Dear Editor,

We would like to thank you and the four referees for handling, respectively reviewing, our manuscript (MS). In line with our responses to the reviewers we carefully revised our MS as outlined below:

- We restructured section 3 by splitting it into two sections. In section 3 we now exclusively introduce the model setups and in section 4 we explain the adaptive modeling in detail. We, furthermore, added a table to section 3 where we summarize the properties of the different model setups and added a new subsection where we explain how we test the model setups.
- We carefully revised the discussion of our MS. This means that we now clearly discuss the limitations of our approach regarding the selected test periods and the chosen similarity metric. We, furthermore, added and discuss the proposed literature of the reviewers and finally have improved the connection of our work to the land surface community.
- We added a second variable, namely Q, to group model states.
- We changed Fig. 8 and added the soil moisture distribution after 48 hrs to highlight that different model elements are again in a similar state after 48 hrs of no rainfall. The latter as well as another new figure in the supplement showing the Shannon entropy of the distributed *model b* underpins that the model states of individual hillslopes are not drifting apart in the summer season. This means that there is no reason to assume that the spatially adaptive *model c* would fail if we would run it for the entire summer season in a fully automated manner.
- We updated the KGE values of the reference model for rainfall event I and II in Tab. 2, which were incorrect due to a wrongly selected too long timing period. This does, however, not change the general pattern of the model results.
- Additionally we add:
    - the measured rainfall at the ground station "Roodt" to Fig. 2b
    - a table with the three components of the KGE to the supplement
    - the location of the distrometers and micro rainfall radars to the supplement
    - a new figure to the supplement showing how different binning widths influence the spatial resolution of the adaptive *model c*
    - added information about the numerical schemes used in CATFLOW
- We removed the appendix and added a supplement to further increase the readably and structure of the MS.

Enclosed we added the revised MS with track chances as well as our detailed responses to the four reviewer comments. Again we would like to thank you and the four referees for handling, respectively reviewing, our MS and look forward for your assessment.

Yours sincerely,

Ralf Loritz, on behalf of the Co-Authors

The role and value of distributed precipitation data for hydrological models

Ralf Loritz[1], Markus Hrachowitz[2], Malte Neuper[1], and Erwin Zehe[1]

[1] Karlsruhe Institute of Technology (KIT), Institute of Water and River Basin Management, Karlsruhe, Germany
[2] Delft University of Technology, Faculty of Civil Engineering and Geosciences, Delft, Netherlands

Correspondence to: Ralf Loritz (Ralf.Loritz@kit.edu)

Abstract

This study investigates the role and value of distributed rainfall for the runoff generation of a mesoscale catchment (20 km^2). We compare ~~the performance of~~ three hydrological ~~models at different periods~~model setups and show that a distributed model setup driven by distributed rainfall ~~yields~~ only ~~to improved~~improves the model performances during certain periods. These periods are dominated by convective summer storms that are typically characterized by higher ~~spatial and~~ spatio-temporal variabilities compared to stratiform precipitation events that dominate ~~the~~ rainfall generation in winter. Motivated by these findings we develop a spatially adaptive model that is capable ~~to~~of dynamically ~~adjust~~adjusting its spatial structure during ~~runtime to represent~~model execution. This spatially adaptive model allows representing the varying ~~importance~~relevance of distributed rainfall within a hydrological model without losing predictive performance compared to a ~~spatially~~fully distributed model. Our results highlight that spatially adaptive modeling ~~might be a promising way to better understand~~has the potential to reduce computational times as well as improve our understanding of the varying ~~relevance~~role and value of distributed ~~rainfall in~~ precipitation data for hydrological models ~~as well as reiterate that it might be one way to reduce computational times. They furthermore show that hydrological similarity concerning the runoff generation does not necessarily mean similarity for other dynamic variables such as the distribution of soil moisture~~.

1 Introduction

"*How important are spatial patterns of precipitation for the runoff generation at the catchment scale?*" – This is a key question for the application of hydrological models that has been addressed in several studies over the ~~last three~~past decades (e.g. *Beven and Hornberger, 1982; Smith et al., 2004; Lobligeois et al., 2014*). A frequently drawn conclusion is that semi-distributed or even lumped models driven by a single precipitation time series often outperform distributed models with respect to their ability to reproduce observed streamflow at the outlet of a catchment (e.g. *Das et al., 2008*). Although the generality of such findings ~~are~~is surely ~~constrained~~limited by the fact that distributed models have more parameters that need to be identified, which makes model calibration much more challenging (*Beven and Binley, 1992; Huang et al., 2019*), they highlight the ability of the hydrological system to dissipate spatial gradients efficiently ~~(e.g. *Obled et al., 1994; Berkowitz and Zehe, 2020*)~~(e.g. *Obled et al., 1994*). This is the case as hydrological processes are strongly dissipative but exhibit despite the non-linearity of surface and subsurface flow processes no chaotic behavior (*Berkowitz and Zehe, 2020*).

In contrast to the above-mentioned finding that hydrological systems can efficiently dissipate spatial gradients, several other studies showed that information about the spatial variability of precipitation can significantly improve the predictive performance of hydrological models. For instance, *Euser et al. (2015)* highlighted that distributed models driven by distributed rainfall could reproduce the observed hydrograph of a 1600 km^2 large catchment in Belgium with higher accuracy compared to spatially lumped model structures. Furthermore, *Woods and Sivapalan (1999)* showed that the interplay between spatial patterns of rainfall and soil saturation can substantially impact the runoff generation of a catchment when they analyzed the dependence of average runoff rates ~~in dependence of~~ on the spatial and temporal variability of the meteorological forcing and the catchment state. The relevance of these spatial patterns is thereby particularly high if the system is close to a threshold where different localized preferential flow processes start dominating (e.g. cracking soils: drying of soil; macropores: occurrence of earthworms) as discussed by *Zehe et al. (2007)*. Spatial averaging of the system state or the meteorological forcing can hence lead to a misrepresentation of relevant spatial patterns, especially at more extreme conditions.

Given the partly contradictory findings present in the literature, it appears reasonable to assume that the relevance of distributed rainfall is changing dynamically over time and depends on the interplay of the prevailing i) system state (e.g. catchment wetness), ii) ~~on~~ the system functional structure, determined by patterns of topography, land-use, and geology, as well as ~~on~~ iii) the strength and spatial organization of the rainfall forcing. In consequence, it seems furthermore rational to hypothesize that also hydrological models should dynamically adapt their spatial structure to the prevailing context thereby reflecting the inherently dynamic nature of hydrological similarity (*Loritz et al., 2018*).

The idea that hydrological models should dynamically allocate their spatial resolution, as well as the associated representation of natural heterogeneity in time, is motivated by our previous work (*Loritz et al. 2018*). In ~~this~~that study, we highlighted that simulations of a distributed model consisting of 105 independent hillslopes were highly redundant to reproduce discharge or catchment storage changes of a mesoscale catchment within one hydrological year. Based on the Shannon entropy as metric we identified periods where a rather small number of representative hillslopes was sufficient because most of them functioned largely similar within the chosen margin of error. However, during other periods up to 32 independent representations of hillslopes were required, which underlines that spatial variability of system properties, such as surface topography or soil types among the hillslopes can exert a stronger influence on the runoff generation at certain times as expected given the findings reported by other studies conducted in the same research environment (e.g. *Fenicia et al., 2016; Loritz et al., 2017*). It can, therefore, be argued that also distributed rainfall and corresponding distributed model structures are only ~~of higher relevance~~important during specific periods, while during other periods a compressed, spatially aggregated model structure may be sufficient. An implementation of such an adaptive spatial model resolution would ensure an appropriate spatial model complexity, defined based on the least amount of details about the system structure (e.g. the variability of topographic gradients) and catchment states that are sufficient to capture the relevant interactions with the spatial pattern of precipitation. Yet it would be as parsimonious as possible to avoid redundant computations, which again could be used to minimize computational costs (~~*Clark et al., 2017*)~~*Clark et al., 2017*).

Moving to the event time scale instead of running continuous simulations is surely one-way to achieve such a dynamical allocation of the model space. This would entail running a set of models that differ with respect to their resolutions in space and time depending on the prevailing structure of the meteorological forcing and wetness state of a landscape. Yet, this introduces multiple new problems, for instance, how to infer the initial conditions of a catchment prior to a rainfall event given the degrees of freedom distributed models can offer (*Beven, 2001*). The latter is of considerable importance particularly during extremes resulting from high-intensity rainfall-runoff events, which can be strongly sensitive to the actual state of the system such as the spatial patterns of macropores (*Zehe et al., 2005*) or of the antecedent soil water content (*Zehe and Blöschl, 2004*).

A different avenue to implement a dynamically changing model resolution is adaptive clustering, as recently demonstrated for a spatially distributed conceptual (top-down) model by *Ehret et al. (2020*). This concept allows for continuous hydrological simulations, which use a higher spatial model resolution only at those time steps when it is necessary. The idea behind adaptive clustering is similar to adaptive time-stepping (e.g. *Minkoff and Kridler, 2006*). However, instead of reducing the time steps during times when large gradients prevail, adaptive clustering increases or decreases the number of independent spatial model elements during times of low or high functional diversity. The general concept behind adaptive clustering is thereby not entirely new to environmental science and is already used for instance in hydrogeology under the term adaptive mesh ~~here~~ with the main focus to increase the resolution of gradients during times of high dynamics ~~(~~by increasing or decreasing the number of nodes (grid elements) in a model (*Berger and Oliger, 1984*). The main difference between the adaptive mesh and adaptive clustering approach is that instead of adjusting the ~~actual~~spatial resolution of the numerical~~ model~~ grid during runtime adaptive clustering changes the number of hydrological response units (HRU) that are used (needed) to represent a catchment. This implies that also the degree of spatial heterogeneity of the catchment state (e.g. the wetness state, energy state, etc.) that is covered by the model is dynamically changing.

While the idea of adaptive clustering is promising as it allows a minimum adequate representations of the spatial variability of a hydrological landscape, it has to our knowledge so far only been tested within a simple top-down model (~~*Ehret et al. (2020)).*~~*Ehret et al., 2020).* It is thus of interest whether such a dynamic clustering is also feasible when using a physically based (bottom-up) model particularly as these models were specifically introduced to explore how distributed system characteristics and driving gradients control hydrological dynamics *(Freeze and Harlan, 1969)*. Here we will hence test and develop an adaptive clustering approach using straightforward physical reasoning and implement it into a distributed bottom-up model. The ~~underlying~~overarching objective of this study is thus to exploit the value of adaptive clustering as a tool to better understand the temporal relevance of distributed precipitation for the runoff generation of a meso-scale catchment and, as by-product, to reiterate that adaptive clustering could potentially be used to reduce computational times as already discussed in detail by *Ehret et al. (2020)*. High computational times are thereby still one of the many reasons why bottom-up are rarely used on larger scales in an spatial explicit manner (*Clark et al., 2017*). For instance, *Hopp and McDonnell (2009)* used the HYDRUS 3D model (e.g. latest version of Hydrus: *Simunek et al., 2016*) and reported computational times ranging from 10 min up to 11 hrs when they simulated water fluxes and state variables at the Panola hillslope (area = 0.001250 km^2 (25 m x 50 m); maximal soil depths = 4 m) for a simulation time of 15 days~~.~~ by changing slope angles, soil depths, storm sizes and bedrock permeability. A meaningful application of bottom-up models at relevant management scales (around 250 km^2 in south Germany e.g. *Loritz, 2019*), without a violation of important physical constraints (e.g. 10^{-2} - 10^{1} m maximum vertical grid size for the Richards equation; *Or et al., 2015; Vogel and Ippisch, 2008*), would thus imply long computational times. This again strongly limits the number of feasible model runs to examine, for instance, different parameter sets (*Beven and Freer, 2001*).

In this study, we therefore test the specific hypothesis ~~if~~that adaptive clustering is a feasible approach to represent the spatial variability of rainfall in a hydrological bottom-up model at the lowest sufficient level of detail without losing predictive performance compared to a fully distributed model. We test this hypothesis by introducing a clustering approach at the example of the model CATFLOW~~, which is~~ applied to the 19.4 km^2 ~~large~~ Colpach catchment using a gridded radar-based quantitative rainfall estimate by addressing the two following research questions:

1. Does the model performance of a spatially aggregated model improve ~~if~~when it is distributed in space and driven by distributed rainfall?
2. Can adaptive clustering be used to distribute a bottom-up model in space that it is ~~capable~~able to represent relevant spatial differences in the system state and precipitation forcing at the least sufficient resolution to avoid being highly redundant as a fully distributed model?

2 Study area, hydrological model and meteorological data

2.1 The Colpach catchment

The 19.4 km^2 Colpach catchment is located in northern Luxembourg and is a headwater catchment of the 256 km^2 large Attert experimental basin (Fig. 1). The prevailing geology of both catchments are Devonian schists of the Ardennes massif which are characterized by shallow, coarse-grained, and highly permeable soils (> 1 m; e.g. *Jackisch et al., 2017; Juilleret et al., 2011*). The steep hills of the Colpach are primarily forested and the elevation of the Colpach ranges from 265 to 512 m a.s.l.. Annual runoff coefficients varied around 50 % ± 7 % for the 2011-2017 period. Precipitation is rather evenly distributed across the seasons (vegetation and winter season), while the runoff generation has a distinct seasonal pattern as around 80 % of the annual discharge is being released between October and March (*Seibert et al., 2017*). The Colpach and its sub-catchments (e.g. Weierbach) have been used as study area in a series of scientific publications. We refer here to *Pfister et al. (2018)*, *Jackisch (2015)* or *Loritz et al. (2017)* for more detailed system description (mean annual precip: 900 – 1000 mm yr^{-1}; mean annual evapotranspiration: 450 – 550 mm yr^{-1}; mean annual discharge: 450 – 550 mm yr^{-1}; land-use: 65 % forest; 23 % agriculture; 2 % others; mean annual temperature: 9.1 °C).

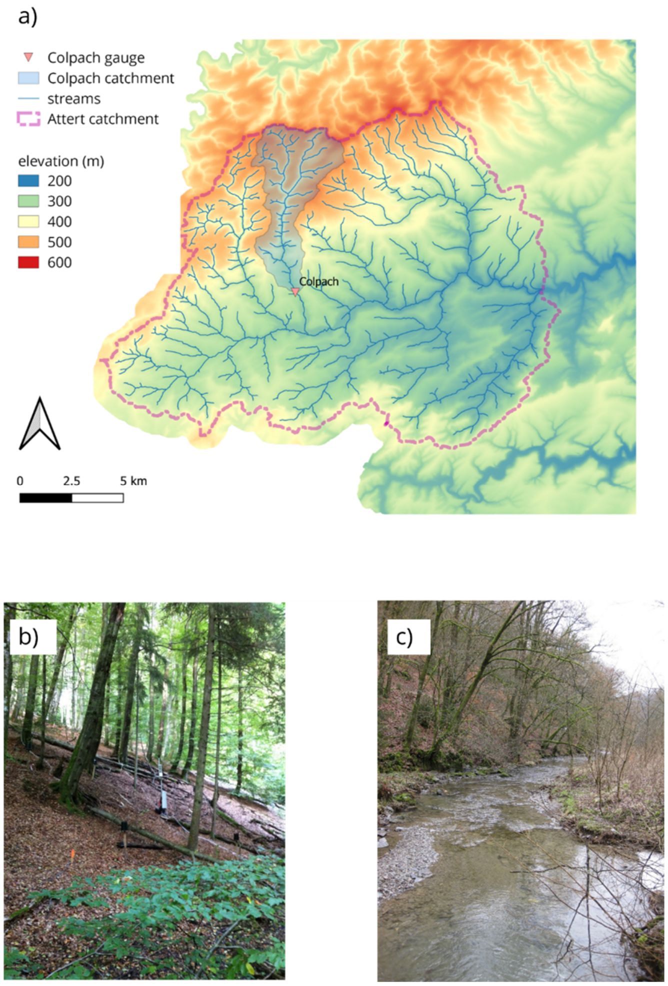


Figure 1. a) map of the Colpach catchment (location northern Luxembourg), b) picture of a typical forested hillslope within the Colpach catchment, c) the Colpach river around 4 km north of the gauging station.

2.2 The CATFLOW model

The key elements of the CATFLOW Model (*Maurer, 1997; Zehe et al., 2001*) are 2d hillslopes which are discretized along a 2-dimensional cross-section using curvilinear orthogonal coordinates. Evapotranspiration is represented using an advanced SVAT (soil–vegetation–atmosphere transfer) approach based on the Penman-Monteith equation, which accounts for tabulated vegetation dynamics, albedo as a function of soil moisture, and the impact of local topography on wind speed and radiation. ~~Soil water dynamics are simulated based on the Darcy-Richards equation and surface runoff is represented by a diffusion wave approximation of the Saint Venant equations using an adaptive time stepping.~~Soil water dynamics are simulated based on the Darcy-Richards equation (solved implicitly, modified Picard iteration; *Celia et al., 1990*) and surface runoff is represented by a diffusion wave approximation of the Saint-Venant equations using adaptive time stepping (solved explicitly, Euler forward starting downslope). Vertical and lateral preferential flow paths are represented as connected pathways containing an artificial porous medium with high hydraulic conductivity and very low retention. The hillslope module is designed to simulate infiltration excess runoff, saturation excess runoff, re-infiltration of surface runoff, lateral water flow in the subsurface, return flow, but cannot handle snowfall or snow accumulation. The latter means that CATFLOW should not be applied if snow is a ~~dominated~~dominant control, which is not the case in the Colpach catchment. The model core is written entirely in FORTRAN77 and the individual hillslopes can be run in parallel on different CPUs to assure low computation times and high performance of the numerical scheme. Up to date model descriptions can be found in *Wienhöfer and Zehe (2014)* or in *Loritz et al. (2017)*.

2.3 ~~Model~~Meteorological forcing and observed discharge

Meteorological input data used here are recorded at a temporal resolution of 1 hr at two official meteorological stations by the "*Administration des Services Techniques de l'Agriculture Luxembourg*" at the locations "*Roodt*" and "*Useldange*". The meteorological station "*Roodt*" measures rainfall within the catchment border (Fig. 2 a) and provided the precipitation input to the model of *Loritz et al. (2017)*. The second station "*Useldange*" is located outside the catchment around 8 km west of the Colpach outlet measures air temperature, relative humidity, wind speed, and global radiation. These data are used as meteorological input (except for precipitation) in all model setups in this study. In other words, this means that all model setups in this study are forced by identical meteorological inputs except for the precipitation data (see section 3.1). Therefore, we cannot account for variations of the wind speed or the temperature within the Colpach catchment. A detailed description and analysis of the meteorological data can be found in *Loritz et al. (2017)*.

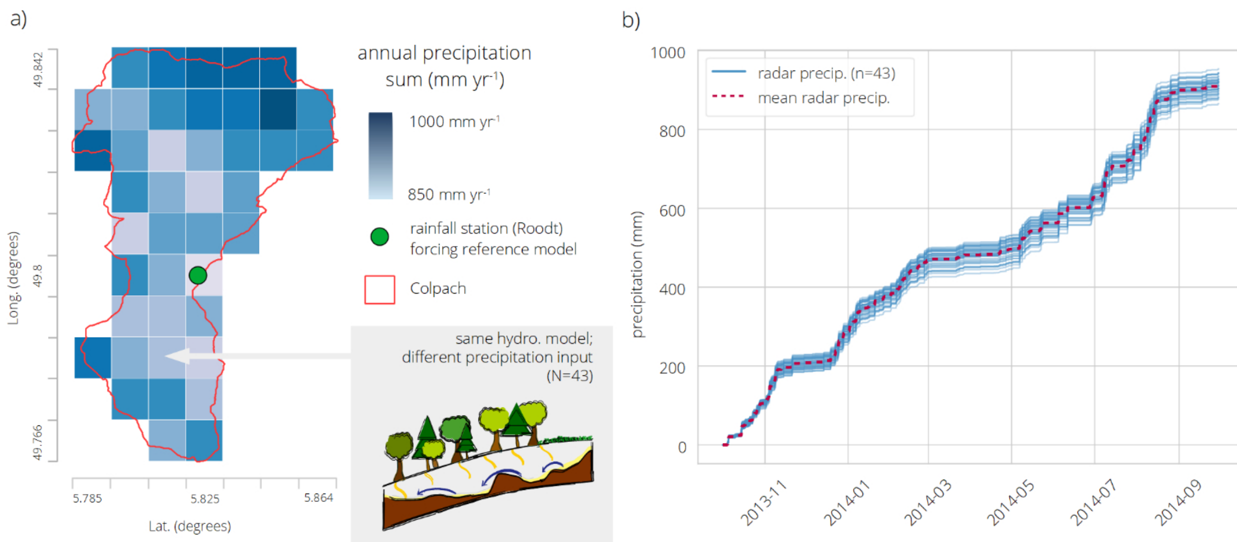
Q~~uality checked discharge~~Discharge observations of the Colpach are provided by the Luxembourg Institute of Science and Technology (LIST) in a 15 min temporal resolution for the hydrological year 2013/14. The data was aggregated to an hourly temporal resolution and ~~transformed~~ to specific discharge given the catchment area of 19.4 km^2.

2.4 Spatially resolved precipitation data

Besides the precipitation data from the ground station located in "*Roodt*", we use a gridded quantitative precipitation estimate, which ~~merges~~is a merged product of two weather ~~radar with~~radars, rain ~~gauge~~gauges, micro rainfall radars and disdrometer observations (Location of the ground measurements in the supplement and in more detail in *Neuper and Ehret, 2019*). The two used radar stations are located 40 to 70 km, respectively 24 to 44 km, away from the ~~study site~~borders of the Attert catchment (Neuheilenbach; Germany, Wideumont, Belgium) and are operated by the German Weather Service (DWD) as well as by the Royal Meteorological Institute of Belgium (RMI). Both distances are within a range that the data can be used at a high-resolution of 1x1 km^2 as the signal is neither degraded by beam spreading nor impacted by partial blindness through cone of silence issues ~~(e.g. *Neuper and Ehret (2019)*). The raw data, 10 min reflectivity data from single pol C-Band Doppler radar, were aggregated to hourly averages as well as~~(*Neuper and Ehret, 2019*). The raw data, 10 min reflectivity data from single polarimetric C-Band Doppler radar, were aggregated to hourly averages and filtered by static, Doppler clutter filters, and bright-band correction following *Hannesen (1998)*. Second trip echoes and obvious anomalous propagation echoes were manually removed, and the corrected data were used to create a pseudo plan position indicator data set at 1500 m above the ground. A more detailed description of how the reflectivity data was transformed to rainfall data, calibrated as well as validated against rain gauges ~~and disdrometers can be found in the appendix.~~, micro rainfall radar and disdrometers can be found in the supplement and in *Neuper and Ehret, (2019).*

The chosen precipitation time series starts on the 1st of October 2013 and ends on the 30th of September 2014. 42 grid cells (1 x 1 km^2) of the precipitation field intersect with more than 50 % of their area with the Colpach catchment and are used in this study (Fig. 2 a). The weather radar measured an area-weighted mean of around 900 mm yr^{-1} in the Colpach catchment for the selected period. This is in accordance with the reported climatic averages

(900 - 1000 mm yr^{-1}) of this region (*Pfister et al., 2017*). The maximum hourly precipitation difference between the grid cells in the study period is 14 mm hr^{-1} (August 2014) and the maximum annual precipitation difference between the grid cells is 95 mm yr^{-1} (Fig. 2 b). Temporally, the precipitation ~~is~~distributes evenly~~ distributed~~ over the year with around 50 % of rainfall in winter and 50 % of rainfall in summer with a short dry spell from mid-March to the end of April. There is a weak correlation between the mean elevation of the grid cells and the annual precipitation sums of 0.43. This implies that precipitation tends to be slightly higher in the northern parts of the catchment that are also characterized by higher altitudes (Fig. 2 a). The measured precipitation time series from the ground station located in "*Roodt*" differs from the mean precipitation of the spatial rainfall field about 30 mm yr^{-1} and around 60 mm yr^{-1} from the exact location in the precipitation grid measured by the weather radar with a tendency of higher rainfall values in especially in the winter season. Why exactly the precipitation observations of the ground station in Roodt differ in this magnitude from the merged product of the weather radar is an interesting research question, however, not the scope of this study.



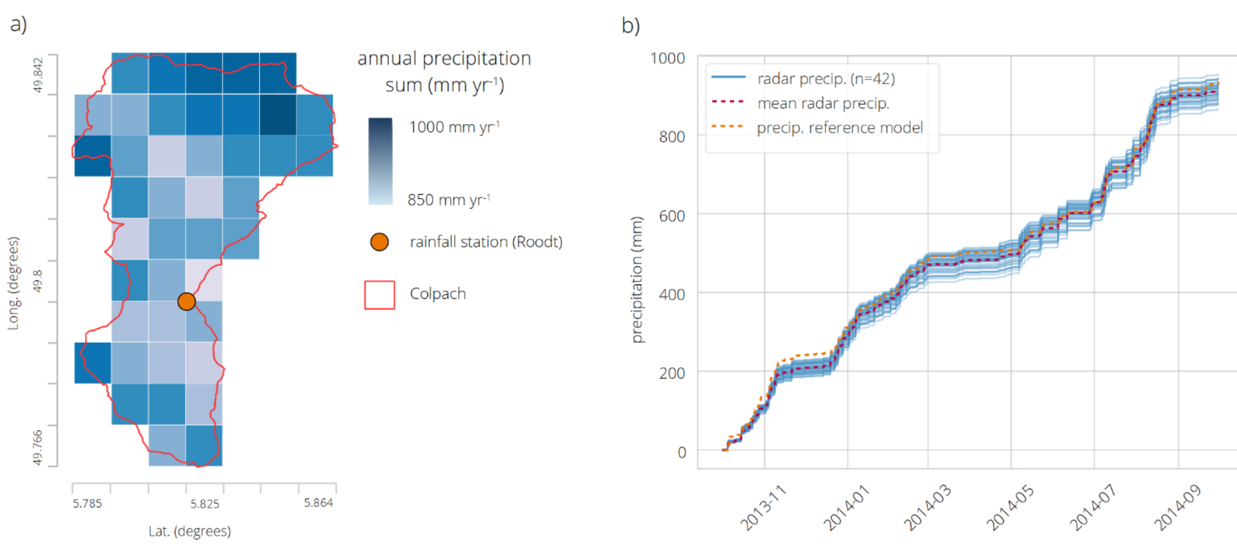


Figure 2. ~~a) annual~~Annual sums of the gridded precipitation field over the Colpach catchment for the hydrological year 2013/14 as well as the location of the rainfall station "Roodt" which is used as precipitation input for the *reference model* (spatial resolution: 1 km^2; coord. system WGS84~~), b) cumulated~~) (panel a). Cumulated precipitation for each grid cell

for the hydrological year 2013/14 of the precipitation field (blue lines~~) and~~), the corresponding mean of the precipitation field (dashed red line) and the precipitation data from the station in Roodt (panel b, dashed orange lines).

3. Modeling approach

In the following ~~section~~sections 3.1 to 3.3, we ~~give~~ introduce three non-adaptive model setups (*reference model*, *model a* ~~short introduction to~~ and *model b*) and a spatially adaptive model setup (*model c*). Details how the model setups are tested are provided in section 3.4 and 3.5. A summary of the different model setups ~~we use in this study and refer to the corresponding subsection for more detailed descriptions of each setup~~can be found in Tab. 1.

~~*Reference model*~~

~~The spatially aggregated *reference model* (section 3.1) was designed and intensively tested in the Colpach catchment in a previous study (*Loritz et al., 2017*). This model serves as benchmark here to a) evaluate the other three models and b) provides the structural basis for them. Moreover, are the model deficits to simulate streamflow in the summer months of the *reference model* discussed in *Loritz et al. (2017)* one of the main motivations of this study (see section 3.2) apart from the finding of *Loritz et al. (2018)* that a suitable model structure needs to adapt its resolution in time.~~

~~*Model a*~~

~~*Model a* (section 3.2) is identical to the *reference model* and hence also spatially aggregated. The only difference between the models is that it is driven by different precipitation data. This precipitation data is the area-weighted mean of the spatially resolved precipitation product described in section 2.4 and measured by a weather radar. The main reason for running *model a* is to exclude that already the quantitative differences between the precipitation data measured at the ground station "*Roodt*" and by the weather rainfall data result in a performance increase and not the spatial variability of the rainfall field.~~

~~*Model b*~~

~~The third model (*model b*; section 3.2) is a distributed version of the *reference model*. *Model b* is thereby distributed based on the resolution of the spatially resolved precipitation data and was designed to examine the role of distributed rainfall on the runoff generation in the Colpach catchment. It represents the Colpach with 42 spatial grids (1 x 1 km²). In each of these grids, we run a model similar to the *reference model*, however, driven with the specific precipitation data measured at this location by the weather radar.~~

~~*Model c*~~

~~Other than the three above mentioned non adaptive models (*reference model*, *model a*, *model b*) we develop a third, spatially adaptive, model (*model c*; section 3.3). This model is capable to dynamically adapt its spatial model structure in time. To dynamically allocate its structure, it uses the spatial variability and the strength of the rainfall forcing as well as its fingerprint the catchment (model) state. The main goal is to show that we can achieve similar simulation results compared to *model b*, however, with a coarser dynamically adapting spatial model structure. We test this model at two selected rainfall-runoff events.~~

~~3.1 Non-adaptive models – The reference model of *Loritz et al. (2017)*~~

3.1 The reference model of Loritz et al. (2017)

All ~~simulations~~model setups in this study are based on a spatially aggregated model structure (*reference model*), developed and extensively tested in the Colpach catchment in a previous study (*Loritz et al., 2017*). The general idea behind the proposed model concept (*representative hillslope*) is that a single bottom-up hillslope model reflects a meaningful compromise between classical top-down and bottom-up models ~~(*Hrachowitz and Clark, 2017; Loritz, 2019*)~~(*Hrachowitz and Clark, 2017; Loritz, 2019*). This is because a representative hillslope resolves the ~~case as it~~effective gradient and resistance controlling water storage and release and allows that macroscopic model parameters can ~~still~~ be derived from available point measurements. The parameters of the model of *Loritz et al. (2017)* were ~~hence~~, for the most part, derived directly from a large amount of field data, and the model was only ~~afterward~~afterwards manually fine-tuned by further exclusively adjusting the spatial macropore density within a few trial and error runs to simulate the seasonal water balance of the Colpach catchment. The model simulations were tested against hourly discharge observations on an annual and seasonal time scale ~~(as well as~~, against discharge observations from a sub-basin of the Colpach~~) and~~, in a different hydrological year, against hourly soil moisture observations (38 sensors in 10 and 50 cm depth), and hourly normalized sap flow velocities (proxy for transpiration; 30 sensors). The developed model structure agreed well with the dynamics of the observables and showed higher model performances as reported in other studies working with different top-down model setups in the same environment (*Wrede et al., 2015*).

To summarize, the *reference model* serves as benchmark here to a) evaluate the other models and b) provides the structural basis for them. None of the other model setups are further calibrated or manually tuned and the only difference between the different model setups is the precipitation data they are driven with and respectively the model resolution. For further details how the *reference model* was setup and tuned we refer to the study of *Loritz et al. (2017).*

3.2 Non-adaptive models ~~– Model~~ a and b

Despite the acceptable annual model performance of the *reference model*, it showed deficits ~~to simulate the~~in simulated runoff response to a series of summer rainfall-runoff events. As discussed in *Loritz et al. (2017)*, one possible explanation for the unsatisfying performance is that summer precipitation in the Colpach catchment is mainly driven by convective atmospheric conditions. These convective precipitation events are characterized by a much smaller spatial extent as well as by higher rainfall intensities compared to the stratiform and frontal precipitation events that dominate during winter (*Neuper and Ehret, 2019*). The insufficient model performance in summer could therefore likely be a consequence of the larger spatial gradients of the rainfall field compared to the winter season that cannot be accounted for in the original model of *Loritz et al. (2017)*. In other words, this entails that a hydrological model, distributed at a sufficiently high spatial resolution, is required to capture the spatial variability of the precipitation field to ~~satisfactorily simulate~~achieve improved simulations of the runoff generation of the Colpach. One goal of this study (first research question) is hence to test the hypothesis whether the performance deficiencies of the representative hillslope model ~~(~~, the *reference model*~~)~~, in summer are mainly caused by the inability of the setup to account for the spatial ~~gradients~~variability of the precipitation field, rather than a result of important structural differences (e.g. soil, land-use, topography) within the Colpach catchment.

To address the first research questions of this study: "*Does the model performance of a spatially aggregated model improve if it is distributed in space and driven by distributed rainfall*" we analyze simulations of two alternative model setups (*model a* and *model b*) additional to the *reference model* from *Loritz et al. (2017~~):~~)*. Both model setups are described in detail below.

3.2.1 Spatially aggregated *model a*

Model a is structurally identical to the *reference model*, however, it is driven by the area-weighted mean of the spatially resolved precipitation data described in section 2.4 ~~(Fig. 2 b).~~ and plotted in Fig. 2 b. We used the area-weighted mean as not all raster cells of the distributed precipitation data are entirely within the borders of the Colpach catchment. This means that that the weight of a grid cells that are not entirely located within the catchment is reduced when we calculated the average according to the percentage areal overlap.

We added *model a* to test if the performance difference between the *reference model* and our distributed *model b* is merely a result of quantitative differences between the different precipitation products measured either by a single ground station or by a weather radar in combination with different ground stations.

3.2.2 Spatially distributed *model b*

M*odel b* is a spatially distributed version of the *reference model*. ~~This means that~~More specifically, here all model parameters of the representative hillslope (*reference model*), as well as all other meteorological variables such as temperature or wind speed, are ~~similar and the only two differences between~~identical to the *reference model* ~~and *model* b is that~~. *However,* *model b* is spatially distributed as well as driven by ~~different~~distributed rainfall data. ~~*Model b*~~This model set-up is ~~thereby~~ distributed on the spatial resolution of the precipitation field similarity as done for instance by *Prenner et al. (2018)* and not following the traditional spatial discretization strategy of CATFOW based on a fixed number of hillslopes, inferred from surface topography or land-use. ~~We justify this assumption based on the model validation in *Loritz et al. 2017* and on a study conducted in the same research environment (*Loritz et al., 2019*)~~*Model b* thus represents the Colpach with 42 spatial grids (1 x 1 km^2). In each of these grids, we run a model identical to the *reference model*, however, driven with the specific precipitation data measured at this location.

We justify this assumption based on the model validation in *Loritz et al. 2017* and on a study conducted in the same research environment (*Loritz et al., 2019*) where we showed that different sub-basins of the Attert ~~basin~~catchment (the Colpach is a headwater catchment of the Attert catchment) have similar specific discharges as long as they are located in the same geological setting and are driven by a similar meteorological forcing (see also section ~~3.3~~4.2).

3.~~2~~3 Spatially adaptive model c

To address the second and main research question of this study: "*Can adaptive clustering be used to distribute a bottom-up model in space that it is able to represent relevant spatial differences in the system state and precipitation forcing at the least sufficient resolution to avoid being highly redundant as a fully distributed model?*" we develop a third adaptive model setup (*model c*). This spatially adaptive model setup is based on the distributed *model b*, however, is able to dynamically adjust its spatial structure in time based on the precipitation forcing, as detailed in the sections 4.1 to 4.3.

3.4 Model ~~analysis~~testing – non adaptive models a and b

We analyze the simulation performances of *model a* ~~and b~~*(spatially aggregated)* and *b (spatially distributed)* by calculating the Kling-Gupta efficiency (KGE; *Kling and Gupta, 2009*) as well as its three components (see supplement) between the hourly discharge simulations of the individual ~~models~~model setups against hourly observed discharge at different time scales (annual, seasonal, event scale). *Model a* and *b* are ~~hence~~ run for the hydrological year 2013-2014 with hourly printout times and differ only concerning the precipitation data they are driven with:

- *Model a*: driven by an area-weighted mean of the spatially resolved precipitation data.
- *Model b*: driven by 42 precipitation time series each reflecting a grid cell of the precipitation field shown in Fig. 2.

To be able to compare the discharge of the spatially aggregated *model a* and the distributed *model b* with the observed discharge of the Colpach catchment and to account for the routing of the water from a specific location to the outlet, we added a simple lag function acting as channel network. ~~The latter is based on the average distance of each grid cell to the outlet of the Colpach assuming a constant flow velocity of 1 m s^{-1}. For *model a*, we simply average all~~The latter is based on the average flow length along the surface topography of each precipitation grid to the outlet of the catchment assuming a constant flow velocity of 1 m s^{-1} (e.g. *Leopold, 1953*). The flow length of each grid is estimated based on 10 m resolved digital elevation model. For the spatially aggregated *model a*, we average all flow distances to the outlet and shift the single discharge simulation accordingly.

3.~~3 Spatially~~5 Model testing – adaptive model ~~– Model~~ c

~~To address the second research question of this study: "*Can adaptive clustering be used to distribute a bottom-up model in space that it is capable to represent relevant spatial differences in the system state and precipitation forcing at the least sufficient resolution to avoid being highly redundant as a fully distributed model?*" we develop a third model setup (*model c*). This spatially adaptive model setup is based on the distributed *model b*, however, is capable to dynamically adjust its spatial structure in time, as detailed in section 3.2.1 to 3.2.3. The underlying adaptive clustering approach is based on straightforward physical arguments on how the spatial and temporal patterns of rainfall control the spatial pattern of the wetness state of a structural similar catchment. By structural similar, we mean that time-invariant properties of the catchment (time-invariant on the scale we are working on) like geology, topography or land use that constrain the state space of a catchment are similarly distributed within potential hydrological sub units of our catchment (e.g. sub basins or hillslopes; see also section 3.2.2).~~ We ~~discuss~~test the spatially adaptive *model c* for two selected rainfall-runoff events, which are characterized by distinctly different precipitation properties. ~~By that, we examine the dynamic relationship~~We chose event I as it has the highest intensity and third-highest spatial variability in summer and event II because it is the event with the longest continuous precipitation in the time series. Both events where picked to represent the spectrum of rainfall events in the summer season. We focus exclusively on the summer season as the distributed *model b* only outperforms the *reference model* in this period indicating that spatially distributed rainfall provides no performance relevant information during winter.

The main goal of the model testing of the spatially adaptive *model c* is to show that we can achieve similar simulation results compared to the fully distributed *model b*, however, with a reduced number of hillslopes (coarser resolution). We therefore calculate not only the KGE between the ~~spatio-temporal patterns of the rainfall forcing~~simulated discharge of *model c* with the observed discharge at event I and ~~its fingerprint the catchment state~~II but also the KGE between the simulated discharge of *model c and* ~~show how they can be represented in a model. Full~~the simulated discharge of *model b* on an hourly basis. A full automation of the proposed adaptive clustering approach and a test on a longer time scale is~~, however,~~ beyond the scope of this study~~. The latter would provide only little more scientific inside (besides being technically challenging) how the variability of rainfall influences the state of a catchment~~ and ~~how this phenomenon can be used to dynamically allocate a model structure in time~~we point towards the study of *Ehret et al. (2020)* who shows the potential of adaptive clustering using a conceptual model also for longer periods.

3~~.3.1~~

Table 1 Summary of the four different model setups

	reference model	spatially aggregated *model a*	spatially distributed *model b*	spatially adaptive *model c*
spatially distributed:	no	no	yes	yes
precipitation forcing:	single precip. time series (ground station)	single precip. time series (weather radar)	distributed precip. (weather radar)	binned distri. precip. (weather radar)
spatially adaptive:	no	no	no	yes
testing period:	hydro. year 13/14, summer season, rainfall event I and II	hydro. year 13/14, summer season, rainfall event I and II	hydro. year 13/14, summer season, rainfall event I and II	rainfall event I and II

4. Spatially adaptive modeling

Spatially adaptive modeling or adaptive clustering is an approach to dynamically adjust the spatial structure of a hydrological model in time offering the possibility to reduce computational times as well as to find an appropriate, time-~~variant~~varying spatial model resolution (*Ehret et al~~.~~., 2020*). The basic idea of adaptive clustering has been motivated within the work of *Zehe et al. (2014)* who stated that functional similarity in a catchment (or in a model) can only emerge if different sub-units are structurally similar (e.g. topography, geology, land-use, etc.), are driven by a similar forcing and are at a similar state. The latter implies that the concept of hydrological similarity, ~~which is~~ frequently used as the basis to discretize a catchment in space ~~(e.g. *Wagener et al., 2007*), cannot be time-invariant but needs to dynamically change in time as corroborated by *Loritz et al. (2018)*. This is the case as the relevance and interaction of different spatial patterns of the catchment structure, state and forcing also vary in time~~(e.g. *Sivapalan et al., 1987*), cannot be time-invariant but needs to dynamically change in time (*Loritz et al., 2018*). This is the case as the relevance of different spatial controls like the topography or pedology of a catchment depend on the prevailing state and forcing (*Woods and Sivapalan, 1999*). A suitable discretization of a catchment into similar functional units needs hence to be time~~-~~ variant and one way to achieve such a dynamic model resolution is spatially adaptive modeling.

Implementing adaptive clustering into a distributed model requires specific decision thresholds that define whether spatial differences in the structure, forcing and state of ~~potential~~ sub-units (e.g. hillslopes, sub-basins, etc.) are so large, that they need a distributed representation. This entails that if differences between the structure, forcing, or state of two or more distributed model elements (here gridded models) are below these thresholds they are by definition similar which ~~means~~entails that they can represent each other's hydrological function. The ~~entire idea~~concept that certain spatial model elements can represent other model elements ~~and hence other areas of a catchment is not~~by no means new and has been used frequently in Hydrology since at least *Sivapalan et al. (1987)* where they introduced the concept of representative elementary areas. The~~ main~~ novelty of adaptive clustering is that hydrological similar model elements are dynamically grouped and split in the runtime of the model instead of running a constant number of ~~functional similar~~model elements for the entire simulation period (*Ehret et al., 2020*).

3~~.3.2~~4.1 Spatially adaptive modeling – similarity assumption

Identifying periods when a ~~given ~~model element~~ or hillslope~~ can represent another one because it functions hydrologically similar is the main challenge of adaptive clustering. ~~In~~For this ~~study~~purpose, we subdivide the precipitation field~~,~~ and the model states at each time step into equally distant bins (bins = groups) and ~~define~~classify those as similar ~~if different precipitation grid cells (forcing) or different gridded hillslope models (states) ~~that occupy the same bin. ~~This implies that ~~If two or more observations or models are hence in the same bin they ~~function similarly~~are by our definition functional similar and can ~~thus ~~represent each other. To give an example, imagine if 50 % of the catchment area receives ~~rainfall~~precipitation of around 1 mm hr^{-1} and 50 % around 2 mm hr^{-1}. In this specific case, we would have two occupied forcing bins (precipitation groups; PB). In the following, we explain ~~our time-invariant similarity assumptions~~how we have chosen the decision thresholds for the system structure~~ as well as our time-variant similarity assumption of the catchment (model) state and the ~~, the precipitation forcing and the model states.

~~*Time invariant similarity of the system structure*~~

~~The first step of our adaptive clustering approach requires the identification of hydrological response units (HRUs) that potentially act similar. A sufficient criterion for this is that their structural setup (e.g. geology, land-use, etc.) and their actual state (e.g. storage) are similar at a given time step. As already mentioned in section 3.2, our previous studies showed that different hydrological sub-units, in this case hillslopes, of the Colpach catchment, can be characterized by similar subsurface characteristics (integral filter properties). This implies a potential similar rainfall-runoff transformation when they are in a similar state. This is supported by our previous work (*Loritz et al., 2017, 2019*) which revealed that a sub-basin of the Colpach catchment (0.45 km^2) and a neighboring catchment (30 km^2) located in the same geological setting have almost identical specific discharges as long as they are at similar states and forced by comparable amounts of precipitation. This implies that the spatial variability of the system structure within the Colpach can be represented by a single spatially aggregated model and all grid cells of the precipitation field can thus be represented by the same model with the same model parameters as long as they are in the same state and driven by the same forcing. This entails, however, also that if we extend our research area to a catchment that is divided, for instance, into two geological settings that function hydrologically differently (regarding their filter properties) we would always need to run at least two structural different models where each of these models represents one of two geological settings.~~

Time invariant similarity of the system structure

The first step of our adaptive clustering approach requires the identification of hydrological response units (HRUs) that potentially behave similar. These similar units are typically identified based on structural properties such as the geological setting, the land-use or the topography. The general idea is that HRUs are grouped together, which share the same controls on gradients and resistances controlling flows of water as long as they are in the same state (*Zehe et al., 2014*). As already mentioned in section 3.2, our previous studies showed that different sub-units of the Colpach catchment, are characterized by similar spatially organized surface and subsurface characteristics (integral filter properties). This entails a potential similar rainfall-runoff transformation when they are in a similar state. The latter is supported by our previous work (*Loritz et al., 2017, 2019*), which revealed that a sub-basin of the Colpach catchment (0.45 km^2) and a neighboring catchment (30 km^2) located in the same geological setting have almost identical specific discharges as long as they are at similar states and forced by comparable amounts of precipitation. We hence assume that all grid cells of the precipitation field can thus be represented by the same model with the same model parameters as long as they are in the same state and driven by the same forcing. This entails, however, also that if we extend our research area to a catchment that is divided, for instance, into two or more geological settings, different dominant land-use or soil type distributions that function hydrologically differently (regarding their integral filter properties) we need to run two or more structurally different models where each of these models represents one of two structural settings. The latter might limit the possibilities to apply this approach on larger scales or in areas with complex structural settings.

Time variant similarity of the precipitation forcing

The second decision threshold we need to identify, defines the minimum difference at which we consider differences in the precipitation field as relevant for the runoff generation. Simply speaking, two ~~structural~~structurally similar hydrological units that are in the same state will only respond differently to an external forcing if the variability in the forcing has exceeded this threshold. Here, we picked a threshold of 1 mm hr^{-1} upon we consider differences between precipitation observations (grid cells) as relevant. We chose this threshold as ~~it represents~~ a reasonable ~~difference~~value upon ~~which~~ we expect ~~that a hydrological landscape element might function differently than another~~differences in ~~a~~hydrologic behaviour in humid ~~environment~~catchment and based on our collective understanding of the Colpach catchment. This means that only if the spatial differences in the precipitation field are above 1 mm hr^{-1} ~~do~~ we drive the spatially adaptive *model c* with different precipitation inputs. The choice of this threshold is important, as it is one of two main controls or parameters of the model resolution of the spatially adaptive model (see supplement B).

Time variant similarity of the catchment state

~~T~~he third assumption is to identify a decision threshold upon which we consider that two model elements are in the same state. This means that we need to select a point in time after a spatially variable rainfall event (> 1 mm hr^{-1}) when two or more ~~models~~model elements in the individual grid cells have ~~“*forgotten*”~~dissipated the differences between them introduced by the ~~interplay of the~~ previous precipitation ~~signal~~input with drainage and ~~evaporation~~evapotranspiration dynamics. Here, we use the change in discharge over time (dQ dt^{-1}; slope of the simulated hydrograph) and the discharge (Q) at a time step to infer similar model states. By that, we expect that two or more gridded models are again in the same state if ~~their~~the individual models estimate runoff ~~simulations change in~~and the slope of the runoff within a ~~similar range (~~0.05 mm hr^{-1}~~).~~ margin. As soon as this is the case and two or more gridded models are in the same state, we average their states (average saturation of each grid cell of the CATFLOW hillslope grid) and by that, aggregate the models back again into a ~~single~~one hillslope. The value of 0.05 mm hr^{-1} for Q and dQ dt^{-1} was picked as it reflects the desired precision of the adaptive *model c*. Similar as in the case of the decision threshold this value needs to be picked carefully. Furthermore, is it important to choose similarity metrics (here dQ dt^{-1} and Q) that are adequately describe the model states during the simulation time.

3~~.3.3~~4.2 Spatially adaptive modeling - model implementation

As stated in section 3.3.2, we classified the entire Colpach catchment as hydrologically similar ~~concerning~~with respect to the runoff generation as long as the different hydrological sub-units of the catchment are in the same state and receive a comparable forcing. This ~~means~~entails that we start the simulation with one gridded hillslope to represent the entire catchment and continue in this mode as long as we have not detected a spatial difference in the precipitation field above the selected threshold of 1 mm hr^{-1} (Fig. 3, t=0). At each time step, we bin the precipitation input of the next time step and determine the number of allocated bins (~~*P*~~*PB* = *number of precipitation bins*). If more than one precipitation bin is occupied (~~*P > 0*~~*PB > 1*) we increase the number of gridded models (*M = no. of running gridded models*) by running the same model in the same initial state, however, driven by different precipitation inputs.

~~Imagine~~Consider a scenario where the Colpach catchment is represented by one hillslope (*S = 1*) and we observe a precipitation event where 50 % of the catchment receives no precipitation, 20 % 7 mm hr^{-1} and 30 % 8 mm hr^{-1} ~~(~~as displayed in an illustrative example in Fig. 3~~,~~ at t=1~~)~~. This would mean that three precipitation bins are allocated (~~*P*~~*PB* = *3*) and ~~hence~~ we need to increase the number of running models ~~also~~ to three (*M = 3*). After running these three models for one time step with the different precipitation inputs, we bin the model states (dQ dt^{-1}; Q). Let us assume we would identify two occupied model state bins, which means that two different model states (*S* = *2*) are needed to represent the spatial variability of catchment states. This could happen if the differences between the 7~~ mm hr^{-1}~~ and 8 mm hr^{-1} rainfall intensity did not result in a significant difference in the discharge simulation of the two corresponding models. Following our approach, we aggregate the two models that are driven by 7 ~~mm hr^{-1}~~ and 8 mm hr^{-1} by averaging their states. We do this by averaging the relative saturation of the corresponding CATFLOW hillslope grids~~, which~~. The latter is straightforward in our study as they have the same width as well as lateral and vertical dimensions. In case that the hillslopes would not be ~~structural~~structurally similar this requires a weighted averaging of soil water contents to avoid a violation of mass conservation. After the aggregation of the two models, we have two model states left (*S* = *2*) each representing 50 % of the catchment area.

If there is no further rainfall occurring we wait until the gradients in system states ~~have been~~are depleted and the two running models have "*forgotten*" the difference in the past forcing and both predict similar dQ dt^{-1} and Q values and eventually aggregate the two models ~~again~~ two one ~~gridded~~hillslope model. If rainfall continues in the next time step *(~~P~~PB > 1)* we need to check which model states (*S*) receive which forcing. For instance, given our hypothetical example, we know that after the last simulation step we needed two model states (*S* = *2*) to represent our catchment. Each of these two states represents 50 % of the area of the catchment. ~~At~~Imaging that at the next time step~~,~~ we observe a precipitation event where 50 % of the catchment receives 8 mm hr^{-1}and the other 50 % 3 mm hr^{-1} (Fig. 3, t = 2). In this case, we have to check if the two model states (*S* = *2*) receive both precipitation inputs of 8 and 3 mm hr^{-1}. Let us assume that one model state is receiving 80 % of the 8 mm hr^{-1} and 20 % of 3 mm hr^{-1} rainfall. The other model 20 % of the 8 mm hr^{-1} and 80 % 3 mm hr^{-1}. In this specific setting, we would need to run four models (*M* = *4*) to account for the spatial variability of the model states and precipitation input, while each of those reflect a different combination of the model state and forcing in different parts of the catchment. At this stage, we again either wait until the internal differences have been dissipated to reduce the number of models or we increase the number of models in case that precipitation with larger spatial variability of ~~P~~PB = 1 is continuing (Fig. 3, t = n). The maximum number of models we could require in our adaptive clustering approach depends on the maximum ~~resolution~~number of ~~the~~ precipitation ~~input upon we divided the Colpach catchment and is~~ grid cells (42 in ~~our~~this study). The highest resolution that the spatially adaptive *model c* can reach in this study is reflected by the spatially distributed *model b*.

4.3~~.3.4~~ Spatially adaptive modeling - model analysis

To test our spatially adaptive *model c* against the observed discharge of the catchment, we route the simulated runoff contributions according to their location to the outlet by assuming a mean flow velocity of water within the channel network of 1 m s^{-1}. However, as the same model can represent different grids with different locations we additionally need to calculate the average flow distances to the outlet of all grids a model is representing and shift the simulation by the average distance accordingly. We then take the area-weighted mean of every simulation at each time step. The performance of the adaptive *model c* is then ~~measured~~quantified by the KGE against the observed discharge and the area-weighted average of the distributed *model b*. The latter addresses our second research question and follows the logic that ~~an appropriate~~a suitable adaptive ~~model~~modelling approach should lead to similar simulations as a fully distributed model, however, with fewer model elements. While we use CATFLOW as a model here, the proposed approach is not restricted to this model and can be used in any hydrological model ~~that distributes a catchment into independent spatial units~~framework that distributes a catchment into independent spatial units. One advantage of CATFLOW (or similar type of models) is that it uses also an adaptive time step procedure making the final model adaptive in space and time. However, if a model represent a landscape in entirely continues manner without a delineation of the landscape into independent sub-units like several 2d surface runoff models an adaptive mesh (numerical grid) is required if the spatial resolution should adapt itself during runtime.

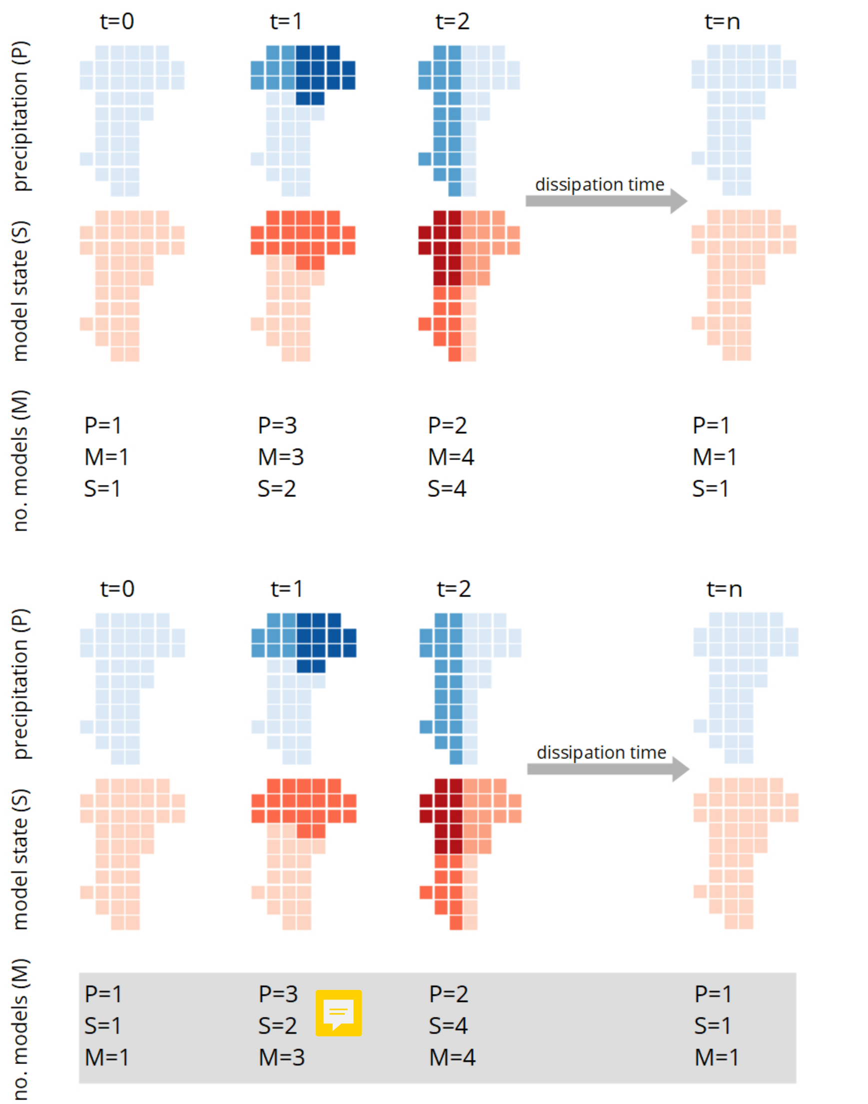


Figure 3. Sketch of the spatial adaptive modeling described in section ~~3.3.3~~4.2. The upper panel shows the precipitation forcing (blue) and the lower panel the model states (red). The numbers below the figures indicate how many precipitation (~~P),~~PB) and model state (S) bins (groups) are occupied and how many models (M) are running at the given time step.

45. Results

In the following section, we investigate the precipitation field and compare the performance of the discharge simulations of the *reference model,* the spatially aggregated *model a*, and distributed *model b* at the annual, seasonal, and event scale by comparing ~~the~~hourly simulations against hourly observed discharge~~ data~~. We, furthermore, present the simulation results of the adaptive ~~modeling~~*model c* for two selected rainfall events ~~and~~including the spatial distribution of the precipitation ~~forcing as well as~~and the model states. Finally, we show the soil moisture distribution of ~~*model c* for rainfall event I~~two hillslope models at different time steps that have received a significant dissimilar precipitation forcing to highlight the importance of the dissipation time scale for adaptive modeling.

45.1 Precipitation characteristics

While rainfall sums are equally distributed between the winter (Oct. – Mar.) and vegetation season (Apr. – Sep.) in the selected hydrological year 2013/14 (Fig. 2 b), the rainfall intensities and the associated standard deviation (here used to measure the spatial variability of the precipitation field) of the precipitation field are in general higher in summer (Fig. 4 a ~~&~~and b). For instance, the five rainfall events with the highest rainfall intensities ~~as well as~~and the highest standard deviation in space were all observed in the summer season. Rainfall intensity and spatial variability are ~~thereby~~ strongly linked to each other which is reflected in their ~~linear~~ correlation of 0.82. The latter is no surprise as convective storms, which dominate the precipitation generation in summer, are typically characterized by higher ~~spatial~~spatio-temporal variabilities and higher rainfall intensities. This finding is ~~surely~~ neither surprising nor limited to the chosen research environment (e.g. *Hrachowitz and Weiler, 2011; Wilson et al., 1979*) but it confirms one of our initial assumptions that rainfall is spatially more diverse in the summer season compared to the winter months in the Colpach catchment.

We selected two rainfall-runoff events to test the adaptive *model c* (Fig. ~~4, *time of the events are indicated by the red horizontal bars*).~~4). We chose the first event as it has the highest rainfall intensity of 19 mm hr^{-1} and the third-highest spatial variability ~~measured~~estimated by the standard deviation of 3.8 mm hr^{-1} in the time series as well as a distinct runoff reaction. Rainfall event I was observed at the beginning of August, lasted for about 5 hrs and the highest spatial differences between the grid cells of 14 mm hr^{-1} was reached right at the beginning of the event (Fig. 5 and 6). The rainfall event I moved from west to east over the catchment and reached its maximum rainfall intensity after ~~approximately~~ 3 hrs. No rainfall had occurred before the event for a period of 102 hrs. ~~We~~It can hence ~~assume~~be assumed that the catchment was in a moderately dry state before the event ~~which is~~ also indicated by soil moisture measurements presented in *Loritz et al. (2017).*

The second rainfall event was selected as it has distinctly different properties (low spatial variability, low intensity, longer duration) ~~when~~as compared to the first event. Event II has a maximum rainfall intensity of 5.8 mm hr^{-1} and a maximum spatial difference between the grid cells of 4 mm hr^{-1}. The event lasted for around 15 hrs~~,~~ making it the longest continuing rainfall in the summer season and there was no rainfall observed 20 hrs before the event but more than 36 mm of rainfall ~~in~~over the ~~previous~~preceding three days. ~~We can~~It seems hence reasonable to assume that the soils in the catchment where rather wet which is again supported by the soil moisture measurements presented in *Loritz et al. (2017).*

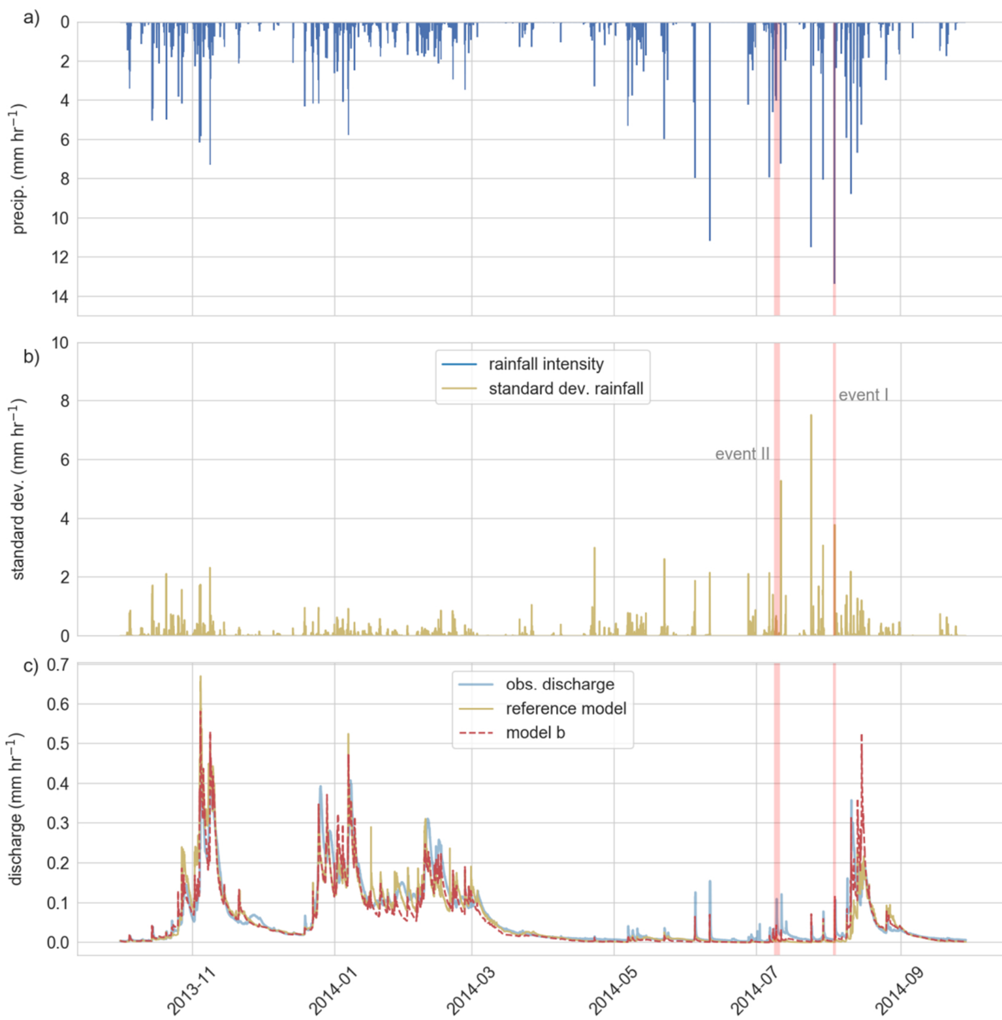


Figure 4. a) average rainfall intensity of the precipitation field (mm hr^{-1}), b) corresponding standard deviation of the precipitation field (mm hr^{-1}), c) observed discharge of the Colpach catchment and the discharge simulation of the reference model as well as of the distributed model b. The two red bars ~~highlight~~display the location of the two selected rainfall-runoff events used to test the adaptive clustering approach.

45.2 Temporal dependency of the model performance

The performances of the four model setups (*reference model, model a, b* and *c*) to simulate the observed discharge of the Colpach catchment ~~measured~~estimated by means of the KGE are shown in Tab. ~~1. If one compares~~2. Comparing the two spatially aggregated models that differ only with respect to their rainfall forcing the *reference model* outperforms *model a* during the winter season and on the annual time scale while m*odel a* has a higher performance ~~in~~during the vegetation season (Apr. – Sep.). Both models are characterized by KGE values ~~larger~~higher than 0.8 in the winter season and for the entire ~~simulation period~~hydrological year while the predictive performance drops in summer and is particularly low for the two rainfall-runoff events resulting even in negative KGE values. The differences between the KGE values (ΔKGE) between the two spatially aggregated models (*reference model* and *model a*) are low in winter, increase in summer, and are the highest for the convective rainfall event I. ~~Here does the *reference model* only have a similar performance as the mean of the discharge time series indicated by a KGE value of -0.41 (please note that it is not zero as in the case when using the Nash-Sutcliff efficiency as shown by *Knoben et al. (2019)*).~~Here ~~does~~ *model a* have only a slightly improved predictive performance as the average discharge of the event indicated by a KGE value of -0.16 (please note that it is not zero as in the case when using the Nash-Sutcliffe efficiency as shown by *Knoben et al., 2019*).

The observed discharge of the Colpach catchment, the discharge simulations of the *reference model* as well as the discharge simulation of the distributed *model b* are presented in Fig. 4 c. ~~The visual~~Visual comparison of the two models shows that the *reference model* has ~~a~~ lower runoff production during summer, which is particularly visible in August and September. Interestingly, the latter cannot be explained by the annual or seasonal precipitation sums as both models are driven by on average similar precipitation sums of around 900 mm yr^{-1} for the entire year and around 450 mm 6 month^{-1} in the summer season. Overall, *model b* has the highest predictive performance ~~measured~~as indicated by ~~means of~~ the KGE in all five test periods (annual, winter, summer, and the two selected rainfall events) ~~if~~when compared to the two spatially aggregated models~~.~~ (*reference model* and *model a*). The absolute differences between the model performances depend again on the selected period. For instance, for the entire simulation period, the *reference model* and *model b* have close to equal KGE values around 0.9 while the differences between the KGE values are ΔKGE = 0.21 in summer ~~ΔKGE = 0.2~~ and for the ~~spatially variable~~ rainfall event ~~I~~II around ΔKGE = 0.~~7~~3.

Although *model b* has the highest KGE values for the two selected rainfall-runoff events, the general model performance is, given the KGE values of 0.29 and 0.1, still relatively low for both runoff events. The low performance can be explained by a general underestimation of the total runoff volume at both events (Fig. 7), while it seems that the shape of the simulated hydrograph is simulated acceptable~~.~~ underpinned by a correlation of 0.72 and 0.86 between the simulation and observation (see supplement for the three components of the KGE). The latter is supported by the fact that ~~the~~ distributed *model b* is ~~capable~~able to simulate the observed double peak at event I. ~~Furthermore, we~~We, furthermore, tested the addition of a direct runoff component by assuming that 10 % of the rainfall is directly added to the channel network instead of falling on the hillslopes. This model extension could be justified by sealed areas within the catchment, by precipitation that directly falls into the stream or on saturated areas like the riparian zone ~~and~~. This rather simply model extension increases the KGE of *model b* from 0.29 to 0.48 at event I. However, we do not update our model here as the main goal of this study is not to perform the best possible rainfall-runoff simulation but to investigate the role of the spatio-temporal patterns of the rainfall for the runoff generation of a mesoscale catchment by introducing the concept of a spatially adaptive hydrological model.

Table 2. Model performances of the four model setups to simulate the observed discharge of the Colpach catchment~~,~~ ~~which are~~ measured ~~by~~ using the Kling-Gupta efficiency (KGE) based on ~~the~~ hourly simulation and observation time steps. Performances are shown for the entire hydrological year (2013/2014), for the winter (Oct. – Mar.) and summer season (Apr. – Sep.) as well as for two selected summer rainfall-runoff events in July and August. The three components of the KGE can be found in the supplement.

	annual performance (KGE)	winter performance (KGE)	summer performance (KGE)	rainfall event I (KGE)	rainfall event II (KGE)
reference model from Loritz et al. (2017)	0.88	0.88	0.52	-0.~~41~~1	-0.~~09~~2
model a (spatially-aggregated)	0.85	0.84	0.65	-0.16	-0.05
model b (distributed model)	0.91	0.89	0.73	0.29	0.1
model c (adaptive model)	-	-	-	0.29	0.1

45.3 Spatially adaptive modeling - simulation results

The upper panel of Fig. 5 shows the binned precipitation field of rainfall event I. The precipitation field was binned based on the chosen bin width of 1 mm hr^{-1}. The rainfall field allocates 0 bins (precipitation groups) at t = 0 *(~~P~~PB = 0)*, 12 bins at t = 1 *(~~P~~PB = 12)*, ~~12~~13 bins at t = 2 *(~~P = 12~~PB = 13)*, 3 bins at t = 3 *(~~P~~PB = 3)* and 2 bins at t = 4 *(~~P~~PB = 2)*. The number of occupied bins indicates the spatial variability of the rainfall event at a given time step and would reach maximum spatial complexity if ~~P~~PB equals 42. This ~~means~~entails that if a high number of bins is allocated the forcing is spatially variable and respectively a higher number of models is needed to represent the spatial variability of the precipitation. The number of bins does, however, not specify how large the gradients are within the spatial precipitation field. For instance, if 50 % of a precipitation field is characterized by a rainfall amount of 20 mm hr^{-1} and the other 50 % by 1 mm hr^{-1} the number of allocated bins is two although the absolute difference between the bins is large.

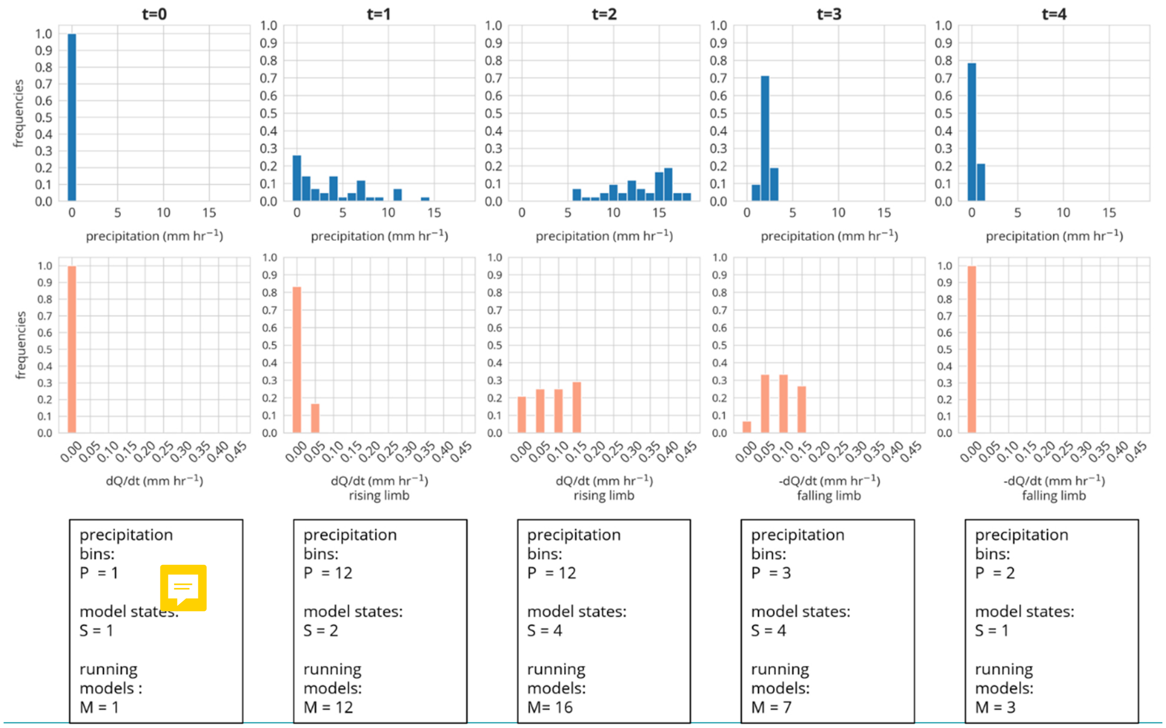
The lower panels of Fig. 5 and Fig. 6 ~~show~~display the binning of the model states (*S*) of the adaptive model for each time step of event I~~.~~ for the similarity measure dQ dt^{-1}. We do not plot the similarity measure Q here as in our specific case Q and dQ dt^{-1} lead to the same classification at both events. However, this does not mean that Q is less relevant as in theory two models could simulate identical dQ dt^{-1} values but very different absolute Q values. This shows that the set of similarity measures should be picked carefully and depend very much on the given modeling task and the research environment.

At t = 0, we run a single model representing the entire catchment with a single model state. At t = 1, the precipitation starts and the spatial field is classified into 12 bins *(~~P~~PB = 12)*. Following our approach, this entails that we need to run 12 models (M = 12) at t = 1 to account for the spatial variability of the rainfall. After one simulation step, we estimate the number of model states by binning the absolute values (Q) and slope (dQ dt^{-1}) of the discharge simulations of the 12 models resulting in two different model states (~~as~~ two model state bins are occupied). Each of these states represents now a different part of the catchment with a different area (Fig. 6, lower panel). For instance, at t = 1 around 76 % of the catchment area is represented by a model in a state where discharge changes below 0.05 mm hr^{-1} and 14 % between 0.05 and 0.1 mm hr^{-1}. At t = 2, the precipitation field has ~~again~~ been classified into ~~12~~13 bins but at this time step, the catchment is represented by two model states from the time step before. This means we need to check which combinations of states and precipitation input occur. In other words, which grids are represented by which state and are forced by which precipitation input. In this specific setting, we need to run 16 models which is lower ~~as~~than the theoretical maximum (2 model states (*S*) x ~~12~~13 precipitation bins *(~~P) = 24~~PB) = maximum of 26* running models (*M*)) as not all model states are driven by all ~~grouped~~binned precipitation inputs. Afterward, we again group the model states (*S = 4*) and continue until t = 4 after which no rainfall occurs and we again represent the entire catchment by a single hillslope model. In total, we were able to reduce the maximum number of gridded models from 42 to a maximum of 16 at rainfall event I and at the second event from 42 to 4 without a predictive performance loss in comparison to the distributed *model b* (Tab. ~~1~~2). The latter is~~, besides the comparison of *model c* with the observed discharge, also~~ shown by the ~~high~~ KGE values between the distributed *model b* and the adaptive *model c* of around 0.98 at both events.

45.4 Spatially adaptive modeling – dissipation of differences

The dissipation timescale (memory timescale) ~~at~~for both events until the different hillslope models have "*forgotten*" the last forcing and are again in the same "*runoff generation state*" is relatively short. ~~More specifically~~Specifically, already after 1 hr of no precipitation at event I and II is the differences between the ~~runoff generation of the~~ hillslope models in *model c* ~~are~~ below the ~~picked~~selected threshold of 0.05 mm hr^{-1}~~.~~ for Q and dQ dt^{-1}. The same is true for the soil moisture distributions in 10-20 and 60-100 cm depth, which is negligibly small at the time step t=4 at event I when the four hillslope models are aggregated. This ~~means~~entails that ~~our *model c* would represent~~ the entire catchment ~~with~~can again be represented by a single hillslope model already shortly after the last rainfall at both events until a new rainfall ~~event (P > 1) occurs.~~ *(PB > 1)* occurs. The latter is supported as the single hillslope from *model c* and the spatially aggregated *model a* are also in a similar state regarding their runoff generation after t=5 at event I.

This picture ~~is quite~~might, however, be different ~~for the soil moisture distribution between the hillslopes, at least in deeper soil layers.~~in certain simulations scenarios. For instance, Fig. 8 ~~shows~~displays the soil moisture distribution of two hillslope models in 10~~ to ~~-20 and 60~~ to ~~-100 cm depth ~~which~~at three time steps during event I (t = 3, t = 24, and t=48) that either ~~has~~have received the highest amount of rainfall measured at a ~~given~~ grid cell ~~at event I~~ (30 mm, 5 hr^{-1}) or the lowest (15 mm, 5 hr^{-1}~~) for two different time steps during and after event I (t = 3 and t = 24; see Fig. 5).~~). Both hillslope models started in ~~a similar~~the same initial model state and ~~Fig. 8 only shows the wetness of the soil matrix. The memory time scale~~the dissipation time of the topsoil correlates ~~thereby quite~~ well with the runoff generation~~ and we observe the~~. The largest ~~difference~~deviation between the "wettest" model, which has received the highest amount of rainfall, and the "driest" model, which has received the lowest amount of rainfall, is at t = 3 shortly after the highest rainfall intensity (see Fig. 5). After 24 hrs, this difference persists but~~ it~~ slowly dissipates and has almost completely disappeared after 48 hrs. In the deeper soil layer, the picture is different. During the event, we see no reaction to the rainfall forcing ~~of~~in the ~~soil matrix and water bypasses these areas through preferential flow paths.~~deeper soil layers. However, 24 hrs after the first rainfall ~~of event I~~ the ~~difference between the~~ two models ~~regarding their soil moisture distributions~~deviate also in deeper layers and the deviation is ~~slowly increasing~~ in a similar range as in the top soil although there was no further rainfall. The ~~latter means that by aggregating the different hillslope models, as done in our adaptive *model c*~~difference in both layers disappears again after ~~only one hour~~48 hrs of no rainfall~~, we delete~~ and the ~~difference between the soil moisture distributions. As we use the mean to aggregate our models, we~~ "wettest" and "driest" model are~~, however, still conserving mass. The question remains of how important these differences are on longer time scales or for the root water uptake.~~ again in a similar state also regarding their soil moisture distributions.



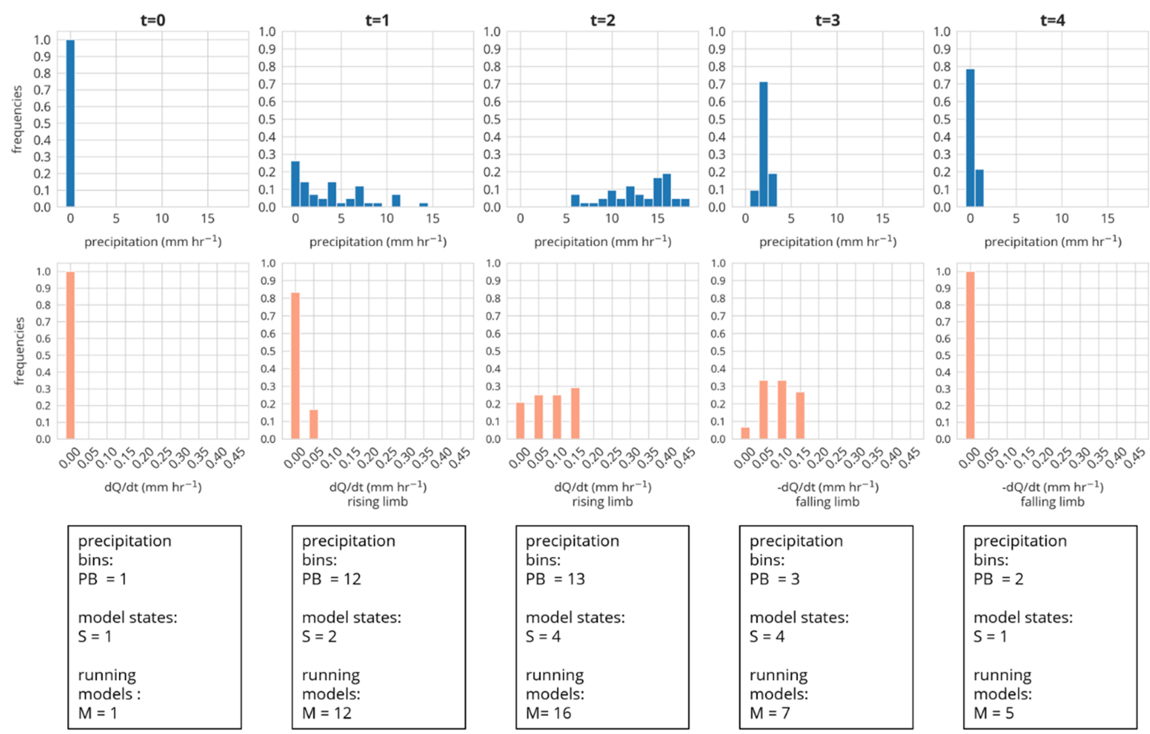


Figure 5. Binned precipitation field (blue) and binned model states (orange) of the adaptive model (t = 0; August 3rd 2014 15:00 CET); ~~P~~PB = no. of allocated precipitation bins, S = no. of allocated model space bins, M = no. of running models at the given time step. The spatial distribution of the precipitation and the model states for event I are displayed in Fig. 6.

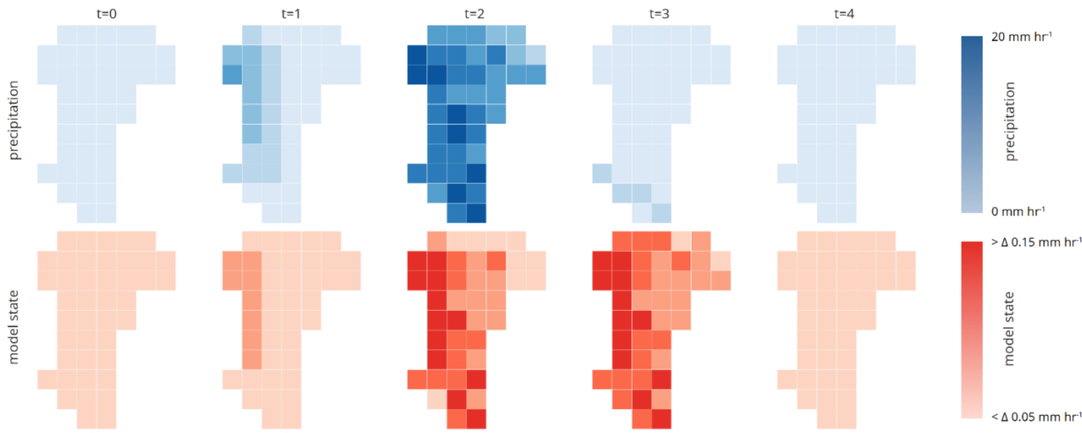


Figure 6. Spatial and temporal distribution of the precipitation field (upper panel) and the corresponding states of the actual model grids used by the adaptive model c (lower panel). The model state is estimated by the slope of the simulated discharge. The corresponding bins (groups) of the precipitation and model states are shown in Fig. 5.

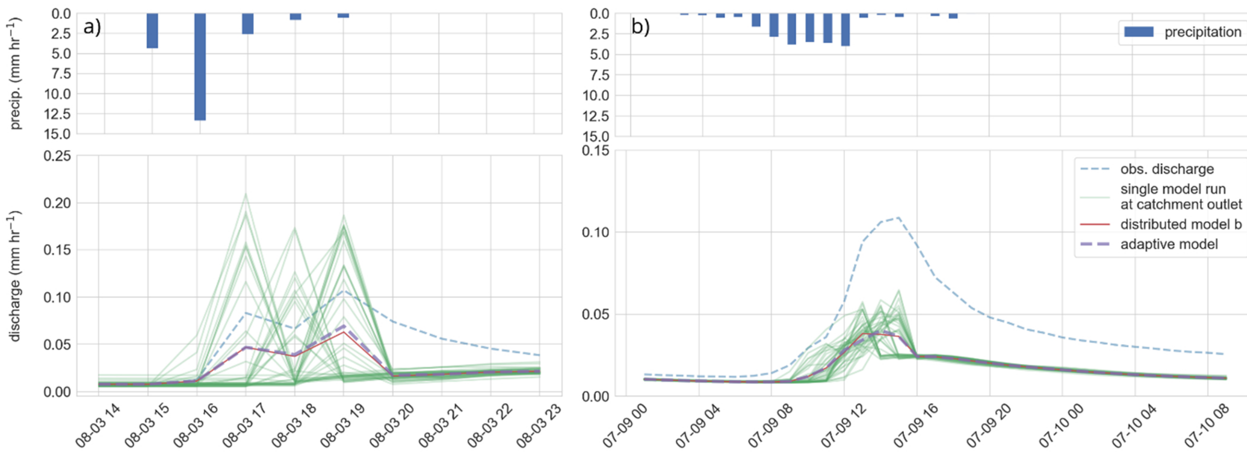
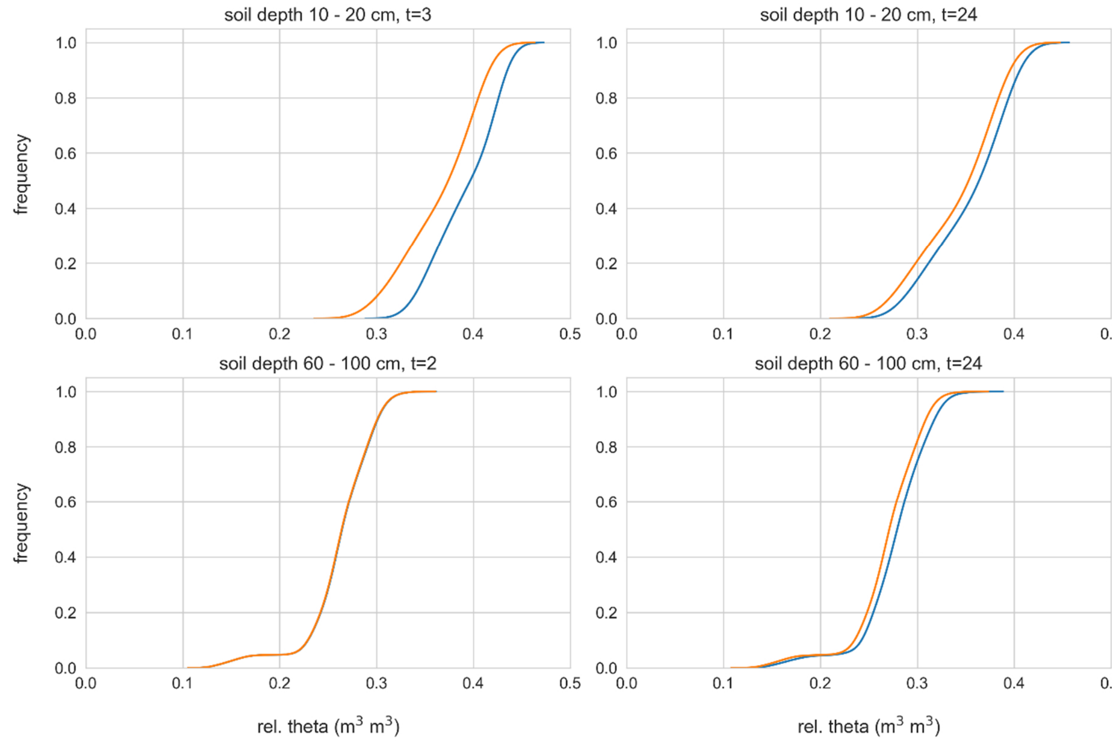


Figure 7. a) rainfall-runoff event I and b) rainfall-runoff event II. Blue bars in the upper panel show the average precipitation of the precipitation field for each time step (mm hr^{-1}). The green curves in the lower panel represent a single gridded model of the distributed model b; red line the area-weighted mean of the distributed model; purple dashed line the area-weighted mean of the adaptive model and dashed blue line the observed specific discharge of the Colpach.



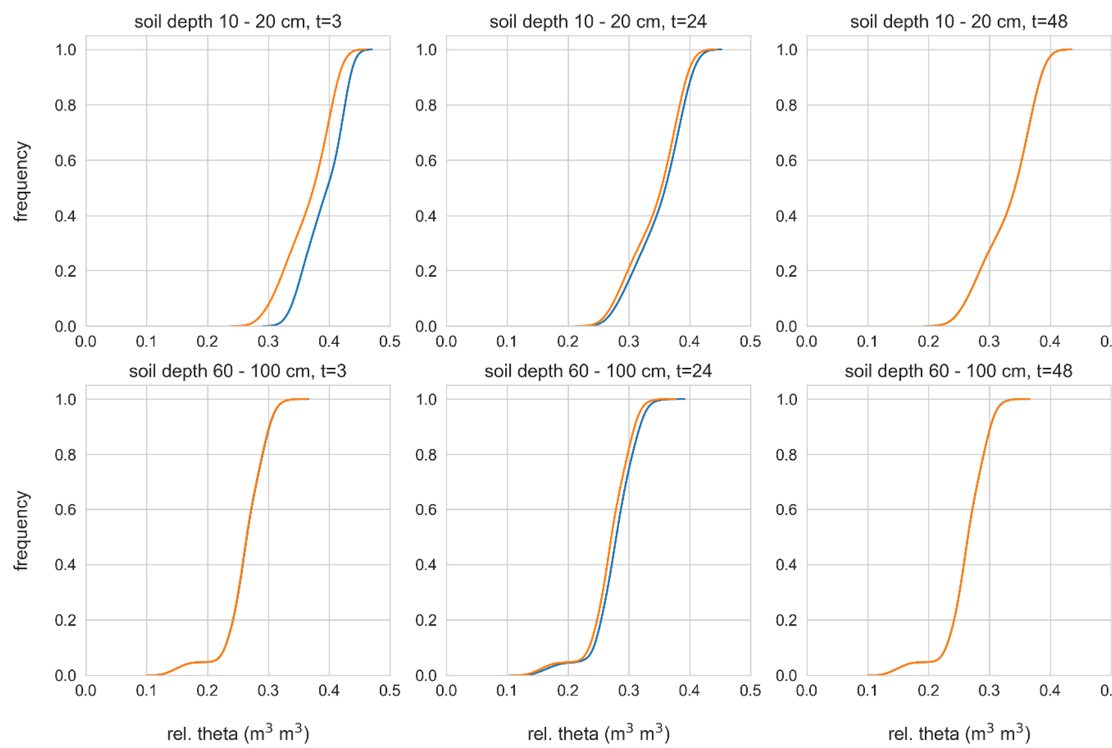


Figure 8. Relative soil moisture distributions for two gridded hillslope models ~~of *model b*~~ that received the lowest (orange curve) respectively the highest (blue curve) amount of rainfall during event I (15 mm 5 hr^{-1} and 30 mm 5 hr^{-1}). Presented for time step t = 3 (during the event~~) and~~), t = 24 (after the event~~).~~), and t = 48 (after the event).

56. Discussion

56.1 The role and value of distributed rainfall for hydrological models

While the three non-adaptive model setups (*reference model, model a,* and *b*) perform equally well with respect to simulate the discharge of the Colpach catchment in the winter season, this is not the case in the summer season where the distributed *model b* has higher KGE values as both spatially aggregated models. This corroborates one of our ~~hypothesis~~hypotheses stating that the predictive performance of the ~~spatially aggregated~~ representative hillslope (*reference model*) introduced by *Loritz et al. (2017)* increases if the model is distributed in space and driven by distributed rainfall. ~~However~~Nevertheless, *model b* has still several deficiencies especially at the two selected rainfall~~-runoff~~ events where it does underestimate the total observed runoff volume~~. This indicates~~ resulting in high correlation values but overall low KGE values. The latter shows that there is ~~still room for further improvements of the hydrological model~~ potential to ~~increase its~~improve the predictive performance of the model beyond only changing the precipitation input for instance by accounting for the sealed areas and forest roads in the catchment.

~~Although the model comparison in this study is rather heuristic (e.g. we used only a single performance metric etc.), the findings highlight that the use of distributed rainfall is recommended during the summer season for simulations of the runoff generation of the Colpach. This insight is consistent with the findings of *Euser et al. (2015)* and *Wilson et al. (1979)* who showed similar results in a 1600 and 42 km^2 large catchment. On the other hand, the winter performance of *model b* did not improve in comparison to the *reference model*. In line with these results, *Obled et al. (1994)* and *Das et al. (2008)* concluded that "*the spatial variability of rainfall, although important, is not sufficiently organized in time and space to overcome the effects of smoothing and dampening when running off through this rural medium sized catchment*.". Given the rather small size of the Colpach catchment and the fact that the use of distributed rainfall increased the performance during the summer, it seems that catchment size might not be the best indicator to decide if or if not a distributed hydrological model driven by distributed rainfall is needed. The higher relevance of spatially distributed precipitation for hydrological modeling in the summer months surely reflects the circumstance that also the average rainstorm size as well as the average rainfall intensities change between the seasons and are on average much higher during summer (*Neuper and Ehret, 2019*). Given the changing meteorological regime, it seems reasonable to link the increasing relevance of distributed rainfall to these changes. The fact that average storm size over a catchment is a key indicator to identify the role of distributed precipitation on the runoff generation of a catchment was also pointed out by *Nicótina et al. (2008)*. As the dominant rainfall generation mechanisms change during the hydrological year in many humid catchments from frontal to convective, it comes from a physical perspective, not as a surprise that also the relevance of distributed rainfall differs between the seasons or even between different rainfall-runoff events.~~

The evaluation of the

Although the model comparison in this study is rather heuristic (e.g. we discuss mainly along a single integrating performance metric, etc.), the findings in this study show that the use of distributed rainfall is at least recommended during the summer season. This contradicts the results of, for instance, *Obled et al. (1994)* who argued that the precipitation over a 71 km^2 large catchment is not sufficiently organized to be relevant for the runoff generation. It is also not in line with the findings of *Nicótina et al. (2008)* who recommend the use of distributed rainfall only in specific scenarios (e.g. infiltration excess) or in catchments above 8000 km^2. Given the improved performance of the distributed *model b* and the rather small size of the Colpach catchment of less than 20 km^2 it seems reasonable to conclude that catchment size alone might not be the best indicator to decide if a distributed hydrological model driven by distributed rainfall is needed or not.

As the dominant rainfall generation mechanisms change during the hydrological year in many catchments from frontal to convective, it comes from a physical perspective, not as a surprise that the increased relevance of distributed rainfall in summer can be linked to the changing meteorological properties. *Ogden and Julien (1993)* argued along these lines when they showed that the spatial distribution of rainfall is only a dominant control in the case that rainfall events have a shorter duration than the average runoff concentration-time of a catchment. Similarly, *Zhu et al. (2018)* reasoned that the spatial patterns of precipitation are less relevant compared to the temporal distribution if the drainage area and therefore typically the concentration-time is decreasing. The question if a structurally similar catchment needs to be represented in a spatially distributed manner depends hence on the spatial and temporal structure of the precipitation as well as on the average concentration-time respectively as proxy on the catchment size.

The changing model performances ~~highlights~~during the two events highlight that the ~~necessary spatial~~required model ~~structure~~and precipitation resolution does not only change between seasons but ~~more importantly~~can vary from~~ rainfall event to~~ rainfall event~~.~~ to rainfall event. This ~~idea ~~has ~~already~~also been ~~highlighted~~argued by *Watts and Calver (1991)*~~. They concluded~~ who stated that "*the finest available definition of rainfall may be desirable for modeling…*" of convective rainfall events while ~~higher~~lower spatial model resolutions ~~did not increase the predictive performance of their models~~are sufficient during spatially and temporal more homogenous often stratiform rainfall events. ~~At first glance contradicting~~In contrast, *Lobligeois et al. (2014)* reported that the distribution of rainfall is in general of higher relevance in certain regions of France when they analyzed 3620 rainfall-runoff events ~~of~~in 181 different mesoscale catchments. However, they also ~~discussed~~argued that a substantial ~~amount~~number of rainfall-runoff events ~~contradicted~~do not match this general pattern~~, which shows~~ showing that the distribution of rainfall can be of high importance~~ in regions,~~ even if the spatial precipitation patterns are usually not a dominant control on the runoff formation in a region. As such "rare" events are frequently linked to extremes~~, which~~ that are in turn beyond the realms of experience on what these landscapes have adapted to, they are of considerable importance despite their low occurrence in time ~~(*Loritz, 2019*).~~(e.g. *Loritz, 2019*). This point is underpinned by the work of *Zhu et al. (2018)* and *Peleg et al. (2017)* who both questioned the common practice to use spatially uniform rainfall based on single or a few rain gauges for performing flood risk assessments especially for longer return periods in rural respectively urban catchments. The proposed spatially adaptive modeling approach could thereby be one way to tackle this issue as it enables continuous physically-based simulations with model structures that adapt to the precipitation forcing.

56.2 Spatially adaptive modeling – as ~~a ~~tool to reduce redundant computations

~~The proposed adaptive modeling approach is promising because the spatial adaptive *model c* performed similarly as the distributed *model b*, although it used a much smaller number of hillslopes. The maximum number of gridded model elements that were necessary to represent the variability of catchment states and precipitation was reduced by a factor of 2.5 for event I. The total gain in computational efficiency is however larger as most of the time less than 16 models are required to represent the catchments runoff generation.~~

The first results of the adaptive modeling approach seem promising as the spatial adaptive *model c* performed similarly as the distributed *model b,* however, using a smaller number of hillslopes. Similar findings were reported, for instance, by *Chaney et al. (2016).* They applied their HRU-based model called "HydroBlocks" in a 610 km^2 large catchment and showed that a compressed, semi-distributed model consisting of 1000 HRUs performed similar compared to a gridded fully distributed model while being two orders of magnitudes faster than the distributed model and requiring only 0.5 gigabytes instead of 250. They concluded that: "… *the spatial patterns of the fully distributed model can be reproduced with a fraction of the computational expense.*" highlighting the potential of approaches like "HydroBlocks" as tools to improve the representation of hydrological processes in large-scale land surface models without drastically increasing the computational times and model complexity.

The main difference between models like "HydroBlocks" and our approach is that HRUs are dynamically reassigned during model execution based on the spatial properties of the precipitation forcing. By that, we try to avoid redundant calculations to reduce computational times similar to *Chaney et al. (2016)* but also try to avoid situations in which we underestimate the spatial variability of the meteorological forcing or the system state in the case that the test period is not representative for certain spatial constellations. The latter can thereby significantly impact hydrological simulations during extreme conditions (e.g. *Zehe et al., 2005; Zhu et al., 2018*). Our results show that the maximum number of gridded models necessary to represent the variability of the catchment states and precipitation elements can be reduced by a factor of 2.5. The total gain in computational efficiency is however larger as in the majority of time fewer than 16 models are required to represent the catchments runoff generation. For instance, during low flow conditions are the spatially aggregated *model a,* all hillslopes of the spatially distributed *model b* and the spatially adaptive *model c* in a similar state and produce hence similar results. In addition, the fact that during the winter season a single representative hillslope (*reference models*) perform close to similar to the distributed *model b* indicates that possibly to save computational times by dynamically adapting the model structure is higher than the factor of 2.5 suggests.

Clark et al. (2017) recognized computational times as ~~mayor~~a major obstacle when using physically-based models for practical applications, as proposed in the landmark publication of ~~*Freeze and Harlan, (1969).*~~*Freeze and Harlan (1969).* The discussion about saving computational times with adaptive clustering is, however, challenging ~~(*Ehret et al., 2020*)~~ as the gain depends on the chosen model approach (e.g. numerical scheme), on the used hardware, the programming language, the compiler, or on the number of printout times ~~of a model~~ to the hard-drive~~. Furthermore, the~~ (*Ehret et al., 2020*). The relevance of saving computational times of, for instance, 10 % depends furthermore on the absolute calculation time of a model and ~~hence~~ whether a model run needs 100 min or 100 d to be completed. A fair comparison would mean to ~~setup~~set up a virtual environment and work under similar conditions, e.g. ~~by~~ using a virtual machine as well as using a fully automated adaptive clustering approach. Both ~~is~~are, however, beyond the scope of this study and we ~~would like to~~ point toward the study of *Ehret et al. (2020*~~)~~*)*, which discusses the potential of adaptive clustering with respect to saving computational times in detail.

56.3 Spatially adaptive modeling – as ~~a~~learning tool to better understand the dissipative nature of a hydrology

In this study, we focus on the potential of adaptive modeling to examine when interactions between a variable precipitation forcing and a variable catchment state cause a variable runoff response and when these differences get "*forgotten*" due to the dissipative nature of hydrological systems. ~~Our results show or reiterate that the~~

~~relevance of distributed rainfall for hydrological modeling is dynamically changing in space and time. One way to account for this dynamically changing relevance is to run distributed models driven by distributed rainfall the entire time at the highest possible spatial resolution. Such an approach, sometimes referred to as hyper-resolution modeling (e.g. *Bierkens et al., 2015*), would avoid cases in which we unnecessarily underestimate the needed (spatial) model complexity of a hydrological model (e.g. *Fenicia et al., 2011b; Höge et al., 2018*). However, this procedure may lead to a strong increase of uncertainty due to an increased number of model parameters (e.g. *Beven, 1989*), result in a general overestimation of the simulated spatial variability due to error propagations within the model as well as increase the number of redundant computations in a majority of the simulation period (*Clark et al., 2017*). The latter implies a vast amount of computations as the natural length scale (grid size) of water flow in the critical zone, which is frequently simulated by using the Darcy-Richards equation, should not exceed a lateral grid size of 10 m and vertical grid size below 1 m in homogeneous soils (*Vogel and Ippisch, 2008*). The same is true for simulating surface runoff with different diffusive wave approaches where typically much higher flow velocities occur compared to the subsurface which again requires high resolutions and small calculation time steps. Hyper-resolution modeling without a delineation of the underlying system in independent sub-units for parallelization is hence up-to-date constrained to rather small~~Our results illustrate that the relevance of distributed rainfall for hydrological modeling is dynamically changing in space and time. One way to account for this dynamically changing importance is to run distributed models driven by distributed rainfall the entire time at the highest possible resolution. Such an approach would avoid cases in which we unnecessarily underestimate the needed (spatial) model complexity of a hydrological model, which again could lead to limited predictive performances (e.g. *Fenicia et al., 2011; Höge et al., 2019; Schoups et al., 2008*). However, this procedure may result in a strong increase of uncertainty due to an increased number of model parameters (e.g. *Beven, 1989*) frequently by an unchanged amount of data for validation (*Melsen et al., 2016*), lead to a general overestimation of the simulated spatial variability due to error propagations and can drastically increase the number of redundant computations (*Clark et al., 2017; Loritz et al., 2018*).The issue of increasing computational times due to redundant calculations is thereby reinforced by the fact that physically-based simulations of hydrological fluxes rely on relatively short natural length scales in time and space. For instance, the water flow in the critical zone, which is frequently simulated using the Darcy-Richards equation, should not exceed a lateral grid size of 10 m and a vertical grid size below 1 m in homogeneous soils (*Vogel and Ippisch, 2008*). The same is true, although on other scales, for simulating surface runoff with derivatives of the Saint-Venant equation but also for conceptual models where the assumption that a few macroscopic water tables can represent the heterogeneity of driving potentials in a landscape is rarely questioned. Even the gridded spatial resolution of 100 m proposed in the comment by *Wood et al. (2011)* for hyper-resolution models seems from a purely physical perspective on hydrological processes questionable given the importance of hillslopes as key building blocks in a hydrological landscape (*Fan et al., 2019*). This is underpinned by the fact that hillslopes in the upper part of the Colpach are barely longer than 100 m but different segments of these hillslopes can vary substantially in their wetness and connections to the river (e.g. *Martínez-Carreras et al., 2016*). Hydrological physically based modeling with top-down or bottom-up models without a delineation of the underlying system in smaller sub-units is hence up-to-date constrained to rather short length scales, at least if applications shall not compromise the underlying physics.

~~Physical constraints of small grid sizes and calculation time steps must not be a dead-end for applying bottom-up models on larger scales. This is because it is often found that different catchments in the same hydrological landscape function similarly despite the overwhelming small scale variability we often observe with point-scale measurements (e.g. *Loritz et al., 2017*). This entails a large potential to transfer information about model states~~

~~from one catchment or hillslope to another~~ Physical constraints, which result in small grid sizes and calculation time steps, must however not be a dead-end for physically based modeling on larger scales. This is because it is frequently found that different catchments in the same hydrological landscape function similarly despite the overwhelming small scale variability we frequently observe on the plot scale (e.g. *Mälicke et al., 2019; Sternagel et al., 2019*). This phenomenon sometimes referred to as spatial organization entails a large potential for hydrological modeling as it allows to transfer information about functional relationships and catchment states from one catchment to another (e.g. *Hrachowitz et al., 2013*) ~~and~~as well as offers the possibility to aggregate structurally similar sub-units ~~of a system~~ and simulate their ~~functioning~~function by a ~~single~~ representative~~,~~ ~~as long as they are in a similar state and driven by a similar forcing~~ (*e.g. Sivapalan et al., 1987; Zehe et al., 2014*)~~. The fact that hydrological systems are highly dissipative but constrained by there structure is thereby the key to explain the feasibility of this dynamic grouping as the unique characteristics of the forcing over an area do not prevail but are depleted or "*forgotten*" in a relatively short time, at least if the focus is on the runoff generation. Specifically, we found during both events that already after 1 hr of no rainfall the spatially adaptive model required only a single hillslope model to represent the diversity in the runoff generation between the models. While this finding is surely constrained by the chosen threshold, the picture is nevertheless quite different in deeper soil layers where the diversity of the rainfall forcing leads even after 24 hrs to increasing differences between the "driest" and "wettest" models. A part of the information about the different meteorological forcings between the two models is hence still stored in the model state after 24 hrs and has not yet been dissipated. The importance of those differences likely depends on the dominant runoff generation process. In the present case, they have a minor impact as *model b* and *c* show similar average baseflow simulations 24 hrs after the rainfall event I and II although *model c* uses only a single hillslope model (difference smaller than 0.001 mm hr^{-1}).~~

. The fact that hydrological systems are highly dissipative (*Loritz et al., 2019*) but constrained by their structural setting is thereby the key to explain the feasibility of this aggregation as the unique characteristics of the forcing over an area do not prevail but are depleted or "*forgotten*" in a relatively short time, at least if the focus is on the runoff generation. Specifically, we found during both tested events that already after 1 hr of no rainfall the spatially adaptive *model c* required only a single hillslope model to represent the runoff generation of the Colpach. While this finding is surely constrained by the chosen thresholds of the two selected similar metrics (dQ dt^{-1} and Q) and the chosen time frame, the picture is underpinned by the soil moisture distributions of the model elements of the spatially adaptive *model c* that are also close to similar at the time step when they are aggregated.

Nonetheless, another virtual experiment showed that there are clear limitations to the proposed approach and the chosen parameters. We could demonstrate that two hillslope models that received significant dissimilar precipitation amounts (>15 mm 5 hr^{-1}) showed differences regarding their soil moisture distributions in 60-100 cm 24 hrs after the last rainfall although the runoff generation at this time step were close to similar. The latter means that there could be specific constellations when we aggregate hillslope models using the chosen similarity measures dQ dt^{-1} and Q and thereby remove relevant information about the different model states from our ensemble. This is the case as we can simulate the same flux by combining different combinations of driving potentials with integral resistance terms, a phenomena that is inherent to all our governing equations and sometimes referred to equifinality in hydrological modeling (*Beven, 1993; Loritz et al., 2019; Zehe et al., 2014*). This highlights that the similarity metrics that are used to group similar models by their state should be chosen with care and need to be adapted to the given research environment and process under study. For instance, in a snow-dominated area we need to group model states not only based on their runoff production but also based on their snow cover. The choice of dQ dt^{-1} and Q in this study seems, however, to be sufficient to identify similar model elements at least as long as we focus on the summer season. This is the case as our hillslope models are all structurally identically and only simulate shallow subsurface storm flow during the entire summer season. We can hence assume that we do have a rather unique relationship between our model states and the chosen similarity metrics $dQ\ dt^{-1}$ and Q. This is underpinned by the fact that the individual hillslope models of the distributed *model b,* which reflect the highest spatial resolution of the spatially adaptive *model c*, are not drifting apart in the chosen summer season. Contrary they are mainly producing redundant simulations already shortly after each rainfall event at least as long as we focus on the summer season (see supplement). The latter means, however, also that the drawn conclusions are not necessarily true for the winter season where we have not tested the adaptive model as the distributed *model b* and spatially aggregated *reference model* perform close to similar. Nevertheless, a test of the proposed spatially adaptive modelling approach on a longer time scale is an interesting task for further research.

While the structure of a catchment constraints its state space, its actual position therein is controlled by the meteorological forcing and by an attracting local thermodynamic equilibrium, a point where all driving gradients are depleted. As larger gradients dissipate faster than smaller ones~~ if~~, as long as they are controlled by the same integral resistance properties, ~~structural~~structurally similar parts of a landscape will converge to the same state and thereby ~~-~~“*forget*“ differences between their past forcing and current state. This convergence leads to the emergence of hydrological similarity in time (*Loritz et al., 2018*) and explains the changing relevance of distributed rainfall~~.~~ within hydrological models. This again is the theoretical ~~ground~~foundation that explains why adaptive modeling works in hydrological systems and not necessarily in meteorological systems as their chaotic nature can amplify state differences on longer time scales, instead of dissipating those (e.g. *Lorenz, 1963*). Our developed adaptive modeling approach is using this straightforward physical reasoning of the causal and dissipative interplay between the precipitation forcing and the catchment state to dynamically allocate its model structure~~.~~ during model execution. It is built upon a well-established concept in hydrology, which ~~states~~assumes that individual observations or model states can represent each other if they are allocated to the same group (e.g. *Wood et al., 1990*). The related bin widths (grouping) can be selected either based on our physical understanding (*Loritz et al., 2018*) or identified based on a statistical analysis of the underlying distribution (of for instance the precipitation data; e.g. *Gong et al., 2014; Scott, 1979*). The general approach is strongly motivated by the idea that a spatially homogeneous field can be compressed to a single time series without losing information about the spatial pattern of rainfall. This is, however, not the case if the spatial field is highly variable where a compression to a single observation reduces the information provided to a hydrological model and hence can average out extremes and potentially relevant spatial constellations (e.g. *Loritz et al., 2018; Weijs et al., 2013)*. Spatially adaptive modeling can, therefore, be used not only as a tool to ~~analyze~~reduce computational times but to analyse the relevance of certain spatial detail in a hydrological model and therefore as ~~well as~~a tool to better understand the dissipative nature of hydrology.

67. Conclusions

~~In this study, we try to better understand the role and value of distributed precipitation data for the runoff generation of a mesoscale catchment. We compare the model performances of three hydrological models at different periods and show that a distributed model driven by distributed rainfall yields only to improved performances during certain periods. We then step beyond this finding and develop a spatially adaptive model that is capable to dynamically adjust its spatial model structure in time. This model is capable to represent the varying importance of distributed rainfall within a hydrological model without losing performance compared to a spatially distributed,~~

~~gridded model. Our results confirm that spatially adaptive modeling might be a) one way to reduce computational times as already shown by *Ehret et al. (2020), b)* can be used to better understand the varying importance of spatial state and forcing differences in hydrological models and c) highlight that similarity between the runoff generation of two hillslopes does not necessarily mean similarity between other state variables (e.g. soil moisture in deeper soil layers).~~

In this study, we ~~try to improve our understanding of~~ the role and value of distributed precipitation data for the runoff generation of a mesoscale catchment. We therefore compare the model performances of three model setups at different periods and show that a distributed model driven by distributed rainfall yields only to improved performances during certain periods. We then step beyond this finding and develop a spatially adaptive model that is able to dynamically adjust its spatial model structure in time. This model is capable to represent the varying importance of distributed rainfall within a hydrological model without losing predictive performance compared to a spatially distributed, gridded model. Our results confirm that spatially adaptive modeling might be one way to reduce computational times in physically-based hydrological simulations as well as be used as tool to better understand the causal and dissipative interplay between a catchment's state and its meteorological forcing.

The main findings of this study are:

1) The importance of distributed rainfall on hydrological modeling is given by the natural variability of rainfall dynamically changing in time. In consequence, there cannot be a time-invariant answer to the question "*How important are spatial patterns of precipitation for the runoff generation at the catchment scale?*" nor to any related question which deals with an "*optimal*" spatial discretization of a hydrological landscape within a model.
2) Spatially adaptive modeling is a feasible way to account for the changing importance of distributed rainfall within a hydrological model and at the same time can be used as a tool to ~~better understand~~improve our understanding of the interplay ~~of the~~between rainfall forcing,~~ the~~ catchment structure, and its state.
3) ~~The tested catchment is organized in a manner that spatial differences between the precipitation forcing are effectively "*forgotten*". This entails that gradients, which drive runoff, are effectively dissipated in a relatively short period. This period might, however, be quite different for other fluxes and state variables depending on the dominant runoff generation process.~~

~~**Appendix A: Detailed description of the distributed rainfall data.**~~

~~The distributed precipitation data used in this study is based on single-polarization C-band Doppler radar measurements. The mainly used radar data is from the radar located in Neuheilenbach, Germany and operated by the German Weather Service (DWD). The raw volume data set has an azimuthal resolution of 1° and a radial resolution of 500 m. The -3dB beamwidth of the antenna is 1°. The radar site is between 40 and 70 km away from the study area. This means that the resolution is yet neither significantly degraded by the beam spreading, nor partial blinded through cone of silence issues. During the period from the 1st of October 2013 to the 27th of March 2014, the radar in Neuheilenbach was out of service due to maintenance issues. We hence used data from a radar located in Wideumont, Belgium in this period. The radar in Wideumont is operated by the Royal Meteorological Institute of Belgium (RMI) and is also a C-band Doppler radar with the same technical specifications as the radar of the DWD. The distance between radar site in Wideumont and the study area is between 24 to 44 km. Thus, the same statements about the resolution, which were made in the case of the data from Neuheilenbach, also apply to the radar data of Wideumont.~~

~~The data was quality controlled and a correction was performed. The particular raw data was at first filtered by a static clutter filter and then also by a Doppler clutter filter. Subsequently, a bright band correction (*Hannesen, 1998*) was applied. Occasional contamination of the data by second trip or anaprop echoes was removed by using approaches of *Bückle (2009)* and *Neuper (2009)*. Specific attenuation corrections were not applied. Furthermore, the data was carefully quality checked by an experienced radar meteorologist and operational weather forecaster, who even spends his spare time watching radar pictures. From the corrected data a pseudo PPI (plan position indicator) data set at 1500m above ground was created and afterward an adequate (based on the synoptic situation) reflectivity-rain rate relation (Z-R relation) was applied to compute the precipitation rate (e.g. *Fabry, 2015*). In the last step, the distributed precipitation fields were checked against quality controlled rain gauges and if necessary manually corrected.~~

3) Hydrological landscapes are organized in a manner that spatial differences within the precipitation forcing are "*forgotten*" or smooth out in often surprisingly short period when rainfall becomes runoff. This entails that gradients that drive runoff are effectively dissipated, which happens frequently under the influence of preferential flow (e.g. *Berkowitz and Zehe, 2020*). The dissipative nature of hydrological processes combined with the observation that structural similar hydrological landscape are also functionally organized alike explains thereby why hydrological similarity must be a time invariant concept and why spatially adaptive modelling is a physical reasonable way to represent hydrological systems with a model.

Data availability. The hydrological model CATFLOW and all simulation results are available from the leading author on request. The rainfall data were provided by fourth author Malte Neuper from the Karlsruhe Institute of Technology. The discharge observations were provided by the Luxembourg Institute of Science and Technology within the "Catchments As Organized Systems (CAOS)" research group (FOR 1598) funded by the German Science Foundation (DFG). Please contact Laurent Pfister or Jean-Francois Iffly.

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

Acknowledgements. This research contributes to the "Catchments As Organized Systems (CAOS)" research group (FOR 1598) funded by the German Science Foundation (DFG ZE 533/11-1, ZE 533/12-1). Laurent Pfister and Jean-Francois Iffly from the Luxembourg Institute of Science and Technology (LIST) are acknowledged for organizing the permissions for the experiments and providing discharge data and the digital elevation model. We also thank the whole CAOS team of phase I & II.

References

Berger, M. J. and Oliger, J.: Adaptive mesh refinement for hyperbolic partial differential equations, J. Comput. Phys., 53(3), 484–512, doi:10.1016/0021-9991(84)90073-1, 1984.

Berkowitz, B. and Zehe, E.: Surface water and groundwater: Unifying conceptualization and quantification of the two “water worlds,” Hydrol. Earth Syst. Sci., 24(4), 1831–1858, doi:10.5194/hess-24-1831-2020, 2020.

Beven, K.: Changing ideas in hydrology — The case of physically-based models, J. Hydrol., 105(1–2), 157–172, doi:10.1016/0022-1694(89)90101-7, 1989.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, Adv. Water Resour., 16(1), 41–51, doi:10.1016/0309-1708(93)90028-E, 1993.

Beven, K.: Dalton Lecture: How Far Can We Go In Distributed Hydrological Modelling? Keith Beven Lancaster University, Computer (Long. Beach. Calif)., 5(Figure 2), 1–12 [online] Available from: http://eprints.lancs.ac.uk/4420/, 2001.

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrol. Process., 6(3), 279–298, doi:10.1002/hyp.3360060305, 1992.

Beven, K. and Freer, J.: A dynamic topmodel, Hydrol. Process., 15(10), 1993–2011, doi:10.1002/hyp.252, 2001.

Beven, K. J. and Hornberger, G. M.: Assessing the effect of spatial pattern of precipitation in modeling stream flow hydrographs, J. Am. Water Resour. Assoc., 18(5), 823–829, doi:10.1111/j.1752-1688.1982.tb00078.x, 1982.

~~Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H. and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next?, Hydrol. Process., 29(2), 310-320, doi:10.1002/hyp.10391, 2015.~~

~~Bückle, J.: Korrektur von Second Trip Echos in Radardaten, Karlsruhe Institute of Technology., 2010.~~

Celia, M. A., Bouloutas, E. T. and Zarba, R. L.: A general mass-conservative numerical solution for the unsaturated flow equation, Water Resour. Res., 26(7), 1483–1496, doi:10.1029/WR026i007p01483, 1990.

Chaney, N. W., Metcalfe, P. and Wood, E. F.: HydroBlocks: a field-scale resolving land surface model for application over continental extents, Hydrol. Process., 30(20), 3543–3559, doi:10.1002/hyp.10891, 2016.

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(7), 3427–3440, doi:10.5194/hess-21-3427-2017, 2017.

Das, T., Bárdossy, A., Zehe, E. and He, Y.: Comparison of conceptual model performance using different representations of spatial variability, J. Hydrol., 356(1–2), 106–118, doi:10.1016/j.jhydrol.2008.04.008, 2008.

~~Ehret, U.,~~ van Pruijssen, R~~. Van,~~., Bortoli, M., Loritz, R., Azmi, E. and Zehe, E.: Adaptive clustering~~: A method to analyze dynamical similarity and to reduce redundancies in~~: reducing the computational costs of distributed (hydrological~~) modeling, , (February), 1 33~~) modelling by exploiting time-variable similarity among model elements, Hydrol. Earth Syst. Sci., 24(9), 4389–4411, doi:~~https://doi.org/~~10.5194/hess-24-4389-2020~~-65~~, 2020.

Euser, T., Hrachowitz, M., Winsemius, H. C. and Savenije, H. H. G.: The effect of forcing and landscape distribution on performance and consistency of model structures, Hydrol. Process., 29(17), 3727–3743, doi:10.1002/hyp.10445, 2015.

~~Fabry, F.: Radar Meteorology, Cambridge University Press, Cambridge., 2015.~~

Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., Brooks, P. D., Dietrich, W. E., Flores, A., Grant, G., Kirchner, J. W., Mackay, D. S., McDonnell, J. J., Milly, P. C. D., Sullivan, P. L., Tague, C., Ajami, H., Chaney, N., Hartmann, A., Hazenberg, P., McNamara, J., Pelletier, J., Perket, J., Rouholahnejad-Freund, E., Wagener, T., Zeng, X., Beighley, E., Buzan, J., Huang, M., Livneh, B., Mohanty, B. P., Nijssen, B., Safeeq, M., Shen, C., Verseveld, W., Volk, J. and Yamazaki, D.: Hillslope Hydrology in Global Change Research and Earth System Modeling, Water Resour. Res., 55(2), 1737–1772, doi:10.1029/2018WR023903, 2019.

Fenicia, F., Kavetski, D. and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47(11), 1–13, doi:10.1029/2010WR010174, ~~2011a~~2011.

~~Fenicia, F., Kavetski, D. and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47(11), 1–13, doi:10.1029/2010WR010174, 2011b.~~

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 Resour. Res., 52(2), 954–989, doi:10.1002/2015WR017398, 2016.

Freeze, R. A. and Harlan, R. L.: Blueprint for a physically-based, digitally-simulated hydrologic response model, J. Hydrol., 9(3), 237–258, doi:10.1016/0022-1694(69)90020-1, 1969.

Gong, W., Yang, D., Gupta, H. V. and Nearing, G.: Estimating information entropy for hydrological data: One-dimensional case, Water Resour. Res., 50(6), 5003–5018, doi:10.1002/2014WR015874, 2014.

Hannesen: Analyse konvektiver Niederschlagssysteme mit einem C-Band Dopplerradar in orographisch gegliedertem Gelände, University of Karlsruhe., 1998.

Höge, M., ~~Wöhling, T~~Guthke, A. and Nowak, W.: ~~A Primer for Model Selection: ~~The ~~Decisive Role of Model Complexity, Water Resour. Res., 54(3), 1688–1715~~hydrologist's guide to Bayesian model selection, averaging and combination, J. Hydrol., 572, 96–107, doi:10.~~1002/2017WR021902, 2018~~1016/j.jhydrol.2019.01.072, 2019.

Hrachowitz, M. and Clark, M. P.: HESS Opinions: The complementary merits of competing modelling philosophies in hydrology, Hydrol. Earth Syst. Sci., 21(8), 3953–3973, doi:10.5194/hess-21-3953-2017, 2017.

Hrachowitz, M. and Weiler, M.: Uncertainty of Precipitation Estimates Caused by Sparse Gauging Networks in a Small, Mountainous Watershed, J. Hydrol. Eng., 16(5), 460–471, doi:10.1061/(ASCE)HE.1943-5584.0000331, 2011.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, a., Gupta, H. V., Hughes, D. a., Hut, R. W., Montanari, a., Pande, S., Tetzlaff, D., Troch, P. a., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. a., Zehe, E. and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrol. Sci. J., 58(6), 1198–1255, doi:10.1080/02626667.2013.803183, 2013.

Huang, Y., Bárdossy, A. and Zhang, K.: Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data, Hydrol. Earth Syst. Sci., 23(6), 2647–2663, doi:10.5194/hess-23-2647-2019, 2019.

Jackisch, C.: Linking structure and functioning of hydrological systems., KIT - Karlsruher Institut of Technology., 2015.

Jackisch, C., Angermann, L., Allroggen, N., Sprenger, M., Blume, T., Tronicke, J. and Zehe, E.: Form and function in hillslope hydrology: in situ imaging and characterization of flow relevant structures, Hydrol. Earth Syst. Sci., 21(7), 3749–3775, doi:10.5194/hess-21-3749-2017, 2017.

Juilleret, J., Iffly, J. F., Pfister, L. and Hissler, C.: Remarkable Pleistocene periglacial slope deposits in Luxembourg (Oesling): pedological implication and geosite potential, Bull. la Société des Nat. Luxemb., 112(1), 125–130, 2011.

Kling, H. and Gupta, H.: On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability, J. Hydrol., 373(3–4), 337–351, doi:10.1016/j.jhydrol.2009.04.031, 2009.

Knoben, W. J. M., Freer, J. E. and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrol. Earth Syst. Sci., 23(10), 4323–4331, doi:10.5194/hess-23-4323-2019, 2019.

Leopold, L. B.: Downstream change of velocity in rivers, Am. J. Sci., 251(8), 606–624, doi:10.2475/ajs.251.8.606, 1953.

Lobligeois, F., Andréassian, V., Perrin, C., Tabary, P. and Loumagne, C.: When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation using 3620 flood events, Hydrol. Earth Syst. Sci., 18(2), 575–594, doi:10.5194/hess-18-575-2014, 2014.

Lorenz, E. N.: Deterministic Nonperiodic Flow, J. Atmos. Sci., 20(2), 130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2, 1963.

Loritz, R.: The role of energy and information in hydrological modeling, Karlsruhe Institute of Technology (KIT)., 2019.

Loritz, R., Hassler, S. K., Jackisch, C., Allroggen, N., van Schaik, L., Wienhöfer, J. and Zehe, E.: Picturing and modeling catchments by representative hillslopes, Hydrol. Earth Syst. Sci., 21(2), 1225–1249, doi:10.5194/hess-21-1225-2017, 2017.

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(7), 3663–3684, doi:10.5194/hess-22-3663-2018, 2018.

Loritz, R., Kleidon, A., Jackisch, C., Westhoff, M., Ehret, U., Gupta, H. and Zehe, E.: A topographic index explaining hydrological similarity by accounting for the joint controls of runoff formation, Hydrol. Earth Syst. Sci., 23(9), 3807–3821, doi:10.5194/hess-23-3807-2019, 2019.

Mälicke, M., Hassler, S~~. K~~., Blume, T., Weiler, M. and Zehe, E.: Soil moisture: variable in space but redundant in time, Hydrol. Earth Syst. Sci~~., 24(5), 2633–2653~~. Discuss., (November), 1–28, doi:10.5194/hess-~~24-2633-2020, 2020~~2019-574, 2019.

Martínez-Carreras, N., Hissler, C., Gourdol, L., Klaus, J., Juilleret, J., François Iffly, J. and Pfister, L.: Storage controls on the generation of double peak hydrographs in a forested headwater catchment, J. Hydrol., 543, 255–269, doi:10.1016/j.jhydrol.2016.10.004, 2016.

Maurer, T.: Physikalisch begründete zeitkontinuierliche Modellierung des Wassertransports in kleinen ländlichen Einzugsgebieten., Karlsruher Institut für Technologie., 1997.

Melsen, L. A., Teuling, A. J., Torfs, P. J. J. F., Uijlenhoet, R., Mizukami, N. and Clark, M. P.: HESS Opinions: The need for process-based evaluation of large-domain hyper-resolution models, Hydrol. Earth Syst. Sci., 20(3), 1069–1079, doi:10.5194/hess-20-1069-2016, 2016.

Minkoff, S. E. and Kridler, N. M.: A comparison of adaptive time stepping methods for coupled flow and deformation modeling, Appl. Math. Model., 30(9), 993–1009, doi:10.1016/j.apm.2005.08.002, 2006.

~~Neuper, M.: Anomale Strahlausbreitung – Prinzip und Fallbeispiele, Karlsruhe Institute of Technology., 2009.~~

Neuper, M. and Ehret, U.: Quantitative precipitation estimation with weather radar using a data- and information-based approach, Hydrol. Earth Syst. Sci., 23(9), 3711–3733, doi:10.5194/hess-23-3711-2019, 2019.

Nicótina, L., Alessi Celegon, E., Rinaldo, A. and Marani, M.: On the impact of rainfall patterns on the hydrologic response, Water Resour. Res., 44(12), 1–14, doi:10.1029/2007WR006654, 2008.

Obled, C., Wendling, J. and Beven, K.: The sensitivity of hydrological models to spatial rainfall patterns: an evaluation using observed data, J. Hydrol., 159(1–4), 305–333, doi:10.1016/0022-1694(94)90263-1, 1994.

Ogden, F. L. and Julien, P. Y.: Runoff sensitivity to temporal and spatial rainfall variability at runoff plane and small basin scales, Water Resour. Res., 29(8), 2589–2597, doi:10.1029/93WR00924, 1993.

Or, D., Lehmann, P. and Assouline, S.: Natural length scales define the range of applicability of the Richards equation for capillary flows, Water Resour. Res., 51(9), 7130–7144, doi:10.1002/2015WR017034, 2015.

Peleg, N., Blumensaat, F., Molnar, P., Fatichi, S. and Burlando, P.: Partitioning the impacts of spatial and climatological rainfall variability in urban drainage modeling, Hydrol. Earth Syst. Sci., 21(3), 1559–1572, doi:10.5194/hess-21-1559-2017, 2017.

Pfister, L., Martínez-Carreras, N., Hissler, C., Klaus, J., Carrer, G. E., Stewart, M. K. and McDonnell, J. J.: Recent Trends in Rainfall-Runoff Characteristics in the Alzette River Basin, Luxembourg, Hydrol. Process., 31(10), 1828–1845, doi:10.1023/A:1005567808533, 2017.

Pfister, L., Hissler, C., Iffly, J. F., Coenders, M., Teuling, R., Arens, A. and Cammeraat, L. H.: Contrasting Hydrologic Response in the Cuesta Landscapes of Luxembourg, in The Luxembourg Gutland Landscape, edited by A. M. Kooijman, L. H. Cammeraat, and A. C. Seijmonsbergen, pp. 73–87, Springer International Publishing, Cham., 2018.

Prenner, D., Kaitna, R., Mostbauer, K. and Hrachowitz, M.: The Value of Using Multiple Hydrometeorological Variables to Predict Temporal Debris Flow Susceptibility in an Alpine Environment, Water Resour. Res., 54(9), 6822–6843, doi:10.1029/2018WR022985, 2018.

Schoups, G., van de Giesen, N. C. and Savenije, H. H. G.: Model complexity control for hydrologic prediction, Water Resour. Res., 44(12), 14, doi:10.1029/2008WR006836, 2008.

Scott, D. W.: On Optimal and Data-Based Histograms, Biometrika, 66(3), 605, doi:10.2307/2335182, 1979.

Seibert, S. P., Jackisch, C., Ehret, U., Pfister, L. and Zehe, E.: Unravelling abiotic and biotic controls on the seasonal water balance using data-driven dimensionless diagnostics, Hydrol. Earth Syst. Sci., 21(6), 2817–2841, doi:10.5194/hess-21-2817-2017, 2017.

Sivapalan, M., Beven, K. J. and Woods, E.: On Hydrologic Similarity, , 23(12), 2266–2278, 1987.

Smith, M. B., Seo, D. J., Koren, V. I., Reed, S. M., Zhang, Z., Duan, Q., Moreda, F. and Cong, S.: The distributed model intercomparison project (DMIP): Motivation and experiment design, J. Hydrol., 298(1–4), 4–26, doi:10.1016/j.jhydrol.2004.03.040, 2004.

Sternagel, A., Loritz, R., Wilcke, W. and Zehe, E.: Simulating preferential soil water flow and tracer transport using the Lagrangian Soil Water and Solute Transport Model, Hydrol. Earth Syst. Sci., 23(10), 4249–4267, doi:10.5194/hess-23-4249-2019, 2019.

Vogel, H.-J. and Ippisch, O.: Estimation of a Critical Spatial Discretization Limit for Solving Richards' Equation at Large Scales, Vadose Zo. J., 7(1), 112–114, doi:10.2136/vzj2006.0182, 2008.

~~Wagener, T., Sivapalan, M., Troch, P. and Woods, R.: Catchment Classification and Hydrologic Similarity, Geogr. Compass, 1(4), 901–931, doi:10.1111/j.1749-8198.2007.00039.x, 2007.~~

Watts, L. G. and Calver, A.: Effects of spatially-distributed rainfall on runoff for a conceptual catchment, Nord. Hydrol., 22(1), 1–14, doi:10.2166/nh.1991.0001, 1991.

Weijs, S. V., van de Giesen, N. and Parlange, M. B.: Data compression to define information content of hydrological time series, Hydrol. Earth Syst. Sci., 17(8), 3171–3187, doi:10.5194/hess-17-3171-2013, 2013.

Wienhöfer, J. and Zehe, E.: Predicting subsurface stormflow response of a forested hillslope – the role of connected flow paths, Hydrol. Earth Syst. Sci., 18(1), 121–138, doi:10.5194/hess-18-121-2014, 2014.

Wilson, C. B., Valdes, J. B. and Rodriguez-Iturbe, I.: On the influence of the spatial distribution of rainfall on storm runoff, Water Resour. Res., 15(2), 321–328, doi:10.1029/WR015i002p00321, 1979.

Wood, E. F., Sivapalan, M. and Beven, K.: Similarity and scale in catchment storm response, Rev. Geophys., 28(1), 1, doi:10.1029/RG028i001p00001, 1990.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A. and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resour. Res., 47(5), 1–10, doi:10.1029/2010WR010090, 2011.

Woods, R. and Sivapalan, M.: A synthesis of space-time variability in storm response: Rainfall, runoff generation, and routing, Water Resour. Res., 35(8), 2469–2485, doi:10.1029/1999WR900014, 1999.

Wrede, S., Fenicia, F., Martínez-Carreras, N., Juilleret, J., Hissler, C., Krein, A., Savenije, H. H. G., Uhlenbrook, S., Kavetski, D. and Pfister, L.: Towards more systematic perceptual model development: a case study using 3 Luxembourgish catchments, Hydrol. Process., 29(12), 2731–2750, doi:10.1002/hyp.10393, 2015.

Zehe, E. and Blöschl, G.: Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions, Water Resour. Res., 40(10), 1–21, doi:10.1029/2003WR002869, 2004.

Zehe, E., Maurer, T., Ihringer, J. and Plate, E.: Modeling water flow and mass transport in a loess catchment, Phys. Chem. Earth, Part B Hydrol. Ocean. Atmos., 26(7–8), 487–507, doi:10.1016/S1464-1909(01)00041-7, 2001.

Zehe, E., Becker, R., Bárdossy, A. and Plate, E.: Uncertainty of simulated catchment runoff response in the presence of threshold processes: Role of initial soil moisture and precipitation, J. Hydrol., 315(1–4), 183–202, doi:10.1016/j.jhydrol.2005.03.038, 2005.

Zehe, E., Elsenbeer, H., Lindenmaier, F., Schulz, K. and Blöschl, G.: Patterns of predictability in hydrological threshold systems, Water Resour. Res., 43(7), doi:10.1029/2006WR005589, 2007.

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., van Schaik, L., 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, Hydrol. Earth Syst. Sci., 18(11), 4635–4655, doi:10.5194/hess-18-4635-2014, 2014.

Zhu, Z., Wright, D. B. and Yu, G.: The Impact of Rainfall Space-Time Structure in Flood Frequency Analysis, Water Resour. Res., 54(11), 8983–8998, doi:10.1029/2018WR023550, 2018.

Reply to Referee #1 Daniel Wright:

Daniel Wright ***(DW): Summary and Recommendation:*** *The authors present a framework for dynamically adapting the level of spatial detail re-solved within a physics-based rainfall-runoff model depending on the spatial variability in precipitation. I found this the be one of the most interesting manuscripts that I've ever reviewed, and commend the authors on this innovative work. Nonetheless, there are some issues that should be addressed before the manuscript is suitable for publication in HESS, and that could help maximize the impact of the work.*

Ralf Loritz (RL): We would like to thank Daniel Wright for his positive comments and the time he invested to review our Manuscript (MS). The revised MS will follow the reviewer's recommendations and include among other things a re-structured section 3 (model introduction) as well as a more extensive discussion about its connection to the land surface modeling community. Furthermore, will we carefully check the references the reviewer recommended and see whether they help us to improve our argumentation.

Major comments:

1. DW: *I believe the discussion could be strengthened by deeper consideration of how this approach would "scale up" to larger watersheds or regions. Part of my reason for encouraging this is that the land surface modeling (LSM) community is at least as concerned as the rainfall-runoff community about model computational demands of long-term/ensemble simulations, and are seeking ways of representing fine-scale (e.g. hillslope and below) over continental-to-global domains. In fact, land surface modeling was the focus of the well-known Wood et al. (2011) hyperresolution modeling opinion piece. In addition, there has been relevant progress in LSM development that the authors should cite. I will mention these below. But in terms of scaling up, the key aspects seem to be acknowledgement that heterogeneity of model parameters will increase with modeled area, while the rainfall spatial coverage will, on average, decrease.*

RL: Thank you for raising this point. We mentioned in our MS in the method section page 14 line 28 to 31: "*This entails, however, also that if we extend our research area to a catchment that is divided, for instance, into two geological settings that function hydrologically differently (regarding their filter properties) we would always need to run at least two structural different models where each of these models represents one of two geological settings.*"

We agree with the reviewer that this section is rather short and will provide a more extensive discussion in a revisited MS (see also the following points). We will also carefully read the proposed references by Woods et al. (2011).

2. DW: *I believe the discussion could also be strengthened by some discussion of how well this approach might fit with specific types of spatial discretizations. It fits quite naturally with hillslope-based models.*

The fit is less clear with gridded or TIN-based models-or at least with high-resolution gridded models in which individual model grids must "communicate" with each other to transmit water via overland or subsurface flow to channels.

RL: This is an important point. Indeed our approach is limited to hydrological models that are based on a division of the landscape into partly independent spatial units (similar to the work of *Chaney et al. 2016*). However, at least in theory, there is no limit on how complex the interaction between these independent sub-units are as long as there is redundancy/similarity when different model elements "*communicate"* with each other. However, the question at what point of model complexity we would still save computational times by reducing redundancy depends on a series of factors (e.g. model, resolution, no. of processes and state variables). We will discuss this in a revisited MS.

DW: *It seems that the computational advantages of the approach might be limited in that case. In addition, models such as GSSHA in which overbank river flow can return to the land surface would have some limits here too. These issues are worth discussing because such models constitute important current directions in physics-based model development.*

RL: From our perspective, it makes much sense to divide a landscape into different building blocks such as hillslopes, sub-basins, etc. (*e.g. Zehe et al., 2014*). This is the case as current physically-based models are still constrained to small areas if they are set up on an appropriate grid size. We see hence no way around dividing a landscape into some kind of independent sub-units and either run models in parallel or/and group similar model elements (dynamically or time-invariant) if we want to work on larger scales. That said, we also believe that it would be rather difficult to implement a spatial adaptive modeling approach in a current model like Delft2d or GSSHA.

We wrote in our MS (Pg. 16 line 10): "*While we use CATFLOW as a model here, the proposed approach is not restricted to this model and can be used in any hydrological model that distributes a catchment into independent spatial units.*". To underpin this point we will discuss the limitation of our approach in more detail (please also see the discussion with the second referee and the third). Again we thank DW for this valuable comment.

3. DW: *While there may be other relevant LSM developments, the one that I am aware of is Hydroblocks (Chaney et al. 2016). While I recommend reading that paper, the basic approach is similar to this manuscript's in that spatial units are grouped into hydrologically similar clusters to reduce the computational demand.*

RL: Thank you very much for pointing us to the study of *Chaney et al. (2016)*. We will examine it carefully.

DW: *The difference is that in Hydroblocks, these clusters are not dynamically reassigned according to time-varying characteristics (unless the developers have recently added that capability). So in fact, your* *approach appears to be superior in some respects. Specifically, within Hydroblocks, since there is no dynamic reassignment, you can never have a cluster that extends beyond the spatial extent of a single precipitation grid cell, which means that their approach loses computational efficiency with higher-resolution precipitation datasets. Your approach thus seems to hold more promise in terms of flexibility to advances in precipitation inputs*.

RL: Interesting comment. We would like to highlight that we only showed that our approach is theoretical and practically feasible. It remains an open question if we could actually save more computational times in comparison to HydroBlocks or similar time-invariant approaches. The question is open for discussion as our approach is also more complicated. We will discuss this in a revised discussion section.

4. DW: *More clear description of what each model does and does not do is needed in Section 3. Specifically, I found it confusing the way that the models are briefly introduced at the beginning of the section, and then discussed further in various subsections. I also find it strange that you have text that is not assigned to specific subsections. It isn't clear why section 3.2.1 is needed...convention is that you don't include subsections unless you have at least 2 or 3 (i.e. 3.2.2, 3.2.3). This section structuring needs rethinking. […] Also, a table that compares the key features and differences of all the models could be effective. I think one think that would really help is to not use "model a", "model b", etc. but some brief descriptive names that actually help the reader understand and recall the differences. More important, I really couldn't figure out from the descriptions what the differences between some models were. I also don't understand the motivation for using a different rainfall dataset for the reference model and model a; this seems unnecessary.*

RL: We stated in our MS on page 11 line 26-29: "*We added model a to test if the performance difference between the reference model and our distributed model b is merely a result of quantitative differences between the different precipitation products measured either by a single ground station or by a weather radar.*"

However, we understand the comment of Daniel Wright and we will restructure section 3 entirely and remove the short introduction of the different models. We will also follow your advice and add a table with the key features of each model.

5. DW: *Zhu et al. (2018) and Peleg et al. (2017) both highlight how distributed rainfall structure is really important in determining flood frequency across a range of scales. Though I normally refrain from suggesting that authors cite my own work, in this case it seems appropriate to highlight these studies, since they do show that for extreme events, rainfall space-time structure is extremely important in determining hydrologic response even at very small scales (see Peleg et al. in particular), and that this importance varies with rainfall magnitude and basin size.*

RL: We have carefully read both publications and they fit nicely into our revisited discussion. Thank you for pointing us towards these two references.

DW: *Along with this, I disagree with the statement on pg. 27: "it seems that catchment size might not be the best indicator to decide if" a distributed model is needed. It probably is the best single indicator, but is still insufficient. I draw a somewhat different conclusion from your work: that a distributed approach is always needed to reap the full benefit of spatially distributed rainfall (at least in locations in which convective rainfall can occur), and that provides motivation for continued developments such as this into ways of handling this need in computationally-efficient ways.*

RL: What we wanted to convey here is that it is the combination of the drainage area and the average size of a typical rainstorm, which is important and not the drainage area alone. For instance, if you wanted to predict the runoff formation in the Colpach catchment in the winter season a spatially aggregated model driven by a single precipitation time series might be sufficient as our results show. This means also that you could invest your limited time and improve for instance the groundwater representation in your model instead of setting up a distributed model. However, if you wanted to make predictions in the summer months our results highlight that you need some sort of a distributed model to be able to capture the spatial variability of the rainfall. This means that only because the Colpach is 20 km^2 we cannot decide if we need a spatially distributed model as the catchment size does not explain how variable its meteorological forcing is. *Nicotina et al. (2008)* argued along these lines and stated that the "*total residence time of a water parcel is often controlled by the travel time within hillslopes, we find that when typical hillslope size is smaller than the characteristic size of rainfall structures (say, a correlation length of rainfall intensity), the rainfall pattern effectively samples all possible residence times and the response of the catchment does not depend on the specific rainfall pattern.*" and the second referee pointed us towards the study of *Ogden and Julien (1993)*. The second reviewer also nicely summarized their key finding: *"only for rainfall with durations shorter than the concentration time of a catchment does the spatial distribution of the rainfall matters, for longer rainfall events only the temporal distribution matters."*. Following these two studies and our own results we would argue that our first research question in our MS: "How important are spatial patterns of precipitation for the runoff generation at the catchment scale?" can only be answered if we combine information about the catchment size (e.g. average hillslope length, concentration time) with information about the meteorological forcing (e.g. intensity, correlation length, velocity). In a revisited MS we will rephrase this paragraph and explain in more detail what we meant by this statement.

DW: *Likewise, I disagree with the statement on pg. 30 line 18-19: compressing rainfall into a single time series isn't so important as the ability to only use as much computational power as is truly needed to solve the problem at hand.*

RL: Again an interesting point you raise here. In our specific setting, compression of precipitation and saving computational power are the same. By compressing the precipitation field to a single time series we also compress our model, minimize redundant calculations, which again means that we save computational power. So, we would argue that we first need to understand (test) how far we can compress our rainfall field without losing predictive performance before we can save computational times in a meaningful manner. The data-based / machine learning community most likely would disagree ☺

6. DW: *Some discussion of implications for calibration would be interesting. Is it necessary to calibrate using a fully distributed model? This would limit the usefulness of this approach in some respects such as automated calibration procedures.*

RL: Typically, one run of the reference model (a single CATLOW hillslope) for a simulation period of one year and hourly printout times takes about 2 - 3 hrs. Assuming that you run your code on a workstation with 32 cores you can run about 400 model setups in 24 hrs. As structurally similar areas are represented by the same model in our approach, testing different model parameters sets should be feasible even in larger areas if the structural properties are not too variable/complex. We will discuss this in a revisited MS.

7. DW: *There are a number of minor grammatical issues that nonetheless cause some distraction from the overall high quality of the manuscript. I will point out some of these below, but it could be worthwhile to have a native English speaker perform a careful proofreading.*

RL: We will carefully proofread the MS once more and would like to highlight that there will be another professional language check by Copernicus if the MS is accepted for publication.

References:

Nicótina, L., Alessi Celegon, E., Rinaldo, A. and Marani, M.: On the impact of rainfall patterns on the hydrologic response, Water Resour. Res., 44(12), 1–14, doi:10.1029/2007WR006654, 2008.

Ogden, F. L. and Julien, P. Y.: Runoff sensitivity to temporal and spatial rainfall variability at runoff plane and small basin scales, Water Resour. Res., 29(8), 2589–2597, doi:10.1029/93WR00924, 1993.

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., van Schaik, L., 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, Hydrol. Earth Syst. Sci., 18(11), 4635–4655, doi:10.5194/hess-18-4635-2014, 2014.

Reply to Anonymous Referee #2:

Anonymous Referee #2 ***(AR2): Summary and Recommendation:*** *The manuscript introduces an adaptive spatial clustering of hydrologic response units (HRU) to cope with the dynamics of the intermittent rainfall by keeping the model as parameter parsimonious (=model states) as possible in terms of reduction of similar-reacting HRUs. The manuscript is well-written and I enjoyed reading it. The introduced clustering is innovative from and fits into the scope of the journal. I have a few moderate and a number of minor comments, which are stated below. My overall recommendation would be a moderate revision to give the authors enough time to solve the open issues. Since I can only choose between minor and major revision, major revision it is.*

Ralf Loritz (RL): We would like to thank the second referee for the time and the effort she/he put into his review. The points she/he raises are relevant and addressing them will help to improve our manuscript. We hope that after this discussion (as well as after we revised our manuscript) all issues she/he raises can be clarified.

Moderate comments:

1. AR2: *The manuscript is about the reduction of the spatial model resolution based on the variety of precipitation as input signal. I'm wondering if there is not an adaption of the temporal resolution required as well since scales in space and time are not independent of each other (see Melsen et al. (2015) and references therein)? Maybe it's not an issue for the small catchment studied here …*

RL: Important comment. The results of the study of *Zhu et al. (2018;* recommended by the first reviewer) highlight that the timing of the precipitation is more important in smaller catchments while it is the spatial pattern in larger catchments. We will discuss this in a revisited MS and carefully read the study of *Melsen et al. (2015).*

AR2: *… but for larger catchments with a small hydrologic variability the numeric stability can be questioned due to the large spatial discretization and the high temporal resolution (e.g. in terms of the Courant-Friedrichs-Lewy condition, Courant et al., 1928). The authors should proof this condition for their model setup and discuss possible issues in the manuscript. An alternative would be to reduce the temporal resolution as well, which would lead to an additional reduction of parameters/computational costs.*

RL: CATFLOW uses an adaptive time-stepping, which means that time steps can be reduced down to seconds depending on the numerical solver. In the presented study the Darcy Richards equation is solved implicitly while the surface runoff is solved explicitly (for more details see also *Zehe et al., 2001*). As the horizontal grid resolution of the CATFLOW hillslope (reference model) is below 1 m, the vertical below 10 cm and time steps are small we have no issue to fulfill the Courant criteria in our model. Nevertheless, you mention an important point here and the fulfillment of physical and numerical constraints were the main motivations of our former "representative hillslope" study (*Loritz et al. 2017*).

CATFLOW hillslopes are typically interconnected by a river network and runoff is routed downstream with a diffusion wave approach (explicitly solved) assuming a prismatic river cross-section and roughness that changes with changing Strahler order. However, the combination of a river network with the raster layout of our adaptive model is not straightforward (although not impossible, for instance by linking the centroid of a raster cell to the closest node of the river network). To make things not more complicated as necessary we decided to use a lag function in this study. This lag function is not solved numerically but shifts the simulated hydrographs in time by a constant velocity. Again we have no issue with the Courant criteria here. The latter is different if we would have used an adaptive mesh approach where the numerical grid is changed during runtime. Here we carefully need to check the courant criteria when we increase the size of the grids. We will discuss this in a revisited MS.

2. AR2: *The authors have selected two events to show the ability of adaptive clustering. The choice of both events seems to be very arbitrary. From Fig. 4 it seems that the resulting runoff peaks are not representative for runoff mechanisms of the catchment. As far as I understand from P13 l8-10 the clustering is carried out manually and not automatically so far, which is the reason why the authors decided for two small events covering only a few time steps. However, I disagree with the hypothesis that "a test on a longer timescale...would provide only little more scientific inside" (P13 l9-10), which is also not proven by the authors.*

RL: Respectfully, we do not agree with the assessment of the reviewer regarding the selection of our rainfall events. We chose event I as it has the highest intensity and third-highest spatial variability in the chosen period. We chose event II because we wanted to test our adaptive model at a rainfall event with a longer duration and lower intensity). From examining the rainfall-runoff events in summer we believe that both events represent the runoff generation in summer well as long as subsurface storm flow is dominant. We agree with the reviewer that our statement is a bit misleading and we will explain better what we mean here. Please see also the discussion with the third and fourth reviewers.

AR2: *I rather expect that the reduction of model parameters due to the adaptive spatial resolution is reduced significantly for long-lasting rainfall events causing a direct runoff response over several days as e.g. in Nov 2014, Jan 2014-Mar2014 and Aug 2014.*

RL: You are correct. The needed spatial model resolution in winter is much lower compared to the chosen summer rainfall-runoff events. This is indicated by the fact that the distributed *model b* and the *reference model* perform almost identical with respect to simulate the observed discharge in the winter season. We would argue that is is difficult to justify the use of a spatially distributed model over a spatially aggregated model if they perform similarly as long as the focus is on an integral response of a system.

AR2: *Another point that can be questioned is snow, which does not cause runoff immediately, but when snow melt begins. How will this be affected/can be incorporated by the adaptive clustering? The impact* *of more complex events than those analysed in the current study has at least to be discussed sufficiently in the manuscript, although an analysis of more events is encouraged to represent the effect of the adaptive clustering on the variety of runoff responses.*

RL: Interesting point. Snow is rare in the Colpach catchment which is fortunate as CATFLOW has no internal snow routine. However, let's assume we would have used a model with a snow routine in an area where snow is a dominant control on the runoff generation. In this specific scenario, we would indeed have to adapt our definition of similarity between the model states. In other words, instead of using the only slope of the simulated hydrograph alone to define similarity, we would also need to check the snow cover before we would group models as functional similar based on their state. Two similar hillslopes would then have the "same" snow cover (given a threshold) as well as the same slope of the hydrograph. We very much like the idea of testing the approach in an area where snow is an important factor. However, for now, we will discuss the limits of choosing a single variable to group model states in a revisited MS.

3. AR2*: The model states are identified by the slope of the resulting runoff curve. However, the slope can be more or less identical for one time step independent of the current runoff situation, e.g. if runoff is reduced in one tile from 25mm to 20mm and in another tile from 10mm to 5mm (which could be the case in a stratiform event with a convective cell inside), the resulting slope is the same, right? So the soil moisture and other storage elements is then "averaged" due to the same model state of both tiles, although both tiles are in completely different hydrologic situations. It would be useful if the authors would comment on that issue or, if I understood it not correctly, clarify the part where I got lost.*

RL: You are correct. In an earlier version of our spatial adaptive model, we used the absolute discharge to identify similar model states. The issue here was that two models could produce the same discharge at a given time step but one model would simulate a rising hydrograph while the other a declining. We hence decided to take the slope of the hydrograph assuming that the model differences would be small given the size of the Colpach catchment, the focus on the summer season and because we only simulate shallow subsurface stormflow. In the case of a stratiform event with a convective cell inside or if we have snow in a catchment our assumption might be violated. Thank you for raising this issue and along your lines, we will add another criterion to our spatially adaptive model. In a revisited MS only model elements which share a similar dQ dt^{-1} (0.05 mm hr^{-1}) as well as Q (0.05 mm hr^{-1}) will be grouped together.

Specific comments:

AR2: *P4 l5-8 The difference is not clear formulated at this point. It becomes clearer while reading the manuscript, but should be communicated concisely at this point.*

RL: We will rephrase this sentence.

AR2: *P7 l27 Where are the disdrometers located? Can they be used to improve the rainfall input for the reference model to achieve a more realistic uniform areal rainfall? If not, could be an increase of rainfall amounts with altitude improve the areal rainfall estimate? The Roodt station is situated in the raster field with the lowest rainfall amounts (Fig 2) and not representable for the catchment. So any correction has to be done to enable a fair comparison between reference model and model a.*

RL: When we were setting up the reference model for our proceeding study (*Loritz et al., 2017*) the only rainfall measurement available at that time was the ground station in "Roodt". As we are aware that the comparison between the *reference model* and *model b* (the distributed model) is not entirely fair as they used different rainfall data we added *model a* to the model ensemble. In a restructure section 3 we will clarify this as well as add the location of the distrometers to the appendix A1.

AR2: *Fig 2 Please add rain gauge data to Fig 2b) to enable a comparison of all rainfall inputs.*

RL: We will add the rainfall data from "Roodt" to Fig2b.

AR2: *P10 l2, p11 l26 area-weighted -> As I understand the areal mean is estimated by the arithmetic mean of the satellite data. How do weights for different areas affect this estimation? This is not clear for me, please rephrase/add the explanation.*

RL: As not all of the 42 raster cells of the distributed rainfall data are entirely within the borders of the Colpach catchment their weight was reduced when we calculated the average precipitation for *model a*. We will rephrase this sentence accordingly.

AR2: *P11 l2 sap flow -> Do the authors mean by sap flow the flow in plants? I can't imagine at this point how the authors applied observations like that in the current study. If so, please describe a bit more detailed, since it is not a conservative measure for model validation and hence of great interest for the community.*

RL: By sap flow we indeed mean sap flow in plants. In our proceeding study (*Loritz et al. 2017*) we compared normalized sap flow velocities against normalized transpiration simulations of CATFLOW to evaluate the transpiration simulation. Sap flow measurements where thereby one of the keys for a successful simulation in the Colpach catchment as they helped us to identify the onset of the vegetation (when the trees started to transpire). The comparison is described in detail in *Loritz et al. (2017*; Figure 12). Although we agree that this might be of great interest for the community we would like to avoid discussing this once more to keep the MS as focused as possible.

AR2: *P12 l23 "average distance of each grid cell to the outlet" -> Should it not be the distance along the flow path/flow direction? So it would be possible that the runoff is assumed to stream upwards in some areas of the catchment- Please rephrase or reconsider.*

RL: For each grid cell of the precipitation field we calculated the average flow length along the surface topography to the outlet of the catchment using an underlying DEM with a 10 m resolution. We used the averaged flow distances in our lag function. We will explain this in more detail in a revised MS.

AR2: *P12 l30 "3.2.1 to 3.2.3" -> "3.3.1 to 3.3.3"*

RL: Following the discussion with the first reviewer Daniel Wright we will restructure section 3 and remove all "subsubsections".

AR2: *P13 l2 "wetness state" Please define this term. It sounds as only soil moisture is included without any additional information, but there is more included, right? If not, why not using the term soil moisture? Section 3.3 and 3.3.1 There are repetitions among the paragraphs, please remove them.*

RL: We will remove the term "wetness state" with the term "catchment state" to make clear that we also mean the shallow groundwater table, soil moisture, etc. We will furthermore restructure section 3 and remove the repetitions.

AR2: *P15 l4 & 21 Both thresholds are catchment size-dependent (as the authors state also later).For other applications it would be useful to introduce a catchment size-dependency to derive these thresholds. This is beyond the scope of the study since it demands a multi-catchment analysis, but the authors should add a small sensitivity analysis by e.g. using ΔP >{0.5, 1, 2} mm/hr as thresholds. This is along with a comment I have for the results section later, but I want to state it already here. In the results discussion it is often mentioned, that the number of parameters is reduced between model b and c, there is no figure illustrating it, although I would imagine it would bean impressive plot with y as KGE over x as the summarized number of model parameters per time step (or on average) for one event. Model reference, a, b, c (ΔP>1mm), c(ΔP>0.5mm) and c(ΔP>2mm) would be the points to show in the diagram. I assume model c would represent a break in the curve (KGE not increasing, while number of model parameters do) and the different thresholds would represent the uncertainty of this approach.*

RL: Interesting comment. Using a typically physically-based model (CATFLOW) and specifically the setup of our model based on field measurements it is kind of difficult to estimate the number of model parameters in our study. However, we will add a plot with the distributed precipitation binned into different thresholds (0.1, 0.5, 1, 2, 5 mm hr^{-1}) to the appendix. Based on this plot we will discuss how the binning will most likely affect our spatial adaptive model. Furthermore, will we discuss that a sensitivity analysis with different thresholds is needed along your line of arguments. Again thank you for this comment.

AR2: *Table 1: As far as I understood the calibration was done only for the reference model, right? Although that seems to be done in a former publication, a brief information about calibration and validation period, objective function and so on is required to interpret the table. For model a, b and c no parameters were changed, so the same parameter set was used throughout the study to enable comparisons? If there was a re-calibration for model c, the reference model and models a and b should be re-calibrated for the events only as well to enable a fair comparison*

RL: Exactly the calibration was done in a former publication exclusively for the reference model. All model parameters remain the same. The only differences between the models are the rainfall data which we use to drive the models as well as their spatial resolution. We will add more details about the calibration in a restructured section 3.

AR2*: Fig. 5: I'm a bit confused here. The authors state P=12 for t=2, but from counting it is P=13 – please double-check (also the number of entries in the following text refer ring to t=2). Additionally, for t=4 M=3 results from P=2 and S=1 – from my understanding the maximum of model states is max(M)=2 in this case, please double-check.*

RL: You are correct. We will check the figure as well as the corresponding text passages. Thank you for checking the figures so carefully.

AR2: *P27 l4-22 This paragraph provides already a good overview of related references. However, from my understanding the reference of Nicotina et al. (2008) concluded that spatial patterns of rainfall are only important for large catchments (8000km2 in their study) for hourly time steps, the correct estimation of areal rainfall is sufficient for smaller catchments. The authors should review this reference again and check their implementation in the current manuscript.*

RL: Thank you for this comment. We were referring to the following section in *Nicotina et al. (2008)*: "*As noted in section 4, this is because the spatial scales of variability of rainfall are very often much larger than the typical hillslope scale. Whenever infiltration excess mechanisms are important, the spatial distribution of areas of intense rainfall may be an important factor in determining the hydrologic response, …* ". In the current MS the use of this reference is indeed misleading and a leftover from an earlier version. We will revisit the corresponding sentence.

AR2: *Also, Ogden and Julien (1993) state that only for rainfall with durations shorter than the concentration time of a catchment the spatial distribution of the rainfall matters, for longer rainfall events only the temporal distribution matters. To highlight the importance of distributed models the authors could also look at Krajewski et al. (1991), Bardossy & Das (2008) or Müller-Thomy et al. (2018)*

RL: Thank you very much for pointing us to these publications we will read them carefully and see if they can help us to improve our argumentation.

Reply to Referee #3 Wouter Knoben:

Wouter Knoben ***(WK): Summary and Recommendation:*** *The authors develop and test a hydrological model that is able to change its spatial complexity in time. In its most simple state, the model represents the Colpach catchment in Luxembourg as a single representative hillslope. In its most complex state, the model would be able to use 42 hillslope elements to simulate the catchment's response to extremely spatially variable rainfall inputs. The model adds hillslope elements based on the spatial complexity of incoming precipitation and removes hillslope elements based on the change of runoff over time. Both processes use a threshold to decide when upscaling or downscaling the model is needed or possible. The authors show that the adaptive model reaches the same KGE scores as a fully distributed model that uses 42 hillslope elements all the time, while the adaptive model needs 16 representative hillslopes at most. This is shown for two short-duration event that occurred during summer.*

I have read this paper with much interest and found it generally easy to read and understand. As models grow more complex, computation times go up and studies such as this could open up great opportunities to reduce computation costs by avoiding redundancy in model calculations. However, I have some questions about the tests and metric the authors use to show that the adaptive model is as good as the fully distributed one. These are outlined below. I've provided additional requests for clarification in the line-by-line comments in the hopes that these are helpful.

Ralf Loritz (RL): We would like to thank Wouter Knoben for the interesting discussion on our Manuscript (MS). WK raises a couple of important and well-thought comments and we hope that after this discussion as well as after we have revisited our MS all open issues can be clarified.

Comments:

1. WK: *My main concern is the choice of using dQ/dt to reduce the number of model elements. Using the change in discharge over time to measure similarity of states can only work if there is a unique relationship between model state and dQ/dt. Given the equifinality in the fluxes-discharge relation that's typically visible in hydrological models (see e.g. Khatami et al., 2020), I think the section that introduces this concept (P16, l17) is not quite clear about why this dQ/dt assumption can be used together with CATFLOW.*

RL: Important comment and also the second reviewer had similar concerns. We will hence add Q as a second variable to group and ungroup model states and improve the discussion on how using a single variable to define similarity between model states will always lead to errors in certain scenarios and following that these variables need to be picked carefully.

WK: *Reading further, the authors address this concern to some extent in section 4.4 (P23, l18). This section however seems to show that CATFLOW does not exhibit such a unique relationship and the model reduces the number of model elements before the groundwater states reach similarity. This does apparently not affect the quality of the simulations much, because the KGE scores in Table 1 seem to indicate the adaptive model is as good as the fully distributed model for the two testing events.*

RL: In the current MS we did not mention that CATFLOW simulates only shallow subsurface stormflow in the entire summer period. This means that we do have a rather unique relationship between our model states and their function. Furthermore, the example in Fig.8 shows two extreme cases where one model receives much more precipitation than the other exactly intending to show the limits of our approach (section 5.3). As discussed in more detail with the fourth reviewer we will show that there is no difference between *model a* and *model c* (also concerning soil moisture in both depths) already shortly after the rainfall stops and when *model c* represents the entire catchment with a single hillslope. Furthermore, by calculating the Shannon entropy of the 42 hydrographs simulated by the spatially distributed *model b* we can see that there is no reason to assume that two models drift apart in the selected time frame. We will disscuss this in a revisted MS.

WK: *Fig. 4 shows that both testing events are selected in the middle of summer, when presumably the catchment is in quite a dry state (catchment state is not mentioned when selection of the two events is discussed on P18, l20 to P19, l6).*

RL: We mention the catchment states for event I and II on page 18 line 26 to 28 and page 19 line 5 to 6. Furthermore, do we refer to our former study where we showed 38 time series of soil moisture in the Colpach catchment in various depths and locations for the same hydrological year.

WK: *The fact that both events are selected during the dry summer could mean that the model can reset itself to mostly empty between the events and as such the long term (seasonal) impacts of not keeping the groundwater states separate cannot be investigated with the current two testing events.*

Equally, the events concern high flows so the impact of differences in slow ground-water states probably do not register in the dQ/dt values during the falling limb of the hydrograph (and thus the adaptive model simplifies itself).

There is the compounding issue that the KGE scores used to calculate the performance of model c are only calculated during the high flow event and that metrics such as KGE are typically not very sensitive to errors in low flows. This means that the parts of the simulation time series where the differences in groundwater states could be seen are both not included in calculation of the KGE score of model c and if they were, the KGE metric might not be able to pick up on any differences.

RL: Again an important point. As already mentioned above we discussed in section 5.3. "*While this finding is surely constrained by the chosen threshold, the picture is nevertheless quite different in deeper soil layers where the diversity of the rainfall forcing leads even after 24 hrs to increasing differences between the "driest" and "wettest" models. A part of the information about the different meteorological forcings between the two models is hence still stored in the model state after 24 hrs and has not yet been dissipated. The importance of those differences likely depends on the dominant runoff generation process. In the present case, they have a minor impact as model …*"

In a revisited MS we will underpin once more that our approach with the current definition of similarity regarding the model states can have significant impacts for long-term simulations however that there is no reason to expect that in our specific case.

WK: *Summarizing the above, I'm not sure that the dQ/dt criterion is entirely appropriate to determine when the adaptive model can reduce its complexity, and I'm equally unsure if the current two testing events would be able to show if the dQ/dt criterion is or is not appropriate. The straightforward solution would be to run model c for the year, add these results to Table 1 and briefly investigate for example the relative contributions of different fluxes to the overall water balance and the model's response to a few precipitation events during winter. Given that the adaptive model should be faster than the fully distributed one, this should not be a large computational burden and it will provide a much more complete impression of the capabilities of the adaptive model.*

RL: We hope that an improved discussion of the limits of the dQ/dt criterion (or any other criterion) as well as the addition of a second similarity measure (Q) will clarify the issues WK raises. We would also like to stress once more that we focus exclusively on the summer season as the distributed *model b* outperforms the *reference model* only in this period and because the meteorological boundary conditions change between the winter (frontal) and summer seasons (convective; Fig. 4 b). Furthermore, did we chose two rainfall-runoff events instead of the entire period as it allows us to analyzes the events in great detail (Event I: Fig. 5, 6, 7) and as our main focus in this study is on the rainfall-runoff interaction and not on low flow conditions. We selected event I because it has the highest rainfall intensity and third-highest spatial variability (highest Shannon entropy) in the selected period and event II because we wanted to test our spatial adaptive model at a summer rainfall event with longer duration. We believe that both events represent the state space of the runoff formation of the Colpach in summer well and see no reason to assume that the *model c* would fail at another rainfall event. A test of the spatial adaptive model for the entire hydrological year (or even for a longer period), in a different environment, with more variables and different thresholds to group and ungroup the model states and maybe even with a different type of model, is indeed desirable. However, to keep the already quite elaborated MS as focused as possible we will focus on improving the discussion with respect to the limits of our approach, add Q as second criteria to define similar model states, add a new figure to the appendix where we show how the thresholds impact the number of precipitation groups (please see the discussion with reviewer 2) and finally plot the soil moisture of *model a* and *c* at the end of event I and II to highlight that both models are in a similar state also with respect to their soil moisture.

Line-by-line comments

WK: *P5, l5. This question seems quite strongly related to the contrasting results in the literature that the authors discuss in the first and second paragraph of the introduction, where they conclude that* *the impact of using a distributed model and/or distributed forcing data is conditional on the catchment under investigation. This research question seems a bit generic in that light, given that only a single model and catchment are being investigated in this work. As is, question 1 seems more like a formality to me (it must be answered with "yes" before Q2 can be investigated) and the main focus of the manuscript seems to be on Q2. Perhaps the manuscript can gain a bit in focus if only the current research question 2 is specified, and the work done to answer the current Q1 is presented as a prerequisite to address the current question 2. For example, "We test this hypothesis by first showing that the model CATFLOW applied to the 19.4 km2Colpach catchment using a gridded radar-based quantitative rainfall estimate improves in performance when it is distributed in space and driven by distributed rainfall. We then address the following research question: "Can adaptive clustering be used to distribute a bottom-up model in space that it is capable to represent relevant spatial differences in the system state and precipitation forcing at the least sufficient resolution to avoid being highly redundant as a fully distributed model?"*

RL: Good idea. We will consider rephrasing this section following your lines.

WK: *P5, l14. Assuming that "> 1 m" refers to soil depth, should it be "< 1 m"?*

RL: No, soils are rather deep in this area and vary between 1 to 2.7 m according to several drillings and electrical resistivity tomography (ERT) measurements.

WK: *P7, l9. Which numerical scheme is used by CATFLOW?*

RL: Darcy-Richards: implicitly solved by a mass conservative modified Picard iteration scheme (Celia et al. 1990); Surface runoff (1d St. Verdant eq.) explicit Euler forward. We will add this information to the MS.

WK: *P7, l20. If possible without using too much space, it might be helpful to the reader to briefly summarize the main findings of Loritz et al. (2017).*

RL: The main findings of this study are summarized on page 10 section 3.1.

WK: P7, l21. What are the outcomes of this quality control?

RL: Manually quality checked by the Luxembourg, Institute of Science and Technology (LIST; no negative values, etc). We will remove the term "quality checked" as it is not necessary here.

WK: *P7, l28. I'm not quite sure I understand why these distances are given as a range if only a single station is concerned. Does this indicate minimum and maximum distance of the catchment bounds to each radar station.*

RL: Exactly. These are the distance to the boundaries of the Attert catchment in which the Colpach is located. We will add this information in a revisited MS.

WK: *P9, l13. I find this sentence a bit hard to follow. Is the part from "apart from..." onwards necessary here? This is already discussed in the introduction.*

RL: We will remove this part.

WK: *P10, l21. Why is the model tested during two events instead of over the full year? How were these events selected?*

RL: Please see the discussion above and the discussion with the second and fourth reviewer.

WK: *P11, l14. The conclusion that a distributed model is needed to account for runoff driven by convective precipitation would be stronger if the authors can (briefly) list which processes are represented at too coarse a scale in the reference model for it to properly deal with convective precipitation.*

WK: *P11, l14. I believe this sentence would be more complete if it also explicitly mentioned that distributed instead of catchment-averaged precipitation data is needed to properly simulate the result of convective precipitation events.*

RL: We wrote on page 11 line 14: *"In other words, this entails that a hydrological model, distributed at a sufficiently high spatial resolution, is required to capture the spatial variability of the precipitation field to satisfactorily simulate the runoff generation of the Colpach"*. We believe that our argumentation is well justified here.

WK: *P11, l27. It would be helpful for the reader to repeat that the only difference between reference model and model a is the choice of precipitation data.*

RL: We wrote in the sentence before the sentence you mention: "*Model a is identical to the reference model, however, driven by the area-weighted mean of the spatially resolved precipitation data described in section 2.4 (Fig. 2 b)."*

WK: *P12, l3. Are these variables similar or identical to those used in the reference model?*

RL: Identical. We will change the word accordingly.

WK: *P12, l4. To clarify, does this mean that model b is run in a gridded fashion with the catchment divided into 42 grids (matching the precipitation grid)? If not, it would be good to clarify this in the text and mention the number of model elements that the precip field similarity approach gives. Line 18 on this page could benefit from a similar clarification.*

RL: Yes, this means that model b is "*divided into 42 grids (matching the precipitation grid)"*. We will consider rephrasing the corresponding sentences.

WK: *P12, l23. Are there some observations that could help support the choice for 1 m/s?*

RL: We will add the reference of *Leopold, (1953)***.** Fig. 1 in this reference shows an average relation of flow velocities and discharge in rivers. Correspondingly we picked an average value of 1 m s^{-1} (2 to 3 feet per second).

***WK:** P14, l29. It might be good to extend this line of reasoning to soil types and vegetation cover, as these are commonly used as model inputs/parameters.*

RL: Agreed. We will rephrase the corresponding sentence.

***WK:** P15, l7. This sentence is quite general (referring to humid environments) and could use a reference. However, if the authors chose 1 mm hr-1 based on their expertise and knowledge about this catchment, then I think it's more accurate (and in no way worse) to phrase this decision along those lines, e.g.: "We chose this threshold as a reasonable value upon which we expect differences in hydrologic behavior, based on our collective understanding of the Colpach catchment."*

RL: Valuable point, we will rephrase this sentence.

***WK:** P17, l10. I think it's import to repeat the similarity condition of dQ/dt here, because for a model that has no unique relation between model state and dQ/dt values this method cannot be applied without accounting for this difference.*

RL: Please see the discussion above and in section 5.3 in our MS.

***WK:** P20, l6. The authors use KGE values in this section and Table 1. I'm not sure to what extent the aggregated value is a useful metric for events that last only a handful of time step. It would be good to at least disaggregate the KGE into its correlation, variability and bias components (e.g. quantify what can be qualitatively estimated from Figure 7) to see if the total KGE scores of the individual models are generated by (roughly) the same types of errors in the simulations.*

RL: Good point. We will add the three components of the KGE in the appendix for each model.

***WK:** P21, l25. "acceptable" is somewhat subjective because no standard of acceptability has been defined. It might be cleaner to simply report the correlation component of the KGE to quantify to what extent the hydrograph shape is simulated.*

RL: Agreed. We will rephrase this term.

***WK:** P21, l26. This trial of a direct runoff component seems somewhat ad-hoc to me. I don't think this adds anything to the manuscript and that it will take more space than is available to properly justify this change. I suggest to remove these sentences.*

RL: Thank you. We will consider removing this sentence.

***WK:** P30, l4-24. These sentences seem as if they would be better placed in the introduction or methodology sections.*

RL: We will rephrase some of these sentences. Please see the discussion with reviewer #1 (Daniel Wright).

Reply to Referee #4 Anna E. Sikorska-Senoner:

Anna E. Sikorska-Senoner ***(AS): Summary and Recommendation:*** *This paper proposed an adaptive modelling as an alternative to a distributed model for representing spatial variability of the catchment and forcing input (precipitation). Such an adaptive modelling should be able to run faster than a distributed model but should provide a similar model performance as its fully distributed version. The manuscript is generally well written and it is easy to follow. The idea of a spatially adaptive model that dynamically adjusts its spatial structure during runtime is indeed very interesting and has a potential for being applied in many (hydrologic) modelling approaches. Yet, I have few major issues that should be addressed before considering this manuscript for a publication in HESS. Thus, I recommend a major revision.*

Ralf Loritz (RL): We would like to thank Anna E. Sikorska-Senoner for her comments and the time she invested to review our Manuscript (MS). We hope that after the discussion as well as after we have revisited our MS all open issues she raises can be clarified.

Comments:

1. AS: *The adaptive model (model c) is tested here only on two rainfall events, which I see as the major weakness of this manuscript. As the strength of this approach should lie in the possibility to apply it to a continuous modelling and not to an event-based modelling. Thus, I think it would be important to demonstrate how the model c works on continuous time series. As this is missing in the current manuscript, we still do not know at the end whether it is a good or a bad option to be used.*

RL: *Model a* and *model c* simulate close to identical hydrographs at the end of both rainfall events when *model c* represents the Colpach catchment again by a single hillslope model. This is also true for the soil moisture distributions, which we did not show in the current MS. This means that the information about the spatial organization of a past rainfall event have already been dissipated closely after the spatial adaptive *model c* represents the catchment by a single hillslope. In other words, there is no difference between *model a* and *c* after this point and we would learn not much by letting *model c* run continuously until the next rainfall event.

Furthermore, as rainfall event II is characterized by one of the longest rainfall durations in summer and event I by the highest intensity and third highest spatial variability we see no reason to expect that the spatial adaptive model will fail at other summer rainfall-runoff events. We think that it is not the length of the simulation that matters here but the fraction of the visited state space (or in other words if your training data set is representative). The latter means that we do not assume that the catchment and the model which represents it will function differently at the other untested events. This is underpinned by the fact that also the 42 model elements in the distributed *model b* do not drift apart. The latter reflects the highest complexity *model c* could reach.

However, we agree that we did not well justify the selection of the two events. Following your comment, we will hence plot the soil moisture distribution of *model a* and *c* for event I and II at the time step when the catchment is again represented by a single hillslope. This will show that there is no difference between the *spatially aggregated model a* and *model c* already shortly after the rainfall stopped. Furthermore, will we improve our discussion regarding the choice of our two rainfall-runoff events. Again, we would like to thank AS for her comment.

2. AS: *The performance metrics of the calibrated (tuned) models should be provided so that the model ability to predict rainfall events could be assessed*

RL: The reference model is the only model which was manual tuned to match the seasonal water balance of the Colpach. This procedure is described in detail in *Loritz et al. (2017)* and in the current MS in section 3.1. The KGE value of the reference model is reported in table 1. We will furthermore add the three components of the KGE as discussed with Wouter Knoben to the appendix.

3. AS: *A model set-up between the model a and b could be very didactical, i.e. having a structure as the model b but using the precipitation input as the model a (the same for each grid cell)*

RL: The only difference between *model a* and *b* is the precipitation input. Running *model b* with the input of *model a* would mean to produce the same hydrograph as *model a* 42 times.

4. AS: *It is not quite clear how the switch between different model setups (i.e. the number of model run in the model c) affect the setup of initial conditions for next runs, which is important to be considered for continuous model simulations but also for simulations of events. More details should be provided on that.*

RL: Please see the discussion above.

5. AS: *It would be also very didactical to see the comparison of the precipitation records from the ground station with the precipitation fields obtained from the gridded data. This is never done in the manuscript and no reason for not doing that is given.*

RL: Agreed. We will add the precipitation from the ground station to Fig. 2b.

6. AS: *The (rather) poor model performance of all tested models' set-ups for two selected events requires some discussion. It appears that none of these model can really capture the dynamics of these two events even if using the distributed model and the distributed rainfall information (with KGE<0.3). Hence, it is even more important to verify the model performance (pkt. 2). An addition of other metrics that focus entirely on the flood event such as peak or time to peak could be here very informative. Given a rather* *poor models' performance, it is difficult to justify the need of developing the adaptive model based on the distributed model if the latter does not provide acceptable simulation results.*

RL: Respectfully, the focus of this MS is not to minimize residuals between an observed quantity and a model simulation. The main goal of this study is to introduce an approach with the goal to setup a spatially adaptive model and equally important underpin this approach with a physical meaning. Furthermore, would we like to highlight that we a) discuss the model performance and how it could be improved on page 21 line 26 to 28 and b) would like to reiterate that the *reference model*, which is the basis of this study, was tested against a series of different variables (sap flow, discharge, water balance, soil moisture, etc.), at different hydrological years, in an additional sub-basins as well as mainly setup based on field observations. We believe that such an evaluation and model-building process underpins the quality as well as the ability of a model to mimic the hydrological dynamic of a landscape sufficiently and maybe even more than adding another performance metric.

As the model was setup to simulate the seasonal water balance we think that the annual performance is quite "good" and we are not surprised that if we zoom into a single event that we loss performance. Furthermore would we like to highlight that the performance metric, which is important here is the KGE between *model b* and *c*, which is 0.98. To improve the interpretability of our model scenarios we will add a second table with the three components of the KGE to the appendix as discussed with Wouter Knoben. Furthermore, will we clearly state that the goal of this study is not to perform a best as possible streamflow simulation.

7. AS: *A fair comparison of all presented models should involve the same metrics, i.e. computation over the same time at a continuous time scale. In this study, different model setups are compared at different scales that makes it difficult to get an overview of their performance.*

RL: We compare and discuss the connection between *model b* and *model c* only for the two events as well as for the corresponding summer period. Respectfully, we do not think that the comparison is unfair.

Detailed comments

1. AS: *Abstract: 'a mesoscale catchment'; a 20-km^2 catchment appears rather small tome than meso-scale.*

RL: Meso-scale: 5 to 1000 km^2, we refer here to the work of *Zehe et al. (2014)* and *Dooge, (1986)*.

2. AS: *Abstract: 'three hydrological models', the model is actually the same but different set-ups are used that span from the averaged model until the distributed model. Please clarify that here.*

RL: This depends on your the definition of the term "model". However, I agree and we will use the term model setups here.

3. AS: *L. 20-21 p. 3: It is not always possible and justified to switch from a continuous model to an event based model. Hence, continuous modelling is often required in many applications.*

RL: Agreed. Could you provide a reference here?

4. AS: *L. 19 p. 4: The tested catchment appears rather small to me. How do you define the cut here for a small/meso-scale catchment?*

RL: We refer here to the work of *Zehe et al. (2014)* and *Dooge, (1986)* which is around 5 to 250 km^2. The definition of organized complexity is that such systems are too complex that we can tread them exclusively in a mechanistic manner but too organized that we can represent them in a purely statistical manner.

5. AS: L. 7-9 p. 5: consider restructuring this sentence.

RL: Thank you. We will consider rephrasing it.

6. AS: *L. 11-13 p. 7: could you add the location of these meteorological stations to the map in fig 1?*

RL: The station "Useldange" is too far away to be added to the map. But its location is provided in the corresponding reference. We will add this information.

7. AS: *L. 15 p. 7: it should be 'and measures...'*

RL: Thank you. Changed.

8. AS: *L. 16 p.7: is there any weighting applied here and what kind of?*

RL: No weighting applied.

9. AS: *L. 28 p. 7: could the locations of radars be also placed in the fig. 1?*

RL: No, they are too far away. However, their location is displayed in the reference provided by *Neuper and Ehret, (2019).* We will add this information.

10. AS: *L. 12-14 p. 8: Why do you compare these values with literature and not with the ground station records from your catchment? Is there any reason that you are not using the precipitation records from the ground station?*

RL: These values represent the climatic averages of the area. We have only data for about 10 to 15 years.

11. AS: *Fig. 2: One would expect that the radar values would be compared with the ground station values as a corresponding mean (Fig. 2b). Could you add these values to the figure?*

RL: Agreed. Will be added.

12. AS: *L. 10-17 p. 9 till 21 p. 10: I am not quite sure if this text is really helpful. After reading these lines, we still do not know how the reference model and other models look like. Maybe you could merge these lines with the sections 3.1-3.3.*

RL: We will restructure section 3. Please see the discussion with Daniel Wright.

13. AS: *L. 19-20 p. 10: I would say that the main goal is to test or verify whether similar model performance can be achieved with the adaptive model as compared to the model b. However, by the comparison that you did we still do not know the answer to this question as the comparison is done only based on two pre-selected events both having rather a poor model performance. Please comment on that and also state why these events were chosen for the comparison (and not others)?*

RL: Please, see the discussion above.

14. AS: *L. 21-22 p. 10: Why do you compare the adaptive model with model b using only these two events and not the entire simulation period? In my opinion, the greatest potential of the adaptive modelling lies in continuous modelling and not in the event-based.*

RL: Please, see the discussion above.

15. AS: *L. 31 p. 10 – l. 1 p. 11: why do you test the model only based on the annual assessment and not on hourly simulations? It is quite surprising because you use the model for assessing the model performance at an event-based scale in the second step, i.e. when comparing different models. I think it is important to report here how the model behaves at an hourly time scale so that one knows what can be expected from the model.*

RL: I am not sure if I have understood that comment correctly. But we tested our models by comparing hourly simulations with hourly observations for one hydrological year.

16. AS: *L. 3 p. 11: Which metrics were used here for assessing that the model performance agreed well with the dynamics of observed values? Can you give some more details on that?*

RL: We use the Spearman rank correlation, the Nash-Sutcliff eff. and the KGE. We refer to the study of *Loritz et al., (2017).*

17. AS: *L. 7-8 p. 11: It is not surprising that the model performs poorly at time series scale if it was evaluated only on an annual basis. Some insights should be given here; why was the model tested at an annual basis if its intention is to predict events?*

RL: Respectfully, the goal of this study is not to perform a best possible streamflow simulation with regards to minimize residuals. If this would be the case we would have picked a more data driven approach.

18. AS: *L. 3 p. 12: similar to what?*

RL: They are the same in all models.

19. AS: *L. 12. p. 12: the model analysis should go after the introduction of all models.*

RL: Agreed. We will restructure section 3.

20. AS: *L. 16-19 p. 12: an additional model between the models a and b would be here very useful, i.e. a model that has a structure as the model b but uses precipitation as in the model a (so it uses the same precipitation for each grid cell). This inclusion could nicely show the added value (or no value) of including a spatial distribution of i) the model and ii) of the precipitation input data.*

RL: Please, see the discussion above.

21. AS: *L. 8-10 p. 13: why exactly? In my opinion, the strength of this approach lies in the possibility to apply it to a continuous modelling and not to an event-based modelling. Thus, I think it would be nice to demonstrate how the model works on continuous time series in terms of the model performance and computational efforts. As such a test is missing in the current manuscript, we still do not know at the end whether it is a good or a bad option to use the adaptive modelling approach... Based on the two events selected, we cannot say much about the value of the adaptive approach as the model performance remains poor for these events (as seen from the Table 1 and fig. 7). If the full simulation is not possible(could you give more details why exactly?), already a simple test with shorter but continuous time series of few months or weeks could provide some more insights on how this approach is really working.*

RL: Please, see the discussion above.

22. AS: *Tab. 1: as the initial idea is to improve the model performance for rainfall events, it appears from the table that the model c and model b have still rather poor model performance for the event I and II. In addition, all models perform poor for these events. Yet, an inclusion of the spatial variability does not improve much the model performance that is still not so good. Thus, it calls a question of the need of such an adaptive inclusion to this spatially distributed model which performance is rather low.... Could you comment on that? A decomposition of KGE into its components would bring more insights on the models' behaviour.*

RL: We will add the three components of the KGE to the appendix.

23. AS: *L. 9-10 p. 15: How many grid cells need to have a difference higher than this threshold to use the model c?*

RL: One.

24. AS: *L. 30-31 p. 15 fig.3: is the re-arranged model running with the same initial conditions of the original model or how do you decide on these initial conditions if you want to increase or decrease the number of M in the subsequent time intervals particularly if a continuous simulation is performed?*

RL: We aggregate their states.

25. AS: *L. 4-6 p. 18: for a fair comparison of different models, you should use the same metrics and the same time periods for evaluation. It is not clear why this is not the case here.*

RL: See discussion above.

26. AS: *Fig. 4: could you add simulations with the model a?*

RL: If we add all simulations the figure is hard to read. However, we will consider your comment when we revisit your MS.

27. AS: *L 8. P. 20: The reference Knoben et al. (2019) is missing in the literature list.*

RL: Thank you we will add this reference.

28. AS: *L. 6-7 p. 21: the performance of KGE below 0 is still rather very poor, which re-quires some further explanations. According to Knoben et al. (2019), simulations can be considered as behavioral if KGE>0.3 (with KGE≈−0.41 for a mean flow benchmark).*

RL: See comments above.

29. AS: *Table 1: the performances (KGE) of all models is rather poor for the events here selected (with KGE between -0.41 and 0.29). As already the model b (distributed)cannot simulate the two events in a good way (as also seen from the fig.7), why would you spend time on developing the adaptive model based on the model b instead if improving the model b or testing different models here? Could you comment or justify that?*

RL: See comments above.

30. AS: *Fig. 7: the simulations with the model a and the reference model should also be added here. Moreover, for both events, all models largely underestimate the events. Could you comment on that?*

RL: In fig. 7 we focus on the comparison of *model b* and *c*.

31. AS: *L 10 p. 26 – l. 2 p. 27: do you have any idea where this large underestimation may come from and how it could be improved?*

RL: Discussed in the MS (page 21 line 26 to 28).

32. AS: *Discussion: I missed some recommendations for other works. When and how would such an adaptive modelling be recommended? How one can set up the adaptive process? And why it is really needed to implement such an adaptive modelling?*

RL: Thank you for this comment. We will revisit the discussion of the MS in this regards.