This paper presents a very interesting work to develop a simple topography-driven and calibration-free runoff generation module. The module works for saturation excess runoff generation mechanism, which prevails in most humid/semi-humid areas and is demonstrated by some experts to operate in some arid areas also. The module was rigorously compared against the corresponding models in HBV and TOPMODEL. The experiments in both data-rich experimental watersheds and MOPEX catchments support the superiority of the new module (called HSC and HSC-MCT). The authors also discuss the deep reason of why type question (why can HSC outperforms calibrated-type module) in the context of ecological evolution theory. The proposed method has a wide implication for hydrological and ecological research.

We thank the Editor's positive comments.

Some minor comments are listed below for authors' reference:

1. P5L125, one or two sentences should add to explain MCT concisely. This term is not a popular one in hydrological literature. No further explanation will hinder the reader's understanding.

More explanation of MCT is added. (L158-161)

2. P8L214, the term of subsurface flow. Quite a few different terms have been used for the flow in soil media. The authors can refer to Markus Weiler and Jeffery MacDonnell (in Encyclopedia of Hydrological Sciences. Edited by M G Anderson.). In my mind, the term subsurface flow could refer to all kinds of flow types occurring in soil media, including soil matrix flow, preferential flow, or others. I understand the authors mean preferential type flow by subsurface flow here.

This sentence is modified (L247-248). And we add more interpretation in the discussion. (L585-590)

3. P16L440, 'interestingly' is not suitable here, because HAND by its definition should not depend on elevation.

Changed. (L477-479)

A simple topography-driven and calibration-free runoff generation module

Hongkai Gao[1,2,3,4]*, Christian Birkel [5,6], Markus Hrachowitz [7], Doerthe Tetzlaff [6], Chris Soulsby [6], Hubert H. G. Savenije[7]

[1] Key Laboratory of Geographic Information Science (Ministry of Education of China), East China Normal University, Shanghai, China.

[2] School of Geographical Sciences, East China Normal University, Shanghai, China.

[3] Julie Ann Wrigley Global Institute of Sustainability, Arizona State University PO Box 875402. Tempe, AZ 85287-5402.

[4] Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, China.

[5] Department of Geography, University of Costa Rica, San José, Costa Rica

[6] Northern Rivers Institute, University of Aberdeen, Scotland.

[7] Water Resources Section, Delft University of Technology, Delft, Netherlands.

*Corresponding to Hongkai Gao (hkgao@geo.ecnu.edu.cn)

Abstract

Reading landscapes and developing calibration-free runoff generation models that adequately reflect land surface heterogeneities remains the focus of much hydrological research. In this study, we report a novel and simple topography-driven runoff generation parameterization – the HAND-based Storage Capacity curve (HSC), that uses a topographic index (HAND, Height Above the Nearest Drainage) to identify hydrological similarity and the extent of saturated areas in catchments. The HSC can be used as a module in any conceptual rainfall-runoff model. Further, coupling the HSC parameterization with the Mass Curve Technique (MCT) to estimate root zone storage capacity (S_{uMax}), we developed a calibration-free runoff generation module HSC-MCT. The runoff generation modules of HBV and TOPMODEL were used for comparison purposes. The performance of these two modules (HSC and HSC-MCT) was first checked against the data-rich Bruntland Burn (BB) catchment in Scotland, which has a long time series of field-mapped saturation area extent. We found that HSC, HBV and TOPMODEL all perform well to reproduce the hydrograph, but the HSC module performs better in reproducing saturated area variation, in terms of correlation coefficient and spatial pattern. The HSC and HSC-MCT modules were subsequently tested for 323 MOPEX catchments in the US, with diverse climate, soil, vegetation and geological characteristics. In comparison with HBV and TOPMODEL, the HSC performs better in both calibration and validation, particularly in the catchments with gentle topography, less forest cover and arid climate. Despite having no calibrated parameters, the HSC-MCT module performed comparably well with calibrated modules, highlighting the robustness of the HSC parameterization to describe the spatial distribution of the root zone storage capacity and the efficiency of the MCT method to estimate S_{uMax}. This novel and calibration-free runoff generation module helps to improve the Prediction in Ungauged Basins and has great potential to be generalized at the global scale.

1 Introduction

Determining the volume and timing of runoff generation from rainfall inputs remains a central challenge in rainfall-runoff modelling (Beven, 2012; McDonnell, 2013). Creating a simple, calibration-free, but robust runoff generation module has been, and continues to be, an essential pursuit of hydrological modellers. Although we have made tremendous advances to enhance our ability on Prediction in Ungauged Basins (PUB) (Sivapalan et al., 2003; Blöschl et al., 2013; Hrachowitz et al., 2013), it is not uncommon that models become increasingly complicated in order to capture the details of hydrological processes shown by empirical studies (McDonnell, 2007; Sivapalan, 2009; Yu et al., 2014). More detailed process conceptualization normally demands higher data requirements than our standard climatological and hydrological networks can provide, leading to more calibrated parameters and a probable increase in model uncertainty (Sivapalan, 2009).

Hydrological connectivity is a key characteristic of catchment functioning, controlling runoff generation. It is a property emerging at larger scales, describing the temporal dynamics of how spatially heterogeneous storage thresholds in different parts of catchments are exceeded to contribute to storm runoff generation and how they are thus “connected to the stream” (e.g. Zehe and Blöschl, 2004; Bracken and Croke, 2007; Lehmann et al., 2007; Zehe and Sivapalan, 2009; Ali et al., 2013; Blume and van Meerveld, 2015). Connectivity is controlled by a multitude of factors (Ali and Roy, 2010), including but not limited to surface (e.g. Jencso et al., 2009) and subsurface topography (e.g. Tromp-van Meerveld and McDonnell, 2006), soils (including preferential flow networks; e.g. Zehe et al., 2006; Weiler and McDonnell, 2007), land cover (e.g. Imeson and Prinsen, 2004; Jencso and McGlynn, 2011; Emanuel et al., 2014), the wetness state of the system (e.g. Detty and McGuire, 2010; Penna et al., 2011; McMillan et al., 2014; Nippgen et al., 2015).

In detailed distributed hydrological bottom-up models, connectivity emerges from the interplay of topography, soil type and water table depth. For example, TOPMODEL (Beven and Kirkby, 1979; Beven and Freer, 2001) uses topographic wetness index (TWI) to distinguish hydrologic similarity; and SHE (Abbott et al. 1986) and tRIBS (Ivanov et al. 2004; Vivoni et al. 2005) use partial differential equations to describe the water movement based on pressure gradients obtained by topography; and the Representative Elementary Watershed (REW) approach divides catchment into a number of REWs to build balance and constitutive equations for hydrological simulation (Reggiani et al., 1999; Zhang and Savenije, 2005; Tian et al., 2008). As the relevant model parameters such as local topographic slope and hydraulic conductivity can, in spite of several unresolved issues for example relating to the differences in the observation and modelling scales (e.g. Beven, 1989; Zehe et al., 2014), be obtained from direct observations, they could *in principle* be applied without calibration.

Zooming out to the macro-scale, top-down models, in contrast, are based on emergent functional relationships that integrate system-internal heterogeneity (Sivapalan, 2005). These functional relationships require parameters that are effective on the modelling scale and that can largely not be directly determined with small-scale field observations (cf. Beven, 1995), thus traditionally determined by calibration. However, frequently the number of observed variables for model calibration is, if available at all, limited to time series of stream flow. The absence of more variables to constrain models results in such models being ill-posed inverse problems. Equifinality in parameterization and in the choice of parameters then results in considerable model uncertainty (e.g. Beven, 1993, 2006). To limit this problem and to also allow predictions in the vast majority of ungauged catchments, it is therefore desirable to find ways to directly infer effective model parameters at the modelling scale from readily available data (Hrachowitz et al., 2013).

The component that is central for establishing connectivity in most top-down models is the soil moisture routine. Briefly, it controls the dynamics of water storage and release in the unsaturated root zone and partitions water into evaporative fluxes, groundwater recharge and fast lateral storm flow generating runoff (Gao et al., 2018a; Shao et al., 2018). The latter of which is critical from the aspect of connectivity. In majority regions, Hortonian overland flow (HOF, i.e. infiltration excess overland flow) is of minor importance(Dunne and Black, 1970; Sklash and Farvolden, 1979; Beven, 2004; Burt and McDonnell, 2015), even in arid regions where often most locally generated HOF is re-infiltrated while flowing on hillslopes (Liu et al., 2012) and never reaches the stream channel network. Thus the term saturation excess flow (SEF) can represent, depending on the model and the area of application, different processes, such as saturation overland flow, preferential flow, flow through shallow, high permeability soil layers or combinations thereof. The interplay between water volumes that are stored and those that are released laterally to the stream via fast, connected flow paths ("connectivity") is in most top-down models described by functions between water stored in the unsaturated root zone ("soil moisture") and the areal proportion of heterogeneous, local storage thresholds that are exceeded and thus "connected" (Zhao et al., 1980). In other words, in those parts of a catchment where the storage threshold is exceeded will generate lateral flows, and can alternatively be interpreted as runoff coefficient (e.g. Ponce and Hawkins, 1996; Perrin and Andreassian, 2001; Fenicia et al., 2007; Bergström and Lindström, 2015). Thus the idea goes back to the variable contributing area concept, assuming that only partial areas of a catchment, where soils are saturated and thus storage thresholds are exceeded, contribute to runoff (Hewlett, 1961; Dunne and Black, 1970; Hewlett and Troendle, 1975). Although originally developed for catchments dominated by saturation overland flow, the extension of the concept to subsurface connectivity, posing that surface and subsurface connectivity are "two sides of the same coin" (McDonnell, 2013), proved highly valuable for models such as Xinanjiang (Zhao et al., 1980), HBV (Bergström and Forsman, 1973; Bergström and Lindström, 2015), SCS-CN (Ponce and Hawkins, 1996; Bartlett et al., 2016), FLEX (Fenicia et al., 2008) and GR4J (Perrin and Andreassian et al., 2001).

Among these models, connectivity is formulated in a general form as C_R=f($S_U(t)$,S_{uMax},β), where C_R is the runoff coefficient, i.e. the proportion of the catchment generating runoff, $S_U(t)$ is the catchment water content in the unsaturated root zone at any time t, S_{uMax} is a parameter representing the total storage capacity in the unsaturated root zone and β is a shape parameter, representing the spatial distribution of heterogeneous storage capacities in the unsaturated root zone. The parameters of these functions are typically calibrated. In spite of being the core component of soil moisture routines in many top-down models, little effort was previously invested to find ways to determine the parameters at the catchment-scale directly from available data. An important step towards understanding and quantifying connectivity pattern directly based on observations was recently achieved by intensive experimental work in the Tenderfoot Creek catchments in Montana, US. In their work Jencso et al. (2009) were able to show that connectivity of individual hillslopes in their headwater catchments is highly related to their respective upslope accumulated areas. Using this close relationship, Smith et al. (2013) successfully developed a simple top-down model with very limited need for calibration, emphasizing the value of "enforcing field-based limits on model parameters" (Smith et al., 2016). Based on hydrological landscape analysis, FLEX-Topo model (Savenije, 2010) can dramatically reduce the need for calibration (Gharari et al., 2014), and hold considerable potential for spatial model transferability without the need for parameter re-calibration (Gao et al., 2014a; H. Gao et al., 2016). In a recent development, several studies suggest that S_{uMax} can be robustly and directly inferred from long term water balance data, by the Mass Curve Technique (MCT)~~, without the need for further calibration~~ (Gao et al., 2014; de Boer-Euser et al., 2016; Nijzink et al., 2016). The MCT~~, that~~ is an engineering method for reservoir design, in which the reservoir size is estimated as a function of accumulated inflow and human water demand. The MCT treats the root zone as a reservoir, and estimates catchment-scale S_{uMax} from measurable hydrometeorological data, without the need for further calibration. This leaves shape parameter β as the only free calibration parameter for soil moisture routines of that form. Topography is often the dominant driver of water movement caused by prevailing hydraulic gradients. More crucially, topography usually provides an integrating indicator for hydrological behavior, since topography is usually closely related with other landscape elements, such as soil vegetation climate and even geology (Seibert et al., 2007; Savenije, 2010; Rempe and Dietrich, 2014; Gao et al., 2014b; Maxwell and Condon, 2016; Gomes, 2016). The Height Above the Nearest Drainage (HAND; Rennó et al., 2008; Nobre et al., 2011; Gharari et al., 2011), which can be computed from readily available digital elevation models (DEM), could potentially provide first order estimates of groundwater depth , as there is some experimental evidence that with increasing HAND, groundwater depths similarly increase (e.g. Haria and Shand, 2004; Martin et al., 2004; Molenat et al., 2005, 2008; Shand et al., 2005; Condon and Maxwell, 2015; Maxwell and Condon, 2016). HAND can be interpreted as a proxy of the hydraulic head and is thus potentially more hydrologically informative than the topographic elevation above sea level (Nobre et al., 2011). Compared with the TWI in TOPMODEL, HAND is an explicit measure of a physical feature linking terrain to water related potential energy for local drainage (Nobre et al., 2011). More interestingly, topographic structure emerges as a powerful force determining rooting depth under a given climate or within a biome (Figure 1), revealed by a global synthesis of 2,200 root observations of >1000 species (Fan et al., 2017). This leads us to think from ecological perspective to use the topographic information as an indicator for root zone spatial distribution without calibrating the β, and coupling it with the MCT method to estimate the S_{uMax}, eventually create a calibration-free runoff generation module.

In this study we are therefore going to test the hypotheses that: (1) HAND can be linked to the spatial distribution of storage capacities and therefore can be used to develop a new runoff generation module (HAND-based Storage Capacity curve, i.e. HSC); (2) the distribution of storage capacities determined by

HAND contains different information than the topographic wetness index; (3) the HSC together with water balance-based estimates of S_{uMax} (MCT method) allow the formulation of calibration-free parameterizations of soil moisture routines in top-down models directly based on observations. All these hypotheses will be tested firstly in a small data-rich experimental catchment (the Bruntland Burn catchment in Scotland), and then apply the model to a wide range of larger MOPEX catchments (Model Parameter Estimation Experiment).

This paper is structured as follows. In the Methods section, we describe two of our proposed modules, i.e. HSC and HSC-MCT, and two benchmark models (HBV, TOPMODEL). This section also includes the description of other modules (i.e. interception, evaporation and routing) in rainfall-runoff modelling, and the methods for model evaluation, calibration and validation. The Dataset section reviews the empirically-based knowledge of the Bruntland Burn catchment in Scotland and the hydrometerological and topographic datasets of MOPEX catchments in the US for model comparison. The Results section presents the model comparison results. The Discussion section interprets the relation between rainfall-runoff processes and topography, catchment heterogeneity and simple model, and the implications and limitations of our proposed modules. The conclusions are briefly reviewed in the Summary and Conclusions section.

2 Methods

Based on our perceptual model that saturation excess flow (SEF) is the dominant runoff generation mechanism in most cases, we developed the HAND-based Storage Capacity curve (HSC) module. Subsequently, estimating the parameter of root zone storage capacity (S_{uMax}) by the MCT method without calibration, the HSC-MCT was developed. In order to assess the performance of our proposed modules, two widely-used runoff generation modules, i.e. HBV power function and TOPMODEL module, were set as benchmarks. Other modules, i.e. interception, evaporation and routing, are kept with identical structure and parameterization for the four rainfall-runoff models (HBV, TOPMODEL, HSC, HSC-MCT, whose names are from their runoff generation modules), to independently diagnose the difference among runoff generation modules (Clark et al., 2008; 2010).

2.1 Two benchmark modules

HBV power function

The HBV runoff generation module applies an empirical power function to estimate the nonlinear relationship between the runoff coefficient and soil moisture (Bergström and Forsman, 1973; Bergström and Lindström, 2015). The function is written as:

$$A_s = \left(\frac{S_u}{S_{uMax}}\right)^{\beta} \tag{1}$$

Where A_s (-) represents the contributing area, which equals to the runoff coefficient of a certain rainfall event; S_u (mm) represents the averaged root zone soil moisture; S_{uMax} (mm) is the averaged root zone storage capacity of the studied catchment; β (-) is the parameter determining the shape of the power function. The prior range of β can be from 0.1 to 5. The S_u - A_s has a linear relation while β equals to 1. And the shape becomes convex while the β is less than 1, and the shape turns to concave while the β is larger than 1. In most situations, S_{uMax} and β are two free parameters, cannot be directly measured at the catchment scale, and need to be calibrated based on observed rainfall-runoff data.

TOPMODEL module

The TOPMODEL assumes topographic information captures the runoff generation heterogeneity at catchment scale, and the TWI is used as an index to identify rainfall-runoff similarity (Beven and Kirkby, 1979; Sivapalan et al., 1997). Areas with similar TWI values are regarded as possessing equal runoff generation potential. More specifically, the areas with larger TWI values tend to be saturated first and contribute to SEF; but the areas with lower TWI values need more water to reach saturation and generate runoff. The equations are written as follow:

$$D_i = \overline{D} + S_{uMax}(\overline{I_{TW}} - I_{TW_i}) \tag{2}$$

$$\overline{D} = S_{uMax} - S_u \tag{3}$$

$$A_s = \sum A_{s\_i}; \quad \text{while } D_i < 0 \tag{4}$$

Where D_i (mm) is the local storage deficit below saturation at specific location (i); $\overline{D}$ (mm) is the averaged water deficit of the entire catchment (Equation 2), which equals to (S_{uMax} - S_u), as shown in Equation 3. I_{TWi} is the local I_{TW} value. $\overline{I_{TW}}$ is the averaged TWI of the entire catchment. Equation 2 means in a certain soil moisture deficit condition for the entire catchment ($\overline{D}$), the soil moisture deficit of a specific location (D_i), is determined by the catchment topography (I_{TW} and I_{TWi}), and the root zone storage capacity (S_{uMax}).

Therefore, the areas with D_i less than zero are the saturated areas (A_{s_i}), equal to the contributing areas. The integration of the A_{s_i} areas (A_s), as presented in Equation 4, is the runoff contributing area, which equals to the runoff coefficient of that rainfall event.

Besides continuous rainfall-runoff calculation, Equations 2-4 also allow us to obtain the contributing area (A_s) from the estimated relative soil moisture (S_u/S_{uMax}), and then map it back to the original TWI map, which makes it possible to test the simulated contributing area by field measurement. It is worth mentioning that the TOPMODEL in this study is a simplified version, and not identical to the original one, which combines the saturated and unsaturated soil components.

2.2 HSC module

In the HSC module, we assume 1) primarily saturation excess flow as the dominant runoff generation mechanismSEF is the dominant runoff generation mechanism, while surface overland flow (SOF) and subsurface flow (SSF) cannot be distinguished; 2) the local root zone storage capacity has a positive and linear relationship with HAND, from which we can derive the spatial distribution of the root zone storage capacity; 3) rainfall firstly feeds local soil moisture deficit, and no runoff can be generated before local soil moisture being saturated.

Figure 2 shows the perceptual HSC module, in which we simplified the complicated 3-D topography of a real catchment into a 2-D simplified hillslope. And then derive the distribution of root zone storage capacity, based on topographic analysis and the second assumption as mentioned in the preceding paragraph. Figure 3 shows the approach to derive the S_u-A_s relation, which are detailed as follows.

I. **Generate HAND map.** The HAND map, which represents the relative vertical distance to the nearest river channel, can be generated from DEM (Gharari et al., 2011). The stream initiation threshold area is a crucial parameter, determining the perennial river channel network (Montgomery and Dietrich, 1989; Hooshyar et al., 2016), and significantly impacting the HAND values. In this study, the start area was chosen as 40ha for the BB catchment to maintain a close correspondence with observed stream network. And for the MOPEX catchments, the stream initiation area threshold is set as 500 grid cells (4.05 km^2), which fills in the range of stream initiation thresholds reported by others (e.g. Colombo et al., 2007; Moussa, 2008, 2009). HAND maps were then calculated from the elevation of each raster cell above nearest grid cell flagged as stream cell following the flow direction (Gharari et al., 2011).

II. **Generate normalized HAND distribution curve.** Firstly, sort the HAND values of grid cells in ascending order. Secondly, the sorted HAND values were evenly divided into *n* bands (e.g. 20 bands in this study), to make sure each HAND band has similar area. The averaged HAND value of each band is regarded as the HAND value of that band. Thirdly, normalize the HAND bands, and then plot the normalized HAND distribution curve (Figure 2b).

III. **Distribute S_{uMax} to each HAND band ($S_{\mathrm{uMax\_i}}$).** As assumed, the normalized storage capacity of each HAND band ($S_{\mathrm{uMax\_i}}$) increases with HAND value (Figure 2c). Based on this assumption, the unsaturated root zone storage capacity (S_{uMax}) can be distributed to each HAND band as $S_{\mathrm{uMax\_i}}$ (Figure 3a). It is worth noting that S_{uMax} needs to be calibrated in the HSC module, but free of calibration in the HSC-MCT module.

IV. **Derive the S_{u} - A_{s} curve.** With the number of *s* saturated HAND bands (Figure 3a-c), the soil moisture (S_{u}) can be obtained by Equation 5; and saturated area proportion (A_{s}) can be obtained by Equation 6.

$$S_{\mathrm{u}} = \frac{1}{n}\left[\sum_{i=1}^{s} S_{\mathrm{uMax\_i}} + S_{\mathrm{uMax\_s}}(n-s)\right] \tag{5}$$

$$A_{\mathrm{s}} = \frac{s}{n} \tag{6}$$

Where $S_{\mathrm{uMax\_s}}$ is the maximum $S_{\mathrm{uMax\_i}}$ of all the saturated HAND bands. Subsequently, the A_{s} - S_{u} curve can be derived, and shown in Figure 3d.

The SEF mechanism assumes that runoff is only generated from saturation areas, therefore the proportion of saturation area is equal to the runoff coefficient of that rainfall-runoff event. Based on the S_{u}-A_{s} curve in Figure 3d, generated runoff can be calculated from root zone moisture (S_{u}). The HSC module also allows us to map out the fluctuation of saturated areas by the simulated catchment average soil moisture. For each time step, the module can generate the simulated root zone moisture for the entire basin (S_{u}). Based on the S_{u}-A_{s} relationship (Figure 3d), we can map S_{u} back to the saturated area proportion (A_{s}) and then visualize it in the original HAND map. Based on this conceptual model, we developed the computer program and created a procedural module. The technical roadmap can be found in Figure 4.

2.3 HSC-MCT module

The S_{uMax} is an essential parameter in various hydrological models (e.g. HBV, Xinanjiang, GR4J), which determines the long-term partitioning of rainfall into infiltration and runoff. Gao et al., 2014a found that S_{uMax} represents the adaption of ecosystems to local climate. Ecosystems may design their S_{uMax} based on the precipitation pattern and their water demand. The storage is neither too small to be mortal in dry seasons, nor too large to consume excessive energy and nutrients. Based on this assumption, we can estimate the S_{uMax} without calibration, by the MCT method, from climatological and vegetation information. More specifically, the average annual plant water demand in the dry season (S_R) is determined by the water balance and the vegetation phenology, i.e. precipitation, runoff and seasonal NDVI. Subsequently, based on the annual S_R, the Gumbel distribution (Gumbel, 1935), frequently used for estimating hydrological extremes, was used to standardize the frequency of drought occurrence. S_{R20y}, i.e. the root zone storage capacity required to overcome a drought once in 20 years, is used as the proxy for S_{uMax} due to the assumption of a "cost" minimization strategy of plants as we mentioned above (Milly, 1994), and the fact that S_{R20y} has the best fit with S_{uMax} . The S_{R20y} of the MOPEX catchments can be found in the map of (Gao et al., 2014a).

Eventually, with the MCT approach to estimate S_{uMax} and the HSC curve to represent the root zone storage capacity spatial distribution, the HSC-MCT runoff generation module is created, without free parameters. It is worth noting that both the HSC-MCT and HSC modules are based on the HAND derived S_u-A_s relation, and their distinction lays in the methods to obtain S_{uMax}. So far, the HBV power function module has 2 free parameters (S_{uMax}, β). While the TOPMODEL and the HSC both have one free parameter (S_{uMax}). Ultimately the HSC-MCT has no free parameter.

2.4 Interception, evaporation and routing modules

Except for the runoff generation module in the root zone reservoir (S_{UR}), we need to consider other processes, including interception (S_{IR}) before the S_{UR} module, evaporation from the S_{UR} and the response routine (S_{FR} and S_{SR}) after runoff generation from S_{UR} (Figure 5). Precipitation is firstly intercepted by vegetation canopies. In this study, the interception was estimated by a threshold parameter (S_{iMax}), set to 2 mm (Gao et al., 2014a), below which all precipitation will be intercepted and evaporated (Equation 9) (de Groen and Savenije, 2006). For the S_{UR} reservoir, we can either use the HBV beta-function (Equation 12), the runoff generation module of TOPMODEL (Equation 2-4) or the HSC module (Section 2.3) to partition precipitation into generated runoff (R_u) and infiltration. The actual evaporation (E_a) from the soil equals to the potential evaporation (E_p), if S_u/S_{uMax} is above a threshold (C_e), where S_u is the soil moisture and S_{uMax} is the catchment averaged storage capacity. And E_a linearly reduces with S_u/S_{uMax}, while S_u/S_{uMax} is below C_e (Equation 13). The E_p can be calculated by the Hargreaves equation (Hargreaves and Samani, 1985), with maximum and minimum daily temperature as input. The generated runoff (R_u) is further split into two fluxes, including the flux to the fast response reservoir (R_f) and the flux to the slow response reservoir (R_s), by a splitter (D) (Equation 14, 15). The delayed time from rainfall peak to the flood peak is estimated by a convolution delay function, with a delay time of T_{lagF}. Subsequently, the fluxes into two different response reservoirs (S_{FR} and S_{SR}) were released by two linear equations between discharge and storage (Equation 19, 21), representing the fast response flow and the slow response flow mainly from groundwater reservoir. The two discharges (Q_{f} and Q_{s}) generated the simulated streamflow (Q_{m}). The model parameters are shown in Table 1, while the equations are given in Table 2. More detailed description of the model structure can be referred to Gao et al., 2014b and 2016. It is worth underlining that the only difference among the benchmark HBV type, TOPMODEL type, HSC, and HSC-MCT models is their runoff generation modules. Eventually, there are 7 free parameters in HBV model, 6 in TOPMODEL and HSC model, and 5 in the HSC-MCT model.

2.5 Model evaluation, calibration, validation and models comparison

Two objective functions were used to evaluate model performance, since multi-objective evaluation is a more robust approach to quantifying model performance with different criteria than a single one. The Kling-Gupta efficiency (Gupta *et al.*, 2009) (I_{KGE}) was used as the criteria to evaluate model performance and as an objective function for calibration. The equation is written as:

$$I_{\mathrm{KGE}} = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\varepsilon-1)^2} \quad (7)$$

Where r is the linear correlation coefficient between simulation and observation; α ($\alpha = \sigma_{\mathrm{m}} / \sigma_{\mathrm{o}}$) is a measure of relative variability in the simulated and observed values, where σ_{m} is the standard deviation of simulated streamflow, and σ_{o} is the standard deviation of observed streamflow; ε is the ratio between the average value of simulated and observed data. And the I_{KGL} (I_{KGE} of the logarithmic flows) (Fenicia et al., 2007; Gao et al., 2014b) is used to evaluate the model performance on baseflow simulation.

A multi-objective parameter optimization algorithm (MOSCEM-UA) (Vrugt et al., 2003) was applied for the calibration. The parameter sets on the Pareto-frontier of the multi-objective optimization were assumed to be the behavioral parameter sets and can equally represent model performance. The averaged hydrograph obtained by all the behavioral parameter sets were regarded as the simulated result of that catchment for further studies. The number of complexes in MOSCEM-UA were set as the number of parameters (7 for HBV, 6 for TOPMODEL and the HSC model, and 5 for HSC-MCT model), and the number of initial samples was set to 210 and a total number of 50000 model iterations for all the catchment runs. For each catchment, the first half period of data was used for calibration, and the other half was used to do validation.

In module comparison, we defined three categories: if the difference of I_{KGE} of model A and model B in validation is less than 0.1, model A and B are regarded as “equally well”. If the I_{KGE} of model A is larger than model B in validation by 0.1 or more, model A is regarded as outperforming model B. If the I_{KGE} of model A is less than model B in validation by -0.1 or less, model B is regarded as outperforming model A.

3 Dataset

3.1 The Bruntland Burn catchment

The 3.2 km^2 Bruntland Burn catchment (Figure 6), located in north-eastern Scotland, was used as a benchmark study to test the ~~models~~model’s performance based on a rich data base of hydrological measurements. The Bruntland Burn is a typical upland catchment in North West Europe (e.g. Birkel et al., 2010), namely a combination of steep and rolling hillslopes and over-widened valley bottoms due to the glacial legacy of this region. The valley bottom areas are covered by deep (in parts > 30m) glacial drift deposits (e.g. till) containing a large amount of stored water superimposed on a relatively impermeable granitic solid geology (Soulsby et al., 2016). Peat soils developed (> 1m deep) in these valley bottom areas, which remain saturated throughout most of the year with a dominant near-surface runoff generation mechanism delivering runoff quickly via micro-topographical flow pathways connected to the stream network (Soulsby et al., 2015). Brown rankers, peaty rankers and peat soils are responsible for a flashy hydrological regime driven by saturation excess overland flow, while humus iron podzols on the hillslopes do not favor near-surface saturation but rather facilitate groundwater recharge through vertical water movement (Tetzlaff et al., 2014). Land-use is dominated by heather moorland, with smaller areas of rough grazing and forestry on the lower hillslopes. Its annual precipitation is 1059 mm, with the summer months (May-August) generally being the driest (Ali et al., 2013). Snow makes up less than 10% of annual precipitation and melts rapidly below 500m. The evapotranspiration is around 400 mm per year and annual discharge around 659 mm. The daily precipitation, potential evaporation, and discharge data range from January 1 in 2008 to September 30 in 2014. The calibration period is from January 1, 2008 to December 31, 2010, and the data from January 1, 2011 to September 30, 2014 is used as validation.

The LiDAR-derived DEM map with 2m resolution shows elevation ranging from 250m to 539m (Figure 6). There are 7 saturation area maps (Figure 7) (May 2, July 2, August 4, September 3, October 1, November 26, in 2008, and January 21, in 2009), measured directly by the “squishy boot” method and field mapping by global positioning system (GPS), to delineate the boundary of saturation areas connected to the stream network (Birkel et al., 2010; Ali et al., 2013). These saturation area maps revealed a dynamic behavior of expanding and contracting areas connected to the stream network that were used as a benchmark test for the HSC module.

3.2 MOPEX catchments

The MOPEX dataset was collected for a hydrological model parameter estimation experiment (Duan et al., 2006; Schaake et al., 2006), containing 438 catchments in the CONUS (Contiguous United States). The longest time series range from 1948 to 2003. 323 catchments were used in this study (see the name list in SI), with areas between 67 and 10,329 km^2, and excluding the catchments with data records <30 years, impacted by snowmelt or with extreme arid climate (aridity index $E_p/P > 2$). In order to analyze the impacts of catchment characteristics on model performance, excluding hydrometeorology data, we also collected the datasets of topography, depth to rock, soil texture, land use, and stream density (Table 3). These characteristics help us to understand in which catchments the HSC performs better or worse than the benchmark models.

Hydrometeorology

The dataset contains the daily precipitation, daily maximum and minimum air temperature, and daily streamflow. The daily streamflow was used to calibrate the free ~~parameters, and~~parameters and validate the models.

Topography

The Digital Elevation Model (DEM) of the CONUS in 90m resolution was download from the Earth Explorer of United States Geological Survey (USGS, http://earthexplorer.usgs.gov/). The HAND and TWI map can be generated from DEM. The averaged elevation and HAND are used to as two catchment characteristics.

Soil texture

In this study, soil texture is synthetically represented by the K factor, since the K factor is a lumped soil erodibility factor which represents the soil profile reaction to soil detachment (Renard et al., 2011). Generally, the soils (high in clay and sand) have low K values, and soils with high silt content have larger K values. The averaged K factor for each catchment was calculated from soil survey information available from USGS (Wolock, 1997).

Land use

Land use data was obtained from National Land Cover Database (NLCD, http://www.mrlc.gov/nlcd.php). Forest plays an essential role in hydrological processes (Gao et al., 2018a), especially for the runoff generation (Brooks et al., 2010). Forest area proportion was utilized as an integrated indictor to represent the impact of vegetation cover on hydrological processes.

Stream density

Stream density (km/km^2) is the total length of all the streams and rivers in a drainage basin divided by the total area of the drainage basin. Stream density data was obtained from Horizon Systems Corporation (http://www.horizon-systems.com/nhdplus/).

Geology

Bedrock is a relative impermeable layer, as the lower boundary of subsurface stormflow in the catchments where soil depth is shallow (Tromp-van Meerveld & McDonnell). The depth to bedrock, as an integrated geologic indicator, was accessed from STATSGO (State Soil Geographic, http://www.soilinfo.psu.edu/index.cgi?soil_data&conus&data_cov&dtb) (Schwarz & Alexander, 1995). The averaged depth to bedrock for each catchment was calculated for further analysis.

4 Results of the Bruntland Burn

4.1 Topography analysis

The generated HAND map, derived also from the DEM, is shown in Figure 6, with HAND values ranging from 0m to 234m. Based on the HAND map, we can derive the S_u-A_s curve (Figure 8) by analyzing the HAND map with the method in Section 2.3. The TWI map of the BB (Figure 6) was generated from its DEM. Overall, the TWI map, ranging from -0.4 to 23.4, mainly differentiates the valley bottom areas with the highest TWI values from the steeper slopes. This is probably caused by the fine resolution of the DEM map in 2 m, as previous research found that the sensitivity of TWI to DEM resolution (Sørensen and Seibert, 2007). From the TWI map, the frequency distribution function and the accumulative frequency distribution function can be derived (Figure 8), with one unit of TWI as interval.

4.2 Model performance

It is found that all the three models (HBV, TOPMODEL, and HSC) can perform well in reproducing the observed hydrograph (Figure 9). The I_{KGE} of the three models are all around 0.66 in calibration, which is largely in line with other studies from the BB (Birkel et al, 2010; 2014). And the I_{KGL} are 0.76, 0.72 and 0.74 for HSC, HBV and TOPMODEL respectively in calibration. While in validation, I_{KGE} of the three models are also around 0.66, while I_{KGL} are 0.75, 0.70 and 0.65 for the three models. Since the measured rainfall-runoff time series only lasts from 2008 to 2014, which is too short to estimate the S_{R20y} (proxy for S_{uMax}) by MCT approach (which needs long-term hydro-meteorological observation data,) the HSC-MCT model was not applied to this catchment.

Figure 8 shows the calibrated power curve by HBV (averaged beta=0.98) with the S_u-A_s curve obtained from the HSC module. We found the two curves are largely comparable, especially while the relative soil moisture is low. This result demonstrates that for the BB catchment with glacial drift deposits and combined terrain of steep and rolling hillslopes and over-widened valley bottoms, the HBV power curve can essentially be derived from the S_u-A_s curve of HSC module merely by topographic information without calibration.

The normalized relative soil moisture of the three model simulations are presented in Figure 9. Their temporal fluctuation patterns are comparable. Nevertheless, the simulated soil moisture by TOPMODEL has larger variation, compared with HBV and HSC (Figure 9).

4.3 Contributing area simulation

The observed saturation area and the simulated contributing area from both TOPMODEL and the HSC are shown in Figure 7, 9, 10. We found although both modules overestimated the saturated areas, they can capture the temporal variation. For example, the smallest saturated area both observed and simulated occurred on July-02-2008, and the largest saturated area both occurred on January-21-2009. Comparing the estimated contributing area of TOPMODEL with the HSC module, we found the results of the HSC correlates better (R^2=0.60, I_{KGE}=-3.0) with the observed saturated areas than TOPMODEL (R^2=0.50, I_{KGE}=-3.4) (Figure 10). For spatial patterns, the HSC contributing area is located close to the river ~~network, and~~network and reflects the spatial pattern of observed saturated area. While TOPMODEL results are more scattered, probably due to the sensitivity of TWI to DEM resolution (Figure 7). The HSC is more discriminating in terms of less frequently giving an unrealistic 100% ~~saturation, and~~saturation and retaining unsaturated upper hillslopes.

5 Results from the MOPEX catchments

5.1 Topography analysis of the Contiguous US and 323 MOPEX catchments

To delineate the TWI map for the CONUS, the depressions of the DEM were firstly filled with a threshold height of 100m (recommended by Esri). The TWI map of the CONUS is produced (Figure S1). Based on the

TWI map of the CONUS, we clipped the TWI maps for the 323 MOPEX catchments with their catchment boundaries. And then the TWI frequency distribution and the accumulated frequency distribution of the 323 MOPEX catchments (Figure S2), with one unit of TWI as interval, were derived based on the 323 TWI maps.

In Figure 11, it is shown that the regions with large HAND values are located in Rocky Mountains and Appalachian Mountains, while the Great Plains has smaller HAND values.~~ Interestingly, t~~ The Great Basin, especially in the Salt Lake Desert, has small HAND values, illustrating its low elevation above the nearest drainage, ~~although their~~despite a high elevation~~s~~ above sea~~s~~ level~~ are high~~. From the CONUS HAND map, we clipped the HAND maps for the 323 MOPEX catchments with their catchment boundaries. We then plot their HAND-area curves, following the procedures of I and II in Section 2.2. Figure 12a shows the normalized HAND profiles of the 323 catchments.

Based on the HAND profiles and the Step III in Section 2.2, we derived the normalized storage capacity distribution for all catchments (Figure 12b). Subsequently, the root zone moisture and saturated area relationship (A_s-S_u) can be plotted by the method in Step IV of Section 2.2. Lastly, reversing the curve of A_s-S_u to S_u-A_s relation (Figure 12c), the latter one can be implemented to simulate runoff generation by soil moisture. Figure 12c interestingly shows that in some catchments, there is almost no threshold behavior between rainfall and runoff generation, where the catchments are covered by large areas with low HAND values and limited storage capacity. Therefore, when rainfall occurs, wetlands response quickly and generate runoff without a precipitation–discharge threshold relationship characteristic of areas with higher moisture deficits. This is similar to the idea of FLEX-Topo where the storage capacity is distinguished between wetlands and hillslopes, and on wetlands, with low storage capacity, where runoff response to rainfall is almost instantaneous.

5.2 Model performance

Overall, the performance of the two benchmark models, i.e. HBV and TOPMODEL, for the MOPEX data (Figure 13) is comparable with the previous model comparison experiments, conducted with four rainfall-runoff models and four land surface parameterization schemes (Duan et al., 2006; Kollat et al., 2012; Ye et al., 2014). The median value of I_{KGE} of the HBV type model is 0.61 for calibration in the 323 catchments (Figure 13), and averaged I_{KGE} in calibration is 0.62. In validation, the median and averaged values of I_{KGE} are kept the same as calibration. The comparable performance of models in calibration and validation demonstrates the robustness of benchmark models and the parameter optimization algorithm (i.e. MOSCEM-UA). The TOPMODEL improves the median value of I_{KGE} from 0.61 (HBV) to 0.67 in calibration, and from 0.61 (HBV) to 0.67 in validation. But the averaged values of I_{KGE} for TOPMODEL are slightly decreased from 0.62 (HBV) to 0.61 in both calibration and validation. The HSC module, by involving the HAND topographic information without calibrating the β parameter, improves the median value of I_{KGE} to 0.68 for calibration and 0.67 for validation. The averaged values of I_{KGE} in both calibration and validation are also increased to 0.65, comparing with HBV (0.62) and TOPMODEL (0.61). Furthermore, Figure 13 demonstrates that, comparing with the benchmark HBV and TOPMODEL, not only the median and averaged values were improved by the HSC module, but also the 25^{th} and 75^{th} percentiles and the lower whisker end, all have been improved. The performance gains on baseflow (I_{KGL}) have been investigated and shown in the supplementary figure S3. These results indicate the HSC module improved model performance to reproduce hydrograph for both peak flow (I_{KGE}) and baseflow (I_{KGL}).

Additionally, for HSC-MCT model, the median I_{KGE} value is improved from 0.61 (HBV) to 0.65 in calibration, and from 0.61 (HBV) to 0.64 in validation, but not as well performed as TOPMODEL (0.67 for calibration and validation). For the averaged I_{KGE} values, they were slightly reduced from 0.62 (HBV) and 0.61 (TOPMODEL) to 0.59 for calibration and validation. Although the HSC-MCT did not perform as well as the HSC module, considering there is no free parameters to calibrate, the median I_{KGE} value of 0.64 (HBV is 0.61) and averaged I_{KGE} of 0.59 (TOPMODEL is 0.61) are quite acceptable. In addition, the 25^{th} and 75^{th} percentiles and the lower whisker end of the HSC-MCT model are all improved compared to the HBV model. Moreover, the largely comparable results between the HSC and the HSC-MCT modules demonstrate the feasibility of the MCT method to obtain the S_{uMax} parameter and the potential for HSC-MCT to be implemented in prediction of ungauged basins.

Figure 14 shows the spatial comparisons of the HSC and HSC-MCT models with the two benchmark models. We found that the HSC performs “equally well” as HBV (the difference of I_{KGE} in validation ranges -0.1 ~ 0.1) in 88% catchments, and in the remaining 12% of the catchments the HSC outperforms HBV (the improvement of I_{KGE} in validation is larger than 0.1). In not a single catchment did the calibrated HBV outperform the HSC. Comparing the HSC model with TOPMODEL, we found in 91% of the catchments that the two models have approximately equal performance. In 8% of the catchments, the HSC model outperformed TOPMODEL. Only in 1% of the catchments (two in the Appalachian Mountains and one in the Rocky Mountains in California), TOPMODEL performed better.

In order to further explore the impact of catchment characteristics on model performance, we used topography (averaged HAND, averaged slope, and averaged elevation), soil (K-factor), land cover (forest area proportion), climate (aridity index), stream density, and geology (depth to rock) information to test the impact of catchment features on model performance. Table 4 clearly shows that compared with HBV, the 39 catchments with better performance have lower HAND values (37m), more gentle slopes (4.0 degree), and smaller forest area (22%); while the elevation, K-factor, aridity index, stream density and depth to rock are almost similar. Also, in the catchments where HSC outperformed TOPMODEL, the catchments have smaller HAND (27m), more gentle slopes (3.6 degree), moderate elevation (469 m), less forest proportion (14%), and more arid climate (aridity index is 1.3). TOPMODEL performs better in only three catchments with larger HAND (193m), steeper slopes (13.5 degree), higher elevation (740 m), more humid climate (aridity index is 0.8), and larger depth to rock (333 cm). In summary, the HSC showed better performance in catchments with gentle topography and more arid climate.

Without calibration of S_{uMax}, as expected, the performance of HSC-MCT module slightly deteriorates (Figure 13). In comparison with HBV, the outperformed percentage reduced from 12% (HSC) to 4% (HSC-MCT), the approximately equal-well simulated catchments dropped from 88% to 79%, and the inferior performance increased from 0% to 17%. Also, in comparison with TOPMODEL, the better performance dropped from 8% (HSC) to 7% (HSC-MCT), the approximately equal catchments reduced from 91% to 72%, and the inferior performance increased from 1% to 21%. The inferiority of the HSC-MCT model is probably caused by the uncertainty of the MCT method for different ecosystems which have different survival strategies and use different return periods to bridge critical drought periods. By using ecosystem dependent return periods, this problem could be reduced (Wang-Erlandsson et al., 2016).

To further explore the reason for the better performance of the HSC approach, we selected the 08171000 catchment in Texas (Figure 14), in which both the HSC module and the HSC-MCT module outperformed the two benchmark modules to reproduce the observed hydrograph (Figure S4). The HBV model dramatically underestimated the peak flows, with I_{KGE} as 0.54, while TOPMODEL significantly overestimated the peak flows, with I_{KGE} as 0.30. The HSC-MCT model improved the I_{KGE} to 0.71, and the HSC model further enhanced I_{KGE} to 0.74.

Since the modules of interception, evaporation and routing are identical for the four models, the runoff generation modules are the key to understand the difference in model performance. Figure S5 shows the HBV β curve and the S_u-A_s curve of the HSC model, as well the TWI frequency distribution. We found that with a given S_u/S_{uMax}, the HBV β function generates less contributing area than the HSC model, which explains the underestimation of the HBV model. In contrast, TOPMODEL has a sharp and steep accumulated TWI frequency curve. In particular, the region with TWI=8 accounts for 40% of the catchment area, and over 95% of the catchment areas are within the TWI ranging from 6 to 12. This indicates that even with low soil moisture content (S_u/S_{uMax}), the contributing area by TOPMODEL is relatively large, leading to the sharply increased peak flows for all rainfall events.

6 Discussion

6.1 Rainfall-runoff processes and topography

We applied a novel approach to derive the relationship between soil moisture storage and the saturated area from HAND. The areas with relatively low HAND values are saturated earlier than areas with higher HAND values, due to the larger storage capacity in higher HAND locations. The outperformance of the HSC over the benchmark HBV and TOPMODEL in gentle sloping catchments indicates that the HSC module likely has a higher realism than the calibrated HBV beta-function and the TWI of TOPMODEL in these regions. Very interestingly, Fan et al., (2017) presented an ecological observation in global scale, and revealed the systematic variation of rooting depth along HAND (Fig.1, in Fan et al., 2017). Since rooting depth can be translated to root zone storage capacity through combination with soil plant-available water (Wang-Erlandsson et al., 2016). This large sample dataset, from ecological perspective, provides a strong support for the assumption of the HSC model on gentle slopes, i.e. the increase of root zone storage capacity with HAND. More interestingly, on excessively drained uplands, rooting depth does not follow the same pattern, with shallow depth and limited to rain infiltration (Fig.1, in Fan et al., 2017). This could explain the inferior performance of HSC model to TOPMODEL in three MOPEX catchments with excessively drained uplands (larger HAND, steeper slope, higher elevation, and deeper depth to rock), where Hortonian overland flow is likely the dominant mechanism, and the HSC assumption likely does not work well. This indicates that comparing with TWI, the HAND is closer to catchment realism distinguishing hydrological similarity in gentle topography catchments. The HSC module assumes SEF as the dominant mechanism. But since in a real catchment different runoff generating processes may act simultaneously in different environments (McDonnell, 2013; Hrachowitz and Clark, 2017). Such SEF dominated catchments, or parts thereof, are typically characterized by a subdued relief and thus gently sloping. In steeper catchments, where the groundwater table is deeper and thus more additional water can be stored in the soil, another conceptual parametrisation would be appropriate.

The FLEX-Topo model (Savenije, 2010) also uses HAND as a topographic index to distinguish between landscape-related runoff ~~processes, and~~processes and has both similarity and differences with the HSC model. The results of the HSC model illustrate that the riparian areas are more prone to be saturated, which is consistent with the concept of the FLEX-Topo model. Another important similarity of the two models is their parallel model structure. In both models it is assumed that the upslope area has larger storage capacity, therefore the upper land generates runoff less and later than the lower land. In other words, in most cases, the local storage is saturated due to the local rainfall, instead of flow from upslope. The most obvious difference between the HSC and the FLEX-Topo is the approach towards discretization of a catchment. The FLEX-Topo model classifies a catchment into various landscapes, e.g. wetlands, hillslopes and plateau. This discretization method requires threshold values to classify landscapes, i.e. threshold values of HAND and slope, which leads to fixed and time-independent proportions of landscapes. The HSC model does not require landscape classification, which reduced the subjectivity in discretization and restricted the model complexity, as well as simultaneously allowing the fluctuation of contributing areas (termed as wetlands in FLEX-Topo).

6.2 Catchment heterogeneity and simple models

Catchments exhibit a wide array of heterogeneity and complexity with spatial and temporal variations of landscape characteristics and climate inputs. For example, the Darcy-Richards equation approach is often consistent with point-scale measurements of matrix flow, but not for preferential flow caused by roots, soil fauna and even cracks and fissures (Beven and Germann, 1982; Zehe and Fluehler, 2001; Weiler and McDonnell, 2007). As a result, field experimentalists continue to characterize and catalogue a variety of runoff processes, and hydrological and land surface modelers are developing more and more complicated models to involve the increasingly detailed processes (McDonnell et al., 2007). However, there is still no compelling evidence to support the outperformance of sophisticated "physically-based" models in terms of higher equifinality and uncertainty than the simple lumped or semi-distributed conceptual models in rainfall-runoff simulation (Beven, 1989; Orth et al., 2015).

But evidence is mounting that a catchment is not a random assemblage of different heterogeneous parts (Sivapalan, 2009; Troch et al., 2013; Zehe et al., 2013), and conceptualising heterogeneities does not require complex laws (Chase, 1992; Passalacqua et al., 2015). Parsimonious models (e.g. Perrin et al., 2003), with empirical curve shapes, likely result in good model performance. Parameter identifiability in calibration is one of the reasons. However, the physical rationale of these parsimonious models is still largely unknown lacking a physical explanation to interpret these empirical curves described by mathematical functions (e.g. Equation 3 in Perrin et al., 2003).

The benefits of the new HSC module are two-fold. From a technical point of view, the HSC allows us to make Prediction in Ungauged Basins without calibrating the beta parameter in many conceptual hydrological models. Furthermore, the HSC module, from a scientific point of view, provides us with a new perspective on the linkage between the spatial distribution patterns of root zone storage capacity (long-term ecosystem evolution) with associated runoff generation (event scale rainfall-runoff generation).

Asking questions of "why" rather than "what" likely leads to more useful insights and a new way forward (McDonnell et al., 2007). The HSC module provides us with a rationale from an ecological perspective to understand the linkage and mechanism between large-sample hillslope ecological observations and the curve of root zone storage capacity distribution (Figure 1, 2, 3). Catchment is a geomorphological and even an ecological system whose parts are related to each other probably due to catchment self-organization and evolution (Sivapalan and Blöschl, 2015; Savenije and Hrachowitz, 2017). This encourages the hope that simplified concepts may be found adequate to describe and model the operation of the basin runoff generation process. It is clear that topography, with fractal characteristic (Rodriguez-Iturbe and Rinaldo, 1997), is often the dominant driver of runoff, as well as being a good integrated indicator for vegetation cover (Gao et al., 2014b), rooting depth (Fan et al., 2017), root zone evaporation and transpiration deficits (Maxwell and Condon, 2016), soil properties (Seibert et al., 2007), and even geology (Rempe and Dietrich, 2014; Gomes, 2016). Therefore, we argue that increasingly detailed topographic information is an excellent integrated indicator allowing modelers to continue systematically represent heterogeneities and simultaneously reduce model complexity. The model structure and parameterization of both HSC and TOPMODEL are simple, but not over simplified, as they capture likely the most dominant factor controlling runoff generation, i.e. the spatial heterogeneity of storage capacity. Hence, this study also sheds light on the possibility of moving beyond heterogeneity and process complexity (McDonnell et al., 2007), to simplify them into a succinct and *a priori* curve by taking advantage of catchment self-organization probably caused by co-evolution or the principle of maximum entropy production (Kleidon and Lorenz, 2004).

6.3 Implications and limitation

The calibration-free HSC-MCT runoff generation module enhances our ability to predict runoff in ungauged basins. PUB is probably not a major issue in the developed world, with abundant of comprehensive measurements in many places, but for the developing world it requires prediction with sparse data and fragmentary knowledge. Topographic information with high spatial resolution is freely available globally, allowing us to implement the HSC model in global scale studies. In addition, thanks to the recent development, testing, and validation of remote sensing evaporation products in large spatial scale (e.g. Anderson et al., 2011; Hu and Jia, 2015), the S_{uMax} estimation has become possible without in situ hydro-meteorological measurements (Wang-Erlandsson et al., 2016). These widely-accessible datasets make the global-scale implementation of HSC-MCT module promising.

Although the new modules perform well in the BB and the MOPEX catchments, we do not intend to propose "a model fits all". The assumption of HSC, to some extent, is supported by large-sample ecological field observation (Fan et al., 2017), but it never means the A_s-S_u curve of HSC can perfectly fit the other existing curves (e.g. HBV and TOPMODEL). Unify all model approaches into one framework is the objective of several pioneer works (e.g. Clark, et al., 2010; Fenicia et al., 2011), but out of the scope of this study. Moreover, while estimating the runoff coefficient by the A_s-S_u relation, rainfall in the early time may cause the increase of S_u/S_{uMax} and runoff coefficient (Moore, 1985; Wang, 2018). Therefore, neglecting this influence factor, HBV (Equation 1), TOPMODEL (Equation 2-4) and HSC (Equation 5-6) theoretically underestimate the runoff coefficient, which needs to be further investigated.

Finally, we should not ignore the limitations of the new module, although it has better performance and modelling consistency. 1) The threshold area for the initiating a stream was set as a constant value for the entire CONUS, but the variation of this value in different climate, geology and landscape classes (Montgomery and Dietrich, 1989; Helmlinger et al., 1993; Colombo et al., 2007; Moussa, 2008) needs to be future investigated. 2) The discrepancy between observed and simulated saturation area needs to be further investigated, by utilizing more advanced field measurement and simultaneously refining the model assumption. To our understanding, there are two interpretations. Firstly, the overestimation of the HSC model is possibly because two runoff generation mechanisms – SOF and the SSF occur at the same time. However, the saturated area observed by the "squishy boot" method (Ali et al., 2013), probably only distinguished the areas where SOF occurred. Subsurface stormflow, also contributing to runoff, cannot be observed by the "squishy boot" method. Thus, this mismatch between simulation and observation probably leads to this saturated area overestimation. The second interpretation might be the different definition of "saturation". The observed saturated areas are places where 100% of soil pore volume is filled by water. But the modelled saturation areas are located where soil moisture is above field capacity, and not necessarily 100% filled with water, which probably also results in the overestimation of saturated areas. Interestingly, in theory the observed saturated area should be within the simulated contributing area, due to the fact that the saturated soil moisture is always larger than field capacity. From this point of view, the observed saturated area is smaller and within the contributing area simulated by HSC, but TOPMODEL missed this important feature. 4) Only the runoff generation module is calibration free, but the interception and response routines still rely on calibration. Although we kept the interception and response routine modules the same for the four models, the variation of other calibrated parameters (i.e. S_{iMax}, D, K_f, K_s, T_{lagF}) may also influence model performance in both calibration and validation. 5) The computational cost of the HSC is more expensive than HBV, and similar to TOPMODEL, due to the cost of preprocessed topographic analysis. But once the S_u-A_s curve is completed, the computation cost is quite comparable with HBV.

7 Summary and conclusions

In this study, we developed a simple and calibration-free hydrological module (HAND-based Storage Capacity curve, HSC) based on a relatively new topographic index (HAND), which is not only an excellent physically-based indictor for the hydraulic gradient, but also represents the spatial distribution of root zone storage capacity supported by large-sample ecological observations. Based on HAND spatial distribution pattern, the soil moisture (S_u) - saturated area (A_s) relation for each catchment was derived, which was used to estimate the A_s of specific rainfall event based on continuous calculation of S_u. Subsequently, based on the S_u-A_s relation, the HSC module was developed. Then, applying the mass curve technique (MCT) approach, we estimated the root zone storage capacity (S_{uMax}) from observable hydro-climatological and vegetation data, and coupled it with HSC to create the calibration-free HSC-MCT module. The HBV and TOPMODEL were used as two benchmarks to test the performance of HSC and HSC-MCT on both hydrograph simulation and ability to reproduce the contributing area, which was measured for different hydrometeorological conditions in the Bruntland Burn catchment in Scotland. Subsequently, 323 MOPEX catchments in the US were used as a large-sample hydrological study to further validate the effectiveness of our proposed runoff generation modules.

In the BB exploratory study, we found that the HSC, HBV and TOPMODEL performed comparably well to reproduce the observed hydrograph. Comparing the estimated contributing area of TOPMODEL with the HSC module, we found that HSC module performed better to reproduce saturated area variation, in terms of the correlation coefficient and spatial patterns. This likely indicates that HAND maybe a better indicator to distinguish hydrological similarity than TWI.

For the 323 MOPEX catchments, HSC improved the averaged validation value of I_{KGE} from 0.62 (HBV) and 0.61 (TOPMODEL) to 0.65. In 12% of the MOPEX catchments, the HSC module outperforms HBV, and in not a single catchment did the calibrated HBV outperform the HSC. Comparing with TOPMODEL, the HSC outperformed in 8% of the catchments, and in only 1% of catchments TOPMODEL has a better performance. Interestingly, we found that the HSC module showed better performance in the catchments with gentle topography, less forest cover, and larger aridity index. Not surprisingly, the I_{KGE} of HSC-MCT model was slightly reduced to 0.59, due to the non-calibrated S_{uMax}, but still comparably well performed as HBV (0.62) and TOPMODEL (0.61). This illustrates the robustness of both the HSC approach to derive the spatial distribution of the root zone storage capacity (β) and the efficiency of the MCT method to estimate the root zone storage capacity (S_{uMax}).

Acknowledgement:

This study was supported by National Natural Science Foundation of China (41801036), National Key R&D Program of China (2017YFE0100700), the Key Program of National Natural Science Foundation of China (No. 41730646), and Key Laboratory for Mountain Hazards and Earth Surface Process, Institute of Mountain Hazards and Environment, Chinese Academy of Sciences (KLMHESP-17-02).

Author contributions:

H.G. and H.H.G.S. designed research; H.G. performed research; C.B., C.S., D.T and H.G. provided data, among which the dynamics of the saturation areas data in the BB was provided by C.B. C.S., and D.T.; H.G. analysed data; C.B. was involved in the interpretation of some of the modelling work in the BB; H.G. M.H, and H.H.G.S. wrote the paper; CS and DT extensively edited the paper, and provided substantial comments and constructive suggestions for scientific clarification.

References:

Anderson, M. C., Kustas, W. P., Norman, J. M., Hain, C. R., Mecikalski, J. R., Schultz, L., González-Dugo, M. P., Cammalleri, C., D'Urso, G., Pimstein, A., and Gao, F.: Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery, Hydrol. Earth Syst. Sci., 15, 223–239, doi:10.5194/hess-15-223-2011, 2011.

Andréassian V, Bourgin F, Oudin L, Mathevet T, Perrin C, Lerat J, Coron L, Berthet L.: Seeking genericity in the selection of parameter sets: Impact on hydrological model efficiency. Water Resources Research 50 (10): 8356–8366, 2014.

Bergström S, Forsman A.: Development of a conceptual deterministic rainfall-runoff model. Hydrology Research 4 (3): 147–170, 1973.

Bergström S, Lindström G.: Interpretation of runoff processes in hydrological modelling—experience from the HBV approach. Hydrological Processes 29 (16): 3535–3545, 2015.

Beven K.: Robert E. Horton's perceptual model of infiltration processes. Hydrological Processes 18 (17): 3447–3460 DOI: 10.1002/hyp.5740, 2004.

Beven K, Freer J.: A dynamic TOPMODEL. Hydrological Processes 15 (10): 1993–2011 DOI: 10.1002/hyp, 2001.

Beven K. 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in Water Resources 16 (1): 41–51 DOI: http://dx.doi.org/10.1016/0309-1708(93)90028-E

Beven K.: Linking parameters across scales: Subgrid parameterizations and scale dependent hydrological models. Hydrological Processes 9 (September 1994): 507–525 DOI: 10.1002/hyp.3360090504.252, 1995.

Beven KJ.: Rainfall–Runoff Models: The Primer, 2012.

Beven K., Germann P.: Macropores and water-flow in soils. Water Resour. Res. 18, 1311–1325, 1982.

Beven KJ, Kirkby MJ.: A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin 24 (1): 43–69 DOI: 10.1080/02626667909491834, 1979.

Beven, K.: Changing ideas in hydrology – the case of physically-based models. J. Hydrol. 105 (1–2), 157–172, 1989.

Birkel C, Tetzlaff D, Dunn SM, Soulsby C.: Towards a simple dynamic process conceptualization in rainfall‐runoff models using multi‐criteria calibration and tracers in temperate, upland catchments. Hydrological Processes 24 (3): 260‐275, 2010.

Birkel, C., Soulsby, C., and Tetzlaff, D.: Conceptual modelling to assess how the interplay of hydrological connectivity, catchment storage and tracer dynamics controls non-stationary water age estimates. Hydrological Processes, DOI: 10.1002/hyp.10414, 2014.

Blöschl G.: Runoff prediction in ungauged basins: synthesis across processes, places and scales. Cambridge University Press, 2013.

Brooks, R. J., Barnard, H. R., Coulombe, R. & McDonnell, J. J.: Ecohydrologic separation of water between trees and streams in a Mediterranean climate. Nature Geoscience 3, 100–104. DOI: 10.1038/ngeo722, 2010.

Budyko MI.: Climate and life, 1971.

Burt TP, McDonnell JJ.: Whither field hydrology? The need for discovery science and outrageous hydrological hypotheses. Water Resources Research 51 (8): 5919–5928 DOI: 10.1002/2014WR016839, 2015.

Chase, CG.: Fluvial landsculpting and the fractal dimension of topography. Geomorphology 5 (1): 39–57 DOI: http://dx.doi.org/10.1016/0169-555X(92)90057-U, 1992.

Clark, MP, Slater, AG, Rupp, DE, Woods, R A., Vrugt, J A., Gupta, H V., Wagener, T, Hay, LE.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models. Water Resources Research 44: 1–14 DOI: 10.1029/2007WR006735, 2008.

Clark, M. P., Kavetski, D. and Fenicia, F.: Pursuing the Method of Multiple Working Hypotheses for Hydrological Modeling. Water Resources Research 47.9: 1–16, 2011.

Colombo, R., Vogt, J. V., Soille, P., Paracchini, M. L., de Jager, A.: Deriving river networks and catchments at the European scale from medium resolution digital elevation data. CATENA 70 (3): 296–305 DOI: http://doi.org/10.1016/j.catena.2006.10.001, 2007.

Condon, Laura E, and Reed M Maxwell. "Evaluating the Relationship between Topography and Groundwater Using Outputs from a Continental-Scale Integrated Hydrology Model." Water Resources Research 51.8 (2015): 6602–6621.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, YM, Habets, F., Hall, A., Hay, L., Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. Journal of Hydrology 320 (1-2): 3–17 DOI: 10.1016/j.jhydrol.2005.07.031, 2006.

Dunne, T., Black, R.D.: Partial area contributions to Storm Runoff in a Small New England Watershed. Water Resources Research 6 (5): 1296–1311, 1970.

Boer-Euser, T., McMillan, H. K., Hrachowitz, M., Winsemius, H. C., and Savenije, H. H. G.: Influence of soil and climate on root zone storage capacity, Water Resour. Res., 52, 2009–2024, doi:10.1002/2015WR018115, 2016.

Fan, Y., Miguezmacho, G., Jobbágy, E. G., Jackson, R. B., & Oterocasal, C.: Hydrologic regulation of plant rooting depth. Proceedings of the National Academy of Sciences of the United States of America, 114(40), 201712381, 2017.

Fenicia, F., Savenije, H.H.G., Matgen, P., Pfister, L.: A comparison of alternative multiobjective calibration strategies for hydrological modeling. Water Resources Research 43 (3): n/a–n/a DOI: 10.1029/2006WR005098, 2007.

Gao, H., Hrachowitz, M., Schymanski, S.J., Fenicia, F., Sriwongsitanon, N., Savenije, H.H.G.: Climate controls how ecosystems size the root zone storage capacity at catchment scale. Geophysical Research Letters 41 (22): 7916–7923 DOI: 10.1002/2014gl061668, 2014a.

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

Gao, H., Hrachowitz, M., Sriwongsitanon, N., Fenicia, F., Gharari, S., Savenije, H.H.G.: Accounting for the influence of vegetation and landscape improves model transferability in a tropical savannah region. Water Resources Research 52 (10): 7999–8022 DOI: 10.1002/2016WR019574, 2016.

Gao, H., Sabo, J.L, Chen, X., Liu, Z., Yang, Z., Ren, Z., Liu, M.: Landscape heterogeneity and hydrological processes: a review of landscape-based hydrological models. Landscape Ecology, DOI: doi.org/10.1007/s10980-018-0690-4, 2018a.

Gao, H., Cai, H., Zheng, D.:. Understand the impacts of landscape features on the shape of storage capacity curve and its influence on flood. Hydrology Research. DOI: Hydrology-D-16-00245R3, 2018b.

Gao, J., Holden, J., Kirkby, M.: The impact of land-cover change on flood peaks in peatland basins. Water Resources Research 52 (5): 3477–3492 DOI: 10.1002/2015WR017667, 2016.

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

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

Gharari, S. On the role of model structure in hydrological modeling: Understanding models, PhD dissertation, 2016

Gomes, G.J.C., Vrugt, J.A., Vargas, E.A.: Toward improved prediction of the bedrock depth underneath hillslopes: Bayesian inference of the bottom-up control hypothesis using high-resolution topographic data. Water Resources Research 52 (4): 3085–3112 DOI: 10.1002/2015WR018147, 2016.

Grabs, T., Seibert, J., Bishop, K., Laudon, H.: Modeling spatial patterns of saturated areas: A comparison of the topographic wetness index and a dynamic distributed model. Journal of Hydrology 373 (1): 15–23, 2009.

De Groen, M.M., Savenije, H.H.G.: A monthly interception equation based on the statistical characteristics of daily rainfall. Water Resources Research 42 (12): n/a–n/a DOI: 10.1029/2006WR005013, 2006.

Gumbel, E. J.: Les valeurs extrêmes des distributions statistiques, Annales de l'institut Henri Poincaré, 5(2), 115–158, 1935.

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

Hargreaves, G.H., Samani, Z.A.: Reference crop evapotranspiration from temperature. Applied engineering in agriculture 1 (2): 96–99, 1985.

Haria, A.H., Shand, P.: Evidence for deep sub-surface flow routing in forested upland Wales: implications for contaminant transport and stream flow generation. Hydrology and Earth System Sciences Discussions 8 (3): 334–344, 2004.

Harte, J.: Toward a synthesis of the Newtonian and Darwinian worldviews. Physics Today 55 (10): 29–34 DOI: 10.1063/1.1522164, 2002.

Helmlinger, K.R., Kumar, P., Foufoula-Georgiou, E.: On the use of digital elevation model data for Hortonian and fractal analyses of channel network. Water Resources Research 29: 2599–2613, 1993.

Hewlett, J.D.: Soil moisture as a source of base flow from steep mountain watersheds. Southeastern Forest Experiment Station, US Department of Agriculture, Forest Service, 1961.

Hewlett, J.D., Troendle, C.A.: Non point and diffused water sources: a variable source area problem. In Watershed Management; Proceedings of a Symposium, 1975.

Homer, C. G., Dewitz, J. A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N. D., Wickham, J. D. & Megown, K.: Completion of the 2011 National Land Cover Database for the conterminous United States-representing a decade of land cover change information. Photogrammetric Engineering and Remote Sensing 81, 345–354, 2015.

Hooshyar, M., Wang, D., Kim, S., Medeiros, S.C., Hagen, S.C.: Valley and channel networks extraction based on local topographic curvature and k‐means clustering of contours. Water Resources Research 52 (10): 8081‐8102, 2016.

Horton, R.E.: The role of infiltration in the hydrologic cycle. Trans. Am. Geophys. Union 14, 446–460, 1933.

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., & Cudennec, C.: . A decade of Predictions in Ungauged Basins (PUB)—a review. Hydrological Sciences Journal 58 (6): 1198–1255 DOI: 10.1080/02626667.2013.803183, 2013.

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

Hu, G. and Jia, L.: Monitoring of evapotranspiration in a semiarid inland river basin by combining microwave and optical remote sensing observations, Remote Sens., 7, 3056–3087, doi:10.3390/rs70303056, 2015.

Iorgulescu, I., Jordan, J-P.: Validation of TOPMODEL on a small Swiss catchment. Journal of Hydrology 159 (1): 255–273 DOI: http://dx.doi.org/10.1016/0022-1694(94)90260-7, 1994.

Kirchner, J.W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research 42 (3): n/a–n/a DOI: 10.1029/2005WR004362, 2006.

Kleidon, A., Lorenz, R.D.: Non-equilibrium thermodynamics and the production of entropy: life, earth, and beyond. Springer Science & Business Media, 2004.

Kollat, J. B., P. M. Reed, and T. Wagener.: When are multiobjective calibration trade‐offs in hydrologic models meaningful?. Water Resources Research 48.3:3520, 2012.

Liang, X., Lettenmaier, D.P., Wood, E,F,, Burges, S,J.: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. Journal of Geophysical Research 99 (D7): 14415 DOI: 10.1029/94JD00483, 1994.

Liu, D., Tian, F., Hu, H., Hu, H.: The role of run-on for overland flow and the characteristics of runoff generation in the Loess Plateau, China. Hydrological Sciences Journal 57 (6): 1107–1117 DOI: 10.1080/02626667.2012.695870, 2012.

Maxwell, R. M., and Condon, L. E.: Connections between Groundwater Flow and Transpiration Partitioning. Science 353.6297: 377 LP – 380, 2016.

McDonnell, J.J., Sivapalan, M., Vaché, K., Dunn, S., Grant, G., Haggerty, R., Hinz, C., Hooper, R., Kirchner, J., Roderick, M.L., Selker, J. and Weiler, M.: Moving beyond heterogeneity and process complexity: A new vision for watershed hydrology. Water Resources Research 43 (7): n/a–n/a DOI: 10.1029/2006WR005467, 2007.

McDonnell, J.J.: Are all runoff processes the same? Hydrological Processes 27 (26): 4103–4111 DOI: 10.1002/hyp.10076, 2013.

Merz, R., Blöschl, G.: Regionalisation of catchment model parameters. Journal of Hydrology 287 (1-4): 95–123 DOI: 10.1016/j.jhydrol.2003.09.028, 2004.

Milly, P. C. D.: Climate, soil water storage, and the average annual water balance, Water Resour. Res., 30(7), 213–2156, 1994.

Molenat, J., Gascuel-Odoux, C., Ruiz, L., Gruau, G.: Role of water table dynamics on stream nitrate export and concentration in agricultural headwater catchment (France). Journal of Hydrology 348 (3): 363–378, 2008.

Molénat, J., Gascuel‐Odoux, C., Davy, P., Durand, P.: How to model shallow water‐table depth variations: the case of the Kervidy‐Naizin catchment, France. Hydrological Processes 19 (4): 901–920, 2005.

Montgomery, D.R., Dietrich, W.E.: Source areas, drainage density, and channel initiation. Water Resources Research 25 (8): 1907–1918, 1989.

Moore, R. J.: The probability-distributed principle and runoff production at point and basin scales, Hydrol. Sci. J., 30, 273-297, 1985.

Moussa, R.: Effect of channel network topology, basin segmentation and rainfall spatial distribution on the geomorphologic instantaneous unit hydrograph transfer function. Hydrological Processes 22 (3): 395–419 DOI: 10.1002/hyp.6612, 2008.

Moussa, R.: Definition of new equivalent indices of Horton-Strahler ratios for the derivation of the Geomorphological Instantaneous Unit Hydrograph. Water Resources Research 45 (9): n/a–n/a DOI: 10.1029/2008WR007330, 2009.

Nobre, A. D, Cuartas, L. A., Hodnett, M., Rennó, C.D., Rodrigues, G., Silveira, A., Waterloo, M., Saleska, S.: Height Above the Nearest Drainage - a hydrologically relevant new terrain model. Journal of Hydrology 404 (1-2): 13–29 DOI: 10.1016/j.jhydrol.2011.03.051, 2011.

Orth, R., Staudinger, M., Seneviratne, S.I., Seibert, J., Zappa, M.: Does model performance improve with complexity? A case study with three hydrological models. Journal of Hydrology 523: 147–159 DOI: http://doi.org/10.1016/j.jhydrol.2015.01.044, 2015.

Passalacqua, P., Belmont, P., Staley, D.M., Simley, J.D., Arrowsmith, J.R., Bode, C.A., Crosby, C., DeLong, S.B., Glenn, N.F., Kelly, S.A., Lague, D., Sangireddy, H., Schaffrath, K., Tarboton, D. G., Wasklewicz, T., Wheaton, J. M.: Analyzing high resolution topography for advancing the understanding of mass and energy transfer through landscapes: A review. Earth-Science Reviews 148: 174–193 DOI: http://doi.org/10.1016/j.earscirev.2015.05.012, 2015.

Pelletier, J.D., Barron-Gafford, G.A., Breshears, D.D., Brooks, P.D., Chorover, J., Durcik, M., Harman, C.J., Huxman, T.E., Lohse, K.A., Lybrand, R., Meixner, T., McIntosh, J. C., Papuga, S. A., Rasmussen, C., Schaap, M., Swetnam, T. L., and Troch, P. A.: Coevolution of nonlinear trends in vegetation, soils, and topography with elevation and slope aspect: A case study in the sky islands of southern Arizona. Journal of Geophysical Research: Earth Surface 118 (2): 741–758 DOI: 10.1002/jgrf.20046, 2013.

Perrin, C., Michel, C., Andréassian, V.: Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. Journal of Hydrology 242 (3-4): 275–301 DOI: 10.1016/S0022-1694(00)00393-0, 2001.

Perrin, C., C. Michel, and V. Andre´assian: Improvement of a parsimonious model for streamflow simulation, J. Hydrol., 279, 275– 289, 2003.

Ponce, V. M., and Hawkins, R. H.: Runoff curve number: Has it reached maturity?, J. Hydrol. Eng., 1(1), 11–19, 1996.

Rempe, D. M., and Dietrich, W. E.: A bottom-up control on fresh-bedrock topography under landscapes, Proc. Natl. Acad. Sci. U. S. A., 111(18), 6576–6581, doi:10.1073/pnas.1404763111, 2014.

Renard, K. G., Yoder, D. C., Lightle, D. T. & Dabney, S. M. Universal soil loss equation and revised universal soil loss equation. Handbook of Erosion Modelling 8, 135–167, 2011.

Rennó, C.D., Nobre, A.D., Cuartas, L.A., Soares, J.V., Hodnett, M.G., Tomasella, J., Waterloo, M. HAND, a new terrain descriptor using SRTM-DEM; mapping terra-firme rainforest environments in Amazonia. Remote Sensing of Environment 112, 3469–3481, 2008.

Rodriguez-Iturbe, I., and A. Rinaldo, Fractal River Basins: Chance and Self-Organization, Cambridge Univ. Press, 547 pp., New York, 1997.

Samaniego, L., Kumar, R., Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resources Research 46 (5): n/a–n/a DOI: 10.1029/2008WR007327, 2010.

Savenije, H. H. G.: HESS Opinions "Topography driven conceptual modelling (FLEX-Topo)", Hydrol. Earth Syst. Sci., 14, 2681–2692, doi:10.5194/hess-14-2681-2010, 2010.

Savenije, H.H.G., 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.

Schaake, J., Cong, S., and Duan, Q.: The US MOPEX data set, IAHS Publ., 307, 9, 2006.

Schwarz, G. E. & Alexander, R. B. State Soil Geographic (STATSGO) Data Base for the Conterminous United States. Open File report 95-449, US Geological Survey, Washington, DC, 1995.

Seibert, J., Stendahl, J., Sørensen, R.: Topographical influences on soil properties in boreal forests. Geoderma 141 (1-2): 139–148 DOI: 10.1016/j.geoderma.2007.05.013, 2007.

Shao, W., Su, Y., and Langhammer, J.: Simulations of coupled non-isothermal soil moisture transport and evaporation fluxes in a forest area. Journal of Hydrology and Hydromechanics, 65, 410–425, 2018

Shand, P., Haria, A.H., Neal, C., Griffiths, K., Gooddy, D., Dixon, A.J., Hill, T., Buckley, D.K., Cunningham, J.: Hydrochemical heterogeneity in an upland catchment: further characterisation of the spatial, temporal and depth variations in soils, streams and groundwaters of the Plynlimon forested catchment, Wales. Hydrology and Earth System Sciences 9 (6): 621–644, 2005.

Sørensen, R., Seibert, J.: Effects of DEM resolution on the calculation of topographical indices: TWI and its components. Journal of Hydrology 347 (1): 79–89 DOI: http://dx.doi.org/10.1016/j.jhydrol.2007.09.001, 2007.

Sivapalan, M., Woods, R.A., Kalma, J.D.: Variable bucket representation of TOPMODEL and investigation of the effects of rainfall heterogeneity. Hydrological processes 11 (9): 1307–1330, 1997.

Sivapalan, M., Takeuchi, K., Franks, S.W., Gupta, V.K., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J.J., Mendiondo, E.M., O'Connell, P.E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., Zehe, E.: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. Hydrological Sciences Journal 48 (6): 857–880 DOI: 10.1623/hysj.48.6.857.51421, 2003.

Sivapalan, M.: The secret to 'doing better hydrological science': change the question! Hydrological Processes 23 (9): 1391–1396 DOI: 10.1002/hyp.7242, 2009.

Sivapalan, M., Blöschl, G.: Time scale interactions and the coevolution of humans and water. Water Resources Research 51 (9): 6988–7022 DOI: 10.1002/2015WR017896, 2015.

Soulsby, C., Birkel, C., Geris, J., Dick, J., Tunaley, C. and Tetzlaff, D.: Stream water age distributions controlled by storage dynamics and non-linear hydrologic connectivity: modelling with high resolution isotope data. Water Resources Research. DOI: 10.1002/2015WR017888, 2015.

Soulsby, C., Bradford, J., Dick, J., McNamara, J.P., Geris, J., Lessels, J., Blumstock, M., Tetzlaff, D.: Using geophysical surveys to test tracer-based storage estimates in headwater catchments. Hydrological Processes 30 (23): 4434–4445 DOI: 10.1002/hyp.10889, 2016.

Sklash, M.G., Farvolden, R.N.: The role of groundwater in storm runoff. Journal of Hydrology 43 (1): 45–65 DOI: http://dx.doi.org/10.1016/0022-1694(79)90164-1, 1979.

Tetzlaff, D., Birkel, C., Dick, J., and C. Soulsby: Storage dynamics in hydropedological units control hillslope connectivity, runoff generation and the evolution of catchment transit time distributions. Water Resources Research, DOI: 10.1002/2013WR014147, 2014.

Tian, F. Q. , Hu, H. P. , & Lei, Z. D.: Thermodynamic watershed hydrological model: constitutive relationship. Science in China Series E: Technological Sciences, 51(9), 1353-1369, 2008. Troch, P. A., Carrillo, G., Sivapalan, M., Wagener, T., Sawicz, K.: Climate-vegetation-soil interactions and long-term hydrologic partitioning: signatures of catchment co-evolution. Hydrology and Earth System Sciences 17 (6): 2209–2217 DOI: 10.5194/hess-17-2209-2013, 2013.

Tromp-van Meerveld, H. J. & McDonnell, J. J.: Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope. Water Resources Research 42. DOI: 10.1029/2004WR003778, 2006.

Van Beek, L.P.H. and M.F.P. Bierkens, The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification, Report Department of Physical Geography, Utrecht University, Utrecht, The Netherlands, http://vanbeek.geo.uu.nl/suppinfo/vanbeekbierkens2009.pdf, 2008.

Vrugt, J. A.: Effective and efficient algorithm for multiobjective optimization of hydrologic models. Water Resources Research 39 (8): 1–19 DOI: 10.1029/2002WR001746, 2003.

Wang, D., Tang, Y.: A one - parameter Budyko model for water balance captures emergent behavior in darwinian hydrologic models. Geophysical Research Letters 41 (13): 4569 - 4577, 2014.

Wang, D.: A new probability density function for spatial distribution of soil water storage capacity leads to SCS curve number method, Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-32, in review, 2018.

Wang-Erlandsson, L., Bastiaanssen, W.G.M., Gao, H., Jägermeyr, J., Senay, G.B., van Dijk, A.I.J.M., Guerschman, J.P., Keys, P.W., Gordon, L.J., Savenije, H.H.G.: Global root zone storage capacity from satellite-based evaporation. Hydrol. Earth Syst. Sci. 20 (4): 1459–1481 DOI: 10.5194/hess-20-1459-2016, 2016.

Weiler, M., McDonnell, J. J.: Conceptualizing lateral preferential flow and flow networks and simulating the effects on gauged and ungauged hillslopes. Water Resour. Res. 43, W03403, 2007.

Wolock, D. M.: STATSGO Soil Characteristics for the Conterminous United States. US Geological Survey, Washington, DC., 1997.

Ye, A., Duan, Q., Yuan, X., Wood, E.F., Schaake, J.: Hydrologic post-processing of MOPEX streamflow simulations. Journal of Hydrology 508: 147–156 DOI: 10.1016/j.jhydrol.2013.10.055, 2014.

Yu, Z., Lu, Q., Zhu, J., Yang, C., Ju, Q., Yang, T., Chen, X., and Sudicky, E. A.: Spatial and temporal scale effect in simulating hydrologic processes in a watershed. Journal of Hydrologic Engineering, 19(1), 99-107, 2014.

Zehe, E., Fluehler, H.: Preferential transport of Isoproturon at a plot scale and a field scale tile-drained site. J. Hydrol. 247, 100–115, 2001.

Zehe, E., Ehret, U., Blume, T., Kleidon, A., Scherer, U., Westhoff, M.: A thermodynamic approach to link self-organization, preferential flow and rainfall-runoff behaviour. Hydrol. Earth Syst. Sci. 17 (11): 4297–4322 DOI: 10.5194/hess-17-4297-2013, 2013.

Zhao, R-J., Zuang, Y., Fang, L., Liu, X., Zhang, Q.: The Xinanjiang model. Hydrological forecasting — Prévisions hydrologiques (129): 351–356, 1980.

Table 1. The parameters of the models, and their prior ranges for calibration. (*S_{uMax} is a parameter in HBV, TOPMODEL and the HSC model, but HSC-MCT model does not have S_{uMax} as a free parameter; ** β is a parameter in HBV model, but not in TOPMODEL, HSC and HSC-MCT models)

Parameter	Explanation	Prior range for calibration
S_{iMax} (mm)	Maximum interception capacity	2
S_{uMax} (mm) *	The root zone storage capacity	(10, 1000)
β (-)**	The shape of the storage capacity curve	(0.01, 5)
C_e (-)	Soil moisture threshold for reduction of evaporation	(0.1, 1)
D (-)	Splitter to fast and slow response reservoirs	(0, 1)
T_{lagF} (d)	Lag time from rainfall to peak flow	(0, 10)
K_f (d)	The fast recession coefficient	(1, 20)
K_s (d)	The slow recession coefficient	(20, 400)

Table 2. The water balance and constitutive equations used in models. (Function (15)* is used in the HBV model, but not used in the TOPMODEL, HSC and HSC-MCT models)

reservoirs	Water balance equations	Constitutive equations
Interception reservoir	$\frac{\mathrm{d}S_i}{\mathrm{d}t} = P - E_i - P_e$ (8)	$E_i = \begin{cases} E_p; S_i > 0 \\ 0; S_i = 0 \end{cases}$ (9)
		$P_e = \begin{cases} 0; & S_i < S_{iMax} \\ P; & S_i = S_{iMax} \end{cases}$ (10)
Unsaturated reservoir	$\frac{\mathrm{d}S_\mathrm{u}}{\mathrm{d}t} = P_\mathrm{e} - E_\mathrm{a} - R_\mathrm{u}$ (11)	$\frac{R_\mathrm{u}}{P_\mathrm{e}} = \left(\frac{S_\mathrm{u}}{S_\mathrm{uMax}}\right)^\beta$ (12)*
		$\frac{E_a}{E_p - E_i} = \frac{S_u}{C_e S_{uMax}}$ (13)
Splitter and Lag function		$R_f = R_u D$ (17); $R_s = R_u(1-D)$ (14)
		$R_{fl}(t) = \sum_{i=1}^{T_{lagf}} c_f(i) \cdot R_f(t-i+1)$ (15)
		$c_f(i) = i / \sum_{u=1}^{T_{lagf}} u$ (16)
Fast reservoir	$\frac{\mathrm{d}S_\mathrm{f}}{\mathrm{d}t} = R_\mathrm{f} - Q_\mathrm{f}$ (17)	$Q_f = S_f / K_f$ (18)
Slow reservoir	$\frac{\mathrm{d}S_s}{\mathrm{d}t} = R_s - Q_s$ (19)	$Q_s = S_s / K_s$ (20)

Table 3. Data source of the MOPEX catchments.

Data	Unit	Resources	Website	Reference
Daily precipitation	mm/d	MOPEX	http://www.nws.noaa.gov/ohd/mopex/mo_datasets.htm	(Duan et al., 2006)
Daily maximum temperature	^{o}C	MOPEX	Same as above	Same as above

Daily minimum temperature	^oC	MOPEX	Same as above	Same as above
Daily runoff	mm/d	MOPEX	Same as above	Same as above
Aridity index	-	MOPEX	Same as above	Same as above
DEM	m	USGS	http://earthexplorer.usgs.gov/	-
Slope	degree	USGS	Same as above	-
K factor of soil	-	USGS	http://water.usgs.gov/GIS/metadata/usgswrd/XML/muid.xml	(Wolock, 1997; Gao et al., 2018)
Percentage of forest cover	%	NLCD	http://www.mrlc.gov/	(Homer et al., 2015; Gao et al., 2018)
Stream density	Km/km^2	Horizon Systems Corporation	http://www.horizon-systems.com/nhdplus/	-
Depth to bedrock	cm	STATSGO	http://www.soilinfo.psu.edu/index.cgi?soil_data&conus&data_cov&dtb	(Schwarz et al., 1995; Gao et al., 2018)

Table 4. Impacts of MOPEX catchment characteristics on model performance (HSC, HBV, and TOPMODEL)

Catchment characteristics	HSC > HBV	HSC ≈ HBV	HSC < HBV	HSC > TOPMODEL	HSC ≈ TOPMODEL	HSC < TOPMODEL
Averaged HAND (m)	37	71	-	27	69	193
Averaged slope (degree)	4.0	5.7	-	3.6	5.6	13.5
Averaged elevation (m)	454	395	-	469	393	740
Averaged K-factor (-)	0.28	0.29	-	0.29	0.29	0.25
Forest proportion (%)	22	43	-	14	43	68
Aridity index (-)	1.1	0.9	-	1.3	0.9	0.8
Stream density (-)	0.72	0.81	-	0.77	0.80	0.83

Averaged depth to rock (cm)	192	219	-	210	215	333

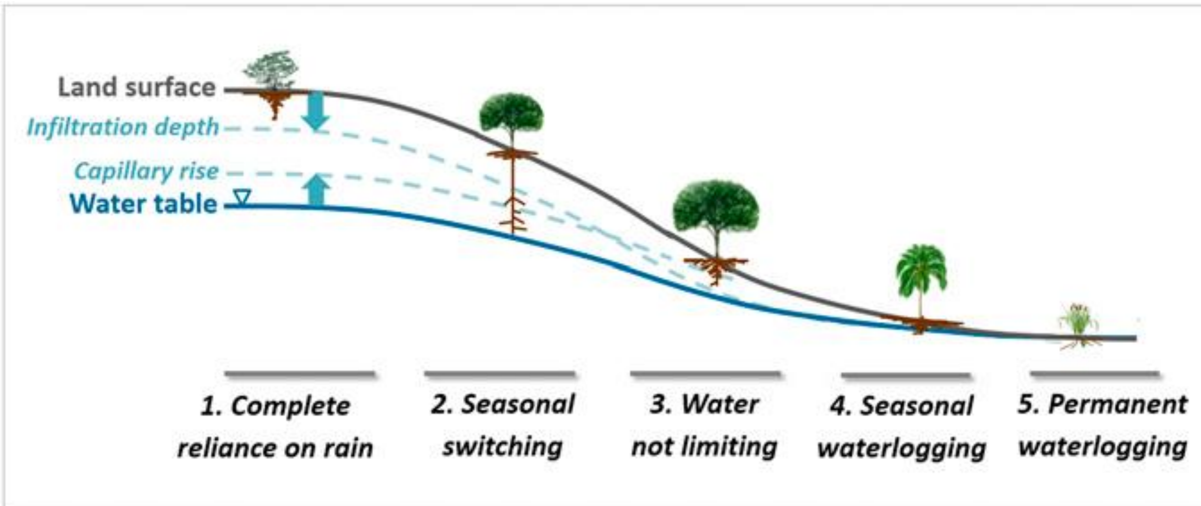

Figure 1. The variation of plant rooting depths along a hillslope profile, showing the impact of HAND (Height Above the Nearest Drainage) on rooting depth. (Taken from Fan et al., 2017 by permission of PNAS)

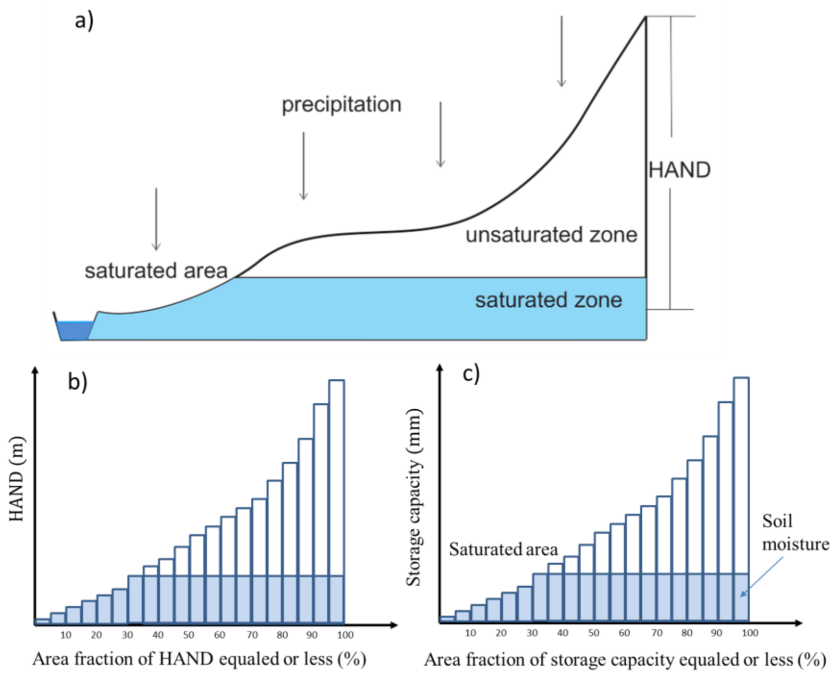

Figure 2. The perceptual model of the HAND-based Storage Capacity curve (HSC) model. a) shows the representative hillslope profile in nature, and the saturated area, unsaturated zone and saturated zone; b) shows the relationship

between HAND bands and their corresponded area fraction; c) shows the relationship between storage capacity-area fraction-soil moisture-saturated area, based on the assumption that storage capacity linearly increases with HAND values.

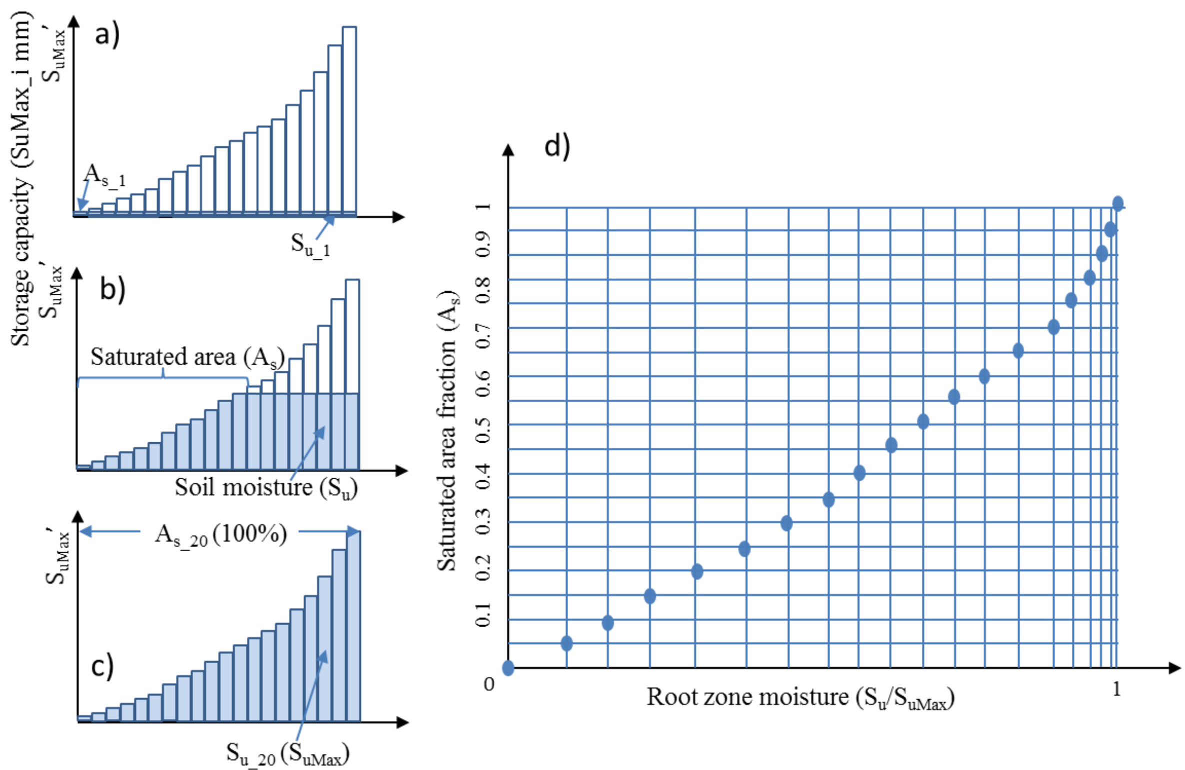

Figure 3. The conceptual model of the HSC model. a), b) and c) illustrate the relationship between soil moisture (S_u) and saturated area (A_s) in different soil moisture conditions. In d), 20 different S_u-A_s conditions are plotted, which allow us to estimate A_s from S_u.

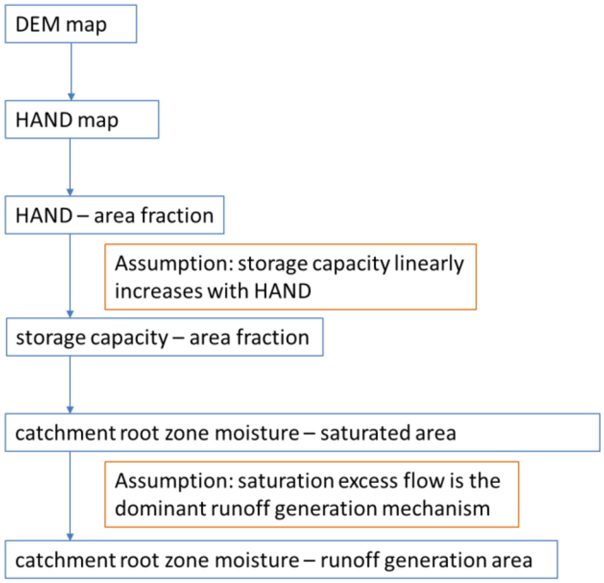

Figure 4. The procedures estimating runoff generation by the HSC model and its two hypotheses.

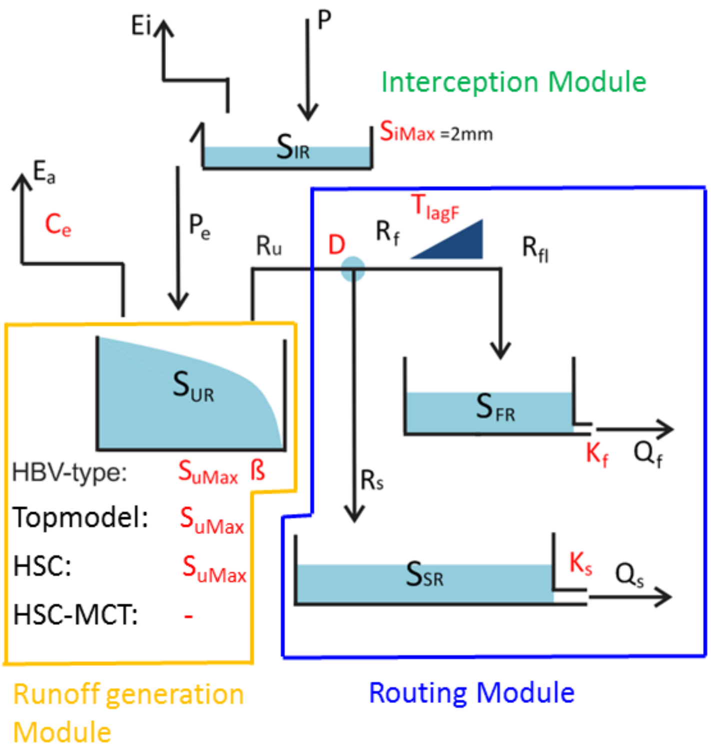

Figure 5. Model structure and free parameters, involving four runoff generation models (HBV-type, TOPMODEL, HSC, and HSC -MCT). HBV-type has S_{uMax} and beta two free parameters; TOPMODEL and HSC models have S_{uMax} as one

free parameter; and HSC-MCT model does not have free parameter. In order to simplify calibration process and make fair comparison, the interception storage capacity (S_{iMax}) was fixed as 2mm.

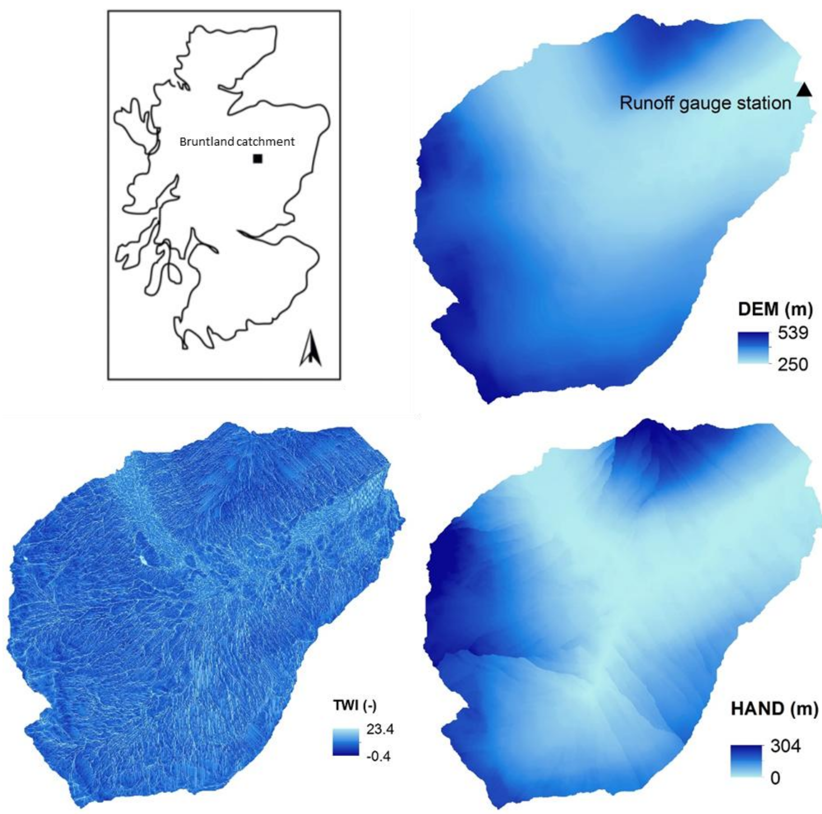

Figure 6. (a) Study site location of the Bruntland Burn catchment within Scotland; (b) digital elevation model (DEM) of the Bruntland Burn catchment; (c) the topographic wetness index map of the Bruntland Burn catchment; (d) the height above the nearest drainage (HAND) map of the Bruntland Burn catchment.

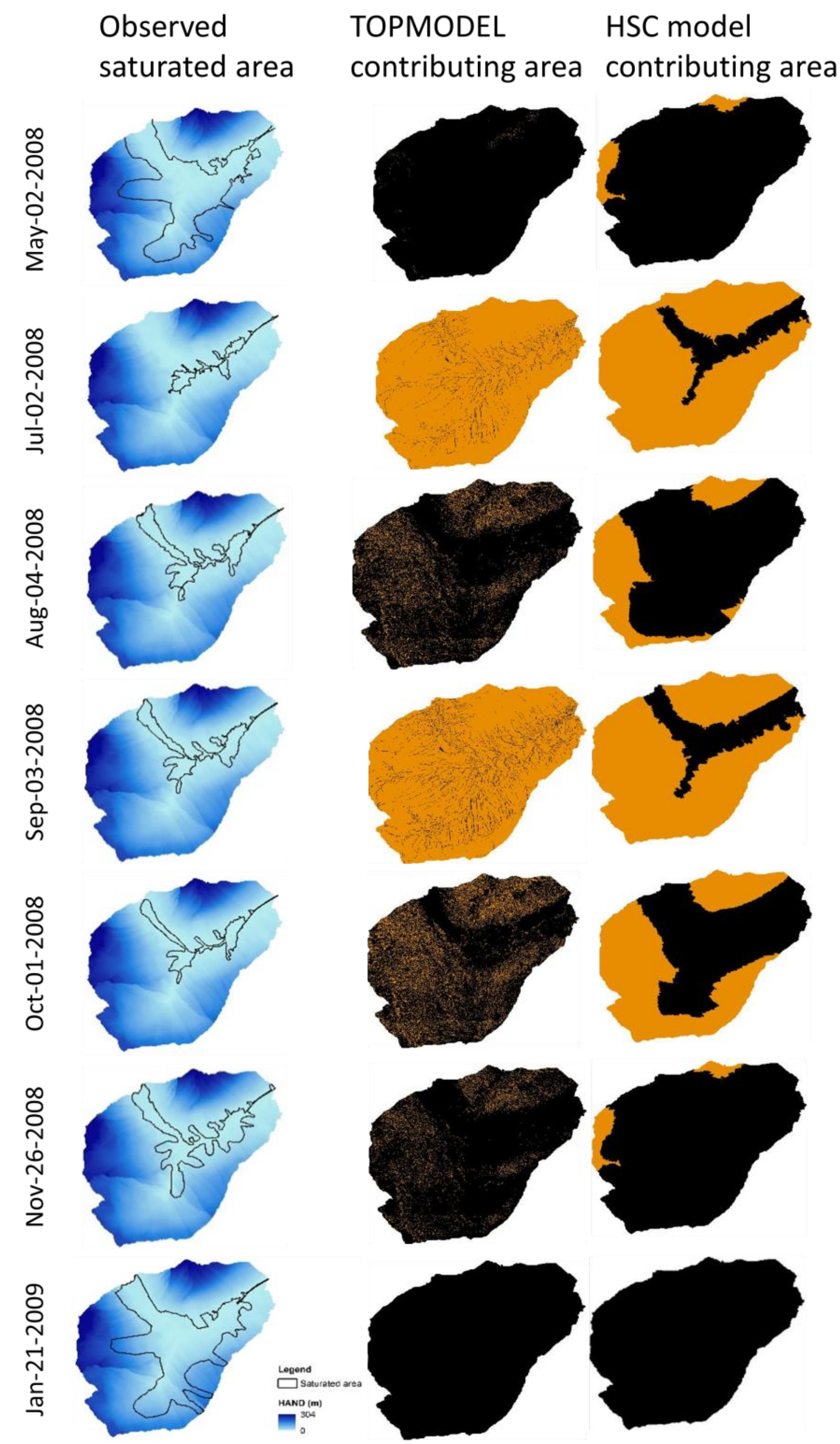

Figure 7. The measured saturated areas and the simulated contributing areas (black) by TOPMODEL and HSC models.

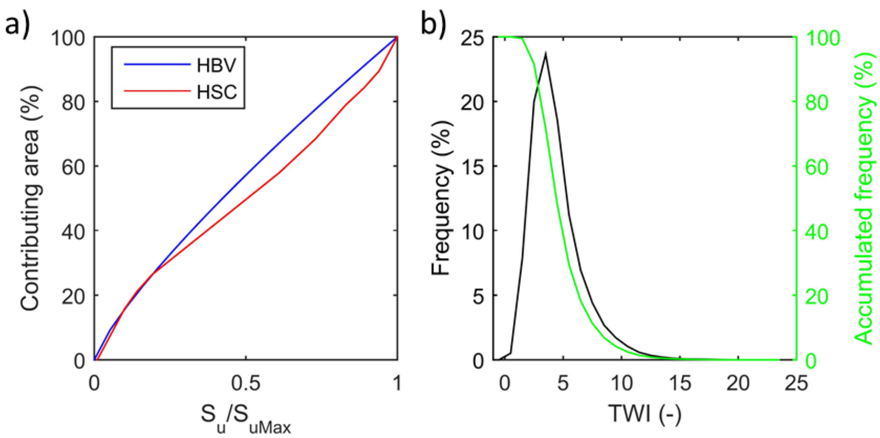

Figure 8. The curves of the beta function of HBV model, and the S_u-A_s curve generated by HSC model (the left figure). The frequency and accumulated frequency of the TWI in the Bruntland Burn catchment (the right figure).

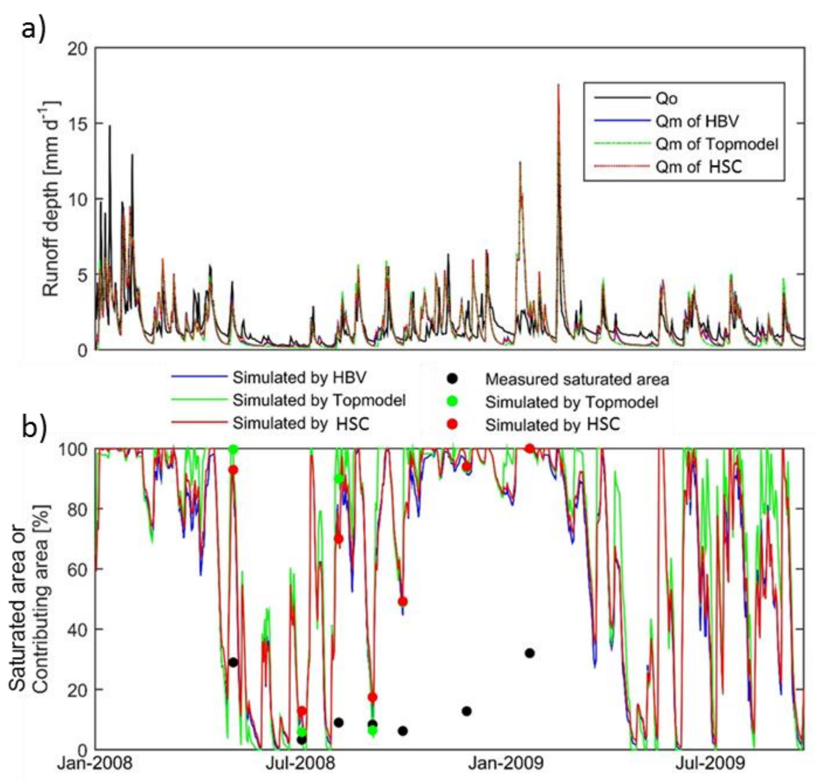

Figure 9. a) The observed hydrograph (Qo, black line) of the Bruntland Burn catchment in 2008. And the simulated hydrographs (Qm) by HBV model (blue line), TOPMODEL (green dash line), HSC model (red dash line); b) the comparison of the observed saturated area of 7 days (black dots) and simulated relative soil moistures, i.e. HBV (blue line), TOPMODEL (green line and dots), HSC (red line and dots).

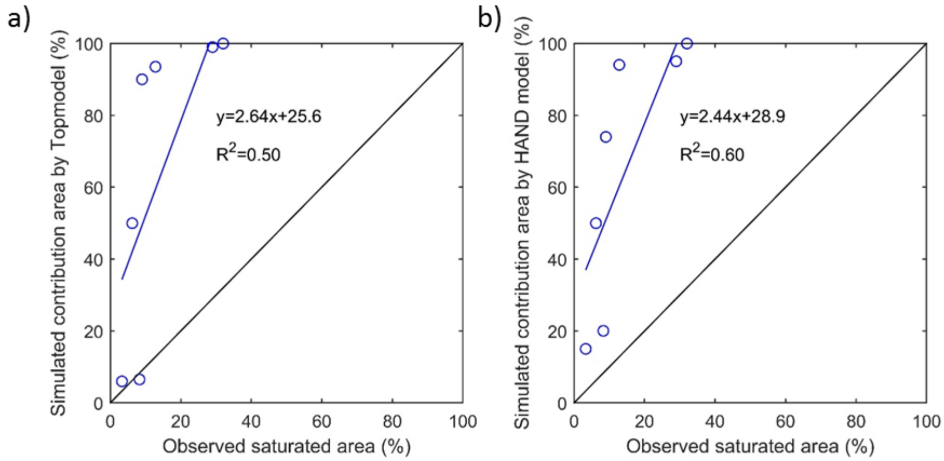

Figure 10. The comparison of the observed saturated area and simulated contributing areas by TOPMODEL and HSC models.

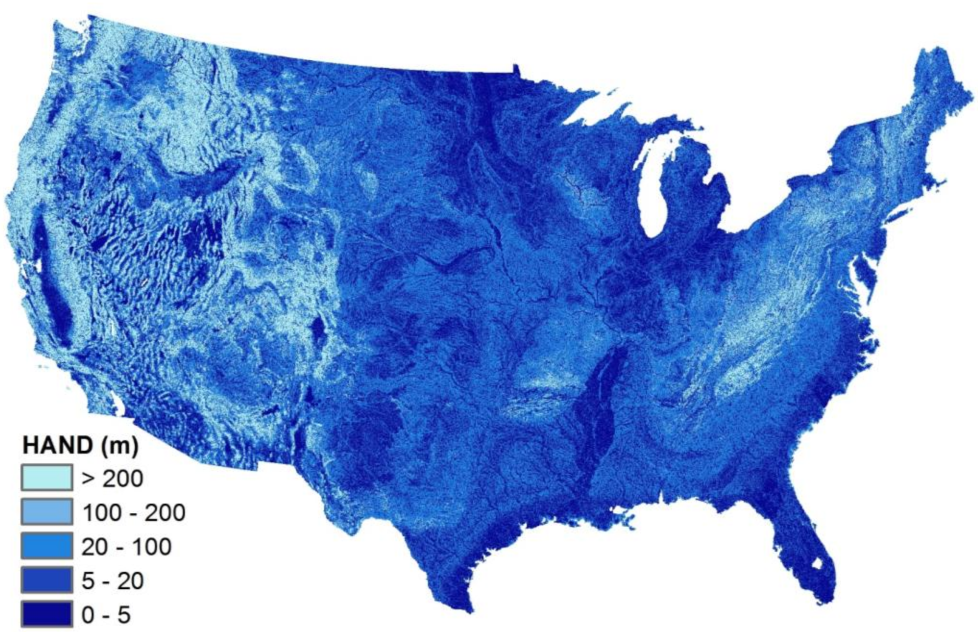

Figure 11. The Height Above the Nearest Drainage (HAND) map of the CONUS.

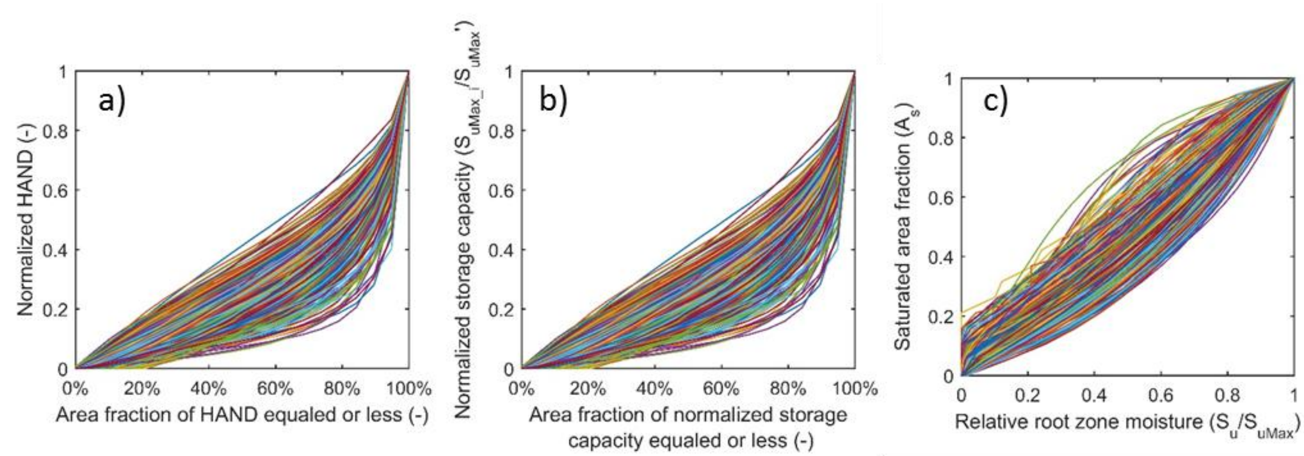

Figure 12. a) The profiles of the normalized HAND of the 323 MOPEX catchments; b) the relations between area fraction and the normalized storage capacity profile of the 323 MOPEX catchments; c) the S_u-A_s curves of the HSC model which can be applied to estimate runoff generation from relative soil moisture for the 323 MOPEX catchment.

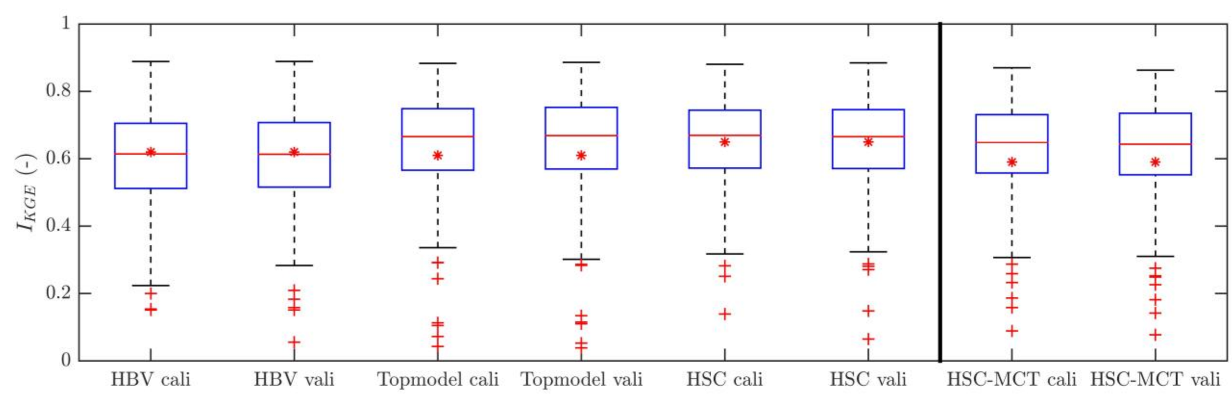

Figure 13. The comparison between the HBV, the TOPMODEL, the HSC, and the HSC-MCT models

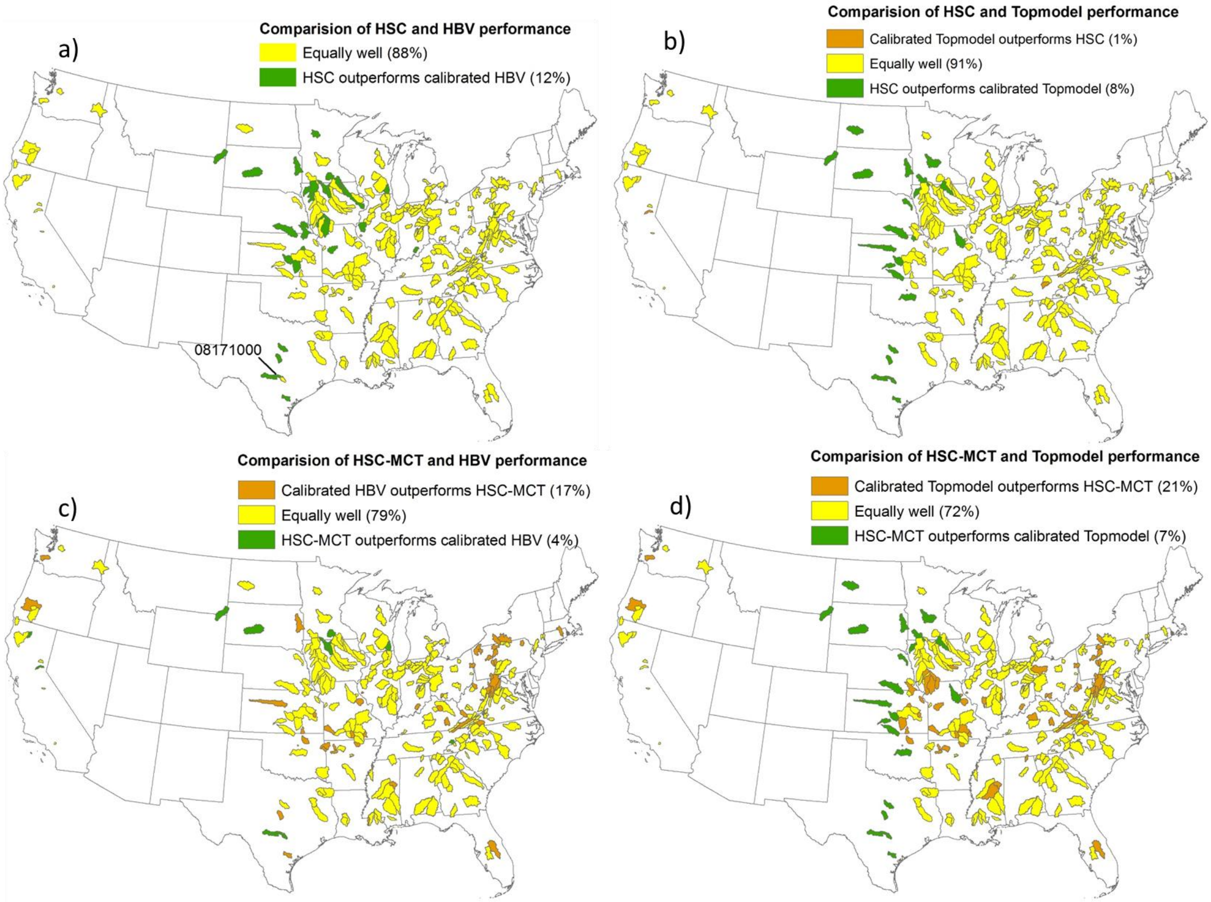

Figure 14. Performance comparison of the HSC and HSC-MCT models compared to two benchmarks models: HBV and TOPMODEL, for the 323 MOPEX catchments.