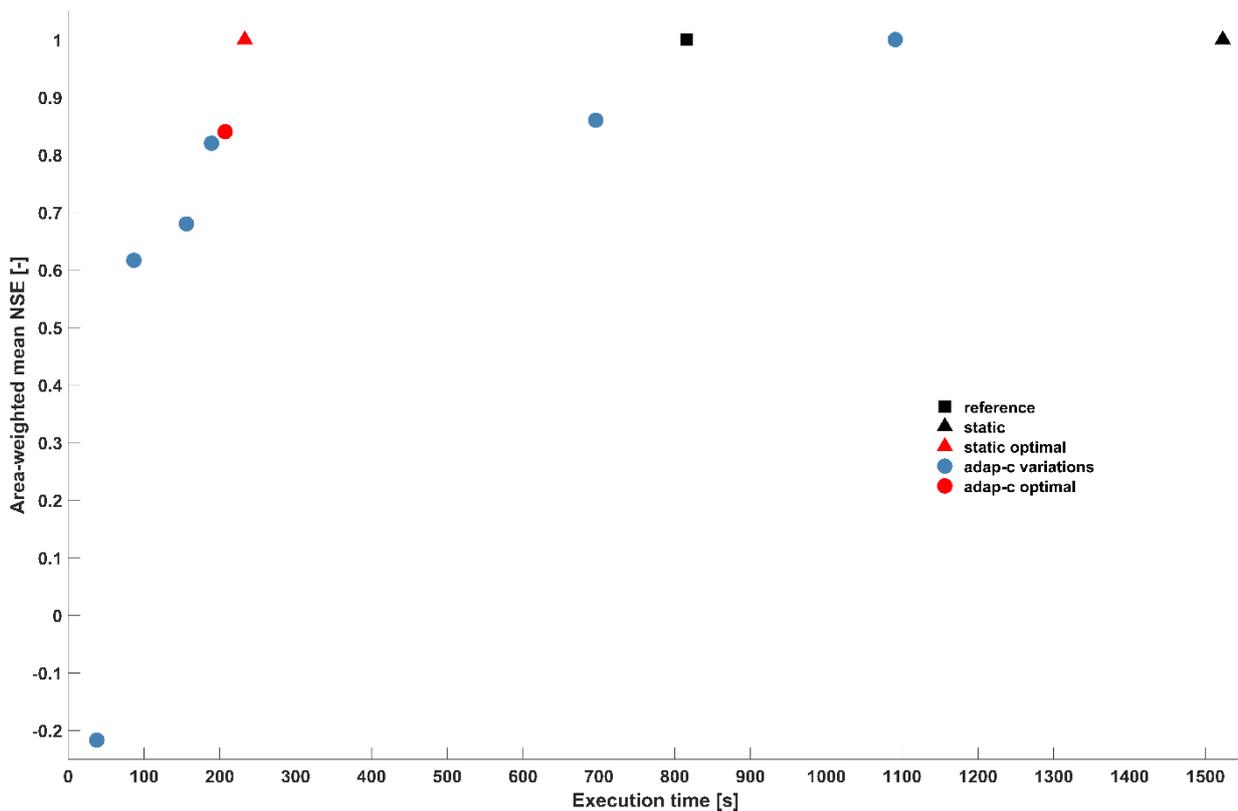
Figure 5. Performance of model runs with respect to effort, measured by execution time, and quality, measured by mean Nash-Sutcliffe efficiency of sub catchment runoff $q_{cat,out}$. 'Reference': Full resolution, no adaptive clustering overhead; 'static': Full resolution, but with adaptive clustering overhead; 'static optimal': Time-invariant optimal clustering, with clusters shown in Table 2; 'adap-c variations': adaptive clustering with various parameter settings; 'adap-c optimal': optimal adaptive clustering with parameters shown in Table 1.

25

Fig. 6 is based on the 'adap-c optimal' model run and gives some insights about the behavior of adaptive clustering. The blue line in panel 'a' shows for each time step the agreement between the inherited and the current clustering (step 'g' in Sect. 2.1 and Sect. 2.3). The related thresholds for starting and ending jumps back in time are indicated by horizontal lines: *sim_crit* is indicated by the lower line, *sim_uncrit* by the upper line (see Table 1). Each time the clustering agreement falls below

5   *sim_crit*, a jump back in time is triggered (red circle in the plot). It goes back to the closest time when the agreement was still above *sim_uncrit* (red dot in in the plot). From this point in time, the model is operated in full distribution (no clusters) until the time of the jump back is reached again. Fig. 6 reveals that the occurrence and the length of jump back periods varies with the hydrological situation. During times of rapidly changing catchment conditions, such as in April, frequent but short jumps back appear, indicating that the inherited clusters are only valid for short periods of time. For periods of low flow, such as in

10   August, no jumps back occur at all: Apparently, the clusters determined in mid-June remain valid until rainfall events at the beginning of September terminate the long period of synchronized drying of all sub catchments. This indicates that updating of the clusters is controlled by changes in the hydro-meteorological situation. For the entire five-year simulation period, overall 165 re-clusterings occur, i.e. on average one every eleven days.

Adaptive clustering increases modeling efficiency by restricting computations to the representatives, but it also comes at a

15   computational cost. This is illustrated by the blue line in Fig. 6, panel 'b', which shows the number of sub catchments for which hydrological processes were calculated at each time step. Normally, this number corresponds to the number of cluster representatives set by parameter *perc_reps* (see Table 1). For the jump back periods it is different: They are visited twice, once when the model is in normal, clustered, forward mode, and once in full resolution when the model is in jump back mode. As a consequence, the total number of sub catchments processed in these times is high. It can even exceed the total

20   number of sub catchments of the model (173, indicated by the red line). Overall, however, the savings prevail: For the entire five-year simulation period, on average 34 of 173 sub catchments were processed per time step, which means a reduction of 80% compared to the 'reference' model run.
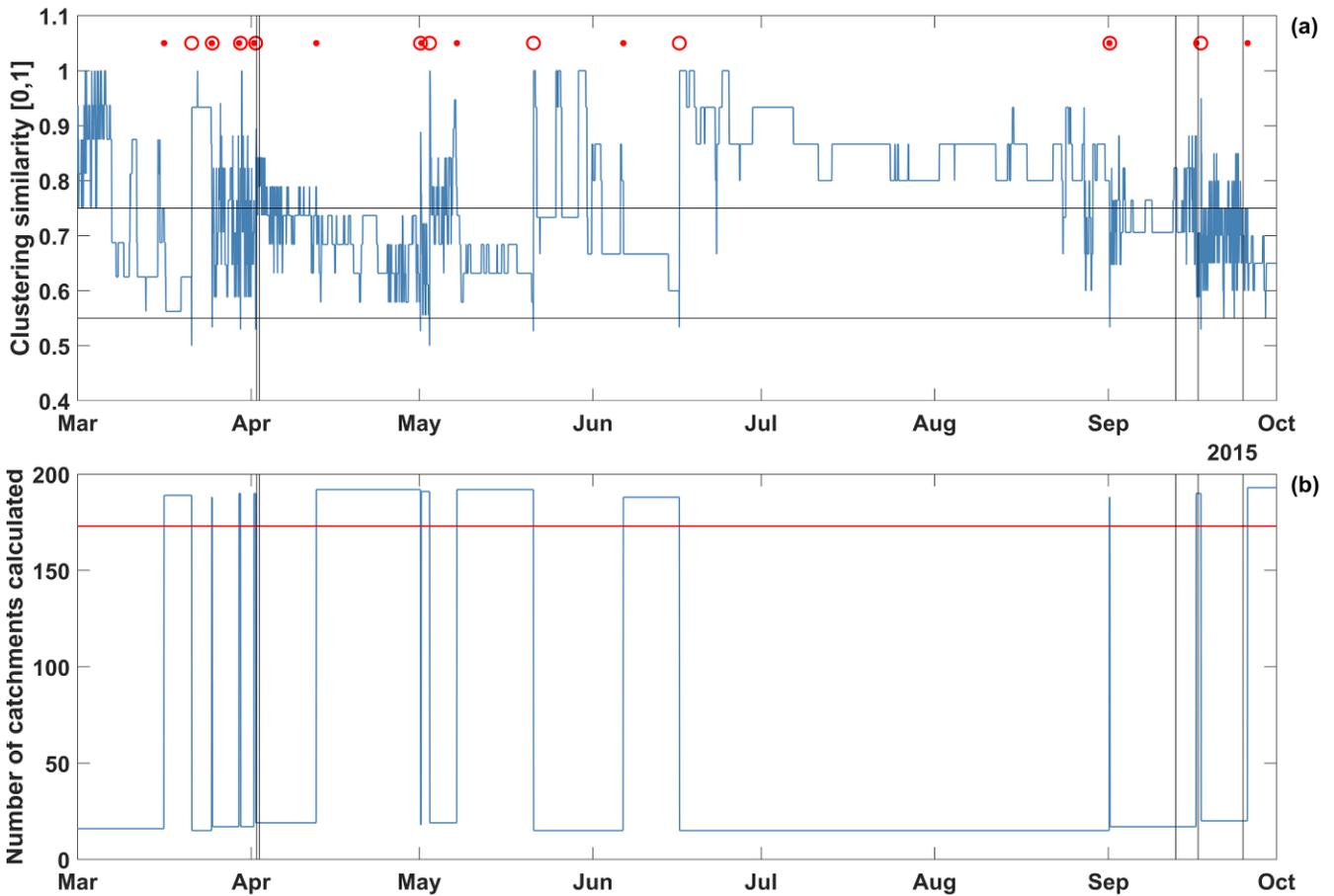
Figure 6. Panel 'a': Time series of agreement between inherited and current clusters, based on the 'adap-c optimal' model run. Black vertical lines indicate times of special interest (compare Fig. 3 and Fig. 4), black horizontal lines indicate the agreement thresholds given by parameters *sim_crit* and *sim_uncrit* (see Table 1). A red circle indicates when a jump back in time was triggered. The jump goes back to the next red dot to the left of each circle. Panel 'b': Number of sub catchments per time step for which hydrological processes were

5   calculated. The red horizontal line indicates the total number of sub catchments (173) in the SHM Attert model.

Adaptive clustering restricts the execution of hydrological processes to a few representative sub catchments. While this seems sufficient to preserve the time-variant behavior of all sub catchments - as indicated by the high NSE values of the adaptive clustering model runs, and the good agreement of the discharge hydrographs in Fig. 3, panel 'b' - the question

10   remains whether spatial patterns of similarity are also preserved. To address this question, spatial maps of sub catchment discharge from the optimized adaptive clustering run are plotted in the right column of Fig. 4. The dates of each plot are identical to those of the 'reference' run in the left column. Comparing the associated maps shows that they largely agree. Further, the main characteristics of sub catchment similarity seem to be preserved by adaptive clustering: Geology is the main and precipitation a secondary control, and the degree of similarity varies over time. But there are also differences: For

15   example, comparing plots 'b' and 'g' reveals a smaller influence of the rainfall pattern in the adaptive clustering case, for plots 'd' and 'i' the opposite is the case. Generally, the overall degree of similarity is larger for the adaptive clustering case, i.e. the number of clusters is always smaller as for the 'reference' case. This is a consequence of mapping states from a few representatives to many recipients in a cluster, which involves averaging and hence an artificial increase of similarity among sub catchments.

20   To summarize: We have tested many variants of adaptive clustering against several benchmark models in terms of computational effort and model quality. For the best variant, computation time was reduced by 75% compared to the full-resolution 'reference' model run, and model quality, expressed by NSE, decreased by 0.16. Compared to the 'static optimal' benchmark, which uses time-invariant clusters, computational savings were much smaller, as the SHM Attert model, due to

17

its simplicity, lends itself well for time-invariant clustering. Analysing the time-variant behaviour of adaptive clustering revealed that re-clustering is linked to changes of the hydro-meteorological situation. Analysing spatial patterns of sub catchment similarity from model runs with and without adaptive clustering revealed that adaptive clustering preserves their main characteristics but shows a tendency for exaggerating similarity.

4 Summary and conclusions

In this paper we proposed and described adaptive clustering as a new way to reduce computational efforts of distributed modelling, while largely maintaining modelling quality. This is done by identifying, in a time-variant manner, similar acting model elements, clustering them and inferring dynamics of all model elements from just a few representatives per cluster.

We started from the observation that hydrological systems generally exhibit spatial variability of their properties, and that this variability is non-negligible if distributed dynamics are of interest, which then requires distributed modelling. We further hypothesized that despite this variability, there is also similarity, i.e. many model elements exist with similar properties, which will exhibit similar internal dynamics and produce similar output when in similar initial states and when exposed to similar forcing. Similarity among model elements is hence not a static but rather a time-variable property dependant on the interplay of these factors, and similarity is also not necessarily limited to contiguous model elements.

Based on these premises we developed adaptive clustering, and provided a proof-of-concept at the example of a distributed, conceptual hydrological model - SHM -, fit to the Attert basin in Luxembourg. Adaptive clustering comprises several steps: Clustering of model elements, choice of cluster representatives, mapping of results from representatives to recipients, and comparison of clusterings over time to decide when re-clustering is required. We explained these steps in general, and its implementation in the SHM Attert model in particular. We used normalized and binned transformations of model states for both clustering and for measuring overall similarity among model elements by Shannon information entropy. Analysing time series of entropy of model states and fluxes revealed that i) for most variables, there is pronounced hydrological similarity among sub catchments, ii) similarity is time-variant, and iii) the spatial patterns of similarity and their variation with time are in accordance with hydrological reasoning. We then evaluated adaptive clustering with respect to both computational gains and losses of model quality against several benchmark models. Compared to a standard, full-resolution model run used as a virtual reality 'truth', computation time could be reduced by 75%, when a decrease of Nash-Sutcliffe efficiency by 0.16 was accepted. Re-clustering of model elements was linked to changes of the hydro-meteorological situation, and was on average carried out once every eleven days.

Our tests and analyses were conducted in the virtual reality of a fully distributed model run, due to a lack of equally comprehensive observations. However, due to the good overall agreement of the model with the available multi-variate observations, we are confident that our main conclusion, namely that adaptive clustering is a promising tool for accelerating distributed modelling of hydrological and other dynamical systems, also holds with respect to real world systems. Additionally, adaptive clustering yields spatial and temporal patterns of similarity among model elements, which can be used for hydrological system analysis. Adaptive clustering integrates and enhances existing methods of time-invariant grouping of model elements, such as lumping or grouped response units (GRU), and it can be applied together with existing methods of exploiting time-variable similarity, such as adaptive gridding or adaptive time-stepping. A limitation of the method lies in the potential violation of conservation laws when mapping results from cluster representatives to recipients.

What's ahead? For the study we selected sub catchment runoff as the single variable for both clustering control and model evaluation. This was mainly based on hydrological reasoning, and clearly other and/or additional variables for clustering control should be tested, and the effect of adaptive clustering on all model states and fluxes should be evaluated. Also, so far cluster representatives were simply chosen by random picking. We expect better performance by a targeted selection of model elements, e.g. those close to the cluster centre. At last, we have tested adaptive clustering at the example of a

relatively simple, conceptual hydrological model with limited internal variability. The performance of adaptive clustering in more advanced models such as MIKE SHE (Abbott et al., 1986), HydroGeoSphere (Brunner and Simmons, 2011; Davison et al., 2018), Noah-MP LSM (Niu et al., 2011), or the Community Land Model CLM (Lawrence et al., 2019) - where computation times are indeed a challenge - remains to be demonstrated. On the one hand, the potential savings by adaptive

5 clustering will increase with the level of process detail in a model. On the other hand, its implementation will become more difficult. For example, existing code may be structured in a way unfavourable for integrating the adaptive clustering functionality; it can be a challenge to combine massive parallel processing with adaptive clustering; and it will be difficult to integrate it into models where many processes act simultaneously at a hierarchy of model elements. However, the same could be said about adaptive gridding and adaptive time stepping, nevertheless they have been implemented very

10 successfully in many advanced earth science models.

Appendix A

A.1 SHM model structure

SHM is a distributed hydrological model, i.e. a catchment is divided into sub catchments which are typically a few square kilometres in size. The water stocks and fluxes in each sub catchment are represented in a conceptualized manner by a set of

15 linked linear reservoirs (see Fig. A1). The choice of the type, number and linkage of reservoirs is based on the insights about the hydrological functioning of the Attert basin and suitable conceptualizations reported by Fenicia et al. (2014) and Fenicia et al. (2016). The model structural elements and all related equations are shown in Fig. A1 and Table A1. The first reservoir represents the unsaturated zone. Precipitation falling onto a sub catchment is divided into direct runoff and soil moisture replenishment as a nonlinear function of current soil moisture (the HBV beta store concept). Evapotranspiration draws water

20 from the unsaturated zone storage. Direct runoff is split by a constant factor and replenishes two linear reservoirs, one representing interflow, the other representing base flow. Runoff from the interflow and base flow reservoirs are added and then enter the river system. The river system is represented by a linear reservoir cascade, where each element represents a river stretch of about one kilometre. The model is coded in Matlab, the numerical scheme is non-iterative forward-in-time, the time-stepping is hourly.

25
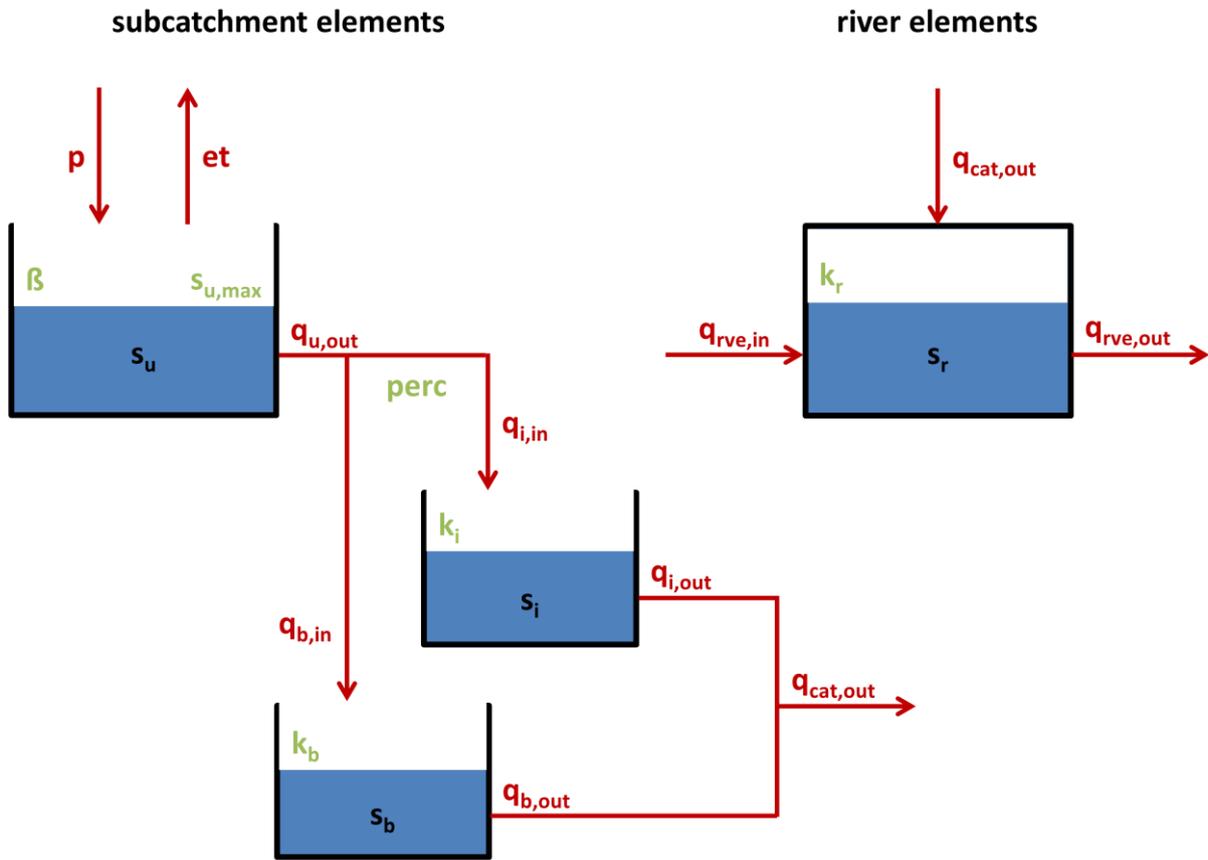
Figure A1. Structural elements, parameters (green), state variables (black) and fluxes (red) of the SHM model.

5

10

15

20

Table A1. SHM model equations and parameters

Number	Equation[a]	Description
(A1)	$vol\_p = p \cdot a$	vol_p: precipitation volume [m³] p: precipitation [m] a: sub catchment area [m²]
(A2)	$\Psi = \left(\dfrac{s_u}{s_{u,max}}\right)^{\beta}$	Ψ: runoff coefficient [-] s_u: storage in unsaturated zone reservoir [m] $s_{u,max}$: maximum storage in unsaturated zone reservoir [m] ß: shape coefficient [-]
(A3)	$q_{u,out} = \Psi \cdot vol\_p$	$q_{u,out}$: runoff from unsaturated zone reservoir [m³]
(A4)	$et = et_{ref} \cdot k_c \cdot k_{\theta}$	et: evapotranspiration [m³] et_{ref}: reference evapotranspiration [m³][b] k_c: vegetation correction factor [-][c] k_{θ}: soil moisture correction factor [-][d]
(A5)	$q_{i,in} = q_{u,out} \cdot perc$	$q_{i,in}$: inflow to interflow reservoir [m³] perc: interflow-baseflow partitioning factor [-]
(A6)	$q_{b,in} = q_{u,out} \cdot (1 - perc)$	$q_{b,in}$: inflow to baseflow reservoir [m³]
(A7)	$q_{i,out} = \dfrac{s_i}{k_i} \cdot a$	$q_{i,out}$: runoff from interflow reservoir [m³] s_i: storage in interflow reservoir [m] k_i: interflow reservoir retention constant [-]
(A8)	$q_{b,out} = \dfrac{s_b}{k_b} \cdot a$	$q_{b,out}$: runoff from baseflow reservoir [m³] s_b: storage in baseflow reservoir [m] k_b: baseflow reservoir retention constant [-]
(A9)	$q_{cat,out} = q_{i,out} + q_{b,out}$	$q_{cat,out}$: runoff from sub catchment [m³]
(A10)	$s_r = s_r + \sum\limits_{i} q_{rve,in,i} + \sum\limits_{j} q_{cat,out,j}$	$s_r(t)$: storage in river element [m³] $q_{rve,in,i}$: inflow from connected river element i [m³] $q_{cat,out,j}$: inflow from connected sub catchment j [m³]
(A11)	$q_{rve,out} = \dfrac{s_r}{k_r}$	$q_{rve,out}$: runoff from river element [m³] k_r: river element retention constant [-]

[a] in all equations, time subscripts t and t-1 are dropped for brevity

[b] evapotranspiration from reference surface (short grass) according to Penman (1956). Equations taken from DVWK (1996), section 5.3.1.

[c] vegetation correction factor as a function of land used and month of the year. Taken from Dunger (2006), Appendix 11. Value range [0.65, 1.3]

[d] adapted from Dunger (2006), section 4.5.8.4, Fig. 37 with the assumptions $k_{\theta} = 0$ for $s_u \leq 0$; $k_{\theta} = 1$ for $s_u \geq 0.8*s_{u,max}$; $k_{\theta} = s_u/s_{u,max}$ for $0 < k_{\theta} < 0.8*s_{u,max}$

A.2 SHM Attert – model setup

Setting up the SHM model for a catchment starts by GIS-based delineation of sub catchments and river elements using a digital elevation model. For the Attert, a 5 m digital elevation model based on LIDAR scans provided by the Luxembourg Institute of Science and Technology (LIST) was used. Each sub catchment was assigned a single land use, based on the Corine Land Cover map provided by the European Environment Agency, and a single geology, based on the 'Carte géologique détaillée 1:25000-1:50000' provided by the Luxembourg geological survey. In the catchment, altogether five

different land use classes and three geological classes occurred. For an overview of geology and land use classes assigned to each sub catchment, please see Table 2.

In the Attert basin, hydrological function is strongly controlled by geology (Fenicia et al., 2016). Therefore, all soil-related model parameters (β, $s_{u,max}$, perc, k_i, k_b) were kept equal for all sub catchments sharing the same geology. The parameter values were determined by calibration (see Appendix A.3). Likewise, all parameters related to evapotranspiration (k_c, k_θ) were kept equal among all sub catchments sharing the same land use class. These were - without calibration - directly inferred from the land use (see Eq. A4 in Table A1). As we set all river elements to be approximately one kilometre in length, we could assign to all 147 of them the same value for k_r, which we determined by calibration.

Running the SHM model requires observed time series of precipitation, air temperature, air relative humidity, wind velocity and global radiation. For precipitation, data from three stations were available. While this is clearly not enough to represent the full spatial variability of precipitation, it nevertheless represents some of it. Each sub catchment was assigned precipitation from a single station using a nearest neighbour approach (see Fig. 2 and Table 2). As the remaining hydro-meteorological variables typically exhibit less spatial variability than precipitation, we used observations from only a single station each (see Table A2 and Fig. 2).

Table A2. Time series data used for model calibration, validation, and operation. All data were available in 1-h resolution for the period 2011/11/01 00:00 – 2016/10/31 23:00. Time reference is MEZ.

ID	Full name	Data type and unit	Catchment area [km²]	Source
ROD_p	Roodt	precipitation [mm]	-	ASTA[a]
RCL_p	Reichlange	precipitation [mm]	-	AGE[b]
USL_p	Useldange	precipitation [mm]	-	ASTA
USL_ta	Useldange	air temperature at 2 m [°C]	-	ASTA
USL_rh	Useldange	air relative humidity at 2 m [%]	-	ASTA
ROD_v	Roodt	wind velocity at 3 m [m/s]	-	LIST[c]
MER_rg	Merl	global radiation [W/m²]	-	ASTA
COL_q	Colpach	discharge [m³/s]	19.5	LIST
WOB_q	Wollefsbach	discharge [m³/s]	4.5	LIST
PLA_q	Platen	discharge [m³/s]	44.2	LIST
USL_q	Useldange	discharge [m³/s]	245	LIST

[a] Administration des services techniques de l'agriculture Luxembourg

[b] Administration de la gestion de l'eau Luxembourg

[c] Luxembourg Institute of Science and Technology

A.3 SHM Attert – model calibration and validation

We applied a multi-criteria calibration approach to ensure good overall performance of SHM Attert, and not just with respect to discharge at the catchment outlet. For calibration we used data from the four-year period 2011/11/01 00:00 – 2015/10/31 23:00; for validation we used the remaining one-year period 2015/11/01 00:00 – 2016/10/31 23:00. We started by joint calibration of all sub catchment parameters on catchment scale, i.e. against observed discharge at the catchment outlet Useldange, and against catchment-averaged observations of soil moisture and evapotranspiration. The unique set of available soil moisture data (observations from 18 sensors in the schist, 11 in the marls, and 19 in the sandstone region, all taken at 50 cm depth) was taken during the CAOS project. There is no direct representation of soil moisture in SHM, we therefore compared normalized and catchment-averaged soil moisture observations against normalized and catchment-averaged

storage in the SHM unsaturated zone reservoir (s_u). While this did not permit quantitative conclusions, it was nevertheless informative in terms of the timing of relative minima and maxima, and the overall shape of the time series. As direct observations of catchment-scale evapotranspiration rates were not available, we used satellite-based estimates provided by EUMETSAT (Trigo et al., 2011) instead and compared them to catchment-averaged evapotranspiration rates of SHM. For each variable we measured model performance by Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970). To measure overall model performance in a single number, we merged the three efficiencies for discharge, soil moisture and evapotranspiration into a single, multi-criteria objective function according to Eq. (A12). The weights assigned to each component were subjectively chosen, mainly based on our evaluation of the quality of the underlying observations. To a lesser degree, the weights also reflect our evaluation of the relative importance of each component for overall model evaluation.

$$NSE_{total} = 0.5 \cdot NSE_{discharge} + 0.3 \cdot NSE_{soil\ moisture} + 0.2 \cdot NSE_{evapotranspiration} \quad\quad (A12)$$

After the first catchment-uniform and multi-criteria estimation of parameters, we refined the estimates of all soil-related model parameters by calibration against three gauges, each representative for a particular geology: Colpach for schist, Wollefsbach for marls, and Platen for sandstone. These parameters (see Table A3) were then assigned to all sub catchments sharing the same geology. After a few iterations of catchment-scale- and geology-specific calibration, we determined the final, distributed parameter sets as shown in Table A3. The main differences among geology-specific parameters appear for the retention behaviour of the interflow and the base flow reservoir (k_i and k_b, respectively), which reflects the geology-specific hydrological functioning of the Attert basin as described by Fenicia et al. (2016): In the schist, dynamics is governed by a combination of two subsurface flow paths, in the marl, fast responses governed by near-surface flow paths prevail, while the sandstone areas are characterized by delayed responses governed by groundwater flow.

The catchment-scale performance measures for both the calibration and the validation period are shown in Table A4, and in Table A5 for performance at the gauges used for geology-specific and for catchment-wide calibration. The model achieves a catchment-scale, multi-objective Nash-Sutcliffe efficiency of 0.73 in the validation period; gauge- or criteria-specific efficiencies range from 0.61 for gauge Wollefsbach to 0.77 for gauge Useldange at the catchment outlet. For a visual comparison of observed and simulated discharge at Useldange in the year 2015, see Fig. 3, panel b, lines 'observed' and 'reference'.

Table A3. Parameters of the SHM Attert found by calibration in the period 2011/11/01 00:00 – 2015/10/31 23:00. The parameters are described in Table A1.

Geology	Target gauge ID	$s_{u,max}$ [m]	β [-]	perc [-]	k_i [-]	k_b [-]	k_r [-]
Schist	COL_q	0.17	5	0.5	44	500	1.1
Marls	WOB_q	0.17	3	0.7	20	3000	1.1
Sandstone	PLA_q	0.17	2	0.05	100	20000	1.1

Table A4. Catchment-scale performance measures (discharge: Nash-Sutcliffe efficiency at gauge Useldange. Soil moisture and evapotranspiration: Nash-Sutcliffe efficiency of catchment averages) of the SHM Attert in the 5-year calibration period (2011/11/01 00:00 – 2015/10/31 23:00) and 1-year validation period (2015/11/01 00:00 – 2016/10/31 23:00). 'Combination' refers to the joint objective function according to Eq. (A12).

Series	Calibration	Validation
Discharge	0.85	0.77
Soil moisture	0.80	0.66
Evapotranspiration	0.58	0.74
Combination	0.78	0.73

Table A5. Gauge-specific performance measures (Nash-Sutcliffe efficiency of discharge) of the SHM Attert in the calibration and validation period. Gauge locations are shown in Fig. 2, catchment sizes in Table A2.

Geology	Gauge ID	Calibration	Validation
Shist	COL_q	0.78	0.65
Marls	WOB_q	0.66	0.61
Sandstone	PLA_q	0.79	0.74
Catchment	USL_q	0.85	0.77

Code availability. The SHM Attert including the adaptive clustering functionality and all code used to conduct the analyses in this paper are publicly available at https://github.com/KIT-HYD/SHM-Attert-Adaptive-Clustering.

Data availability. The precipitation data of stations Roodt and Useldange, the air temperature, relative humidity and global radiation data are publicly available from the Administration des services techniques de l'agriculture Luxembourg ASTA at http://www.agrimeteo.lu/ (last access: 2019/08/20). The precipitation and discharge data at station Reichlange are available upon request from the Administration de la gestion de l'eau Luxembourg AGE at https://www.inondations.lu/ (last access: 2019/08/20). All other discharge data and the wind velocity data are available upon request from Luxembourg Institute of Science and Technology LIST at https://www.list.lu/ (last access: 2019/08/20). The EUMETSAT-based LSA-SAF evapotranspiration products are publicly available from http://landsaf.ipma.pt (last access: 2019/08/20). The soil moisture data are available upon request from Theresa Blume (blume@gfz-potsdam.de) and Markus Weiler (markus.weiler@hydrology.uni-freiburg.de). The digital elevation model is available upon request from LIST. The 2012 Corine Land Cover data are publicly available from the Copernicus sites of the European Environment Agency EEA at http://land.copernicus.eu/pan-european/corine-land-cover/clc-2012/view (last access: 2019/08/20). The geological maps are available upon request from the Service géologique, Administration des ponts et chaussées Luxembourg at http://www.geologie.lu/geolwiki/index.php/Cartes_g%C3%A9ologiques (last access: 2019/08/20).

Author contributions. UE coded the SHM model and the adaptive clustering algorithm and wrote the manuscript. UE, RvP, RL, EA and EZ developed the adaptive clustering concept together. RvP set up the SHM Attert model, MB calibrated and validated it.

Competing interests. The authors declare that they have no conflict of interest.

References

Abbott, M.B., Bathurst, J.C., Cunge, J.A., O'Connell, P.E., and Rasmussen, J.: An introduction to the European Hydrological System — Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system". Journal of Hydrology. 87 (1–2): 45–59, 1986. doi:10.1016/0022-1694(86)90114-9.

Aydogdu, A., Carrassi, A., Guider, C. T., Jones, C., and Rampal, P.: Data assimilation using adaptive, non-conservative, moving mesh models, Nonlinear Processes in Geophysics, 26, 175-193, 10.5194/npg-26-175-2019, 2019.

Bacon, D. P., Ahmad, N. N., Boybeyi, Z., Dunn, T. J., Hall, M. S., Lee, P. C. S., Sarma, R. A., Turner, M. D., Waight, K. T., Young, S. H., and Zack, J. W.: A dynamically adapting weather and dispersion model: The Operational Multiscale Environment Model with Grid Adaptivity (OMEGA), Monthly Weather Review, 128, 2044-2076, 10.1175/1520-0493(2000)128<2044:Adawad>2.0.Co;2, 2000.

Berger, M., and Oliger, J.: Adaptive mesh refinement for hyperbolic partial differential equations. Journal of Computational Physics, 53(3), 484-512. https://doi.org/10.1016/0021-9991(84)90073-1, 1984.

Bergström, S.: Development and application of a conceptual runoff model for Scandinavian catchments, SMHI Report RHO 7, Norrköping, 134 pp, 1976.

Beven, K. J., and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrological Sciences Bulletin, 24, 43-69, 10.1080/02626667909491834, 1979.

Binley, A., Beven, K., and Elgy, J.: A physically based model of heterogeneous hillslopes: 2. Effective hydraulic conductivities, Water Resources Research, 25, 1227-1233, 10.1029/WR025i006p01227, 1989.

Booij, M.: Determination and integration of appropriate spatial scales for river basin modelling, Hydrol. Process., 17, 2581-2598, 10.1002/hyp.1268, 2003.

Brunner, P., and Simmons, C. T.: HydroGeoSphere: A Fully Integrated, Physically Based Hydrological Model, 2011. http://doi.org/10.1111/j.1745-6584.2011.00882.x

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijlenhoet, R., Bennett, K. E., Pauwels, V. R. N., Cai, X., Wood, A. W., and Peters-Lidard, C. D.: The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism, Hydrol. Earth Syst. Sci., 21, 3427-3440, 10.5194/hess-21-3427-2017, 2017.

Cover, T. M. and Thomas, J. A.: Elements of information theory, John Wiley & Sons, New York, 1991.

Davison, J. H., Hwang, H.-T., Sudicky, E. A., Mallia, D. V., and Lin, J. C.: Full Coupling Between the Atmosphere, Surface, and Subsurface for Integrated Hydrologic Simulation, Journal of Advances in Modeling Earth Systems, 10, 43-53, 10.1002/2017ms001052, 2018.

Dehotin, J., and Braud, I.: Which spatial discretization for distributed hydrological models? Proposition of a methodology and illustration for medium to large-scale catchments, Hydrol. Earth Syst. Sci., 12, 769-796, 10.5194/hess-12-769-2008, 2008.

Dunger, V.: Entwicklung und Anwendung des Modells BOWAHALD zur Quantifizierung des Wasserhaushalts oberflächengesicherter Deponien und Halden. Habilitationsschrift an der Fakultät für Geowissenschaften, Geotechnik und Bergbau der TU Bergakademie Freiberg, DOI http://dx.doi.org/10.23689/fidgeo-668, 2006.

DVWK: Ermittlung der Verdunstung von Land- und Wasserflächen. – DVWK-Merkblätter 238, 135 S., 1996.

Fenicia, F., Kavetski, D., Savenije, H. H. G., and Pfister, L.: From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, Water Resources Research, 52, 954-989, 10.1002/2015wr017398, 2016.

Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., and Freer, J.: Catchment properties, function, and conceptual model representation: is there a correspondence?, Hydrol. Process., 28, 2451-2467, 10.1002/hyp.9726, 2014.

Flügel, W-A.: Hydrological Response Units (HRUs) as modeling entities for hydrological river basin simulation and their methodological potential for modeling complex environmental process systems. Die Erde 127,42–62, 1996.

Flügel, W.A.: Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. Hydrological Processes, 9(3-4), pp.423-436, 1995.

Gharari, S., Clark, M. P., Mizukami, N., Knoben, W. J. M., Wong, J. S., and Pietroniro, A.: Flexible vector-based spatial configurations in land models, Hydrol. Earth Syst. Sci. Discuss., 2020, 1-40, 10.5194/hess-2020-111, 2020.

Hundecha, Y., and Bardossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, Journal of Hydrology, 292, 281-295, 10.1016/j.jhydrol.2004.01.002, 2004.

Juilleret, J., Iffly, J-F., Hoffmann, L., and Hissler, C.: The potential of soil survey as a tool for surface geological mapping: a case study in a hydrological experimental catchment (Huewelerbach, Grand-Duchy of Luxembourg), Geologica Belgica [En ligne], volume 15, number 1-2, 36-41, 2012. URL : https://popups.uliege.be:443/1374-8505/index.php?id=3496.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resources Research, 42, 5, 10.1029/2005wr004362, 2006.

Kollet, S. J., Maxwell, R. M., Woodward, C. S., Smith, S., Vanderborght, J., Vereecken, H., and Simmer, C.: Proof of concept of regional scale hydrologic simulations at hydrologic resolution utilizing massively parallel computer resources, Water Resources Research, 46, 7, 10.1029/2009wr008730, 2010.

Kouwen, N., Soulis, E.D., Pietroniro, A., Donald, J. and Harrington, R.A.: Grouped response units for distributed hydrologic modeling. Journal of Water Resources Planning and Management, 119(3), pp.289-305, 1993.

Kuhn, H. W.: The Hungarian method for the assignment problem, Naval Research Logistics Quarterly, 2(1-2), 83-97, 1955.

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P. J., Li, F., Li, H., Lombardozzi, D., Riley, W. J., Sacks, W. J., Shi, M., Vertenstein, M., Wieder, W. R., Xu, C., Ali, A. A., Badger, A. M., Bisht, G., van den Broeke, M., Brunke, M. A., Burns, S. P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J. B., Flanner, M., Fox, A. M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L. R., Lipscomb, W. H., Lu, Y., Pandey, A., Pelletier, J. D., Perket, J., Randerson, J. T., Ricciuto, D. M., Sanderson, B. M., Slater, A., Subin, Z. M., Tang, J., Thomas, R. Q., Val Martin, M., and Zeng, X.: The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty, Journal of Advances in Modeling Earth Systems, 11, https://doi.org/10.1029/2018MS001583, 2019.

Liu, H., Tolson, B. A., Craig, J. R., and Shafii, M.: A priori discretization error metrics for distributed hydrologic modeling applications, Journal of Hydrology, 543, 873-891, https://doi.org/10.1016/j.jhydrol.2016.11.008, 2016.

Loritz, R., Gupta, H., Jackisch, C., Westhoff, M., Kleidon, A., Ehret, U., and Zehe, E.: On the dynamic nature of hydrological similarity, Hydrol. Earth Syst. Sci., 22, 3663-3684, 10.5194/hess-22-3663-2018, 2018.

5 Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., and Uijlenhoet, R.: Representation of spatial and temporal variability in large-domain hydrological models: case study for a mesoscale pre-Alpine basin, Hydrol. Earth Syst. Sci., 20, 2207-2226, 10.5194/hess-20-2207-2016, 2016.

Miller, C. T., Abhishek, C., and Farthing, M. W.: A spatially and temporally adaptive solution of Richards' equation, Advances in Water Resources, 29, 525-545, 10.1016/j.advwatres.2005.06.008, 2006.

10 Minkoff, S. E., and Kridler, N. M.: A comparison of adaptive time stepping methods for coupled flow and deformation modeling, Appl. Math. Model., 30, 993-1009, 10.1016/j.apm.2005.08.002, 2006.

Munkres, J.:Algorithms for the Assignment and Transportation Problems, Journal of the Society for Industrial and Applied Mathematics, 5(1), 32-38, 1957.

Nash, J. E., and Sutcliffe, J. V.: River flow forecasting through conceptual models part i -- a discussion of principles, Journal
15 of Hydrology, 10, 282-290, 1970.

Neuper, M., and Ehret, U.: Quantitative precipitation estimation with weather radar using a data- and information-based approach, Hydrol. Earth Syst. Sci., 23, 3711-3733, 10.5194/hess-23-3711-2019, 2019.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1.
20 Model description and evaluation with local-scale measurements, Journal of Geophysical Research: Atmospheres, 116, 10.1029/2010jd015139, 2011.

Pain, C. C., Piggott, M. D., Goddard, A. J. H., Fang, F., Gorman, G. J., Marshall, D. P., Eaton, M. D., Power, P. W., and de Oliveira, C. R. E.: Three-dimensional unstructured mesh ocean modelling, Ocean Model., 10, 5-33, 10.1016/j.ocemod.2004.07.005, 2005.

25 Penman, H. L.: Evaporation: An introductory survey. Netherlands J. Agric. Sci.1:9-29, 87-97,151-153, 1956.

Pettway, J. S., Schmidt, J. H., and Stagg, A. K.: Adaptive meshing in a mixed regime hydrologic simulation model, Computational Geosciences, 14, 665-674, 10.1007/s10596-010-9179-1, 2010.

Pfister, L., McDonnell, J. J., Wrede, S., Hlúbiková, D., Matgen, P., Fenicia, F., Ector, L., and Hoffmann, L.: The rivers are alive: on the potential for diatoms as a tracer of water source and hydrological connectivity, Hydrol. Process., 23, 2841-
30 2845, 10.1002/hyp.7426, 2009.

Pfister, L., Wagner, C., Vansuypeene, E., Drogue, G., and Hoffmann, L.: Atlas climatique du grand-duché de Luxembourg, Luxembourg: Musée National d'Histoire Naturelle, Société des naturalistes luxembourgeois, Centre de Recherche Public, edited by: Ries, C. and Lippmann, G., Administration des Services Techniques de l'Agriculture, Luxembourg, 2005.

Pfister, L., Humbert, J., and Hoffmann, L.: Recent trends in rainfall-runoff characteristics in the Alzette river basin,
35 Luxembourg, Climatic Change, 45, 323–337, https://doi.org/10.1023/A:1005567808533, 2000.

Reggiani, P., Sivapalan, M., and Majid Hassanizadeh, S.: A unifying framework for watershed thermodynamics: balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics, Advances in Water Resources, 22, 367-398, https://doi.org/10.1016/S0309-1708(98)00012-8, 1998.

Schulz, K., Seppelt, R., Zehe, E., Vogel, H. J., and Attinger, S.: Importance of spatial structures in advancing hydrological
40 sciences, Water Resources Research, 42, 4, 10.1029/2005wr004301, 2006.

Šimunek, J., Šejna, M., and Van Genuchten, M. T.: Hydrus 2-D software package for simulating the two-dimensional movement of water, heat, and multiple solutes in variably saturated media. Riverside, U.S. Salinity Laboratory, Agricultural Research Service ARS, U.S. Department of Agriculture USDA, 1999.

Singh, V. P.: Entropy Theory and its Application in Environmental and Water Engineering, John Wiley & Sons, Ltd, Chichester, Sussex, UK, 2013. ISBN: 978-1-119-97656-1.

Teyssier, R.: Cosmological hydrodynamics with adaptive mesh refinement. A new high resolution code called RAMSES, Astronomy and Astrophysics, 385, 337-364, 2002.

5 Trigo, I. F., Dacamara, C. C., Viterbo, P., Roujean, J.-L., Olesen, F., Barroso, C., Camacho-de-Coca, F., Carrer, D., Freitas, S. C., García-Haro, J., Geiger, B., Gellens-Meulenberghs, F., Ghilain, N., Meliá, J., Pessanha, L., Siljamo, N., and Arboleda, A.: The satellite application facility for land surface analysis, International Journal of Remote Sensing, 32, 2725-2744, 10.1080/01431161003743199, 2011.

Wood, E.F., Sivapalan, M., Beven, K. and Band, L.: Effects of spatial variability and scale with implications to hydrologic
10   modeling. Journal of Hydrology, 102(1-4), pp.29-47, 1988.

Zehe, E., and Sivapalan, M.: Threshold behaviour in hydrological systems as (human) geo-ecosystems: manifestations, controls, implications, Hydrology and Earth System Sciences, 13, 1273-1297, 10.5194/hess-13-1273-2009, 2009.

Zehe, E., Maurer, T., Ihringer, J. and Plate, E.: Modeling water flow and mass transport in a loess catchment. Physics and Chemistry of the Earth Part B-Hydrology Oceans and Atmosphere 26 (7-8), 487-507, 2001.

15 Zehe, E., Ehret, U., Pfister, L., Blume, T., Schröder, B., Westhoff, M., Jackisch, C., Schymanski, S. J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, J., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V., and Kleidon, A.: HESS Opinions: From response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments, Hydrology and Earth System Sciences, 18, 4635-4655, 10.5194/hess-18-4635-2014, 2014.

20