How can expert knowledge increase the realism of conceptual hydrological models? A case study based on the concept of dominant runoff process in the Swiss Pre-Alps

Manuel Antonetti[1,2], Massimiliano Zappa[1]

5 [1] Swiss Federal Research Institute WSL, Birmensdorf, 8903, Switzerland
   [2] Department of Geography, University of Zurich, Zurich, 8057, Switzerland

Correspondence to: Manuel Antonetti (manuel.antonetti@wsl.ch)

Abstract. Both modellers and experimentalists agree that using expert knowledge can improve the realism of conceptual hydrological models. However, their use of expert knowledge differs for each step in the modelling procedure, which

10 involves hydrologically mapping the dominant runoff processes (DRPs) occurring on a given catchment, parameterising these processes within a model, and allocating its parameters. Modellers generally use very simplified mapping approaches, applying their knowledge in constraining the model by defining parameter and process relational rules. In contrast, experimentalists usually prefer to invest all their detailed and qualitative knowledge about processes in obtaining as realistic spatial distribution of DRPs as possible, and in defining narrow value ranges for each model parameter.

15 Runoff simulations are affected by equifinality and numerous other uncertainty sources, which challenge the assumption that the more expert knowledge is used, the better will be the results obtained. To test to which extent expert knowledge can improve simulation results under uncertainty, we therefore applied a total of 60 modelling chain combinations forced by five rainfall datasets of increasing accuracy to four nested catchments in the Swiss Pre-Alps. These datasets include hourly precipitation data from automatic stations interpolated with Thiessen polygons and with the Inverse Distance Weighting

20 (IDW) method, as well as different spatial aggregations of Combiprecip, a combination between ground measurements and radar quantitative estimations of precipitation. To map the spatial distribution of the DRPs, three mapping approaches with different levels of involvement of expert knowledge were used to derive so-called process maps. Finally, both a typical modellers' top-down setup relying on parameter and process constraints, and an experimentalists' setup based on bottom-up thinking and on field expertise were implemented using a newly developed process-based runoff generation module (RGM-

25 PRO). To quantify the uncertainty originating from forcing data, process maps, model parameterisation, and parameter allocation strategy, an analysis of variance (ANOVA) was performed.

The simulation results showed that: (i) the modelling chains based on the most complex process maps performed slightly better than those based on less expert knowledge; (ii) the bottom-up setup performed better than the top-down one when simulating short-duration events, but similarly to the top-down setup when simulating long-duration events; (iii) the

30 differences in performance arising from the different forcing data were due to compensation effects; and (iv) the bottom-up setup can help identify uncertainty sources, but is prone to overconfidence problems, whereas the top-down setup seems to

1

accommodate uncertainties in the input data best. Overall, modellers' and experimentalists' concept of "model realism" differ. This means that the level of detail a model should have to accurately reproduce the DRPs expected must be agreed in advance.

1 Introduction

5   Applying "expert knowledge" (i.e. someone's acquaintance as a result of study and experience) in hydrology, as in any other natural science, is crucial for linking observations and laws governing a given system, such as a catchment. This process usually involves formulating and testing hypotheses about how the system functions (Savenije, 2009). At the root of this scientific reasoning, two opposing ways of using expert knowledge can be identified: the top-down and the bottom-up approaches. The first can be traced back to the Greek philosopher Plato (428 – 348 BC), who was trying to link general

10   theories about the functioning of complex systems to measurable observations. A "bottom-up" approach involves extrapolating general theories from given observations, and can be attributed to Plato's student Aristotle (384 – 322 BC). These two approaches have been applied in nearly all scientific disciplines, e.g. in mathematics (Cellucci, 2013), economics (Böhringer and Rutherford, 2008) and neuroscience (Gilbert and Li, 2013), as well as hydrology. Thus one type of hydrological scientist, the experimentalist or "wet" hydrologist, tries to understand catchment functioning through extended

15   field investigations, whereas the conceptual modeller or "dry" hydrologist tends to develop theories at the catchment scale and successively tries to validate them against measurements (Seibert and McDonnell, 2002).

Both modellers and experimentalists agree on the importance of expert knowledge for improving the realism of hydrological models, e.g. by forcing the model to reproduce the processes observed in the catchment. In recent years, several process-oriented approaches have been developed, of which the concept of dominant runoff process (DRP, for list of abbreviations

20   see Table A1) is one (Blöschl, 2001). It relies on the hypothesis that, among the different runoff generation mechanisms that may occur at a given location (Hortonian overland flow HOF, saturation overland flow SOF, subsurface flow SSF, and deep percolation DP), one, the DRP, will be dominant over the others. Based on this concept, the following process-based modelling chain has been proposed (Clark et al., 2015): (i) reading the landscape, identifying and classifying the processes, (ii) developing a proper parameterisation to reflect our perceptions of the processes observed and (iii) allocating the

25   parameter values of these parameterisations (Fig. 1).

Experimentalists and modellers disagree, however, on how to implement their expert knowledge in each of these steps. For example, Schmocker-Fackel et al. (2007) applied the two philosophies to hydrological classifications using DRPs and claimed: "[…] These top-down approaches try to identify homogeneous landscape units. The assumption is that the hydrological response will also be homogeneous. By contrast, in bottom-up approaches, runoff formation is investigated on

30   the plot scale and then units with the same runoff forming process are identified" (Schmocker-Fackel et al., 2007). Examples of such bottom-up mapping approaches can be found in Markart et al. (2004); Smoorenburg (2015), Scherrer AG (2006),

Scherrer and Naef (2003) and Tilch et al. (2006), and of top-down mapping approaches in Gao et al. (2014), Gharari et al. (2011) and Fenicia et al. (2016).

The distinction between top-down and bottom-up is not sharp and different interpretations of the two philosophies have been applied in hydrological modelling. For example, Hrachowitz and Clark (2017) maintain bottom-up models correspond to physically-based models, where the conservations laws on mass, momentum and energy are applied. In contrast, top-down models are conceptual models. With regard to the level of modelling detail, Nalbantis et al. (2011) linked monometric approaches, where some components are examined in detail and other ones are only roughly described, to the bottom-up philosophy and the holistic approach, when all components are modelled with the same degree of detail, to the top-down one. Sivapalan et al. (2003), in contrast, classify approaches according to the scale considered: if the modelling is performed first at the small scale of e.g. HRU, or hillslopes, and then the results are scaled up to the catchment scale, it can be defined as bottom-up, whereas lumped models developed directly at the catchment scale can be defined as top-down. The definition of Sivalapan et al. (2003) also works well with the concepts of model parameterisation and parameter allocation. For example, in a classical bottom-up exercise, parameter ranges are narrowed and/or model parameterisations are proposed based on catchments properties, expert knowledge and possibly inferences from measurements. By following a top-down approach, expert knowledge can be used instead to define relational rules between the parameters and fluxes of different landscape classes. In this way, the model is forced to behave according to the modeller's perception of the catchment functioning and the parameter space can be reduced so that no calibration is necessary (Bahremand, 2016; Gharari et al., 2014).

Both approaches have strengths and weaknesses when implementing expert knowledge in process-based hydrological modelling. Bottom-up mapping approaches are often considered to require much data (Hümann and Müller, 2013; Müller et al., 2009), whereas top-down classification approaches are considered too coarse to detect the spatial distribution of processes with enough accuracy (Antonetti et al., 2016). Top-down models and parameterisations may be too simplistic to depict the spatial variability of runoff processes within a given catchment and, therefore, require calibration (e.g. Fatichi et al., 2016), whereas physically-based models may be too data demanding and not flexible enough to cope with emergent patterns at large scales (Beven, 2000).

Several attempts have been made to combine bottom-up and top-down philosophies (e.g. Klemeš, 1983; Sivapalan et al., 2003), and Hrachowitz and Clark (2017) in particular stress the need to merge forces. Similarly, Clark et al. (2017) ask: "How can we combine different perspectives on hydrologic modelling to advance the quest for physical realism?". Related questions concern the level of detail needed to reproduce the observed dynamics and pattern and how much detail the available data warrants for a meaningful parameterisation of the chosen process representation (Clark et al., 2015). Clark et al. (2016) note that the structure of the model should reflect that of the landscape. They claim that focussing on the extent to which space accounting models are limited by the available data helps testing the mapping theories and, consequently, improves how well landscape details are represented in models.

Several frameworks have been proposed for testing working hypothesis (e.g. Best et al., 2011; Fenicia et al., 2011; Kraft et al., 2011), but few addressed these questions and explicitly consider ways of implementing expert knowledge in hydrological models. For example, McMillan et al. (2011) developed a set of diagnostic tests based on field data to formulate recommendations for model building. Contextually, Clark et al. (2011) used the modelling framework FUSE (Clark et al., 2008) to allow a proper model structure to be selected based on these recommendations. However, the use of flow data to formulate the recommendations restricts the application of this method to ungauged basins (Hrachowitz et al., 2013). In addition, both the proposed recommendations and the FUSE framework are applicable exclusively at the lumped catchment scale. As a further development of FUSE, Clark et al. (2015) developed the SUMMA approach to provide a framework for both modellers and experimentalists to test alternative model discretisations, parameterisations, and numerical schemes. Nalbantis et al. (2011) compared a bottom-up and a top-down modelling approach with a focus on catchments with high human impact. Fenicia et al. (2016) combined bottom-up (i.e. "distributed" and scaled-up) and "top-down" (i.e. conceptual) approaches to test several hypotheses about process representations and hydrological controls.

Our study is intended to explore how different ways of implementing expert knowledge in hydrological modelling can affect simulation results with a specific focus on floods. In particular, we investigated: (i) Whether the use of more expert knowledge during the mapping phase improves hydrological simulations. (ii) Under which conditions (event type, catchment characteristics) satisfying results can be reached without drawing much on expert knowledge during the mapping phase? (iii) How uncertainty in forcing data and in the initial conditions influences and/or interacts with the simulation results? (iv) How the model setup, i.e. the parameterisation approach and the parameter allocation strategy, affects the results?

To address these questions we produced so-called process maps of a mesoscale catchment in the Swiss Pre-Alps using three mapping approaches derived with different levels of involvement of experts. The effects of the differences between the process maps on runoff simulations were investigated with two different setups of the newly developed PROcess-based Runoff Generation Module (RGM-PRO; Antonetti et al., 2017), which was forced with input data of varying quality. Finally, an analysis of variance (ANOVA) was performed to quantify the uncertainty arising from forcing data, process maps, model parameterisation and parameter allocation strategy.

2 Methods and Data

2.1 Study Area and Process Maps

We performed simulations on the Emme catchment up to Emmenmatt (445 km^2), which is located in the Pre-Alps mainly in Canton Bern and, on the eastern side, in Canton Lucern (Fig. 2). Its elevation ranges from 638 to 2213 m a.s.l.. About half of the catchment (52%) is covered by meadows, and the remaining part by forests (44%) or settlements (4%). The upper part of the catchment is characterised by Flysch and Cretaceous, whereas Freshwater and Marine Molasse and, to a lesser extent, Moraine dominate the lower part of the basin. Three additional runoff gauging stations can be found in Eggiwil (Emme

catchment, 125 km^2), Langnau i.E. (Ilfis catchment, 184 km^2) and Trubschachen (Trueb catchment, 55 km^2), and their measurements were used for this study to evaluate the performance of the models.

The study catchments were mapped according to three approaches with different levels of expert involvement and differing in terms of the data and the time required for mapping (Table 1 and Fig. 3). The simplest mapping approach includes solely topographical information and distinguishes three landscape classes, i.e. wetland, hillslope and plateau, by combining the Height Above the Nearest Drainage (HAND) descriptor (Rennó et al., 2008) and slope (Gharari et al., 2011). These classes are supposed to be a proxy for saturation overland flow (SOF), subsurface flow (SSF), and deep percolation (DP). The expert knowledge involved in this top-down mapping approach is restricted to verifying the classification criteria. Process maps based on the Gharari et al.'s (2011) approach were derived with a resolution of 25 m and are referred to as GH11 maps (Fig. 3a). Müller et al. (2009) developed classification criteria that take into account the topography (slope), land use and permeability of the geological substratum where again expert knowledge is only required for verification phase. This results in nine output classes, where, beside the DRP, information on the process intensity is provided with a number from "1" (almost immediate reaction) to "3" (strongly delayed reaction). To reduce the number of resulting classes, the DRPs were reclassified into five different runoff types (RTs) according to the intensity of the contribution to runoff (Table 2). As the classification was developed by optimising the classification criteria against a reference map, the method can be also seen as top-down. The resulting process maps with a resolution of 25 m are referred to as MU09 maps (Fig. 3b).

Such simplistic, top-down mapping approaches have been criticised by experimentalists for finding no direct relationships between the runoff coefficient and slope (e.g. Scherrer, 1997). The third mapping approach we used is based on the experimentalist approach introduced by Schmocker-Fackel et al. (2007) and Margreth (2010), which has already been used in, for instance, Antonetti et al. (2016) and Antonetti et al. (2017). Basically, the approach consists of the following steps.

(1) All the available information about a given catchment, including its topography, land use, vegetation, soil, geology, and hydrogeology, is collected and the classification algorithm adapted accordingly.

(2) Small test areas are identified and manually mapped according to Scherrer AG (2006).

(3) The parameter values of the algorithm are identified by comparing the automatically derived map with that derived manually on the test area.

(4) Locations where estimations are not straightforward are verified with a field survey and, where necessary, adjustments are carried out.

(5) Step (4) is reiterated until the process map is considered to be consistent with reality.

Expert knowledge plays a crucial role in this bottom-up method, as all the experimentalists' detailed and qualitative knowledge about processes can be drawn on in the mapping. In addition to the same nine DRP classes used by Müller et al. (2009), the original method of Schmocker-Fackel et al. (2007) allows areas where water is artificially drained (D1-3) to be identified, provided that maps of tile drain systems are available. As these were not available for our study catchments, the original 12 DRP classes get reduced to nine, and the same reclassification criteria as for Müller et al.'s (2009) approach were

used (Table 2). Maps derived with the Schmocker-Fackel et al.'s (2007) method with a resolution of 2 m are referred to as SF07 maps (Fig. 3c).

2.2 The Runoff Generation Module RGM-PRO

The implementation of a physically-based hydrological model was beyond the scope of this study even though the goal was to combine bottom-up and top-down approaches at each step in the modelling chain. This could be seen to go against the definition of bottom-up model favoured by Hrachowitz and Clark (2017) and others, who associate it with physically-based. The concepts "bottom-up" and "top-down" can, however, be interpreted differently even if applied to the same topic and some researchers recommend using a semi-distributed conceptual model to accommodate the features of a catchment efficiently (Savenije and Hrachowitz, 2017). To perform the hydrological simulations for this study the newly developed conceptual PROcess-based Runoff Generation Module (RGM-PRO) was therefore used (Antonetti et al., 2017).

RGM-PRO has a grid based discretisation and was applied with a grid size of 500 m. It is able to take into account the sub-grid variability of the output classes of the process maps (Fig. 4). The model is structured so that a specific combination of storages can be defined for each output class of a given hydrological classification, with one storage system for the plant-available soil moisture (SSM), one for the runoff generation (SUZ) controlled by four free parameters, and a third for groundwater (SLZ; cf. Gurtz et al., 2003; Viviroli et al., 2009b). The separation of rainfall between the storage of plant-available soil moisture and the runoff generation module is controlled by a non-linearity parameter (BETA) fixed here at a value of 3 (Viviroli et al., 2009a). In SUZ, the storage times for overland flow (K0H) and subsurface flow (K1H) regulate the generation of the runoff. A threshold (SGRLUZ) determines the separation between overland and subsurface flow, whereas a maximum percolation rate (CPERC) controls the percolation to the groundwater storage. This is divided into one quick-leaking and two slow-leaking storages controlled by three parameters (SLZ1MAX, CG1H, and K2H). For a more detailed description of the groundwater storage system, see Viviroli et al. (2009b) and Schwarze et al. (1999). This basic structure can then be adapted according to the features of the output classes of a given hydrological classification.

2.2.1 Model initialisation

The initial conditions can significantly affect simulation results, especially in a forecasting context (Liechti et al., 2013). For example, in a study about the uncertainties involved in operational flood forecasting chains in an alpine Swiss catchment, Zappa et al. (2011) found that uncertainty in initial conditions lasts for the first 48 hours, but is almost negligible compared with the uncertainty originating from meteorological data. To investigate to which extent the initial wetness conditions of a catchment affect simulation results with the event-based RGM-PRO, information on the plant-available soil moisture (SSM) is assimilated from quasi-operational grid-based simulations of PREVAH with a resolution of 500 m (Zappa et al., 2014). Each simulation was started at least one day before the beginning of the rainfall event and sufficiently far from possible previous events, so that it was possible to assume that no overland flow and no subsurface flow was occurring in the first time step. Consequently, SUZ was set equal to 0. At the beginning of the simulations, however, the soil moisture value

6

simulated with the PREVAH hydrological system was assigned to each output class of the corresponding cell. Alternatively, as the spatial variability of the soil moisture is higher than the model resolution (500 m), the hydrological downscaling technique described in Blöschl et al. (2009) and used in Antonetti et al. (2017) was implemented. The technique relies on three assumptions: (i) the soil moisture pattern at the smaller scale is time invariant, which allows the process maps to be used as fingerprint; (ii) the spatial variance of the soil moisture at the smaller scale is linked with the one at the larger scale by a scaling theory taken from literature (Blöschl et al., 2009); and (iii) the soil moisture is mass conserving. After the soil moisture was downscaled to a resolution of 25 m, it was successively re-aggregated to obtain an average value for each output class and for each grid cell. Although no expert knowledge is directly involved in this step, the influence of the downscaling technique on the results was still investigated.

2.3 Parameterisation and Parameter Allocation Strategies

Our investigation focussed on floods, where the main processes to be parameterised are the runoff generation within the catchment, the runoff concentration to the drainage net and runoff routing in the stream channel. According to Sivapalan et al.'s (2003) definition, in a bottom-up modelling experiment these three steps are generally parameterised in an explicit manner in the model (Fig. 5). For example, runoff concentration can be taken into account by using a lag function, a linear storage or a combination of them (e.g. Nash, 1957). In a similar way, runoff routing can be considered with a hydraulic approach (for a review, see Heatherman, 2008) or a simpler method such as linear storage in the so-called hydrological approach (e.g. Dyck and Peschke, 1995). Conversely, in a top-down configuration, runoff generation, concentration and routing do not necessarily have to be treated separately (Fig. 5). In both the bottom-up and top-down parameterisations, a consistent parameter allocation strategy was implemented as described in the following sections.

2.3.1 Bottom-up setup: A priori definition of parameter ranges

For the bottom-up setup, RGM-PRO was configured as in Antonetti et al. (2017). The main catchment was first subdivided into several sub-catchments up to 2 km^2 in area. The runoff concentration to the outlet of each sub-catchment was therefore explicitly modelled for both overland and subsurface flow. For overland flow, the flow times were calculated using a semi-hydraulic approach (Schulla, 1997), and for subsurface flow a hydrological approach (i.e. a linear storage with one single parameter GS1H) was used. Considered the size (from 0.5 km^2 up to 2 km^2) of the sub-catchments, into which the main catchment was subdivided, an initial range between 1h and 3h was considered to be plausible for the storage constant governing the concentration of subsurface flow (GS1H; Antonetti et al., 2017). The flow times for the runoff routing in the channel were calculated with a Strickler coefficient of 30 $m^{1/3}$ s^{-1} (Schulla, 1997).

For the allocation of parameters, plausible value ranges were defined a priori for each parameter of RGM-PRO based on the results of sprinkling experiments, on physical properties of soils, and on expert knowledge (Table 2, see Antonetti et al., 2017). Also, idealised response curves were defined for each runoff type. These curves are idealised results from the

sprinkling experiments and represent the expected behaviour of the correspondent runoff type in terms of intensity to runoff contribution (for details see Antonetti et al., 2017). By optimising the value ranges against the generalised response curves for each runoff type, they were further narrowed before being applied to the catchments. As the response curves refer exclusively to the total runoff, the parameter ranges were defined in a manner that allows overland flow and subsurface flow to be partitioned in different ways, provided that the total contribution to runoff reflects that of the corresponding response curve. The number of output classes of the process map by Gharari et al. (2011) differs from that of the process maps used in Antonetti et al. (2017) for the identification of plausible parameter ranges. However, by comparing the landscape classes and runoff types on two catchments on the Swiss Plateau using similarity measures, Antonetti et al. (2016) found out that the most similar pairs were wetland-RT1, hillslope-RT3, and plateau-RT5. The same initial ranges of these runoff types were therefore assigned to the corresponding landscape class accordingly.

2.3.2 Top-down setup: Parameter and Process Constraints

The storage constants for overland flow (K0H) and subsurface flow (K1H) in a top-down approach are expected to represent all three steps of the runoff process described above, i.e. runoff generation, concentration and routing, as in the PREVAH hydrological model (Viviroli et al., 2009a). For the parameter allocation, the initial ranges were defined for each parameter and each output class of the hydrological classification according to Viviroli et al. (2009b), who identified a range of suitable values for each parameter of PREVAH for flood predictions in ungauged mesoscale Swiss catchments (Table 3).

In addition, the model parameter were forced to respect the following constraints (Eq. 1 and Eq. 2):

$$\vartheta_{RT1} < \vartheta_{RT2} < \vartheta_{RT3} < \vartheta_{RT5} < \vartheta_{RT5} \tag{1}$$

$$\vartheta_{WETLAND} < \vartheta_{HILLSLOPE} < \vartheta_{PLATEAU} \tag{2}$$

Where ϑ represents each free parameter of RGM-PRO, namely SGRLUZ, K0H, K1H, and CPERC. For those parameters of RGM-PRO physically similar to those of FLEX-Topo, the same constraints as those imposed by Gharari et al. (2014) were defined for the three landscape classes wetland, hillslope, and plateau. For example, the threshold for the activation of overland flow SGRLUZ was forced to be lower for wetlands, which have a lower storage capacity than the two other landscape classes of the GH11 maps. Similarly, the storage times for both overland and subsurface flow were set to be higher for plateaus than for hillslopes, which were in turn set higher than those for wetlands. The only exception was the storage time for the subsurface flow K1H for wetland (SOF) and plateau (DP). This was set at 1000 h as no subsurface flow was expected there according to hydrologists' understanding of SOF and DP. Similarly, the maximum percolation rate CPERC was forced to be higher for plateaus than for hillslope and wetlands. As the overland flow is expected to be faster than subsurface flow independent of the landscape class, the constraint between the two storage times were defined as follows:

$$K0H_i < K1H_i \tag{3}$$

Where the subscript i refers to the output classes of the GH11 maps, namely wetland, hillslope, and plateau. Following the same reasoning, parameter constraints were defined for the five runoff types of the SF07 and MU09 maps, i.e. RT1-5 (Eq. 1

and Eq. 3). One process constraint in addition to the parameter constraints was defined, namely that the specific peak runoff ($qmax$) should be higher for faster runoff types (Eq. 4):

$$qmax_{RT1} > qmax_{RT2} > qmax_{RT3} > qmax_{RT4} > qmax_{RT5} \tag{4}$$

or for landscape classes (Eq. 5):

$$qmax_{WETLAND} > qmax_{HILLSLOPE} > qmax_{PLATEAU} \tag{5}$$

Randomly selected parameter sets satisfying the parameter constraints were used to perform the simulations with the top-down setup. After the simulations, the runs also satisfying the process constraint were then used for the model evaluation, whereas the other runs were discarded (Gharari et al., 2014).

2.4 Experimental Design

To address the research questions, a total of 60 modelling chain combinations consisting of a given dataset of forcing data, a DRP map, and a model setup were designed (Fig. 6). To investigate the interaction between expert knowledge and quality of forcing data, meteorological data with increasing levels of accuracy were used. Precipitation data from five automatic stations in or close to the basin with a hourly resolution were interpolated based on Thiessen polygons (Thiessen, 1911) and following an Inverse Distance Weighting (IDW) method (Isaaks and Srivastava, 1989) with the power parameter p set equal to 2. In addition, the Combiprecip product (Sideris et al., 2014), a combination of ground measurements and radar quantitative estimations of precipitation, was used. To gradually increase the degree of realism, different spatial aggregations of Combiprecip were introduced. First, for each time step, the average precipitation intensity was distributed all over the main basin (CPC.mean). In the next configuration (CPC.mean.subc), the average precipitation intensity was calculated for and assigned to the corresponding sub-catchment. Finally, the Combiprecip data were used directly as they were delivered by MeteoSwiss. A total of six events were simulated with each modelling chain combination (Table 4). According to the flood type classification of Sikorska et al. (2015), three of them can be classified as short-duration events, and the remaining three as long-duration events. The event in August 2005 was also considered in this study even though no data from the automatic meteorological stations were available, as it was by far the largest flood event to have taken place in the last decades in Switzerland (Hegg et al., 2008).

At the beginning of each simulation, for each grid cell, the spatially distributed soil moisture data from PREVAH simulations were either directly assigned to each output class, i.e. runoff type or landscape class, or first downscaled (section 2.2.1) and successively re-aggregated to obtain an averaged value for each output class from the process map. The three mapping approaches of increasing complexity described in section 2.1 were used to map the spatial distribution of the DRP areas. Finally, the two parameterisations of section 2.3 were applied, each with its own parameter allocation strategy. For the modelling chain combinations based on the bottom-up setup, 10 different combinations of parameter values were randomly selected within the ranges defined a priori (section 2.3.1) to gain insights into the parameter uncertainty. For each modelling chain based on the top-down setup, a Monte Carlo simulation with 100 runs was performed for the same reason. To make

9

comparison fairer, however, only the first ten combinations satisfying the process constraint were considered. For both setups, the value distribution within each range was assumed to be uniform.

The modelling chain combinations forced with the best quality and most realistic data, i.e. those driven with Combiprecip data and hydrological downscaled soil moisture data, were treated as the benchmark modelling chains.

5 Runoff simulations were evaluated with the Kling Gupta Efficiency (KGE; Gupta et al., 2009):

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \qquad (6)$$

This allows not only the correlation between the simulated and measured runoff (r) to be taken into account, but also the ratio between the standard deviation of the simulated runoff and that of the measured runoff (α), and the ratio of the mean simulated to the mean observed discharge β. Furthermore, to quantify any potential overconfidence problems with the model

10 setups, two factors were adapted from Abbaspour et al. (2009): the P-factor and the R-factor. The P-factor is the fraction of the measured runoff enveloped by the uncertainty band originating from the different runs of the Monte Carlo simulations, expressed here by the average distance between the highest (Q_{sim}^{max}) and lowest (Q_{sim}^{min}) value of simulated discharge for each time step (Eq. 7):

$$U_{Q_{sim}} = \frac{1}{n} \sum_{i=1}^{n} \left(Q_{sim,i}^{max} - Q_{sim,i}^{min} \right) \qquad (7)$$

15 where n refers to the total number of time steps. The R-factor is the average width of the uncertainty band $U_{Q_{sim}}$ divided by the standard deviation of the measured runoff $\sigma_{Q_{meas}}$.

$$Rfactor = \frac{U_{Q_{sim}}}{\sigma_{Q_{meas}}} \qquad (8)$$

Ideally, the P-factor is equal to 1, meaning that the observed hydrograph is bracketed by the model parameter uncertainty, whereas the R-factor tends to be zero, i.e. the simulation has the smallest uncertainty band.

20 Finally, to obtain insights into which uncertainty source contributes most to the total predictive uncertainty, an analysis of variance (ANOVA) was carried out. Compared to other sensitivity analysis methods, ANOVA was found to yield the most robust results without much computational efforts (Tang et al., 2007). ANOVA is based on the assumption that the uncertainty of an environmental system can be explained by the output variance generated by different effects, and has already been used to assess uncertainty, for instance, in climate impact projections (Addor et al., 2014; Bosshard et al., 2013;

25 Köplin et al., 2013) and agro-hydrological applications (Moreau et al., 2013). ANOVA helps to clarify the question of how much of the available expert knowledge is worth feeding into a hydrological classification, given the unavoidable uncertainty linked with the input data. Assuming that all the chain components have an effect on the variability of the simulation performance ΔKGE, the following effect model was used (Eq. 9):

$$\Delta KGE = \overline{KGE} + ID_a + IC_b + PM_c + PP_d + I_{abcd} + \varepsilon_{abcd} \qquad (9)$$

30 Where $\overline{KGE}$ represents the mean performance of the modelling chain combinations, ID_a is the main effect of the input data (a = THI, IDW, CPC.mean, CPC.mean.subc, CPC), IC_b is the main effect related to the initial conditions (b = with and without hydrological downscaling), PM_c is related to the process maps with increasing amount of expert knowledge (c =

GH11, MU09, and SF07), and PP_d to the parameterisation and parameter allocation approaches (d = bottom-up, and top-down). I_{abcd} represents the interactions between the main factors, that is the effect of a factor that depends on the effects of one or more other factors, and ε_{abcd} the residual error. The method assumes independence between the different levels (e.g., between the different process maps). Each effect is checked for its representativeness and only those with a p-value lower than 0.05 are taken into account (Chambers et al., 1992). For the description and calculation of each single effect we refer to Bosshard et al. (2013), whereas a comprehensive description of the ANOVA method is given e.g. in von Storch and Zwiers (1999).

3 Results

Using the benchmark modelling chain (i.e. Combiprecip and downscaled initial soil moisture data) and varying the process maps produced different results on the catchments investigated, depending on the model setup (i.e. parameterisation and parameter allocation strategy) used. For example, in the Emme catchment up to Emmenmatt during the rainfall events of August 2005 (Fig. 7a) and September 2012 (Fig. 7b), the modelling chain based on the SF07 map simulated best the runoff peaks for the bottom-up setups, whereas the discharge volume was reproduced satisfactorily with all the process maps. However, irrespective of the process map used, the runoff peaks were simulated with a certain delay, and the falling limb of the hydrograph was overestimated, especially for the short-duration event. With the top-down setup, the modelling chain based on the GH11 maps reproduced the runoff peaks better than the other process maps, whilst the runoff volume was slightly underestimated, independent of the process map used.

The results for the other simulated events in the catchments investigated were analysed to gain further insights into the effects of using process maps with different involvement of expert knowledge (Fig. 8). With regard to the short-duration events (Fig. 8a), the bottom-up outperformed the top-down setup in all the catchments investigated with the exception of the Trueb sub-catchment, where none of the configurations reached satisfying results. Concerning the bottom-up configuration, SF07 maps performed best six times, i.e. slightly more often than the MU09 maps (four times), whereas GH11 never performed better than any of the other process maps. In contrast, when performed with the top-down parameterisation, the GH11 map obtained on average better results than the SF07 map, which, in turn, performed slightly better than MU09 map. With respect to the long-duration events (Fig 8b), the performance difference between the two parameterisations was minimal on the main catchment (Emmenmatt), and on the Emme up to Eggiwil, whereas the combinations based on the bottom-up setup performed better than those based on the top-down setup on the Ilfis. None of the two parameterisations outperformed the other one on the Trueb sub-catchment, as they performed best once each. Similarly to what was observed for the short-duration events, none of the process maps outperformed the others within the bottom-up parameterisation. With regard to the top-down setup, the results obtained with the GH11 maps were on average better than those obtained with the other process maps on Emmenmatt and Eggiwil, whereas the MU09 maps performed best on the Ilfis sub-catchment. Again, no clear trend emerged on the Trueb sub-catchment. Over all, the performance spread between different runs of the same

11

Monte Carlo simulation was considerably higher for the top-down than for the bottom-up configuration. Among the combinations based on the top-down experiment, the parameter uncertainty was found to be higher for GH11 maps than for the other process maps.

A visual inspection of the hydrographs in Fig. 9 shows that feeding the modelling chains with rainfall data spatially interpolated with Thiessen polygons has a considerable effect on the runoff peaks and, consequently, on the simulated runoff volume. However, no effect was detected for the falling limb of the hydrographs. Both model setups systematically underestimate the runoff at the gauging station of Trueb, independent of the process map used.

More generally, forcing the modelling chains with rainfall data of lower quality generally decreased the model performance (Fig. 10), moderately for the main catchment and more markedly for Eggiwil and for the Ilfis sub-catchments. The Trueb sub-catchment is an exception, as the use of rainfall data of lower quality increased the model performance nearly everywhere, independent of the process map used. Averaging the Combiprecip data over the whole catchment (CPC.mean) had the lowest impact on the simulated runoff, irrespective of the parameterisation approach and process map used. In contrast, using data interpolated with IDW and Thiessen polygons led on average to considerable performance losses, irrespective of the model parameterisation, especially for short-duration events. The performance losses for short-duration events were higher for the bottom-up than for the top-down setup, whereas their magnitude was similar among the two setups for long-duration events. The most pronounced performance changes were found in the Trueb sub-catchment with the bottom-up setup forced with Combiprecip data averaged over the sub-catchments. The choice of process map appeared to have little effect.

Uncertainty significantly increased with the decrease in size of the sub-catchments according to the analysis of variance (ANOVA), whereas the most important source of uncertainty was the parameterisation and parameter allocation strategy (Fig. 11a). The smallest source of uncertainty was the hydrological downscaling technique, which was found to be responsible for a slight improvement in simulation skills (Fig. A1). The influence of the process maps also increases with decreasing catchment size. However, when considering the two model configurations separately, the main uncertainty source varies depending on the catchment considered (Fig. 11b-c). With regard to the bottom-up experiment, the interaction between input data and process maps was found to be the largest source of uncertainty in the main catchment (Emmenmatt) and in the Ilfis sub-catchment. In the Eggiwil sub-catchment, the hydrological downscaling techniques and the input data were responsible for the largest uncertainties, whereas, on the Trueb sub-catchment, the process maps accounted for most of the differences in performance. Concerning the top-down setup, the input data were responsible for the largest variance in the main catchment and in the Eggiwil and Ilfis catchments, whereas the process maps were increasingly responsible for uncertainty with decreasing size of the sub-catchments.

12

4 Discussion

The main purpose of this study was to test different implementations of expert knowledge in a process-based hydrological modelling framework, following the basic assumption that combining top-down and bottom-up thinking can improve flood predictions and potentially be applied in poorly gauged areas. Methods of different complexity were therefore tested for each step in the modelling process, including hydrological mapping, model parameterisation and parameter allocation. We wanted to find out whether the use of detailed expert knowledge during the mapping phase can improve simulation results, and how different levels of process knowledge interact with the model parameterisation and parameter allocation strategy when they are forced by precipitation products of different quality. In the following sections, we discuss what light our findings shed on the research questions.

4.1 Can more expert knowledge in the mapping phase increase model performance?

We tested the hypothesis that a more complex mapping approach leads to better simulation results with a benchmark modelling chain forced with the best grid-based rainfall data available in real-time for the whole of Switzerland, that is the Combiprecip product (Sideris et al., 2014). Recently, Antonetti et al. (2016) speculated on the added value of using as much of the available expert knowledge as possible for the hydrological classification. Our findings showed, the hypothesis can only be confirmed for the bottom-up setup, where the modelling chain combinations based on the most complex mapping approach (SF07) resulted in, on average, the highest performances in the study catchments. Conversely, no clear performance improvement was obtained by using SF07 maps with the top-down setup, irrespective of the event type considered. The best performances obtained with by the top-down setup and the GH11 map are most likely attributable to the lower number of classes in the GH11 approach (three instead of five), which allowed the model to be more flexible and consequently the hydrographs to be better reproduced, but not necessarily for the right reason (Kirchner, 2006). In fact, the exclusive use of topographical information for the DRP mapping combined with the top-down setup has been shown to work only in the main catchment and in the sub-catchment of Eggiwil. This suggests that combining the mapping method of Gharari et al. (2011) and the parameter allocation strategy of Gharari et al. (2014) is potentially worthwhile for specific types of catchment, especially those topography-controlled, whereas in other basins more complex mapping approaches need to be used (e.g. on Ilfis and Trueb). Fenicia et al. (2016) similarly found that a catchment classification based on geology led to better results than a classification based on HAND in the Attert catchment in Luxemburg.

The results obtained with the simplified mapping approaches (MU09 and GH11) were, on average, only slightly lower than those obtained with the SF07 maps. Therefore, as the effort needed to derive the simplified maps is substantially lower, using one of the two top-down mapping approaches investigated here may be the best choice in terms of cost-benefit. However, this conclusion is not acceptable from an experimentalist point of view. The results may seem acceptable at the gauging stations, but the local representation of the DRP mapped would most likely differ from that expected by an experimentalist. Topography alone cannot furnish information about the storage and infiltration capacity of soils, as Scherrer et al. (2007)

13

pointed out. Therefore, the two top-down mapping approaches tend to overestimate the runoff contribution of steep slope and underestimate it on flat areas (Antonetti et al., 2016).

Modellers and experimentalists need to agree on what they mean by realism, and how much detail hydrologists should provide to achieve it. An exact reproduction of processes at the plot scale (e.g. exact localisation of macropores etc.) is of course unfeasible due to lack of data, and even knowledge, and the high computational effort such a level of detail would require (Beven, 2001, 2000; Semenova and Beven, 2015; Weiler and McDonnell, 2004). No experimentalist would therefore expect this level of detail from a process-based model at the catchment scale, and even if one had the knowledge and computational efficiency, there would still be unknown unknowns to deal with (Di Baldassarre et al., 2016). In our opinion, the hydrological community should aspire to develop models able to reproduce processes in a "realistic" way (i.e. in agreement with the experimentalists' expectation) at least at the sub-catchment or, even better, at the hillslope scale. This should be a feasible goal, especially considering how new measurements techniques continue to be developed and existing ones refined (Savenije and Hrachowitz, 2017). Such high requirements will probably challenge the validity of simplified mapping approaches and highlight the added value of the more complex ones. The availability of measured data for smaller sub-catchments, where the results of the mapping approaches differed greatly (e.g. in the upper part of the Eggiwil sub-basin), could have better emphasised the potential added value given by more accurate process maps. Future research will address this topic.

4.2 Bottom-up versus top-down model setup

Which model setup was more efficient in modelling the catchment systems investigated in this study? To answer this question, the model parameterisations and the parameter allocation strategies used are addressed separately.

The low performances of the top-down setups in simulating the short-duration events probably depend on the parameterisation approach chosen. The coupled parameterisation of runoff generation, concentration, and routing could well be responsible for the insufficiently fast reaction to high precipitation intensity, as, for instance, fast subsurface flow is basically not allowed to occur. With the bottom-up parameterisation, the underestimation of the falling limb of the hydrograph highlighted by the visual inspection of the hydrographs of Fig. 9 is ascribable to the poor representation of the runoff concentration by the bottom-up setup. However, the adaptation of the model structure, e.g. by introducing a function for the explicit consideration of the time lag due to the processes of runoff concentration and routing, was beyond the scope of this study.

Concerning the parameter allocation strategies, the very same low performances reached by the top-down setup during short-duration events could be also related to the modellers' tendency to set relational rules among parameter and fluxes of different classes. Although the definition of parameter and process constraints force the model to behave according to the modeller's perception of the catchment functioning, the parameter space defined by the initial parameter ranges of Viviroli et al. (2009b) was apparently still too large to ensure high performances with only 100 Monte Carlo runs. On the other hand, the bottom-up parameter allocation strategy led to overconfidence problems, as the measured runoff was only partially

enveloped by the uncertainty bands defined by the different runs of the Monte Carlo simulation (Fig. A2). This is directly ascribable to the definition of very narrow initial ranges for each parameter (Antonetti et al., 2017).

Considering the KGE deviations arising from the use of different forcing data furnished further insights into the setups tested here. The lower KGE deviations observed for the top-down setup showed that it can cope better than the bottom-up setup with uncertainties in the input data as it allows parameter values that can compensate for biases in the input data to be selected. This also explains the larger performance spreads reached by the modelling chains based on the top-down setup, as not all the parameter sets fulfil the requirements for compensating a biased forcing.

The bottom-up setup is therefore suitable for identifying uncertainty sources. Once the extent and distribution of DRPs on a given catchment corresponds to the experimentalist's perception, which may still be biased, and once, for each output class of a process map, a proper parameterisation has been chosen, any remaining deviations of the simulated hydrograph from the measured hydrograph can be explained as arising from uncertainties either in the forcing data or in the measured discharge data.

4.3 Expert knowledge under uncertainty

The assumption that more reliable input data would have led to expert knowledge being more effectively applied in hydrological classification was investigated by varying the forcing data of the different modelling chain combinations. No clear trend was however identified among the different process maps. Even using the CombiPrecip data used for the benchmark modelling chain, which provide the best spatially distributed estimation of rainfall data available in real-time for the whole Switzerland, led to considerable uncertainties, especially with short-duration events, due to its spatially resolution (1 km^2) and problems linked with radar images (see also Antonetti et al., 2017). When the input data are of low quality (e.g. interpolated with simple approaches like IDW and Thiessen polygons), the way model performance can change is symptomatic of the presence of compensation effects within the model. For example, the largest deviations, which occurred in the Trueb sub-catchment, are attributable to the meteorological station on Napf, which is located at 1404 m a.s.l.. It only makes sense to regionalise the values from mountain stations if an elevation factor is taken into accounting, otherwise it may, as here, lead to a local overestimation of the precipitation and, consequently, of the discharge (Sevruk and Mieglitz, 2002; Sevruk, 1997).

Over the years, instead of refining the process maps by drawing on more knowledge in the mapping phase, the opposite occurred, and the uncertainty in the input data was used as an excuse for removing complexity from hydrological classifications. For example, Müller et al. (2009) developed their mapping approach based exclusively on information about topography, geology, and land use in order to simplify the method of Schmocker-Fackel et al. (2007), which is in turn a simplification of the manual mapping approach developed by Scherrer and Naef (2003) and is based on all the information available about a basin. Only two years later, Gharari et al. (2011) introduced a further classification approach based exclusively on topography. This oversimplification risk could be avoided by defining better the minimal criteria for "realism" a model should fulfil before claiming that it had improved realism.

15

4.4 Quantifying uncertainty sources

The analysis of variance (ANOVA) on the catchments investigated showed that the uncertainty linked with parameterisation and parameter allocation strategies was always at least comparable quantitatively with that originating from the input data. For the sub-catchments investigated, it was even greater. This suggests that the step in the modelling process in question has the highest potential for improvement. For two of the four catchments investigated, the uncertainty originating from the process maps was found to be comparable with that arising from the different input data. This means that, up to a certain catchment size, a proper mapping of processes is as important as the availability of reliable input data. The soil moisture data assimilated from PREVAH simulations could also represent an important source of uncertainty. Performing a virtual experiment where the catchments were assumed to be completely saturated at the beginning of each event led to large overestimations of the initial peaks during an event (Fig. A3). However, with a view to an operational application of RGM-PRO, the data from the PREVAH simulations used in this study represent the best grid-based estimation of soil moisture available in real time (Horat, 2017). Using of soil moisture data from other grid-based models was beyond the scope of this study.

Results from the ANOVA also showed a considerable increase in uncertainty with decreasing size of the sub-catchments. This was also found by Hellebrand et al. (2011), who attributed it to a wrong choice of the calibration catchment. The poor performances of the bottom-up setup in the Trueb sub-catchment, which originated the large uncertainty shown in Fig. 11, can, however, be attributed to the low quality of the measured discharge data. The measurement accuracy of the gauging station there has already been questioned in another study (Scherrer AG, 2012), and may of course compromise the potential benefits of using more complex process maps. Checking the rating curve of the gauging stations was, however, beyond the scope of this study.

4.5 Limitations of this study

Some aspects to be investigated during future research include working towards a more thorough modelling system by investigating not only the runoff formation process but also other fluxes that can dominate in a basin such as evapotranspiration and interception. Investigating the influence of expert knowledge on the parameterisation of these processes was beyond the scope of this study, but could represent a direction for future research. We restricted our modelling to an event-based runoff generation module because the SF07 maps and the MU09 maps had been developed with a focus on floods. The simulation time step of one hour for investigations on floods is limiting especially when simulating short-duration events (Steinbrich et al., 2016). Sideris et al. (2014) proposed a disaggregation scheme for the generation of precipitation estimates with a resolution of five and ten minutes, but this involves still large uncertainties, and the hourly aggregated data was found to produce higher skill scores in the validation phase. We therefore only included hourly forcing in this study. The equations governing the storage behaviour were solved with an explicit Euler scheme, which has already been found to be responsible for uncertainty in other studies due to the numerical approximations involved (Kavetski and

16

Clark, 2010). To address this issue, an adaptive number of sub-hourly integration steps was introduced according to the intensity of water reaching the upper-zone runoff storage SUZ.

No soft data from experimentalists' campaigns was used to inform or validate our model. This approach was demonstrated to be valuable to pursue the dialogue between modellers and experimentalists (Seibert and McDonnell, 2002). For the evaluation of the modelling chain combinations, we used the KGE metric exclusively instead of multiple validation criteria suggested by several authors (e.g. Güntner et al., 1999; Krause et al., 2005; Moussa and Chahinian, 2009; Seibert and McDonnell, 2002; Uhlenbrook and Leibundgut, 2002; Weiler and McDonnell, 2007). The KGE is, however, a comprehensive objective function that takes into account both peak and volumetric errors. It was therefore considered suitable for event-based model evaluation. The assumption of independence between different levels in the ANOVA is not fulfilled, as e.g. three datasets use the same forcing (i.e. Combiprecip), and the two model setups are based on the same runoff generation module (i.e. RGM-PRO). Finally, to generalise the findings of this study, the number of catchments and events investigated should be increased considerably. For example, investigating catchments with contrasting reactions to heavy rainfall should provide more support for using more complex mapping approaches to identify the extent and distribution of DRPs.

5 Conclusions

Recent calls to combine bottom-up and top-down reasoning to improve the realism of process-based hydrological models were what motivated this study. We wanted to obtain insights into how to best use expert knowledge, given unavoidable uncertainties. First, we investigated how applying different degrees of expert knowledge in landscape classification affects the final outcome of hydrological simulations. We compared two different setups (i.e. parameterisation and parameter allocation strategies): the first is based on experimentalists' (bottom-up) reasoning, and the second is driven by a modellers' (top-down) thinking. We then looked at how performance varied with different levels of uncertainty in the forcing data before finally quantifying the fraction of variance explained by each uncertainty source.

The main findings of the study were:

- Using complex process maps with high involvement of expert knowledge adds little potential value due to large uncertainties occurring even with the best forcing data available in real-time and in the measured discharge data. Performance using simplified mapping approaches was also satisfactory, especially for long-duration events.

- The bottom-up setup performed better on average than the top-down setup in the catchments investigated, independent of the process map used. The top-down setup was able to accommodate biases in the precipitation data at the expense of exactly identifying sources of uncertainty. Conversely, the bottom-up setup can be used diagnostically to identify uncertainty sources, but had overconfidence problems due to an overly narrow a priori definition of parameter ranges.

17

- The uncertainty linked with the process maps and, consequently, the importance of a realistic representation of the spatial distribution of processes, increased with decreasing size of the catchments.

In conclusion, modellers and experimentalists need to reach agreement on what they mean by "model realism", especially concerning the level of detail. In our opinion, a catchment scale model should be able to reflect the real distribution of dominant runoff processes up to the hillslope scale. More accurate process maps can help to achieve this goal.

Data Availability

MU09 maps, GH11 maps and RGM-PRO are available from Manuel Antonetti (manuel.antonetti@wsl.ch), and the PREVAH soil moisture estimations from Massimiliano Zappa (massimiliano.zappa@wsl.ch). The GIS data used for deriving the MU09 maps and GH11 maps can be obtained undo license from the Federal Office of Topography swisstopo, whereas the SF07 maps were provided by Scherrer AG and SoilCom GmbH (contact the authors for help in accessing them). The runoff data is available from the Swiss Federal Office for the Environment and the Canton of Bern, and the precipitation data from the Swiss Federal Office of Meteorology and Climatology MeteoSwiss (free of charge for scientific purposes).

Appendices

See Table A1, Fig. A1, Fig. A2, and Fig. A3.

Competing interests

The authors declare that they have no conflict of interest.

References

Abbaspour, K. C., Faramarzi, M., Ghasemi, S. S. and Yang, H.: Assessing the impact of climate change on water resources in Iran, Water Resour. Res., 45(10), doi:10.1029/2008WR007615, 2009.

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R. and Seibert, J.: Robust changes and sources of uncertainty in

18

the projected hydrological regimes of Swiss catchments, Water Resour. Res., 50(10), 7541–7562, doi:10.1002/2014WR015549, 2014.

Antonetti, M., Buss, R., Scherrer, S., Margreth, M. and Zappa, M.: Mapping dominant runoff processes: an evaluation of different approaches using similarity measures and synthetic runoff simulations, Hydrol. Earth Syst. Sci., 20(7), 2929–2945,

5 doi:10.5194/hess-20-2929-2016, 2016.

Antonetti, M., Scherrer, S., Kienzler, P. M., Margreth, M. and Zappa, M.: Process-based Hydrological Modelling: The Potential of a Bottom-Up Approach for Runoff Predictions in Ungauged Catchments, Hydrol. Process., doi:10.1002/hyp.11232, 2017.

Bahremand, A.: HESS Opinions: Advocating process modeling and de-emphasizing parameter estimation, Hydrol. Earth

10 Syst. Sci., 20(4), 1433–1445, doi:10.5194/hess-20-1433-2016, 2016.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Menard, C., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E. M., Boucher, O., Cox, P. M., Grimmond, C. S. B. and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model\ndescription. Part 1: Energy and water fluxes, Geosci. Model Dev., doi:10.5194/gmd-4-677-2011 <http://dx.doi.org/10.5194/gmd-4-677-2011>, 2011.

15 Beven, K.: How far can we go in distributed hydrological modelling?, Hydrol. Earth Syst. Sci., 5(1), 1–12, doi:10.5194/hess-5-1-2001, 2001.

Beven, K. J.: Uniqueness of place and process representations in hydrological modelling, Hydrol. Earth Syst. Sci., 4(2), 203–213, doi:10.5194/hess-4-203-2000, 2000.

Blöschl, G.: Scaling in hydrology, Hydrol. Process., 15(4), 709–711, doi:10.1002/hyp.432, 2001.

20 Blöschl, G., Komma, J. and Hasenauer, S.: Hydrological downscaling of soil moisture, Final Rep. to H-Sat via Austrian Cent. Inst. Meteorol. Geodyn., 1–64 [online] Available from: http://hsaf.meteoam.it/documents/reference/HSAF_VS_38_TUWIEN-final-report.pdf, 2009.

Böhringer, C. and Rutherford, T. F.: Combining bottom-up and top-down, Energy Econ., 30(2), 574–596, doi:10.1016/j.eneco.2007.03.004, 2008.

25 Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M. and Sch??r, C.: Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections, Water Resour. Res., 49(3), 1523–1536, doi:10.1029/2011WR011533, 2013.

Cellucci, C.: Top-Down and Bottom-Up Philosophy of Mathematics, Found. Sci., 18(1), 93–106, doi:10.1007/s10699-012-9287-6, 2013.

30 Chambers, J. M., Freeny, A. and Heiberger, R. M.: Analysis of variance; designed experiments, in Statistical Models in S, edited by J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole., 1992.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. a., Vrugt, J. a., Gupta, H. V., Wagener, T. and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44(12), n/a–n/a, doi:10.1029/2007WR006735, 2008.

19

Clark, M. P., McMillan, H. K., Collins, D. B. G., Kavetski, D. and Woods, R. A.: Hydrological field data from a modeller's perspective: Part 2: Process-based evaluation of model hypotheses, Hydrol. Process., 25(4), 523–543, doi:10.1002/hyp.7902, 2011.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. a, Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J. and Rasmussen, R. M.: A unified approach for process-based hydrologicmodeling: 1. Modeling concept, Water Resour. Res., 51(4), 1–17, doi:10.1002/2015WR017200.A, 2015.

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Freer, J. E., Arnold, J. R., Moore, R. D., Istanbulluoglu, E. and Ceola, S.: Improving the theoretical underpinnings of process-based hydrologic models, Water Resour. Res., 52(3), 2350–2365, doi:10.1002/2015WR017910, 2016.

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijenhoet, R., Bennet, 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. Discuss., 1–14, doi:10.5194/hess-2016-693, 2017.

Di Baldassarre, G., Brandimarte, L, and Beven, K. The seventh facet of uncertainty: wrong assumptions, unknowns and surprises in the dynamics of human–water systems. Hydrological Sciences Journal Vol. 61, Iss. 9, 2016.

Dyck, S. and Peschke, G.: Die Abflusskonzentration im Gewässernetz, in Grundlagen der Hydrologie, p. 536, Verlag für Bauwesen, Berlin., 1995.

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., Downer, C. W., Camporese, M., Davison, J. H., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M. and Tarboton, D.: An overview of current applications, challenges, and future trends in distributed process-based models in hydrology, J. Hydrol., 537, 45–60, doi:10.1016/j.jhydrol.2016.03.026, 2016.

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, 2011.

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., doi:10.1002/2015WR017398, 2016.

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S. and Savenije, H. H. G.: Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China, Hydrol. Earth Syst. Sci., 18(5), 1895–1915, doi:10.5194/hess-18-1895-2014, 2014.

Gharari, S., Hrachowitz, M., Fenicia, F. and Savenije, H. H. G.: Hydrological landscape classification: Investigating the performance of HAND based landscape classifications in a central European meso-scale catchment, Hydrol. Earth Syst. Sci., 15(11), 3275–3291, doi:10.5194/hess-15-3275-2011, 2011.

Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H. and Savenije, H. H. G.: Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration, Hydrol. Earth Syst. Sci., 18(12), 4839–4859, doi:10.5194/hess-18-4839-2014, 2014.

Gilbert, C. D. and Li, W.: Top-down influences on visual processing, Nat. Rev. Neurosci., 14(5), 350–363,

doi:10.1038/nrn3476, 2013.

Güntner, A., Uhlenbrook, S., Seibert, J. and Leibundgut, C.: Multi-criterial validation of TOPMODEL in a mountainous catchment, Hydrol. Process., 13(11), 1603–1620, doi:10.1002/(SICI)1099-1085(19990815)13:11<1603::AID-HYP830>3.0.CO;2-K, 1999.

5   Gupta, H. V., Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1-2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Gurtz, J., Zappa, M., Jasper, K., Lang, H., Verbunt, M., Badoux, A. and Vitvar, T.: A comparative study in modelling runoff and its components in two mountainous catchments, Hydrol. Process., 17(2), 297–311, doi:10.1002/hyp.1125, 2003.

10   Heatherman, W. J.: Flood Routing on Small Streams: A Review of Muskingum-Cunge, Cascading Reservoirs, and Full Dynamic Solutions, University of Kansas., 2008.

Hegg, C., Bezzola, G. and Koschni, A.: Ereignisanalyse Hochwasser 2005 in der Schweiz, in Proc. of the XI International Congress Interpraevent 2008, Dornbirn, vol. 2, pp. 27–38., 2008.

Hellebrand, H., Müller, C., Matgen, P., Fenicia, F. and Savenije, H.: A process proof test for model concepts: Modelling the

15   meso-scale, Phys. Chem. Earth, 36(1-4), 42–53, doi:10.1016/j.pce.2010.07.019, 2011.

Horat, C.: Operational Applications of a Process-based Runoff Generation Module in the Emme and Ticino Areas, ETHZ., 2017.

Hrachowitz, M. and Clark, M.: HESS Opinions: The complementary merits of top-down and bottom-up modelling philosophies in hydrology, Hydrol. Earth Syst. Sci. Discuss., 1–22, doi:10.5194/hess-2017-36, 2017.

20   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.

25   Hümann, M. and Müller, C.: Improving the GIS-DRP Approach by Means of DelineatingRunoff Characteristics with New Discharge Relevant Parameters, ISPRS Int. J. Geo-Information, 2, 27–49, doi:10.3390/ijgi2010027, 2013.

Isaaks, E. H. and Srivastava, R. M.: An Introduction to Applied Geostatistics, Oxford University Press, New York. [online] Available from: https://app.knovel.com/web/toc.v/cid:kpAIAG000U/viewerType:toc/root_slug:an-introduction-applied (Accessed 10 February 2017), 1989.

30   Kavetski, D. and Clark, M. P.: Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, Water Resour. Res., 46(10), doi:10.1029/2009WR008896, 2010.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42(3), 1–5, doi:10.1029/2005WR004362, 2006.

Klemeš, V.: Conceptualization and scale in hydrology, J. Hydrol., 65(1-3), 1–23, doi:10.1016/0022-1694(83)90208-1, 1983.

Köplin, N., Schädler, B., Viviroli, D. and Weingartner, R.: The importance of glacier and forest change in hydrological climate-impact studies, Hydrol. Earth Syst. Sci., 17(2), 619–635, doi:10.5194/hess-17-619-2013, 2013.

Kraft, P., Vaché, K. B., Frede, H. G. and Breuer, L.: CMF: A Hydrological Programming Language Extension For Integrated Catchment Models, Environ. Model. Softw., 26(6), 828–830, doi:10.1016/j.envsoft.2010.12.009, 2011.

5   Krause, P., Boyle, D. P. and Base, F.: Comparison of different efficiency criteria for hydrological model assessment, Adv. Geosci., 5, 89–97, doi:10.5194/adgeo-5-89-2005, 2005.

Krebs, P., Armbruster, M. and Rodi, W.: Numerische Nachklärbecken-Modelle, KA - Wasserwirtschaft, Abwasser, Abfall, 47, 985–999, 2000.

Liechti, K., Panziera, L., Germann, U. and Zappa, M.: The potential of radar-based ensemble forecasts for flash-flood early 10   warning in the southern Swiss Alps, Hydrol. Earth Syst. Sci., 17(10), 3853–3869, doi:10.5194/hess-17-3853-2013, 2013.

Margreth, M., Naef, F. and Scherrer, S.: Weiterentwicklung der Abflussprozesskarte Zürich in den Waldgebieten, Zurich., 2010.

Markart, G., Kohl, B., Sotier, B., Schauer, T., Bunza, G. and Stern, R.: Provisorische Geländeanleitung zur Abschätzung des Oberflächenabflussbeiwertes auf alpinen Boden-/Vegetationseinheiten bei konvektiven Starkregen (Version1.0), Vienna., 15   2004.

McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M. and Woods, R. A.: Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, Hydrol. Process., 25(4), 511–522, doi:10.1002/hyp.7841, 2011.

Moreau, P., Viaud, V., Parnaudeau, V., Salmon-Monviola, J. and Durand, P.: An approach for global sensitivity analysis of a complex environmental model to spatial inputs and parameters: A case study of an agro-hydrological model, Environ. 20   Model. Softw., 47, 74–87, doi:10.1016/j.envsoft.2013.04.006, 2013.

Moussa, R. and Chahinian, N.: Comparison of different multi-objective calibration criteria using a conceptual rainfall-runoff model of flood events, Hydrol. Earth Syst. Sci., 519–535, doi:Hydrol. Earth Syst. Sci., 13, 519–535, 2009, 2009.

Müller, C., Hellebrand, H., Seeger, M. and Schobel, S.: Identification and regionalization of dominant runoff processes − a GIS-based and a statistical approach, Hydrol. Earth Syst. Sci., 13(6), 779–792, doi:10.5194/hess-13-779-2009, 2009.

25   Naef, F., Scherrer, S., Thoma, C., Weiler, W. and Fackel, P.: Die Beurteilung von Einzugsgebieten und ihren Teilflächen nach der Abflussbereitschaft unter Berücksichtigung der landwirtschaftlichen Nutzung − aufgezeigt an drei Einzugsgebieten in Rheinland-Pfalz., 2000.

Nalbantis, I., Efstratiadis, A., Rozos, E., Kopsiafti, M. and Koutsoyiannis, D.: Holistic versus monomeric strategies for hydrological modelling of human-modified hydrosystems, Hydrol. Earth Syst. Sci., 15(3), 743–758, doi:10.5194/hess-15-30   743-2011, 2011.

Nash, J. E.: The form of the instantaneous unit hydrograph, in International Association of Hydrological Sciences General Assembly, Toronto, pp. 114–121., 1957.

Nijzink, R. C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., Schäfer, D, Savenije, H. G, and Hrachowitz, M.: The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in

distributed hydrological models. Hydrology and Earth System Sciences, 20(3), 1151–1176. doi:10.5194/hess-20-1151-2016, 2016.

Rennó, C. D., Nobre, A. D., Cuartas, L. A., Soares, J. V., Hodnett, M. G., Tomasella, J. and Waterloo, M. J.: HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia, Remote Sens. Environ., 112(9), 3469–3481, doi:10.1016/j.rse.2008.03.018, 2008.

Savenije, H. H. G.: HESS Opinions "The art of hydrology," Hydrol. Earth Syst. Sci., 13(2), 157–161, doi:10.5194/hess-13-157-2009, 2009.

Savenije, H. H. G. and Hrachowitz, M.: HESS Opinions "Catchments as meta-organisms – a new blueprint for hydrological modelling," Hydrol. Earth Syst. Sci., 21(2), 1107–1116, doi:10.5194/hess-21-1107-2017, 2017.

Scherrer AG: Bestimmungsschlüssel zur Identifikation von hochwasserrelevanten Flächen, Mainz., 2006.

Scherrer AG: Massgebende Hochwasserabflüsse an der Ilfis und an verschiedenen Seitenbächen., 2012.

Scherrer, S.: Abflussbildung bei Starkniederschlägen - Identifikation von Abflussprozessen mittels künstlicher Niederschläge, ETH Zürich., 1997.

Scherrer, S. and Naef, F.: A decision scheme to indicate dominant hydrological flow processes on temperate grassland, Hydrol. Process., 17(2), 391–401, doi:10.1002/hyp.1131, 2003.

Scherrer, S., Naef, F., Faeh, A. O. and Cordery, I.: Formation of runoff at the hillslope scale during intense precipitation, Hydrol. Earth Syst. Sci., 11(2), 907–922, doi:10.5194/hess-11-907-2007, 2007.

Schmocker-Fackel, P., Naef, F. and Scherrer, S.: Identifying runoff processes on the plot and catchment scale, Hydrol. Earth Syst. Sci., 11(2), 891–906, doi:10.5194/hess-11-891-2007, 2007.

Schulla, J.: Hydrologische Modellierung von Flussgebieten zur Abschätzung der Folgen von Klimaäderungen., 1997.

Schwarze, R., Dröge, W. and Opherden, K.: Regional analysis and modelling of groundwater runoff components from catchments in hand rock areas, IAHS Publ. no. 254, 221–232, 1999.

Seibert, J. and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, Water Resour. Res., 38(11), 23, 1–14, doi:10.1029/2001WR000978, 2002.

Semenova, O. and Beven, K.: Barriers to progress in distributed hydrological modelling, Hydrol. Process., 29(8), 2074–2078, doi:10.1002/hyp.10434, 2015.

Sevruk, B.: Regional dependency of precipitation-altitude relationship in the swiss alps, Clim. Change, 36, 355–369, doi:10.1023/A:1005302626066, 1997.

Sevruk, B. and Mieglitz, K.: The effect of topography, season and weather situation on daily precipitation gradients in 60 Swiss valleys, in Water Science and Technology, vol. 45, pp. 41–48., 2002.

Sideris, I. V., Gabella, M., Erdin, R. and Germann, U.: Real-time radar-rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland, Q. J. R. Meteorol. Soc., 140(680), 1097–1111, doi:10.1002/qj.2188, 2014.

Sikorska, A. E., Viviroli, D. and Seibert, J.: Flood-type classification in mountainous catchments using crisp and fuzzy

decision trees, Water Resour. Res., 51(10), 7959–7976, doi:10.1002/2015WR017326, 2015.

Sivapalan, M., Blöschl, G., Zhang, L. and Vertessy, R.: Downward approach to hydrological prediction, Hydrol. Process., 17(11), 2101–2111, doi:10.1002/hyp.1425, 2003.

Smoorenburg, M.: Flood behavior in alpine catchments examined and predicted from dominant runo_ processes. Diss. ETH No. 23010, ETHZ., 2015.

Steinbrich, A., Leistert, H. and Weiler, M.: Model-based quantification of runoff generation processes at high spatial and temporal resolution, Environ. Earth Sci., 75(21), 1423, doi:10.1007/s12665-016-6234-9, 2016.

Tang, Y., Reed, P., Wagener, T. and van Werkhoven, K.: Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, Hydrol. Earth Syst. Sci., 11(2), 793–817, doi:10.5194/hess-11-793-2007, 2007.

Thiessen, A. H.: Precipitation averages for large areas, Mon. Weather Rev., 39, 1082 – 1084, 1911.

Tilch, N., Zillgens, B., Uhlenbrook, S., Leibundgut, C., Kirnbauer, R. and Merz, B.: GIS-gestützte Ausweisung von hydrologischen Umsatzräumen und Prozessen im Löhnersbach-Einzugsgebiet (Nördliche Grauwackenzone, Salzburger Land), Österreichische Wasser- und Abfallwirtschaft, 58(9-10), 141–151, doi:10.1007/BF03164495, 2006.

Uhlenbrook, S. and Leibundgut, C.: Process-oriented catchment modelling and multiple-response validation, Hydrol. Process., 16(2), 423–440, doi:10.1002/hyp.330, 2002.

Viviroli, D., Zappa, M., Gurtz, J. and Weingartner, R.: An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools, Environ. Model. Softw., 24(10), 1209–1222, doi:10.1016/j.envsoft.2009.04.001, 2009a.

Viviroli, D., Mittelbach, H., Gurtz, J. and Weingartner, R.: Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part II: Parameter regionalisation and flood estimation results, J. Hydrol., 377(1), 208–225, doi:10.1016/j.jhydrol.2009.08.022, 2009b.

Von Storch, H. and Zwiers, F. W.: Statistical Analysis in Climate Research, Cambridge University Press., 1999.

Weiler, M. and McDonnell, J.: Virtual experiments: A new approach for improving process conceptualization in hillslope hydrology, J. Hydrol., 285(1-4), 3–18, doi:10.1016/S0022-1694(03)00271-3, 2004.

Weiler, M. and McDonnell, J. J.: Conceptualizing lateral preferential flow and flow networks and simulating the effects on gauged and ungauged hillslopes, Water Resour. Res., 43(3), 1–13, doi:10.1029/2006WR004867, 2007.

Zappa, M., Jaun, S., Germann, U., Walser, A. and Fundel, F.: Superposition of three sources of uncertainties in operational flood forecasting chains, Atmos. Res., 100(2-3), 246–262, doi:10.1016/j.atmosres.2010.12.005, 2011.

Zappa, M., Bernhard, L., Spirig, C., Pfaundler, M., Stahl, K., Kruse, S., Seidl, I. and Stähli, M.: A prototype platform for water resources monitoring and early recognition of critical droughts in Switzerland, Proc. Int. Assoc. Hydrol. Sci., 364, 492–498, doi:10.5194/piahs-364-492-2014, 2014.

Table 1. List of the hydrological classifications used in this study, the data they require, the number of output classes used, and, in brackets, the number of output classes with the original approach. Adapted from Antonetti et al. (2016).

Abbr.	Authors	Topography	Land use	Geology	Soil maps	Extensive field investigations	Number of output classes
GH11	Gharari et al. (2011)	X					3
MU09	Müller et al. (2009)	X	X	X			5(9)
SF07	Schmocker-Fackel et al. (2007)	X	X	X	X	X	5(12)

Table 2. Reclassification of DRPs in runoff types according to their contribution to runoff (HOF = Hortonian Overland Flow; SOF = Saturation Overland Flow; SSF = Subsurface Flow; DP = Deep percolation). 1 represents an almost immediate reaction, 2 a slightly delayed one and 3 a greatly delayed one. Adapted from Naef *et al.* (2000).

Runoff type	DRP	Runoff intensity
RT 1	HOF1/2, SOF1	Fast
RT 2	SOF2, SSF1	Slightly delayed
RT 3	SSF2	Delayed
RT 4	SOF3, SSF3	Greatly delayed
RT 5	DP	Not contributing

5

Table 3. Parameter ranges for the bottom-up and top-down model configurations.

Bottom-up		Runoff type					Landscape class		
		RT1	RT 2	RT 3	RT 4	RT 5	Wetland	Hillslope	Plateau
BETA	[-]					3			
SGRLUZ	[mm]	0-40	40-100	40-100	100-200	200-400	0-40	40-100	200-400
K0H	[h]	0.05-0.4	0.05-0.4	0.05-0.4	0.05-0.4	0.05-0.4	0.05-0.4	0.05-0.4	0.05-0.4
K1H	[h]	1000	0.5-2	2-4	2-4	1000	1000	2-4	1000
CPERC	[mm h^{-1}]	0.1	0.1	0.1-0.5	0.5-5	5-50	0.1	0.1-0.5	5-50
GS1H	[h]					1-3			

Top-down		Runoff type					Landscape class		
		RT 1	RT 2	RT 3	RT 4	RT 5	Wetland	Hillslope	Plateau
BETA	[-]					3			
SGRLUZ	[mm]	0-10	5-20	15-50	20-100	80-200	0-30	20-40	30-50
K0H	[h]	1-30	1-30	1-30	1-30	1-30	1-30	1-30	1-30
K1H	[h]	10-60	10-60	10-60	10-60	10-60	10-60	10-60	10-60
CPERC	[mm h^{-1}]	0.04-0.2	0.04-0.2	0.04-0.2	0.04-0.2	0.04-0.2	0.04-0.2	0.04-0.2	0.04-0.2

Table 4. Start and end of the simulated events. IDW = Inverse Distance Weighting, THI = Thiessen Polygons.

Abbreviation	Simulation start	Simulation end	Event type according to Sikorska et al. (2015)	Specific peak runoff measured at Emmenmatt [m^3 s^{-1} km^{-2}]	No. of ground rain stations available for IDW and THI
Aug10	29.07.2010	31.07.2010	Short-duration	0.48	2
Sep12	11.09.2012	13.09.2012	Short-duration	0.40	5
Aug14	11.08.2014	12.08.2014	Short-duration	0.61	5
Aug05	19.08.2005	24.08.2005	Long-duration	1.08	-
Jun12	07.06.2012	15.06.2012	Long-duration	0.19	5
May16	11.05.2016	15.05.2016	Long-duration	0.34	5

Abbreviation	Simulation start	Simulation end	Event type according to Sikorska et al. (2015)	Specific peak runoff measured at Emmenmatt [m^3 s^{-1} km^{-2}]	No. of ground rain stations available for IDW and THI

Table A1. List of abbreviations used in this study.

Abbreviation	Long name/description
ANOVA	ANalysis Of VAriance
BETA	Non-linearity parameter for infiltration module
CG1H	Storage time for quick baseflow
CPC	Combiprecip
CPC.mean	Combiprecip precipitation data averaged over the whole catchment
CPC.mean.subc	Combiprecip precipitation data averaged over the whole corresponding sub-catchment
CPERC	Maximum percolation rate
DP	Deep Percolation
DRP	Dominant Runoff Process
GH11	Mapping approach after Gharari et al. (2011)
GS1H	Storage time for concentration of subsurface flow
HAND	Height Above the Nearest Drainage
HD	Hydrological Downscaling
HOF	Hortonian Overland Flow
IDW	Precipitation interpolated with the Inverse Distance Weighting method
K0H	Storage time for overland flow
K1H	Storage time for subsurface flow
K2H	Storage time for slow baseflow
KGE	Klingt-Gupta Efficiency
MU09	Mapping approach after Müller et al. (2009)
P	Precipitation
PREVAH	PREecipitation-Runoff-EVApotranspiration HRU Model
Q	Discharge
R	Routing
RC	Runoff Concentration
RG	Runoff Generation
RGM-PRO	PROcess-based Runoff Generation Module
RT	Runoff Type
SF07	Mapping approach after Margreth et al. (2010) and Schmocker-Fackel et al. (2007)
SGRLUZ	Thresold for overland flow
SLZ	Lower zone runoff storage
SLZ1MAX	Maximal content of the quick baseflow storage
SOF	Saturation Overland Flow
SSF	SubSurface Flow
SSM	Soil moisture storage
SUZ	Upper zone runoff storage
THI	Precipitation interpolated with Thiessen polygons
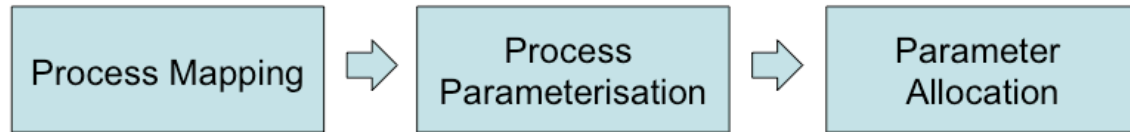
Figure 1: The three main steps for process-based flood predictions and the differences between the bottom-up (bottom) and top-down (top) approaches.
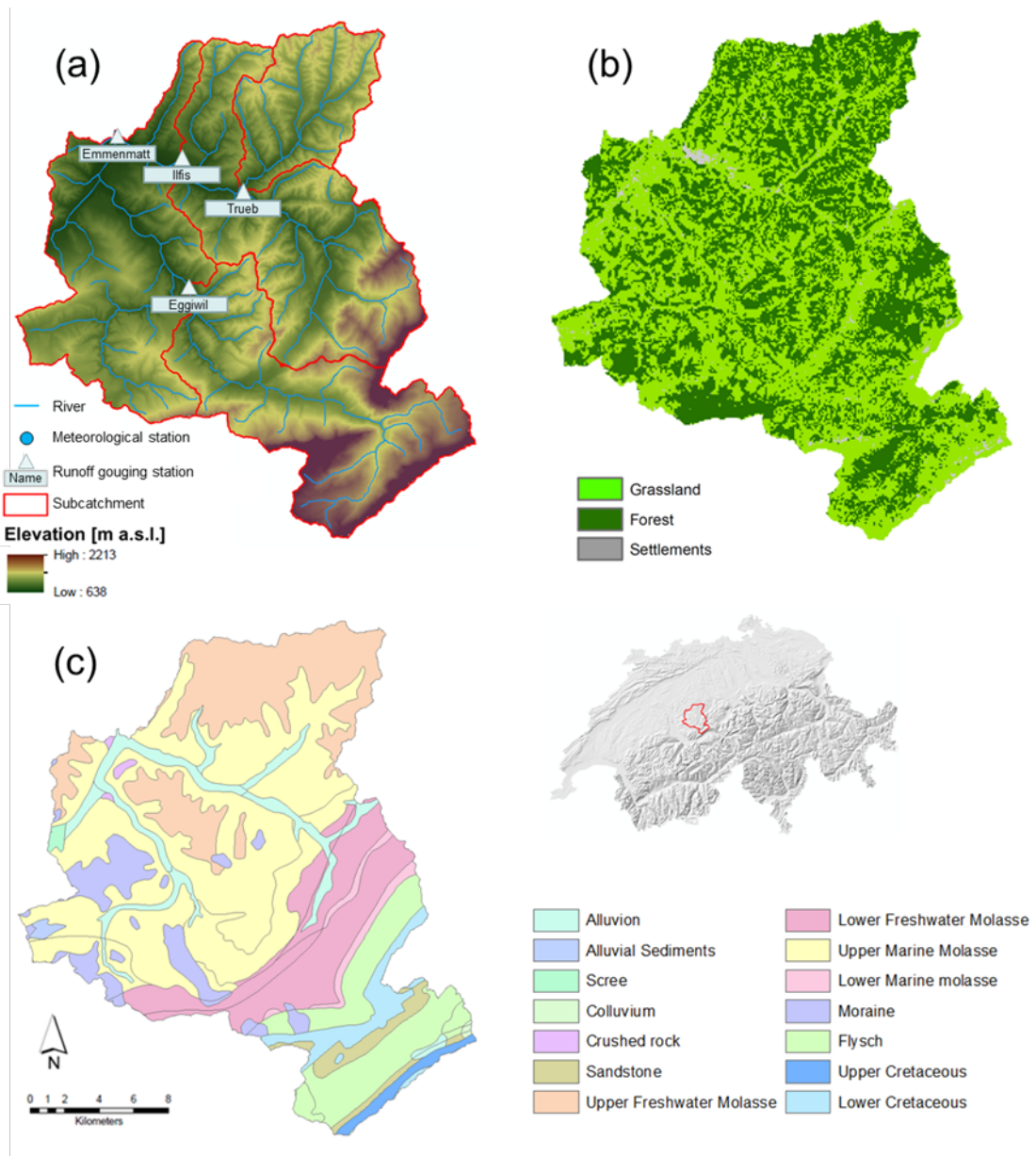
Figure 2: Maps of the Emme catchment, Switzerland. (a) Digital terrain model (25m resolution), river network and location of the runoff gouging stations; (b) land-use map (100 m resolution); (c) geology map. Data: BFS GEOSTAT/Federal Office of Topography swisstopo (DV033492.2).
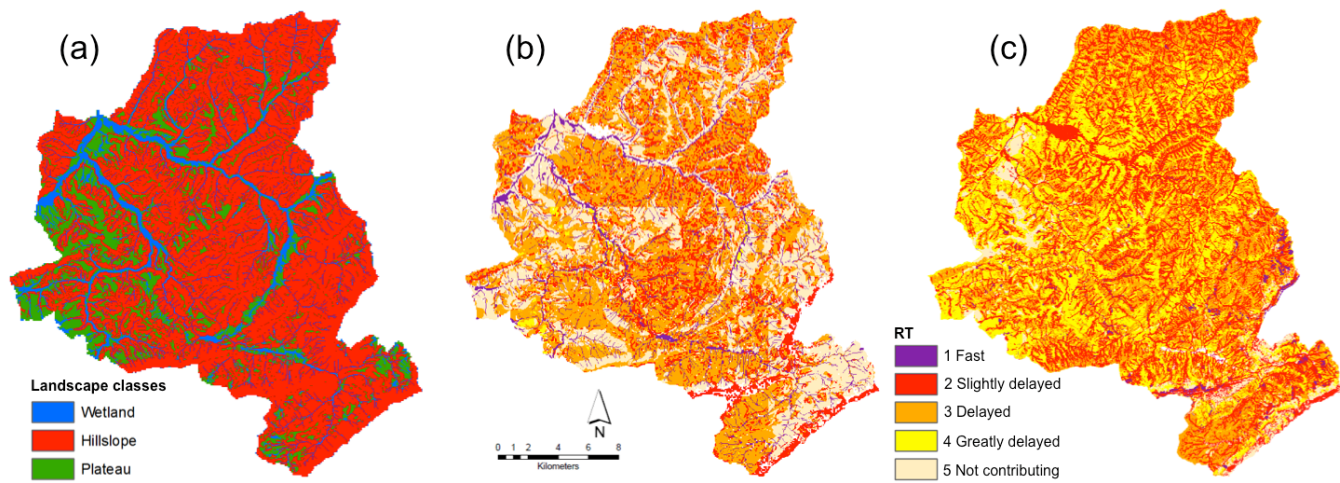
5

Figure 3: Process maps for the Emme catchment map according to (a) Gharari et al. (2011), (b) Müller et al. (2009), and (c) Schmocker-Fackel et al. (2007). RT = runoff type.

5
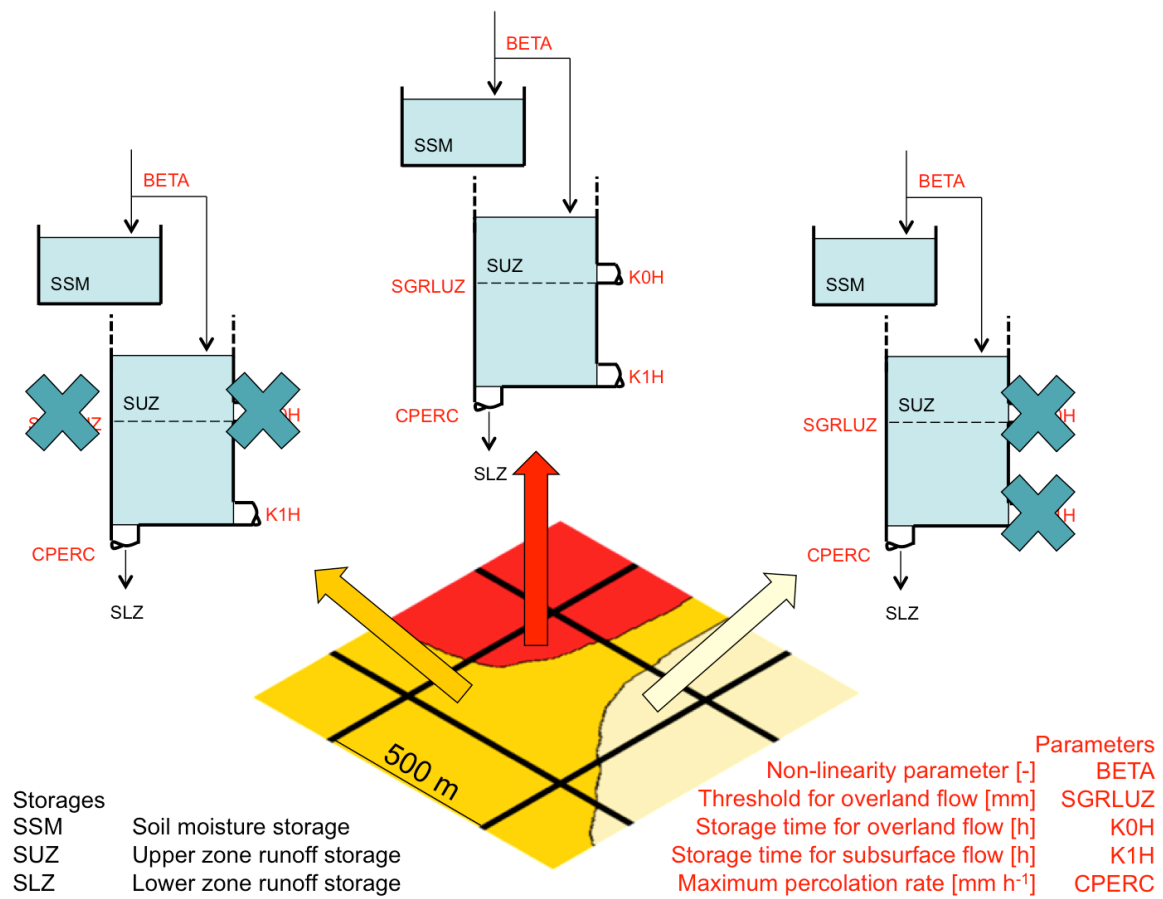
Figure 4. Schematic representation of the spatial discretisation and structure of RGM-PRO. For each class of a given process map, a specific storage system can be defined. For list of abbreviations see Table A1. Similar distributed model structures can be found e.g. in Gharari et al. (2014), Fenicia et al. (2016), Nijzink et al. (2016).

5
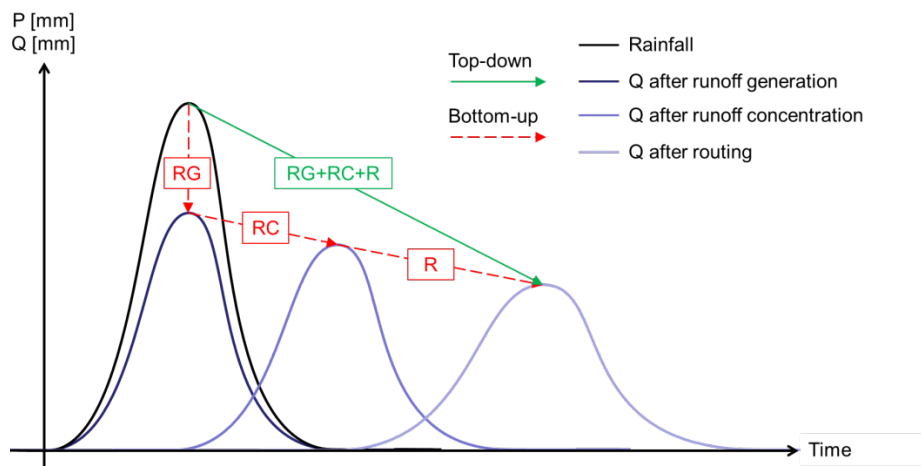
Figure 5. Representation of runoff generation (RG), runoff concentration (RC) and routing (R) in the bottom-up (red) and in the top-down (green) setups. Adapted from Krebs et al. (2000).
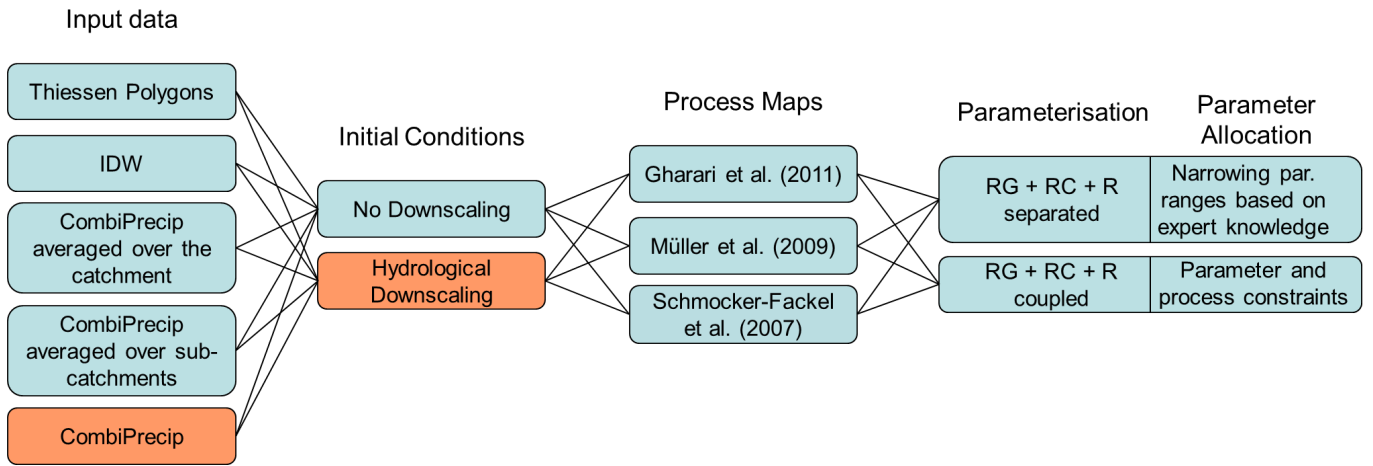
Figure 6. Diagram of the modelling chain combination performed for this study. The components with an orange background form the benchmark modelling chain. IDW = Inverse Distance Weighting; RG = runoff generation; RC = runoff concentration; R = routing.

5
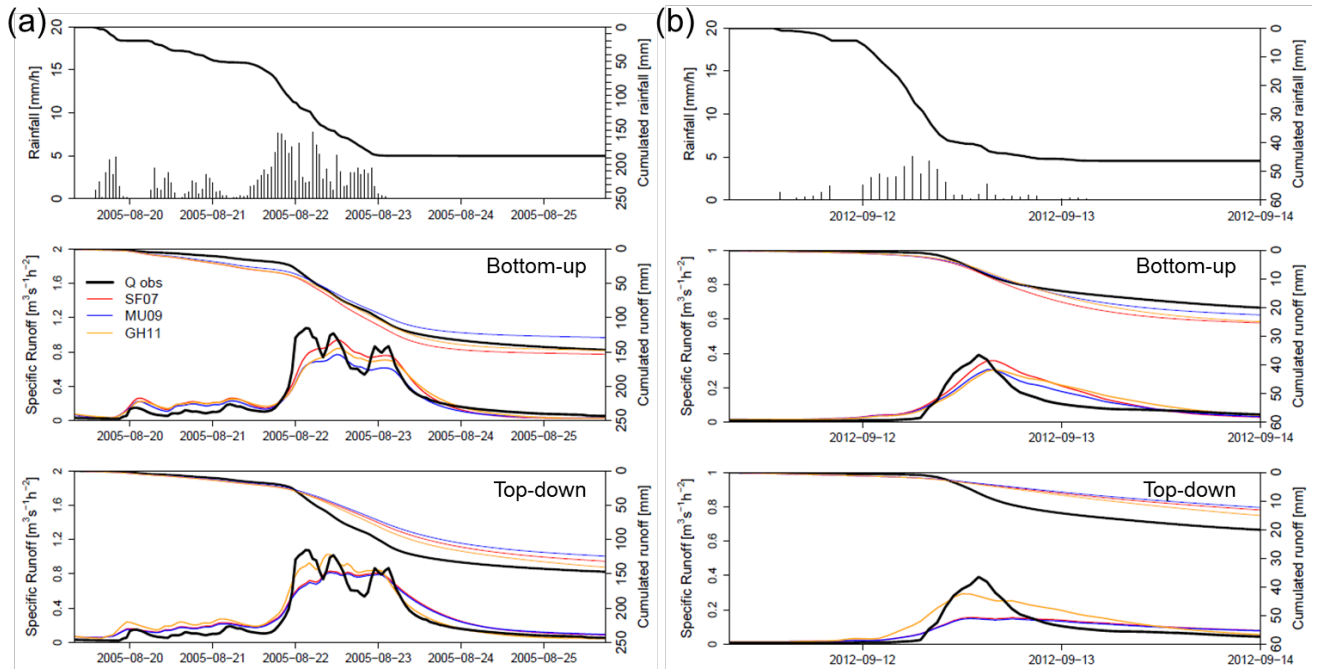
Figure 7. Simulated runoff for the Emme catchment up to Emmenmatt during the long-duration event of August 2005 (a), and during the short-duration event of September 2012 (b), obtained from the different process maps and model parameterisations. The simulated hydrographs refer to the first run of the Monte Carlo simulation performed with the corresponding modelling chain combination. The SF07 map reproduced best the peak runoff with the bottom-up setup, whereas the GH11 map outperformed the other maps with the top-down setup.

36
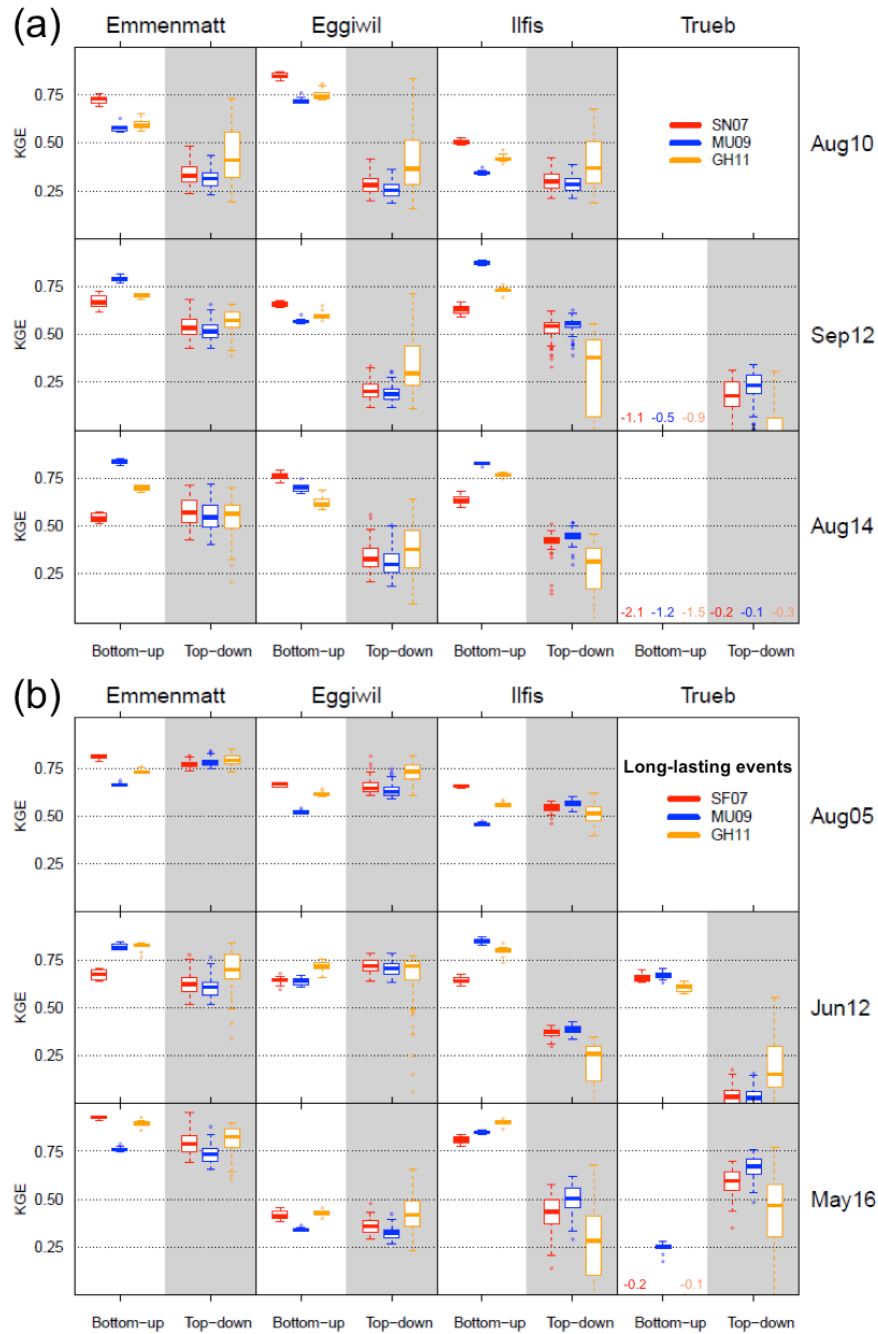
Figure 8. Results from the short-duration (a) and from the long-duration events simulated on the catchments investigated using the benchmark modelling chain. The boxplots represent the simulation results of the bottom-up (white background) and of the top-down (grey background) parameterisations, whereas the coloured borders represent the different mapping approaches. In case of negative performances, the median values are shown. Overall, the bottom-up performed better than the top-down setup during short-duration events, whereas no preference was found for long-duration events.
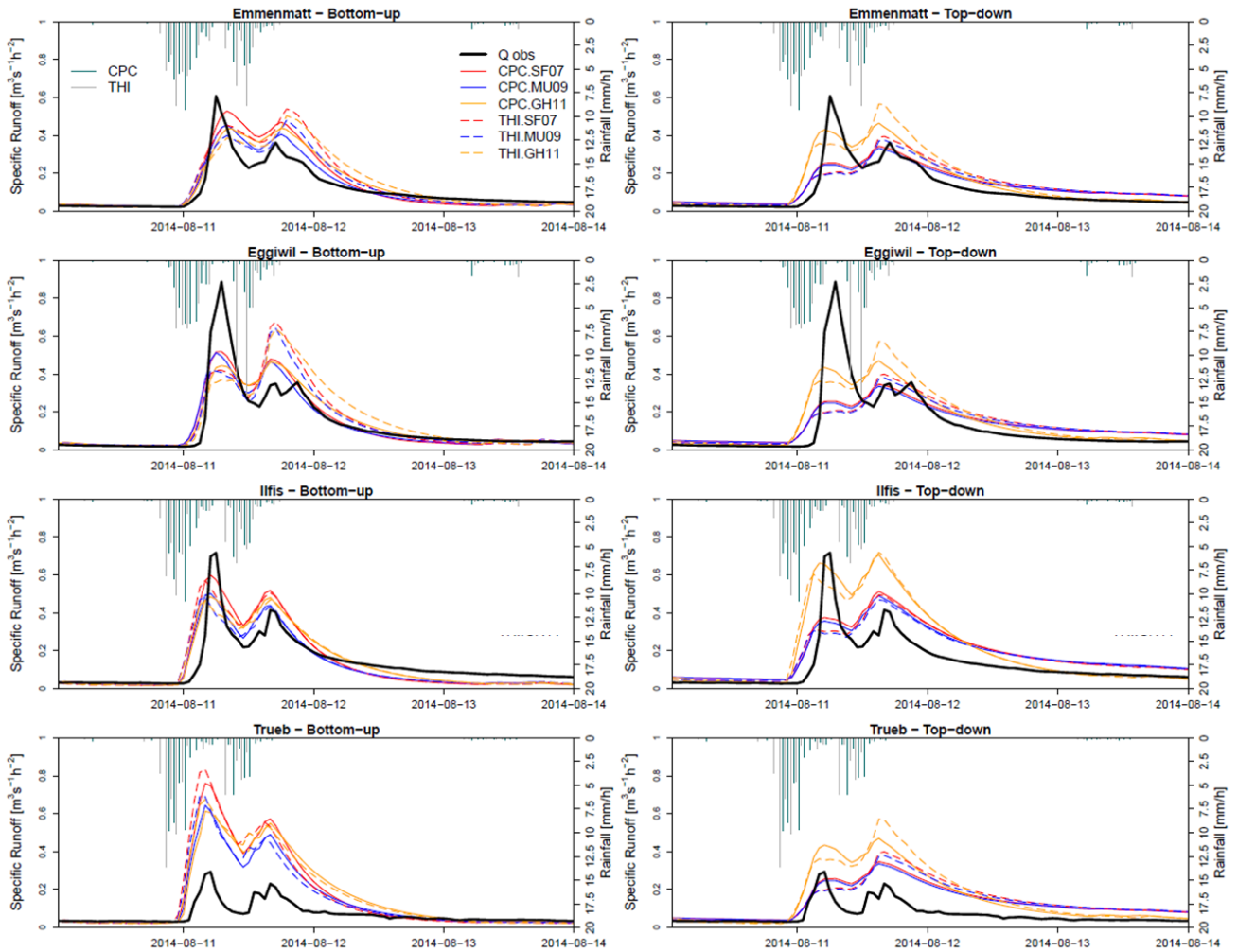
Figure 9. Simulated runoff for the four study catchments during the long-duration event of May 2016, obtained from different input data (CPC = Combiprecip; THI = Thiessen polygons), process maps (SF07, MU09 and GH11), and model setups (bottom-up and top-down). The simulated hydrographs refer to the first run of the Monte Carlo simulation performed with the corresponding modelling chain combination. Errors linked with the input data (e.g. the overestimation of the second runoff peak at Emmenmatt and Eggiwil due to a higher input signal) can be distinguished from those more clearly linked with the model parameterisation.
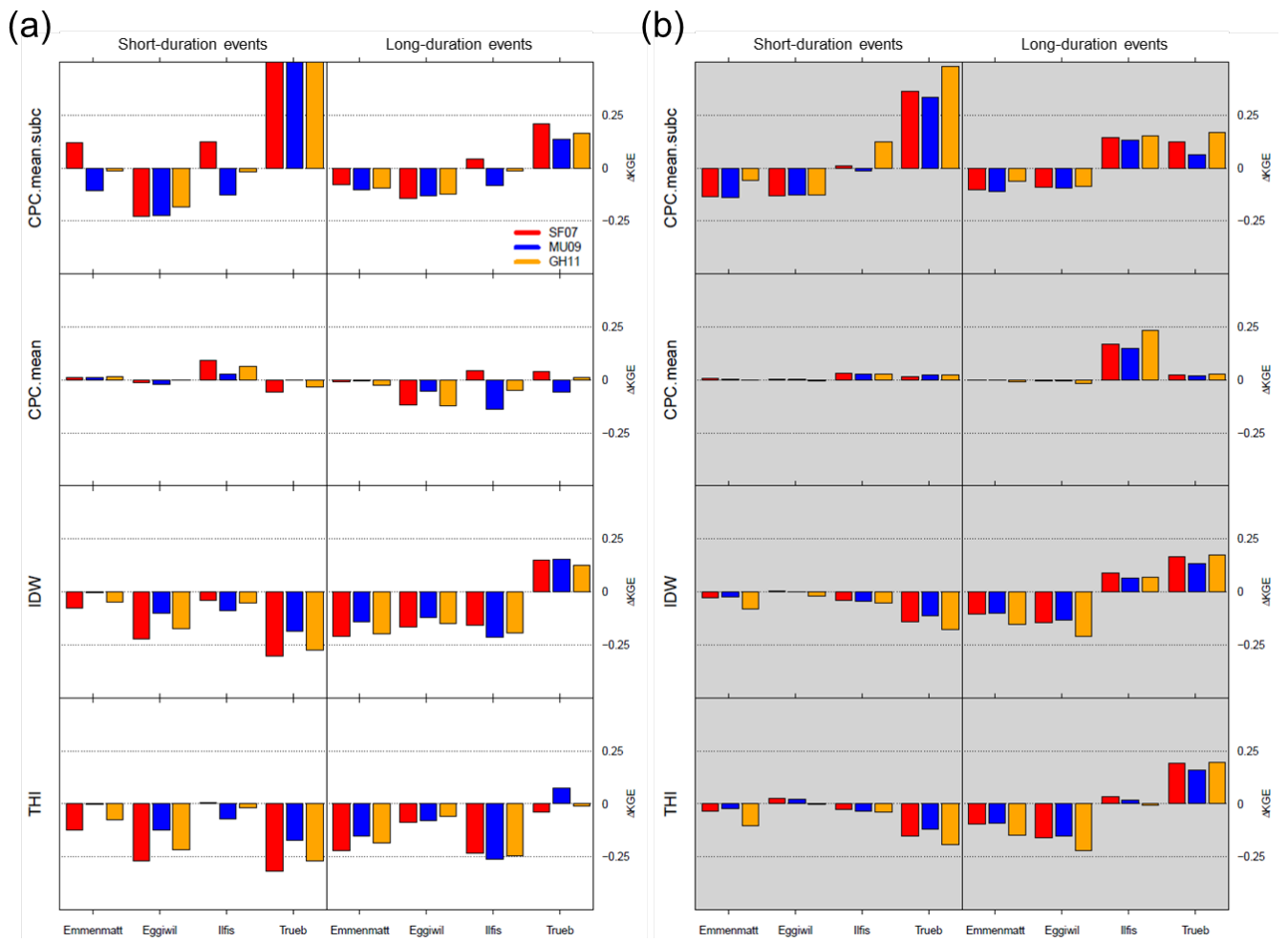
Figure 10. Averaged KGE deviations from the benchmark modelling chain (i.e. driven by Combiprecip data) obtained with the bottom-up (a) and top-down (b) configurations. Each block corresponds to a specific modelling chain based on the rainfall data reported on the left (CPC = Combiprecip; IDW = inverse distance weighting; THI = Thiessen polygons), whereas the displayed event types are reported at the top. The bars represent the average performance difference obtained from Monte Carlo runs for each of the four study areas, whilst the colour of the bars represent the different mapping approaches. Overall, the performance deviations were higher for the bottom-up than for the top-down setup.
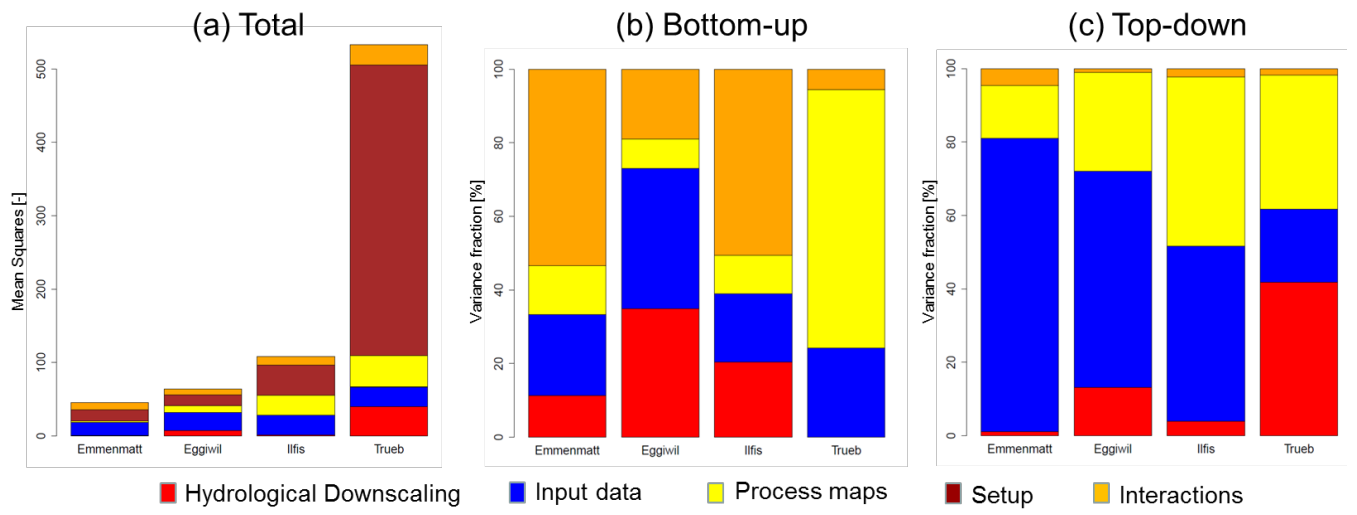
Figure 11. Decomposition of the model performance (KGE) variance at the four gauging stations for all the modelling chain combinations (a), as well as for those based on the bottom-up (b) and top-down (c) configurations. Total uncertainty increases with ingin size of the catchments.
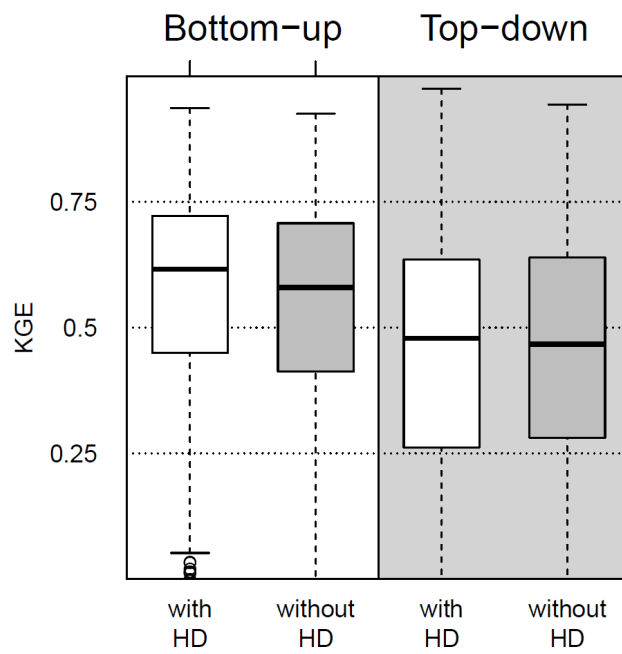
5

Figure A1. Comparison of the simulation results obtained with and without hydrological downscaling (HD) of the initial conditions by the modelling chains based on either the bottom-up or the top-down configuration. HD slightly increased both the best and average performance of the model setups.
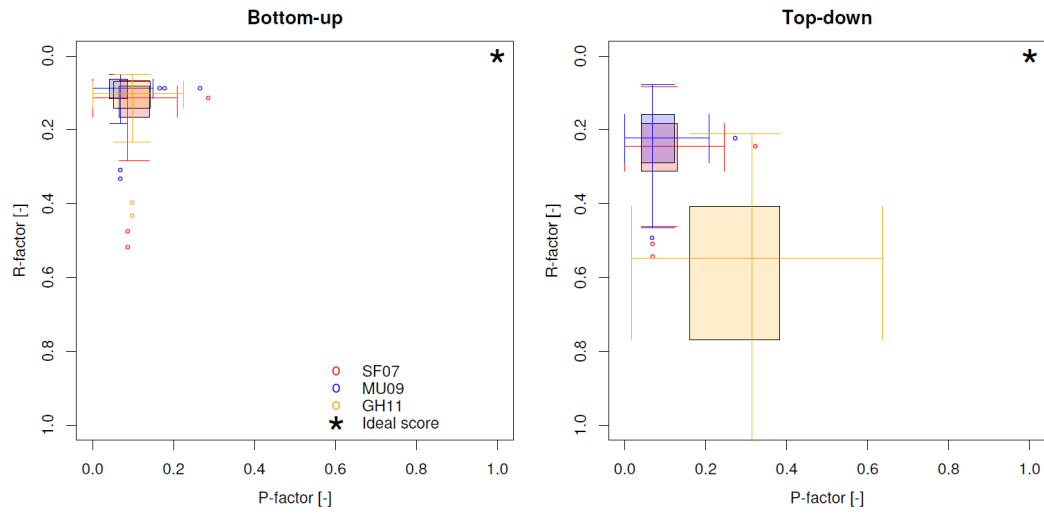
5

Figure A2. Values of P-factors (x axis) and R-factors (y axis) calculated for the different process maps with the bottom-up and top-down setups. The ideal score (i.e. P-factor = 1 and R-factor = 0) is represented with a black asterisk. Whilst the process maps performed similarly with the bottom-up setup, the observed runoff was best bracketed by simulations obtained with the GH11 maps and the top-down setup, but at the expense of a wider uncertainty band (i.e. lower R-factors).
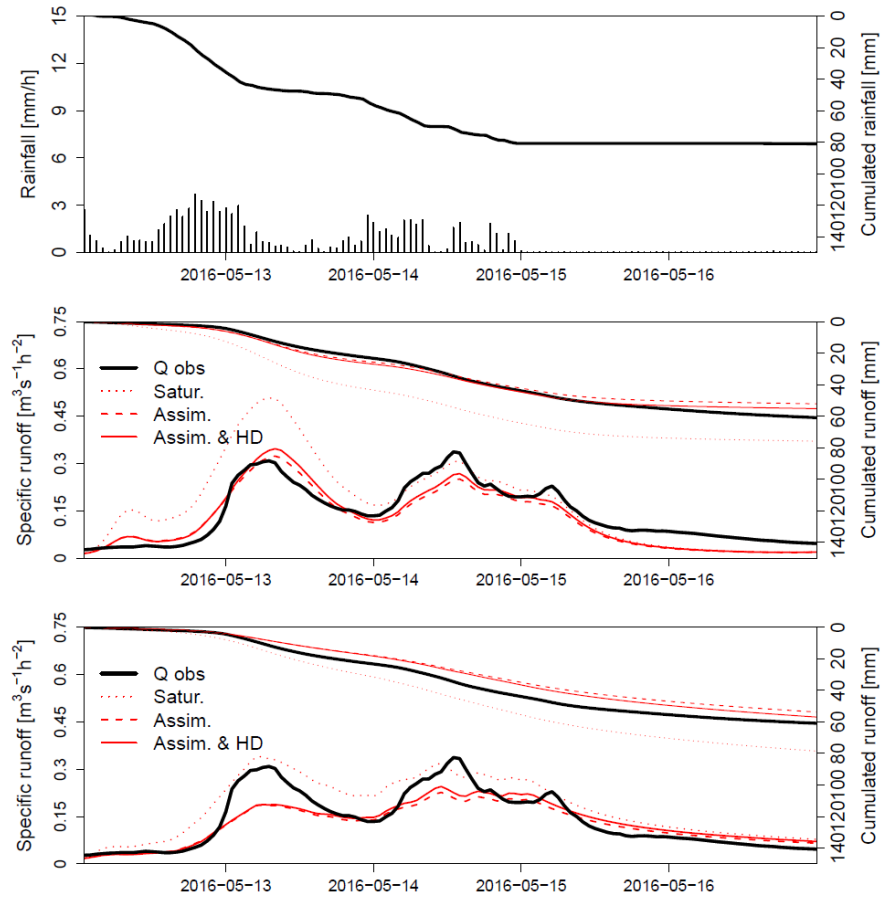
Figure A3. Influence of the soil moisture initial conditions on the simulated runoff for the Emme up to Emmenmatt during the long-duration event of May 2016 obtained with the bottom-up (a) and top-down (b) setup. The simulated hydrographs refer to the first run of the Monte Carlo simulation performed with SF07 map and with Combiprecip as forcing. The saturated initial conditions led to a significant overestimation of runoff at the beginning of the simulations, whereas the hydrological downscaling barely affected the simulation results.