Hydrology and
Earth System
Sciences

Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China

H. Gao[1], M. Hrachowitz[1], F. Fenicia[1,2], S. Gharari[1,3], and H. H. G. Savenije[1]

[1]Water Resources Section, Delft University of Technology, Delft, the Netherlands
[2]Eawag: Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland
[3]Department of Environment and Agro-Biotechnologies, Centre de Recherche Public – Gabriel Lippmann, Belvaux, Luxembourg

Correspondence to: H. Gao (h.gao-1@tudelft.nl)

Abstract. Although elevation data are globally available and used in many existing hydrological models, their information content is still underexploited. Topography is closely related to geology, soil, climate and land cover. As a result, it may reflect the dominant hydrological processes in a catchment. In this study, we evaluated this hypothesis through four progressively more complex conceptual rainfall-runoff models. The first model (FLEX^L) is lumped, and it does not make use of elevation data. The second model (FLEX^D) is semi-distributed with different parameter sets for different units. This model uses elevation data indirectly, taking spatially variable drivers into account. The third model (FLEX^{T0}), also semi-distributed, makes explicit use of topography information. The structure of FLEX^{T0} consists of four parallel components representing the distinct hydrological function of different landscape elements. These elements were determined based on a topography-based landscape classification approach. The fourth model (FLEX^T) has the same model structure and parameterization as FLEX^{T0} but uses realism constraints on parameters and fluxes. All models have been calibrated and validated at the catchment outlet. Additionally, the models were evaluated at two sub-catchments. It was found that FLEX^{T0} and FLEX^T perform better than the other models in nested sub-catchment validation and they are therefore better spatially transferable. Among these two models, FLEX^T performs better than FLEX^{T0} in transferability. This supports the following hypotheses: (1) topography can be used as an integrated indicator to distinguish between landscape elements with different hydrological functions; (2) FLEX^{T0} and FLEX^T are much better equipped to represent the heterogeneity of hydrological functions than a lumped or semi-distributed model, and hence they have a more realistic model structure and parameterization; (3) the soft data used to constrain the model parameters and fluxes in FLEX^T are useful for improving model transferability. Most of the precipitation on the forested hillslopes evaporates, thus generating relatively little runoff.

1 Introduction

Topography plays an important role in controlling hydrological processes at catchment scale (Savenije, 2010). It may not only be a good first-order indicator of how water is routed through and released from a catchment (Knudsen et al., 1986), it also has considerable influence on the dominant hydrological processes in different parts of a catchment, which could be used to define hydrologically different response units (Savenije, 2010). As an indicator for hydrological behaviour, topography is also linked to geology, soil characteristics, land cover and climate through co-evolution (Sivapalan, 2009; Savenije, 2010). Thus, information on other features can to some extent be inferred from topography. However, the information provided by topography is generally underexploited in hydrological models although it is, explicitly or implicitly, incorporated in many models (e.g. Beven and Kirkby, 1979; Knudsen, 1986; Uhlenbrook et al., 2004).

As a typical lumped topography-driven model, TOP-MODEL (Beven and Kirkby, 1979) uses the topographic

wetness index (TWI) (Beven and Kirkby, 1979), which is a proxy for the probability of saturation of each point in a catchment, to consider the influence of topography on the occurrence of saturated overland flow (SOF). Similarly, the Xinanjiang model (Zhao, 1992) implicitly considers the influence of topography in its soil moisture function, as the curve of its tension water capacity distribution can be interpreted as topographic heterogeneity. Conceptually, both models implicitly reflect the variable contributing area (VCA) concept. Although the topography-aided VCA representation is present in many models, experimental evidence has shown that its underlying assumptions may not always hold (Western et al., 1999; Spence and Woo, 2003; Tromp-van Meerveld and McDonnell, 2006). In view of these limitations, and in spite of their often demonstrated suitability, there is an urgent need to explore new and potentially more generally applicable ways to incorporate topographic information in conceptual hydrological models.

Other types of topographically driven hydrological models are distributed physically based models, which use topography essentially to define flow gradients and flow paths of water (Refsgaard and Knudsen, 1996). The limitations of this kind of "bottom-up" model approach include the increased computational cost, and maybe more importantly, the unaccounted scale effects (Abbott and Refsgaard, 1996; Beven and Germann, 2013; Hrachowitz et al., 2013b). Knudsen et al. (1986) developed a semi-distributed, physically based model, which divides the whole catchment into several distinct hydrological response units defined by the catchment characteristics such as meteorological conditions, topography, vegetation and soil types. Although this type of model was shown to work well in case studies, approaches like this suffer from their intensive data requirement and complex model structure, similar to modelling approaches based on the dominant runoff process concept (Grayson and Blöschl, 2001; Scherrer and Naef, 2003). The main problem with physically based models appears to be that by breaking up catchments into small interacting cells, patterns present in the landscape are broken up as well. The question is how to make use of landscape diversity, and the related hydrological processes, while maintaining the larger scale patterns and without introducing excessive complexity.

The recently suggested topography-driven conceptual modelling approach (FLEX-Topo) (Savenije, 2010), which attempts to exploit topographic signatures to design conceptual model structures as a means to find the simplest way to represent the complexity and heterogeneity of hydrological processes, forms a middle way between parsimonious lumped and complex distributed models, and represents the subject of this study. In the framework of FLEX-Topo, topographic information is regarded as the main indicator of landscape classes and dominant hydrological processes. A valuable key for hydrologically meaningful landscape classification is the recently introduced metric HAND (Height Above the Nearest Drainage) (Rennó et al., 2008; Nobre et

al., 2011; Gharari et al., 2011), which is a direct reflection of hydraulic head to the nearest drain. Consequently, within a flexible modelling framework (Fenicia et al., 2008, 2011), different model structures can be developed to represent the different dominant hydrological processes in different landscape classes. Note that FLEX-Topo is not another conceptual model but rather a modelling framework to make more exhaustive use of topographic information in hydrological models and it can in principal be applied to any type of conceptual model.

Model transferability is one of the important indicators in testing model realism (Klemeš, 1986). Although many hydrological models, both lumped and distributed, frequently perform well in calibration, transferring them and their parameter sets into other catchments, or even into nested sub-catchments, remains problematic (Pokhrel and Gupta, 2011). There are several reasons for this: uncertainty in the data, insufficient information provided by the hydrograph or an unsuitable model structure which does not represent the dominant hydrological processes or their spatial heterogeneity sufficiently well (Gupta et al., 2008). Various techniques to improve model transferability have been suggested in the past (Seibert and McDonnell, 2002; Uhlenbrook and Leibundgut, 2002; Khu et al., 2008; Hrachowitz et al., 2013b; Euser et al., 2013; Gharari et al., 2013b), and it became clear that successful transferability critically depends on appropriate methods to link catchment characteristics to model structures and parameters or in other words to link catchment form to hydrological function (Gupta et al., 2008).

In this study, the FLEX-Topo modelling strategy (Savenije, 2010) is applied and tested with a tailor-made hydrological model for a cold, large river basin in northwest China. A lumped conceptual model with lumped input data (FLEX^L) and a semi-distributed model with semi-distributed input data and the different parameters for different units (FLEX^D) are used as benchmarks to assess the additional value of topography-driven semi-distributed modelling (FLEX^{T0}) and the value of soft data in constraining model behaviour (FLEX^T). The models are used as tools for testing different hypotheses within a flexible modelling framework (Fenicia et al., 2008, 2011). The objectives of this study are thus (1) to develop a topography-driven semi-distributed conceptual hydrological model (FLEX^T, FLEX^{T0}), based on topography-driven landscape classification and our understanding of the catchments, and to compare it to model set-ups with less process heterogeneity (FLEX^L, FLEX^D) and (2) to assess the differences in transferability of both model structures and parameters of the tested models to two uncalibrated nested sub-catchments in the study basin, thereby evaluating the predictive power and the realism of the individual model set-ups.
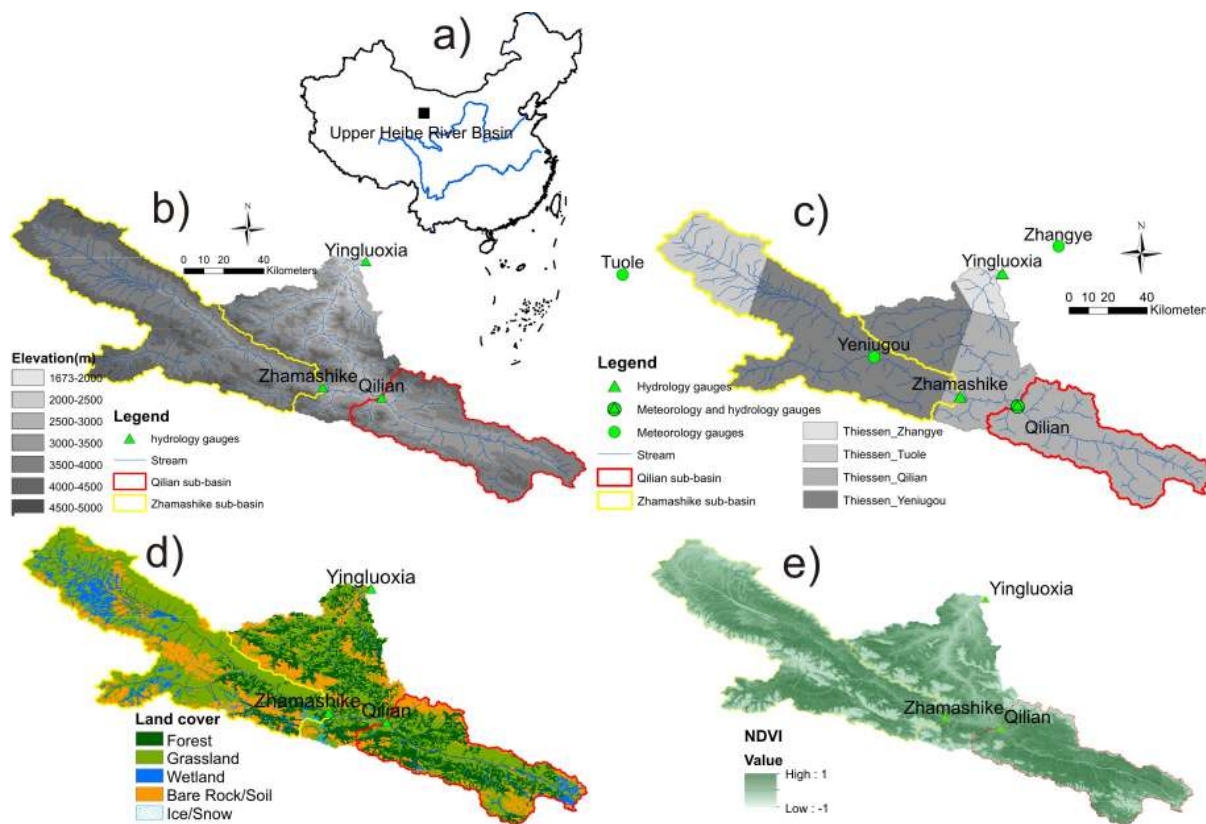
Figure 1. (a) Location of the Upper Heihe in China; **(b)** DEM of the Upper Heihe with its runoff-gauging stations, meteorological stations, streams and the outline of two sub-catchments; **(c)** meteorological stations and associated Thiessen polygons, the different grayscale indicating different long-term annual average precipitation (the darker the shade of gray, the more precipitation: Zhangye is 131 mm a^{-1}; Tuole is 293 mm a^{-1}; Qilian is 394 mm a^{-1}; Yeniugou is 413 mm a^{-1}); **(d)** land cover map of the Upper Heihe; **(e)** averaged NDVI map for the summer of 2002.

2 Study site

The Upper Heihe River basin (referred to as Upper Heihe) is part of the second largest inland river in China which, from its source in the Qilian Mountains, drains into two lakes in the Gobi Desert. The Upper Heihe is located in the southwest of Qilian Mountain in north-western China (Fig. 1a). It is gauged by the gauging station at Yingluoxia, with a catchment area of 10 000 km^2. Two sub-catchments are gauged separately by Zhamashike and Qilian (Fig. 1b). The elevation of the Upper Heihe ranges from 1700 to 4900 m (Fig. 1b). The mountainous headwaters, which are the main runoff-producing region and relatively undisturbed by human activities, are characterized by a cold desert climate. Long-term average annual precipitation and potential evaporation are about 430 and 520 mm a^{-1}. Over 80 % of the annual precipitation falls from May to September. Snow normally occurs in winter but with a limited snow depth, averaging between 4 and 7 mm a^{-1} of snow water equivalent for the whole catchment (Wang et al., 2010). The Thiessen polygons of four meteorological stations in and around the Upper Heihe are shown in Fig. 1c. The soil types are mostly mountain

straw and grassland soil, cold desert, chernozemic soil and chestnut-coloured soil. Land cover in the Upper Heihe is composed of forest (20 %), grassland (52 %), bare rock or bare soil (19 %) and wetland (8 %), as well as ice and permanent snow (0.8 %) (Fig. 1d).

The Upper Heihe has been the subject of intensive research since the 1980s (Li et al., 2009). A number of hydrological models have been previously applied in this cold mountainous watershed (Kang et al., 2002; Xia et al., 2003; Chen et al., 2003; Zhou et al., 2008; Jia et al., 2009; Li et al., 2011; Zang et al., 2012). Because of limited water resources and the increasing water demand of industry and agriculture, the conflict between human demand and ecological demand in the lowland parts of the Heihe River has become more and more severe. As the main runoff-producing region for the Heihe River, the Upper Heihe is thus essential for the water management of the whole river system.
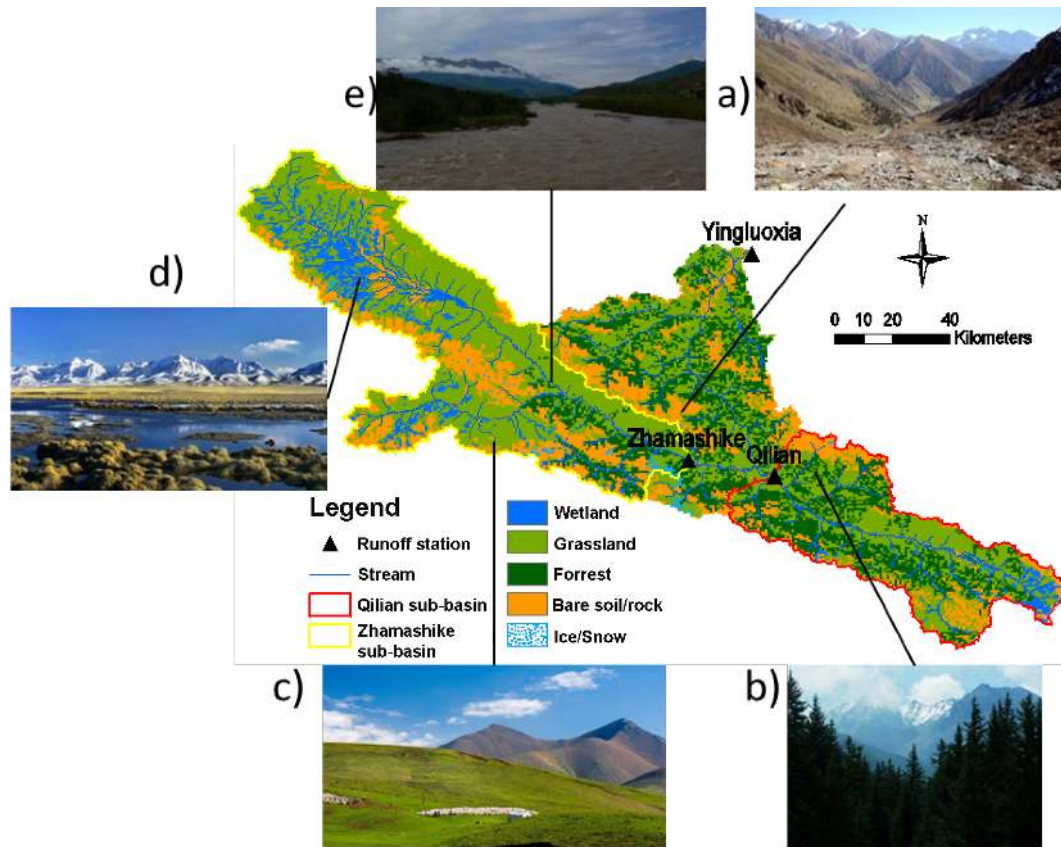
Figure 2. Characteristic landscapes in different locations in the Upper Heihe. **(a)** shows the bare-soil/rock-covered hillslope; **(b)** shows the forest-covered hillslope; **(c)** shows the grass-covered hillslope; **(d)** shows the wetland and terrace; **(e)** shows the muddy river.

The landscapes and the perceptual model of the Upper Heihe

Figure 2 illustrates different characteristic landscape elements in the Upper Heihe which were used to guide model development. Five characteristic landscapes can be identified in the Upper Heihe: bare-rock mountain peaks, forested hillslopes, grassland hillslopes, terraces and wetlands. Typically, above a certain elevation, the landscape is covered by bare soil/rock (Fig. 2a) or permanent ice/snow. At lower elevations, north-facing hillslopes tend to be covered by forest (Fig. 2b), while the bottom of hillslopes and south-facing hillslopes are, in contrast, dominantly covered by grass (Fig. 2c). Terraces, which are irregularly flooded in wet periods and have comparably low terrain slopes, are mostly located between channels and hillslopes, and are typically covered by grassland (Fig. 2d). Wetlands consist of meadows and open water, located in the bottom of the valleys (Fig. 2d).

This information was the basis for the development of a perceptual model of the Upper Heihe, which synthesizes our understanding of catchment hydrological behaviour. Typically, on bare soil/rock, interception can be considered negligible due to the absence of significant vegetation cover. The bare-soil/rock landscape at high elevations is further characterized by a thin soil layer, underlain by partly weathered bedrock, with higher permeable debris slopes at lower elevations. On the rock and the thin soil, the dominant lateral runoff processes are Hortonian overland flow (HOF) and SOF. Part of the localized overland flow may re-infiltrate, thereby feeding the debris slope and groundwater, while the rest of surface runoff, characterized by elevated sediment loads, is routed into streams. The picture of turbid water in a channel (Fig. 2e), illustrates the presence of soil erosion, which is likely to be caused by HOF and SOF in the bare-soil/rock hillslopes. On the grass- and forest-covered hillslopes, subsurface storm flow (SSF) is considered to be the dominant hydrological process as a result of the presence of efficient subsurface drainage networks that were created by biological and geological activity, significantly influencing the hillslope runoff yield mechanism (Beven and Germann, 2013). It can be expected that the forest-covered hillslopes are characterized by higher interception capacities and transpiration rates than the grassland hillslopes due to the larger leaf area index (LAI) and deeper root zone. Typically, the dominant hydrological process of wetlands and terraces is SOF due to the shallow groundwater levels and related limited additional storage capacity. For the same

Table 1. Summary of four the meteorological stations in and close to the Upper Heihe.

Station	Elevation (m)	Latitude (°)	Longitude (°)	Thiessen area ratio (%)	Precipitation (mm a^{-1})	Annual average daily temperature (°C)	Potential evaporation (mm a^{-1})
Zhangye	1484	38.93	100.43	4	131	7.1	804
Yeniugou	3320	38.42	99.58	43	413	−3.2	392
Qilian	2788	38.18	100.25	40	394	0.8	513
Tuole	3368	38.80	98.42	13	293	−3.0	421

Table 2. Catchment characteristics of the entire Upper Heihe and two sub-catchments, Qilian and Zhamashike.

	Latitude (°)	Longitude (°)	Elevation range (m)	Average elevation (m)	Area (km^2)	Discharge (mm a^{-1})
Yingluoxia (Upper Heihe)	38.80	100.17	1673–4918	3661	10 009	145
Qilian (east tributary)	38.19	100.24	2704–4835	3535	2924	142
Zhamashike (west tributary)	38.23	99.98	2819–4840	3990	5526	124

reason, evaporative fluxes in the wetlands can be assumed to be energy rather than moisture constrained and thus close to potential rates. Further, given the short distance to the channel network, the lag times for runoff generation in wetlands can be considered negligible on a daily timescale.

3 Data

3.1 Data set

Meteorological data were available on a daily basis from four stations in and around the Upper Heihe (1959–1978), while daily runoff data were available for the main outlet of the basin at Yingluoxia (1959–1978) and two nested sub-catchments, Qilian (1967–1978) and Zhamashike (1959–1978). The meteorological data, as the forcing data of the hydrological models, included daily precipitation and daily mean air temperature. Because only data from four meteorological stations were available, the Thiessen polygon method (Fig. 1c) was applied to spatially extrapolate precipitation and temperature. A summary of meteorological data is given in Table 1. The basic information of the three basins is listed in Table 2. Potential evaporation was estimated by the Hamon equation (Hamon, 1961), which is based on daily average temperature.

The 90 m × 90 m digital elevation model (DEM) of the study site (Fig. 1b) was obtained from http://srtm.csi.cgiar.org/ and used to derive the local topographic indices HAND, slope and aspect. The normalized difference vegetation index (NDVI) map (Fig. 1e) was derived from cloud-free Landsat Thematic Mapper (TM) maps in the summer of 2002,

which were obtained from US Geological Survey EarthExplorer (http://earthexplorer.usgs.gov/). The land cover map (Fig. 1d) was made available by the Environmental and Ecological Science Data Center for West China. The pictures shown in Fig. 2 as soft data were downloaded from Google Earth.

3.2 Distribution of forcing data

The elevation of the Upper Heihe ranges from 1674 to 4918 m with only four meteorological stations in or around the catchment, covering an area of 10 000 km^2. In addition, the meteorological stations in the Upper Heihe River are all located at relatively low elevations in the valley bottoms, which are easily accessible for maintenance but potentially unrepresentative (Klemeš, 1990). Precipitation and temperature data were thus adjusted using empirical relationships.

The entire catchment was thus first discretized into four parts by the Thiessen polygon method. Each Thiessen polygon was then further stratified into seven elevation zones with steps of 500 m. Annual precipitation (Eq. 1) was assumed to increase linearly with elevation increase, according to empirical relationships for the region obtained from literature (Wang, 2009):

$$P_{aj} = P_a + (h_j - h_0) C_{pa}, \qquad (1)$$

where P_a (mm a^{-1}) is the annual observed precipitation, P_{aj} (mm a^{-1}) is the annual extrapolated precipitation in elevation h_j (m), h_0 (m) is the elevation of the meteorological station and $C_{pa} = 0.115$ mm (m a)$^{-1}$ (Wang, 2009) is the precipitation lapse rate. However, Eq. (1) is only suitable for annual precipitation extrapolation, but not suitable for daily

Table 3. Water balance and constitutive equations used in FLEX^L.

Reservoirs	Water balance equations	Constructive equations
Snow	$\frac{dS_w}{dt} = P_s \; M$ (4)	$M = \begin{cases} F_{DD}\,(T - T_t) & \text{if} \quad T > T_t \\ 0 & \text{if} \quad T <= T_t \end{cases}$ (5)
Interception	$\frac{dS_i}{dt} = P_r - E_i - P_{tf}$ (6)	$E_i = \begin{cases} E_p; \; S_i > 0 \\ 0; \; S_i = 0 \end{cases}$ (7) $P_{tf} = \begin{cases} 0; \; S_i < I_{max} \\ P_r; \; S_i = I_{max} \end{cases}$ (8)
Unsaturated reservoir	$\frac{dS_u}{dt} = P_e\,(1 - C_r) - E_a$ (9)	$E_a = (E_p - E_i)\,\min\left(\frac{S_u}{S_{uMax}\,C_e}, 1\right)$ (10) $C_r = 1 - (1 - S_u/S_{uMax})^{\beta}$ (11) $R_u = P_e\,C_r$ (12)
Splitter and lag function		$R_f = R_u\,D$; (13) $\quad R_s = R_u(1 - D)$ (14) $R_{fl}(t) = \sum_{i=1}^{T_{lag}} c(i) \cdot R_f(t - i + 1)$ (15) $c(i) = i / \sum_{u=1}^{T_{lag}} u$ (16)
Fast reservoir	$\frac{dS_f}{dt} = R_{fl} - Q_{ff} - Q_f$ (17)	$Q_{ff} = \max\,(0, \, S_f - S_{ftr})\,/K_{ff}$ (18) $Q_f = S_f/K_f$ (19)
Slow reservoir	$\frac{dS_s}{dt} = R_s - Q_s$ (20)	$Q_s = S_s/K_s$ (21)

precipitation because there are many days without precipitation within one year. Daily elevation-adjusted precipitation was therefore derived from Eq. (1) with the following expression:

$$P_j = P\,\frac{P_{aj}}{P_a} = P\,\frac{P_a + (h_j - h_0)\,C_{pa}}{P_a} = P\left(1 + \frac{C_{pa}}{P_a}\,(h_j - h_0)\right), \quad (2)$$

where P (mm d^{-1}) is the observed daily precipitation and P_j (mm d^{-1}) is the daily extrapolated precipitation in elevation. Equation (3) is used to extrapolate the daily mean temperature:

$$T_j = T - C_t\,(h_j - h_0). \quad (3)$$

T (°C) is the observed daily mean temperature; T_j (°C) is the elevation-corrected daily mean temperature; C_t (°C m^{-1}) is the environmental temperature lapse rate, which is set to 0.006 °C m^{-1} (Gao et al., 2012). The potential evaporation was estimated in each elevation zone using the elevation-corrected temperature.

4 Modelling approach

In this study four conceptual models of different complexity were designed and tested: a lumped model (FLEX^L); a model with a semi-distributed model structure, consisting of four structurally identical, parallel components with different parameter sets for each component (FLEX^D); a topography-driven semi-distributed model without soft data (FLEX^{T0});

and FLEX^T with the same model structure as FLEX^{T0} but constrained by expert knowledge. All models are a combination of reservoirs, lag functions and connection elements linked in various ways to represent different hydrological functions constructed with the flexible modelling framework SUPERFLEX (Fenicia et al., 2011).

4.1 Lumped model (FLEX^L)

The lumped model (FLEX^L) (Fig. 3) has a similar structure to earlier applications of the FLEX model (Fenicia et al., 2008), and it comprises five reservoirs: a snow reservoir (S_w), an interception reservoir (S_i), an unsaturated reservoir (S_u), a fast-response reservoir (S_f) and a slow-response reservoir (S_s). A lag function represents the lag time between storm and flood peak. The elevation-corrected Thiessen-polygon-averaged precipitation, temperature and potential evaporation are used as forcing data. Note that this model structure was developed based on the perceptual model of the basin: the snow cover in winter shown in Fig. 2d indicates the necessity of the snow reservoir; the forest cover in Fig. 2b indicates the necessity of the interception reservoir; the unsaturated reservoir is important to separate the precipitation into runoff and unsaturated reservoir storage; the response reservoirs (S_f, S_s) are included to represent the fast and slow-response in hydrological models. The equations are shown in Table 3. On the other hand, including the elevation correction of precipitation and temperature in this lumped model is
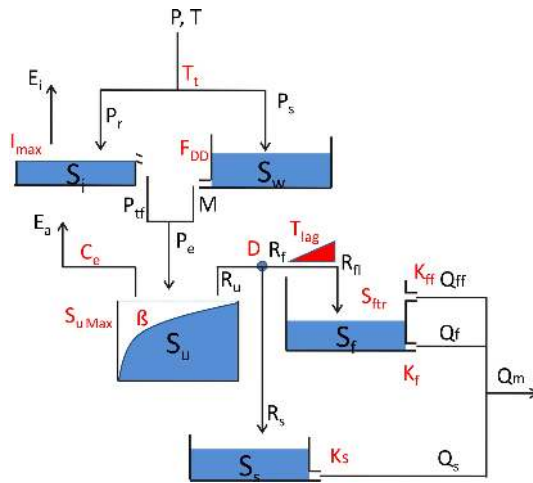
Figure 3. The lumped model structure FLEX$^{\text{L}}$.

also based on our knowledge of the large catchment area and elevation differences.

4.1.1 Snow and interception routine

Precipitation can be stored in snow or interception reservoirs before the water enters the unsaturated reservoir. Basically, the snow routine plays an important role in winter and spring while interception becomes more important in summer and autumn. Here it is assumed that interception happens during rainfall events when the daily air temperature is above the threshold temperature (T_{t}; the units of the parameters are listed in Tables 5 and 6) and there is no snow cover, i.e. typically in summer. When the average daily temperature is below T_{t}, precipitation is stored as snow cover, which normally occurs in winter. When there is snow cover and the temperature is above T_{t}, the effective precipitation (P_{e}; hereinafter the unit of fluxes is mm d^{-1}) is equal to the sum of rainfall (P) and snowmelt (M), conditions normally prevailing in early spring and early autumn. Note that snowmelt water is conceptualized to directly infiltrate into the soil, thus effectively bypassing the interception store. In other words, interception and snowmelt never happen simultaneously. Their respective activation is controlled by air temperature, precipitation and the presence of snow cover.

The snow routine was designed as a simple degree-day model as successfully applied in many conceptual models (Seibert, 1997; Uhlenbrook et al., 2004; Kavetski and Kuczera, 2007; Hrachowitz et al., 2013a; Gao et al., 2012). As shown in Eqs. (4) and (5), M is the snowmelt, S_{w} (the unit of storage is mm) is the storage of snow reservoir, dt (d) is the discretized time step and F_{DD} is the degree day factor, which defines the melted water per day per Celsius degree above T_{t}.

The interception evaporation E_{i} was calculated by potential evaporation (E_{p}) and the storage of interception

reservoir (S_{i}), with a daily maximum storage capacity (I_{max}) (Eqs. 6, 7, 8).

4.1.2 Soil routine

The soil routine, which is the core of hydrological models used in this study, determines the amount of runoff generation. In this study, we applied the widely used beta function of the Xinanjiang model (Zhao, 1992) to compute the runoff coefficient for each time step as a function of the relative soil moisture. In Eq. (11), C_{r} (−) indicates the runoff coefficient, S_{u} is the soil moisture content, S_{uMax} is the maximum soil moisture capacity in the root zone and β is the parameter describing the spatial process heterogeneity in the study catchment. In Eq. (12), P_{e} indicates the effective rainfall and snowmelt into the soil routine; R_{u} represents the generated flow during rainfall events. In Eq. (13), R_{f} indicates the flow into the fast-response routine; D is a splitter to separate recharge from preferential flow. In Eq. (14), R_{s} indicates the flow into the groundwater reservoir. In Eq. (10) S_{u}, S_{uMax} and potential evaporation (E_{p}) were used to determine actual evaporation E_{a}; C_{e} indicates the fraction of S_{uMax} above which the actual evaporation is equal to potential evaporation, here set to 0.5 as previously suggested by Savenije (1997); otherwise E_{a} is constrained by the water available in S_{u}.

4.1.3 Response routine

Equations (15) and (16) were used to describe the lag time between storm and peak flow. $R_{\text{f}}(t − i + 1)$ is the generated fast runoff in the unsaturated zone at time $t − i + 1$, T_{lag} is a parameter which represents the time lag between storm and fast runoff generation, $c(i)$ is the weight of the flow in $i − 1$ days before and $R_{\text{fl}}(t)$ is the discharge into the fast-response reservoir after convolution.

The linear-response reservoirs, representing a linear relationship between storage and release, are applied to conceptualize the discharge from the surface runoff reservoir, fast-response reservoirs and slow-response reservoirs. In Eq. (18), Q_{ff} is the surface runoff, with timescale K_{ff}, active when the storage of the fast-response reservoir exceeds the threshold S_{ftr}. In Eqs. (19) and (21), Q_{f} and Q_{s} represent the fast and slow runoff; S_{f} and S_{s} represent the storage state of the fast and the groundwater reservoirs; K_{f} and K_{s} are the timescales of the fast and slow runoff, respectively, while Q_{m} is the total modelled runoff from the three individual components.

4.2 Model with semi-distributed forcing data (FLEX$^{\text{D}}$)

In order to test the influence of model complexity on model performance and model transferability, another benchmark model (FLEX$^{\text{D}}$) based on FLEX$^{\text{L}}$ was developed. The four parallel model structures of FLEX$^{\text{D}}$ are identical to FLEX$^{\text{L}}$, but they are run with independent parameter sets (Fig. 4) and semi-distributed input data (see Sect. 3.2), resulting in

distributed accounting of the four Thiessen polygons and the seven elevation bands in the study area (Fig. 4). In total, there are 48 parameters for these four Thiessen polygons.

4.3 Topography-driven, semi-distributed models (FLEX^{T0} and FLEX^T)

Based on the perceptual model of the Upper Heihe (see Section "The landscapes and the perceptual model of the Upper Heihe"), the hypotheses that different observable landscape units are associated with different dominant hydrological processes was tested by incorporating these units into hydrological models.

4.3.1 Landscape classification

In this study, HAND (Rennó et al., 2008; Nobre et al., 2011; Gharari et al., 2011), elevation, slope and aspect (Fig. 5) were used for deriving a hydrologically meaningful landscape classification. The stream initiation threshold for estimating HAND was set to 20 cells (0.16 km^2), which was selected to maintain a close correspondence between the derived stream network and that of the topographic map. The HAND threshold value for distinction between wetland and other landscapes was set at 5 m, similar to what was used in earlier studies (Gharari et al., 2011). If HAND is larger than 5 m, but the local slope is less than 0.1, the landscape element defines terrace as a landscape unit that connects hillslopes with wetlands. The most dominant landscape in the Upper Heihe, however, is the hillslope, which has been further separated into three subclasses according to HAND, absolute elevation, aspect and vegetation cover (Fig. 5). Thus, hillslopes above 3800 m and with HAND > 80 m, typically characterized by bare soil/rock, have been accordingly defined as bare-soil/rock hillslopes. At elevations between 3200 and 3600 m and aspect between 225 and 135°, or at elevations below 3200 m and aspect between 270 and 90°, hillslopes in the Upper Heihe are generally forested (Jin et al., 2008) and thus have been defined as forest hillslopes. The remaining hillslopes were defined as grassland hillslopes. From the classification map (Fig. 6b), it can be seen that the landscape classification is similar to the independently obtained land cover map (Fig. 6a) except for the area of wetland, due to different definitions between the land cover map and our classification. Note that wetland and terrace landscape classes have been combined (Fig. 6b), because the area proportion of wetlands varies over time, while terraces may be flooded at times, which can be described by the VCA concept. This combination is unlikely to reduce realism and makes the model simpler. Consequently, the NDVI map has been averaged in accordance with this classification (Fig. 6c).

4.3.2 FLEX^{T0} and FLEX^T model structures

Based on the landscape classification and the perceptual models for each landscape, different model structures to

Table 4. Proportion of different landscape units in the study catchments.

	Wetland/ terrace (%)	Grassland hillslope (%)	Forest hillslope (%)	Bare soil/rock (%)
Heihe	36	31	17	16
Qilian	33	28	25	14
Zhamashike	46	33	6	15

represent the different dominant hydrological processes were assigned to the four individual landscape classes (Table 4). The four model structures ran in parallel, except for the groundwater reservoir (Fig. 7). The snowmelt process was considered in all landscapes using the same method as described in Sect. 4.1.1.

In the bare-soil/rock class, HOF (R_{HB}), caused by the comparatively low infiltration capacity compared to vegetation-covered soils on hillslopes, is controlled by a threshold parameter (P_t) (Eq. 22). HOF only occurs when the daily effective precipitation (P_{eB}) is larger than P_t:

$$R_{HB} = \max\left(P_{eB} - P_t,\ 0\right). \tag{22}$$

SOF (R_{SB}), caused by limited storage capacity of the rather shallow soils at high elevations, happens when the amount of water in the unsaturated reservoir exceeds the storage capacity (S_{uMaxB}). Deep percolation from bare soil/rock into groundwater (R_{pB}) is controlled by the relative soil moisture (S_{uB}/S_{uMaxB}) and maximum percolation (P_{ercB}):

$$R_{pB} = P_{ercB} \frac{S_{uB}}{S_{uMaxB}}. \tag{23}$$

The actual evaporation (E_{aB}) is estimated by potential evaporation (E_{pB}) and relative soil moisture (SuB/SuMaxB), which is the same as the calculation of R_{pB} by P_{ercB} and SuB/SuMaxB in Eq. (23). The generated surface runoff on the bare soil/rock is separated into the water re-infiltrating (R_{rB}) while flowing on the higher permeable debris slopes and the water directly routed to the channel (R_{ffB}) by a separator (D_B). As in FLEX^D, the lag times are characterized by different lengths in the individual components. The response process of the surface runoff is controlled by a linear reservoir.

The grassland and forest hillslopes have the same model structure as FLEX^L, due to their similar runoff-producing mechanisms, but are characterized by different parameter values for interception (I_{maxGH} and I_{maxFH}) and unsaturated zone processes (S_{uMaxGH}, S_{uMaxFH} and $\beta_{GH} \beta_{FH}$), reflecting different land cover and root zone depth. The lag times of grassland and forest hillslopes are the same as in the bare-soil/rock landscape elements. The generated runoff from grassland and forest hillslopes is represented by Q_{fGH} and Q_{fFH}. In the wetland/terrace landscape element, SOF
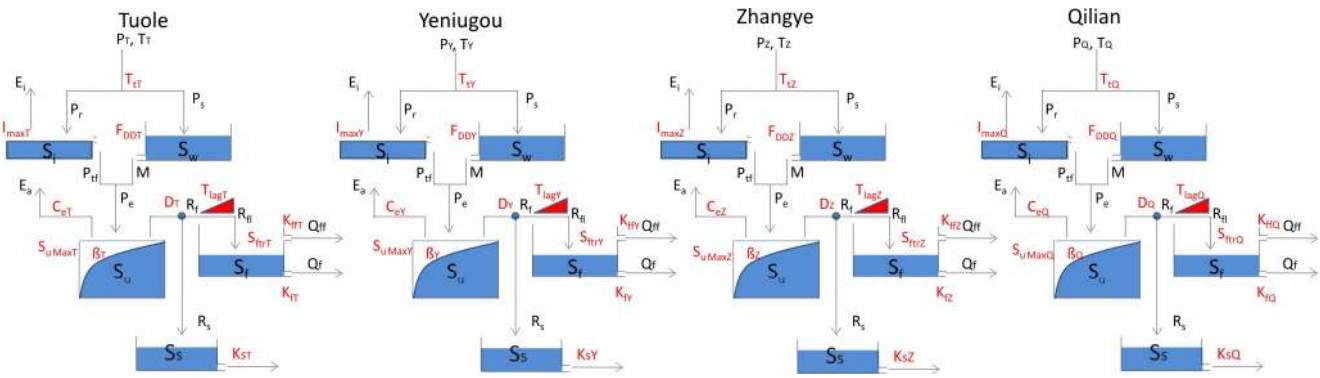
Figure 4. The semi-distributed model structure FLEX^D, consisting of four parallel components, representing one Thiessen polygon each. Each of the parallel components is characterized by an individual parameters set.
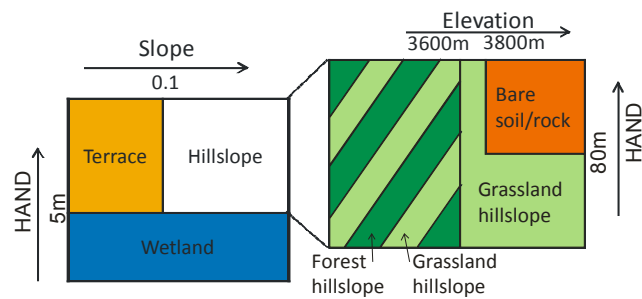

Figure 5. Summary of the landscapes classification criteria. The numbers (5 m, 80 m, 3600 m and 0.1) are the criteria for different landscape classes, e.g. < 5 m for wetland and > 5 m for terrace and hillslope.

(Q_{rW}) is conceptualized as the dominant hydrological process due to the shallow groundwater and resulting limited storage capacity. Additionally capillary rise (C_R) is represented by a parameter (C_{Rmax}) indicating a constant amount of capillary rise. The calculation method of effective rainfall and actual transpiration is the same as for grassland and forest hillslopes. The lag time of storm runoff in wetland is neglected due to the, on average, comparatively close distance of wetlands/terraces to the channel. The groundwater (Q_s) was assumed to be generated from one single aquifer in the catchment and represented by a lumped linear reservoir. In total FLEX^T requires 25 parameters. The final simulated runoff is equal to the sum of runoffs from all landscape elements according to their areal proportions (Fig. 7).

4.4 Model calibration

4.4.1 Objective functions

To allow for the model to adequately reproduce different aspects of the hydrological response, i.e. high flow, low flow and the flow duration curve, and thereby increase model realism, a multi-objective calibration strategy was adopted in this study, using the Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) of the hydrographs (I_{NS}) to evaluate the model performance during high flow, the NSE of the flow duration curve (I_{NSF}) to evaluate the simulated flow frequency and the NSE of the logarithmic flow (I_{NSL}) which emphasizes the lower part of the hydrograph.

4.4.2 Calibration method

The groundwater recession parameter (K_s) is not treated as a free calibration parameter but it was rather obtained directly from the observed hydrograph using a master recession curve approach (MRC) (Fenicia et al., 2006). Therefore, K_s was fixed at 90 (d) to avoid its interference with other processes. Together with fixing C_e this results in 10, 40 and 23 free calibration parameters for FLEX^L, FLEX^D and FLEX$^{T0/T}$, respectively.

The MOSCEM-UA (Multi-Objective Shuffled Complex Evolution Metropolis-University of Arizona) algorithm (Vrugt et al., 2003) was used as the calibration algorithm to find the Pareto-optimal fronts of the three objective functions. There are three parameters to be set for MOSCEM-UA: the maximum number of iterations, the number of complexes and the number of random samples that is used to initialize each complex. For the FLEX^L model the number of iterations was set to 50 000, the number of complexes to 10 and the number of random samples to 1000. To account for increase model complexity, these MOSCEM-UA parameters of the FLEX^D were set to 50 000, 40 and 3200; those of FLEX^T to 50 000, 23 and 2300. The uniform prior parameter distributions of FLEX^L and FLEX^D are listed in Table 5 and the ones of FLEX^{T0} and FLEX^T are given in Table 6.

4.4.3 Constraints on parameters and fluxes in FLEX^T

Guided by our perceptual understanding of the study catchment in the Section on "The landscapes and the perceptual model of the Upper Heihe" and the NDVI map (Fig. 6c), a set of realism constraints for model parameters and simulated
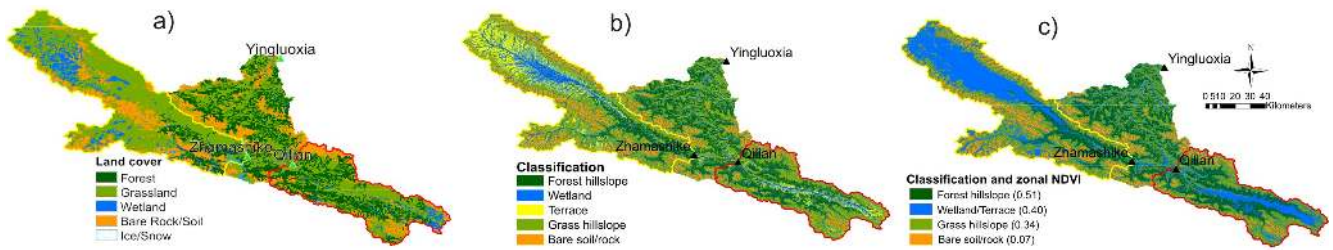
Figure 6. Comparison of land cover **(a)**, landscape classification maps **(b, c)** and the NDVI in each land cover **(c)**.
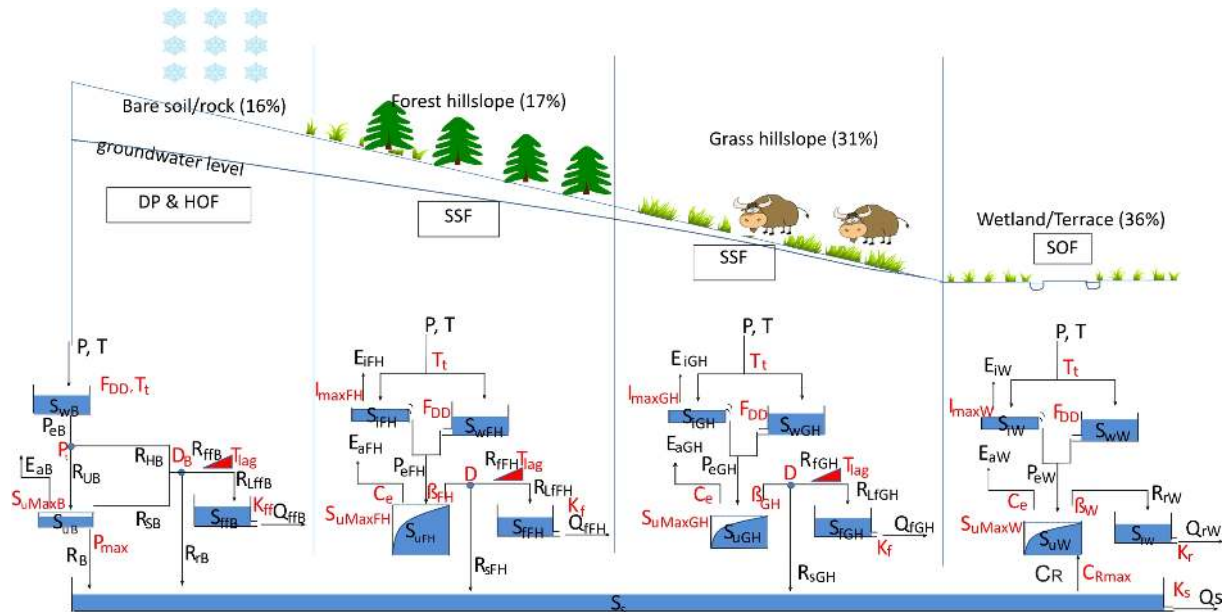

Figure 7. Perceptual model and parallel model structures of FLEX^T for the Upper Heihe.

Table 5. Uniform prior parameter distributions of the FLEX^L and FLEX^D models.

Parameter	Range	Parameters	Range
F_{DD} (mm/(d °C^{-1}))	(1, 8)	K_s (d)	90
T_t (°C)	(−2.5, 2.5)	S_{ftr} (mm)	(10, 200)
I_{max} (mm d^{-1})	(0.1, 5)	K_{ff} (d)	(1, 5)
S_{uMax} (mm)	(50, 1000)	K_f (d)	(1, 20)
β (−)	(0.1, 5)	T_{lag} (d)	(0, 5)
D (−)	(0, 1)	C_e (−)	0.5

fluxes was developed, similar to what was recently suggested by Gharari et al. (2013a). Parameter sets and model simulations that do not respect these constraints were regarded as non-behavioural parameters and rejected during calibration. The motivation is that by reducing unrealistic parameter combinations, the predictive uncertainty of a model may reduce, although the performance during calibration may be slightly decreased. More specifically, the parameters related to interception evaporation and transpiration were

constrained based on expert knowledge (Table 7). It was assumed that the interception threshold in the forest class (I_{maxFH}) needs to be larger than in the grassland (I_{maxGH}) and wetland/terrace classes (I_{maxW}) due to the increased interception capacity of forests. In addition, the root zone depth of forest hillslopes (S_{uMaxFH}) should be deeper than in grassland (S_{uMaxGH}). Furthermore, the root zone depth of wetland/terrace (S_{uMaxW}) and bare soil/rock (S_{uMaxB}) are assumed to be shallower than those of hillslopes. The timescale of groundwater (K_s) was assumed to be the highest due to its slow recession process, while the timescales of SOF in the wetland/terrace class (K_r) and the surface runoff in the bare-soil/rock hillslope class (K_{ff}) were defined to be the shortest, with the timescale of SSF (K_f) on forest and grassland hillslope assumed to be in-between. Additional soft performance constraints were introduced to avoid unreasonable trade-offs among fluxes in different landscapes (Table 7). It is assumed that the combined annual average evaporation and transpiration in forest ($E_{iFH} + E_{aFH}$ – the unit is mm a^{-1}) should be larger than in grassland ($E_{iGH} + E_{aGH}$) due to the higher vegetation cover in forests (see the NDVI map in Fig. 6c).

Table 6. Uniform prior parameter distributions of the FLEX^{T0} and FLEX^T model.

Parameters in all 3 models	Range	Parameters for bare soil/rock	Range	Parameter for forest hillslope	Range	Parameter for grass hillslope	Range	Parameter for wetland	Range
F_{DD} (mm/(d °C^{-1}))	(1, 8)	S_{uMaxB} (mm)	(5, 500)	I_{maxFH} (mm d^{-1})	(1, 10)	I_{maxGH} (mm d^{-1})	(0, 10)	I_{maxW} (mm d^{-1})	(0.1, 10)
T_t (°C)	(−2.5, 2.5)	P_{maxB} (mm d^{-1})	(0.1, 10)	S_{uMaxFH} (mm)	(100, 1000)	S_{uMaxGH} (mm)	(50, 1000)	S_{uMaxW} (mm)	(5, 1000)
P_t (mm d^{-1})	(5, 35)	D_B (−)	(0, 1)	β_{FH} (−)	(0.1, 5)	β_{GH} (−)	(0.1, 5)	β_W (−)	(0.1, 5)
K_f (d)	(1, 20)	K_{ff} (d)	(2, 50)					C_{Rmax} (mm d^{-1})	(0.01, 2)
K_s (d)	90							K_r (d)	(1, 9)
D (−)	(0, 1)								
T_{lagT} (d)	(0, 5)								
T_{lagY} (d)	(0, 5)								
T_{lagQ} (d)	(0, 5)								

Table 7. The soft data to constrain the automatic calibration.

Soft parameter constraint based on perceptual realism	Soft performance constraint based on NDVI map
$I_{maxFH} > I_{maxGH}$	$E_{iFH} + E_{aFH} > E_{iGH} + E_{aGH}$
$I_{maxFH} > I_{maxW}$	$E_{iW} + E_{aW} > E_{iGH} + E_{aGH}$
$S_{uMaxFH} > S_{uMaxGH} > S_{uMaxW}$	$E_{iGH} + E_{aGH} > E_{aB}$
$S_{uMaxFH} > S_{uMaxGH} > S_{uMaxB}$	$E_{aFH} > E_{aGH}$
$K_s > K_f > K_{ff}$	
$K_s > K_f > K_r$	

Similarly, the annual average evaporative fluxes from wetland/terrace ($E_{iW} + E_{aW}$) are assumed to be higher than from the grassland as the latter are more moisture constrained. The evaporative water loss from the bare-soil/rock class (E_{aB}) is the lowest, due to its sparse vegetation cover, limited near-surface storage and lowest temperatures. Furthermore, transpiration in forest (E_{aFH}) should be expected to be higher than in grassland (E_{aGH}) because more water is used for biomass production and deeper roots allow access to a larger pool of water.

4.5 Model evaluation

Model evaluation is usually limited to calibration followed by split-sample validation (Klemeš, 1986). Frequently, split-sample validation can result in satisfactory model performance as the model is trained by data from the same location in the preceding calibration period. On the basis of successful split-sample validation, models and their parameterizations are then often considered acceptable for predicting the rainfall-runoff response at the given study site. It has in the past, however, been observed that many models with adequate split-sample performance failed to reproduce hydrographs even in the nested sub-basins of the calibrated basin (e.g. Pokhrel and Gupta, 2011). In this study, we therefore applied the calibrated models of different complexity and degrees of input data distribution together with their calibrated parameter sets to two nested catchments to test the models' transferability and thus the ability to reproduce the hydrological response in catchments they have not explicitly been trained for. This kind of nested sub-catchment validation can, even if it is not an entirely independent validation in the sense of a proxy-basin test (Klemeš, 1986), give crucial information on the process realism and the related predictive power of a model. In this study the hydrological data at the main outfall Yingluoxia (1959–1968) were used for model calibration. Subsequently the model was tested by a split-sample test at the main outfall (1969–1978) and its two nested stations: Qilian (1967–1978) and Zhamashike (1959–1978).

5 Results

5.1 Results of FLEX^L and FLEX^D

Table 8, as well as Figs. 8a and 9a, illustrate that FLEX^L performed quite well in the calibration period with respect

Table 8. The averaged results of all the points on the Pareto front of three objective functions of the three models $FLEX^L$, $FLEX^D$, $FLEX^{T0}$ and $FLEX^T$ in calibration, split-sample and nested sub-catchments validation.

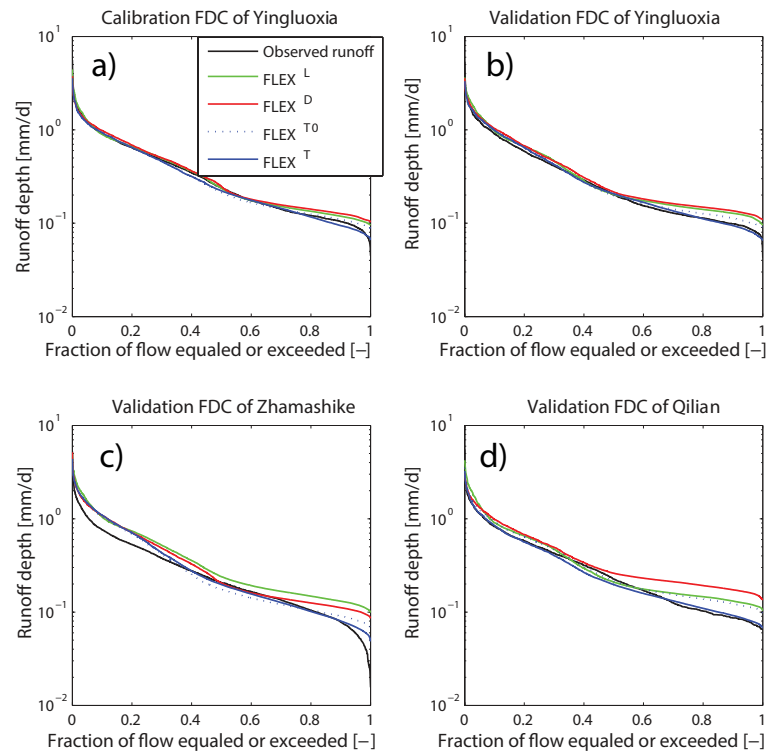|	Calibration			Split-sample validation			Zhamashike validation			Qilian validation		
	I_{NS}	I_{NSF}	I_{NSL}	I_{NS}	I_{NSF}	I_{NSL}	I_{NS}	I_{NSF}	I_{NSL}	I_{NS}	I_{NSF}	I_{NSL}
$FLEX^L$	0.82	0.99	0.87	0.79	0.95	0.78	0.54	0.79	0.56	0.56	0.87	0.59
$FLEX^D$	0.81	0.99	0.84	0.80	0.96	0.83	0.56	0.82	0.60	0.38	0.67	0.44
$FLEX^{T0}$	0.82	0.99	0.88	0.80	0.97	0.84	0.60	0.89	0.70	0.68	0.90	0.71
$FLEX^T$	0.80	0.98	0.84	0.78	0.95	0.82	0.65	0.92	0.74	0.71	0.96	0.75

Figure 8. Calibration **(a)** and validation results for the flow duration curve of four models ($FLEX^L$ (lumped model), $FLEX^D$ (semi-distributed model with different parameter sets for different Thiessen polygons), $FLEX^{T0}$ (FLEX-Topo model without constraints) and $FLEX^T$ (FLEX-Topo model with constraints)) for the entire Upper Heihe **(b)** and the two tested sub-catchments **(c, d)**. The curves make use of the average value of the parameter sets on the Pareto-optimal front.

to both the hydrograph and the flow duration curve (FDC). In spite of a somewhat reduced performance, the lumped model was also able to reproduce the major features of the catchment response in the split-sample validation (Table 8, Figs. 8b and 9b). Testing the model's potential in an uncalibrated part of the catchment as if the uncalibrated parts were ungauged basins (Sivapalan et al., 2003; Hrachowitz et al., 2013b), the performance of $FLEX^L$ was far from satisfactory (Table 8, Figs. 8c and d, 9c and d). The validation hydrographs in two tested sub-catchments, which are in the same period as the split-sample validation, are shown in Fig. 9c and d. One interesting observation is that the large

precipitation event at the end of the warm season in 1970 did not generate a flood peak in all three catchments, a characteristic that cannot be adequately reproduced by $FLEX^L$ (Fig. 9b–d). Similarly, the sub-catchment FDCs (Fig. 8b–d) indicate that $FLEX^L$, while in general mimicking the FDC of the entire catchment well, poorly represents the low flow characteristics of the two sub-catchments.

In the next step the influence of distributing the model forcing and complexity on the model performance was tested using the $FLEX^D$ model set-up (Fig. 4). It was found that although the modelled hydrograph and FDC reflect the observed response dynamics similarly well to $FLEX^L$ (Figs. 8a
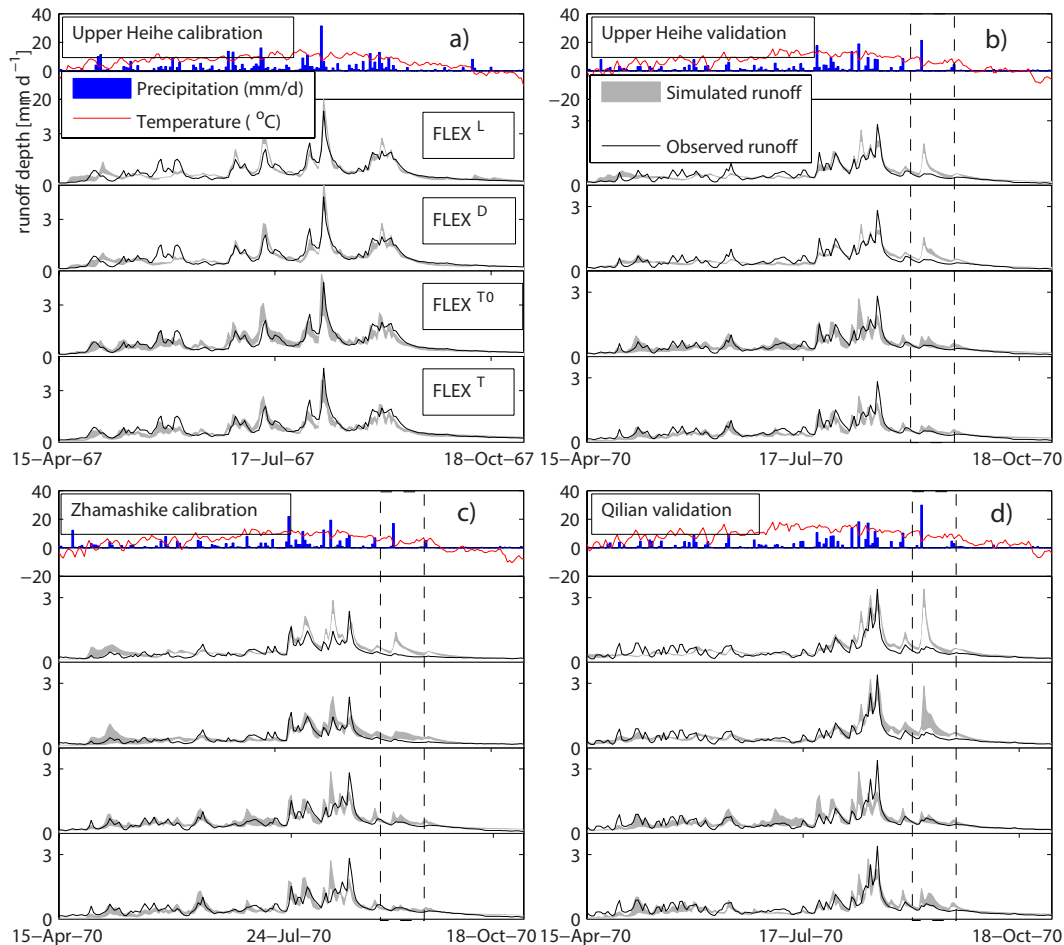
Figure 9. Comparison between observed values (black line) and the envelope of all modelled Pareto-optimal hydrographs (grey shaded area) of four models – FLEX^L (lumped model), FLEX^D (semi-distributed model with different parameter sets for different Thiessen polygons), FLEX^{T0} (FLEX-Topo model without constraints) and FLEX^T (FLEX-Topo model with constraints) – in the calibration period **(a)**, split-sample validation **(b)** and sub-catchments validation **(c, d)**. Precipitation (blue bars) and temperature (red line) are also shown. The dashed box indicates the September 1970 storm event.

and 9a), the objective functions indicate a slightly poorer calibration performance (Table 8), in spite of two additional parameters in each parallel model component accounting for differences in channel routing lag times. In this study, the reason is potentially related to the uncertainty in the precipitation–elevation relationship and the inappropriate soil moisture distribution dictated by the Thiessen polygons. In the split-sample and Qilian sub-catchment validation (Table 8, Fig. 8b and d), FLEX^D does not add value to the results of the lumped FLEX^L model which is mostly related to the adverse effects of increased equifinality. However, the validation results in Zhamashike improve with FLEX^D, with increased NSE values for both FDC and the hydrograph (Table 8), although base flow is still not reproduced well (Fig. 8c). This indicates that at least for the Zhamashike sub-catchment the distributed precipitation is more representative than catchment-averaged precipitation or that the results

are more sensitive to the heterogeneity of precipitation and temperature input than in the Qilian sub-catchment. In addition, hydrograph inspection revealed that the mismatch of observed and modelled runoff, generated by the large precipitation event at the end of the warm season, was not captured well by the FLEX^D either (Fig. 9b–d), although there was a slight improvement (see the reasons in Sect. 6.1). The results of FLEX^D indicate that increased model complexity alone, without deeper consideration of the underlying processes, does not result in a better model transferability.

5.2 Results of FLEX^{T0} and FLEX^T

From Table 8 and Figs. 8 and 9, it can be seen that the FLEX^{T0} set-up (i.e. no constraints) results in a similar performance to FLEX^L and FLEX^D in calibration and validation, but outperforms the latter two when tested in the two sub-catchments. After adding the constraints, we found that the

Table 9. The simulated results of FLEX^T.

Bare soil/rock (18 %)		Forest hillslope (17 %)		Grassland hillslope (31 %)		Wetland/terrace (36 %)	
P (mm a^{-1})	481	P (mm a^{-1})	431	P (mm a^{-1})	431	P (mm a^{-1})	410
E_{aB} (mm a^{-1})	174	E_{iFH} (mm a^{-1})	125	E_{iGH} (mm a^{-1})	58	E_{iW} (mm a^{-1})	87
R_{pB} (mm a^{-1})	115	E_{aFH} (mm a^{-1})	257	E_{aGH} (mm a^{-1})	205	E_{aW} (mm a^{-1})	220
R_{rB} (mm a^{-1})	107	R_{sFH} (mm a^{-1})	15	R_{sGH} (mm a^{-1})	63	C_R (mm a^{-1})	21
Q_{ffB} (mm a^{-1})	77	Q_{fFH} (mm a^{-1})	26	Q_{fGH} (mm a^{-1})	101	Q_{rW} (mm a^{-1})	122

performance of FLEX^T in calibration and split-sample validation, as expected (because of the reduced degrees of freedom) and as indicated by the objective functions (Table 8), is slightly lower than the performance of other models. However, in sub-catchment validation, the performance of FLEX^T is significantly better than the other models. This indicates that both the model structure and the constraints on parameters and fluxes in FLEX^T improve model transferability.

In Table 9, the fluxes modelled by FLEX^T, which are the average values of all the results obtained by the parameter sets on the Pareto-optimal fronts, of each landscape class for the entire catchment are given. The water balances of the individual landscape units illustrate clearly the distinct dominant hydrological functions of these individual units as a priori defined by the modeller's perception of the system. Specifically, the precipitation on bare soil/rock is 481 mm a^{-1} (18 % proportion of the entire catchment precipitation), 174 mm a^{-1} (6 %) evaporates and 115 mm a^{-1} (4 %) infiltrates into cracks and eventually percolates to the groundwater. Overland flow produces 74 mm a^{-1} (3 %), while 112 mm a^{-1} (4 %) is generated as subsurface flow on shallow soil. A total of 107 mm a^{-1} (4 %) of the locally generated overland flows re-infiltrates into groundwater (107 mm a^{-1} (4 %)) due to the high permeability of debris slopes at the foot of the mountains while the remaining water (77 mm a^{-1} (3 %)) is routed to the stream network with considerable sediment loads. Precipitation on the forest hillslopes is 431 mm a^{-1} (17 %), 125 mm a^{-1} (5 %) of which is intercepted by and evaporates from canopy and forest floor; 257 mm a^{-1} (10 %) is modelled as transpiration, while only 26 mm a^{-1} (1 %) and 15 mm a^{-1} (0.6 %), respectively, contribute to fast runoff and groundwater recharge, highlighting the dominant evaporation function of this landscape class. The precipitation on the grassland hillslopes is 431 mm a (31 %), of which 58 mm a^{-1} (4 %) is intercepted, 205 mm a^{-1} (15 %) is transpired, 101 mm a^{-1} (7 %) generates fast runoff and 63 mm a^{-1} (5 %) recharges the groundwater. For the wetland/terrace, precipitation is 410 mm a^{-1} (34 %) and in addition around 21 mm a^{-1} (1.7 %) is contributed from groundwater as capillary rise. 87 mm a^{-1} (7 %) is intercepted; 220 mm a^{-1} (19 %) is consumed by transpiration and 122 mm a^{-1} (10 %) contributes to the fast runoff. These results underline the importance of wetlands and terraces for peak flow generation. Finally, the

groundwater discharge is 51 mm a^{-1} (12 %), which accounts for 37 % of the total runoff. In total, the modelled runoff depth is 143 mm a^{-1}, which is close to the observed runoff (141 mm a^{-1}). For the simulated total evaporation, it is interesting to find that the ratio between forest and wetland is 1.24, which is close to their NDVI ratio (1.27). Similarly, the ratio between water loss to atmosphere in wetland and grassland is 1.21, which is also close to their NDVI ratio (1.20). These results support the hypothesis that allowing for hydrologically meaningful process heterogeneity in models while imposing realism constraints can produce flux dynamics that are adequate reflections of reality, increasing our confidence that these models give "the right answers for the right reasons" (Kirchner, 2006).

Figure 8 shows significant improvement in the reproduction of the FDCs by FLEX^{T0} and FLEX^T. FDCs are an emergent catchment property (cf. Sivapalan, 2006) and therefore a critical characteristic of system function which a model should be capable to reproduce (Westerberg et al., 2011). In particular, the increased skill of FLEX^T to reproduce the FDCs indicates that the flexibility introduced by considering changing proportions of hydrologically distinct landscape elements for different catchments and applying realism constraints in FLEX^T has the potential to substantially increase the predictive power for flow characteristics, especially the low flows (Fig. 8, Table 8). The importance of the expert-knowledge-based constraints imposed on FLEX^T is illustrated by its increased ability to reproduce the FDCs as compared to FLEX^{T0}. The importance of accounting for an adequate level of hydrologically meaningful process heterogeneity is further illustrated by Fig. 9, which shows the envelopes of observed and modelled hydrographs (based on all the parameter sets on the Pareto-optimal front) of three model structures, in calibration, split-sample validation and sub-catchment validation. The intense precipitation event in September 1970, which was observed at all four precipitation gauges and thus not a local event, did not generate a significant peak at any of the flow-gauging stations in this study. However, neither FLEX^L nor FLEX^D were able to accommodate this lack of response. Only the increased process heterogeneity in the FLEX^{T0} and FLEX^T models allowed an adequate representation of this event. For a closer analysis, the modelled hydrograph components from the different landscape elements are shown in Fig. 10. It can be
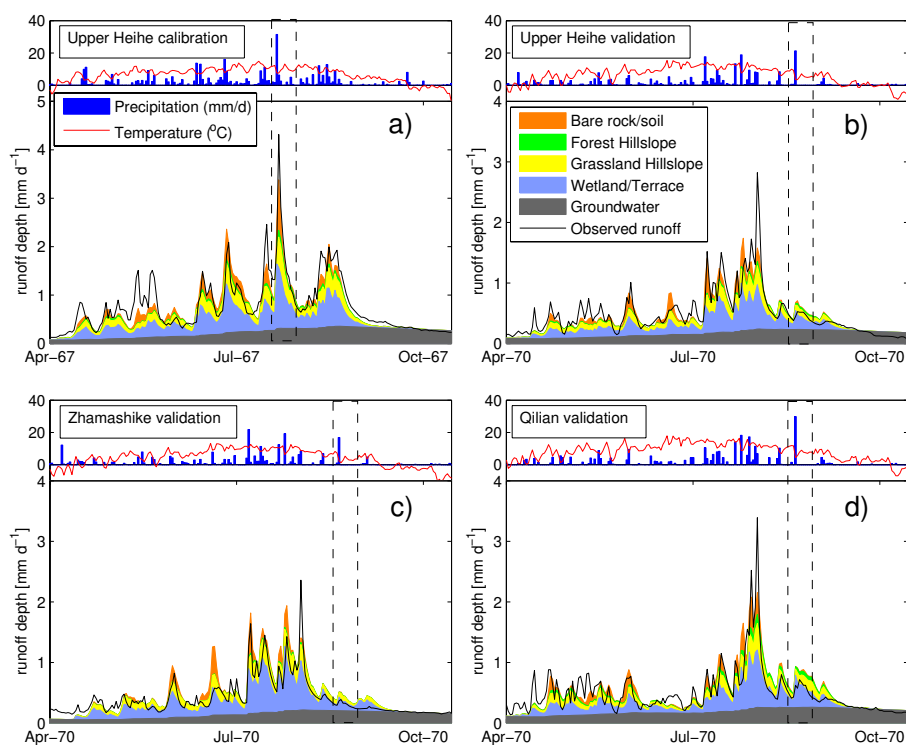
Figure 10. The hydrograph components of the calibration **(a)**, split-sample validation **(b)** and sub-catchments validation **(c, d)**, of the FLEX^T model (using the average value of the parameter sets on the Pareto-optimal front).

seen that the response to storm events is generally dominated by the flows generated in wetlands/terraces. Connectivity of hillslopes and bare-soil/rock landscapes however, is typically established with some delay and the magnitude of their contributions to stream flow, in particular during relatively dry periods, are significantly lower than that of wetlands/terraces (Fig. 10a). This is well illustrated for the September 1970 event shown in the dashed box in Fig. 10b–d: while to some degree all the landscapes eventually contributed to the runoff, the wetland/terrace responded directly to the storm whereas the limited response of hillslopes and bare soil/rock contributed to the peak flow later.

6 Discussion

6.1 Why did FLEX^T perform better than FLEX^L and FLEX^D?

Some clarification can be achieved by comparing the observed precipitation duration curves (PDC) and FDCs. From Fig. 11a it can be concluded that the entire Upper Heihe receives the lowest catchment-average precipitation input both in the original forcing data and the elevation-corrected precipitation, while being characterized by the largest runoff yield (Fig. 11b). The Qilian sub-catchment, in contrast, receives the largest amount of precipitation, but with lower runoff yield. The Zhamashike sub-catchment is characterized

by similar precipitation input as the entire Upper Heihe, but exhibits much higher peak flows and lower base flows than both the entire catchment and the Qilian sub-catchment. These distinct catchment hydrological functions are difficult to reconcile in one lumped model, representing a specific rainfall-runoff relationship. Moving to a different catchment or maybe only even to a sub-catchment is likely to change the relative proportions of landscapes, thus leading to a misrepresentation of the lumped process heterogeneity and thus reduced model performance in the new catchment. A semidistributed approach like FLEX^T, in contrast to FLEX^L and FLEX^D, offers more flexibility in adapting the model to the ensemble of processes in a more realistic way other than the lumped ones trained by adjusting it to the hydrograph, which most likely oversimplifies the catchment heterogeneity. This underlines the increased importance and benefit of more detailed, yet flexible expert-knowledge-guided process representations compared to focusing on mere parameter calibration of lumped models.

The potential reason for overestimating the runoff in the Zhamashike sub-catchment for FLEX^L and FLEX^D (Fig. 8c) is that these two models do not adequately represent the increased importance of evaporation from wetland/terrace. Similarly, the reason for overestimating flow in the Qilian sub-catchment (Fig. 8d) is that these two models cannot accommodate the increased evaporation of forests as much of the Upper Heihe, for which the models were calibrated, is
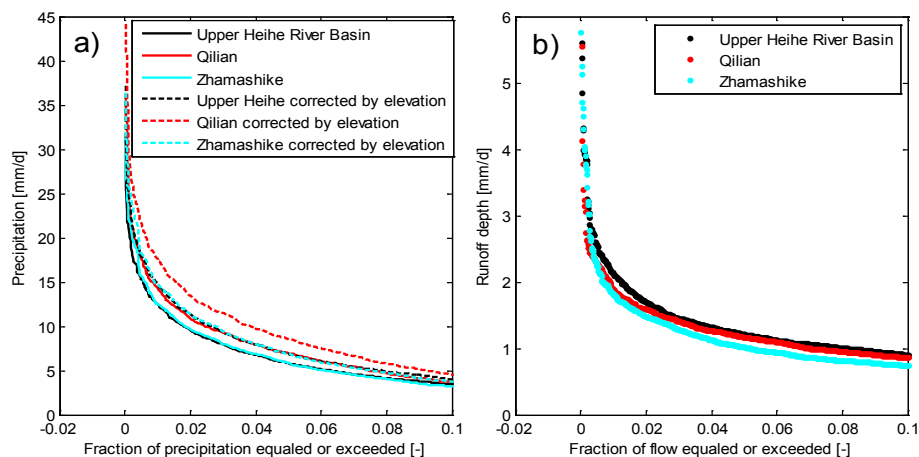
Figure 11. The comparison of three observed precipitation duration curves (**a**) and flow duration curves (**b**).

covered by grassland hillslope and bare soil/rock, characterized by lower evaporation rates than the other landscapes. On the other hand, both $FLEX^L$ and $FLEX^D$ overestimate the baseflow (Fig. 8c and d). This can potentially be linked to neglecting capillary rise in the wetland/terrace, which influences both the baseflow and the evaporation of this landscape element. When capillary rise in $FLEX^T$ is considered, the groundwater feeds the unsaturated reservoir in the wetland, which not only reduces the base flow but also increases the amount of water available for transpiration and eventually evaporation. This hydrological process is especially important in the Zhamashike sub-basin, where higher peak flows and lower base flow happen simultaneously.

The results support the potential of $FLEX^T$ and its parameterization to be spatially and scalewise better transferable than lumped model structures, such as $FLEX^L$ or semi-distributed models such as $FLEX^D$, which do not explicitly allow for changing proportions of landscape units with distinct hydrological function (see Sect. 5.2). In summary, the $FLEX^T$ model set-up, informed by topography, divided the catchments into four topographic subunits, representing different dominant hydrological process ensembles. This kind of modelling strategy allowed enough flexibility to capture the different functional behaviours of the three study catchments simultaneously (Table 8, Figs. 9–11).

6.1.1 Specific rainfall/snowfall-runoff events

Some modelled events, such as the rainfall/snowfall–runoff event in Fig. 10a, also illustrate that $FLEX^T$ is generating internal flow dynamics that better reflect the modeller's perception of the catchment processes. It can be seen that $FLEX^T$ could reproduce the instantaneous response of the wetlands to the storm and the delayed and limited response of other landscapes. As discussed above, another interesting event was observed in September 1970. During that storm the highest daily precipitation of the available record was

observed, while, however, producing only a relatively insignificant runoff peak, both in the Upper Heihe and its subcatchments (Fig. 9). Both $FLEX^L$ and $FLEX^D$ failed to adequately reproduce this event and modelled a much larger peak flow. Significant precipitation measurement error can be excluded, as the event was observed in similar magnitudes at all gauges available for this study. The failure of $FLEX^L$ and $FLEX^D$ to adequately respond to this event can be linked to several reasons. Firstly, the lumped accounting of the snowmelt in $FLEX^L$ can partly be the reason because the lumped model does not consider the change of temperature and then the type of precipitation with elevation. The lumped model treats the precipitation as rainfall in the entire catchment when the average daily air temperature is above the rain/snow temperature threshold. However, there could be snowfall in high elevation zones, when the catchment average temperature is slightly above the threshold temperature. Likewise, there could be rainfall in lower elevation zones when the catchment average temperature is below the threshold. The temperature record (Fig. 9) clearly shows the low average air temperature on the same day as the large precipitation event. This could partly explain the limited runoff response to this specific storm event, as the modelled results obtained from $FLEX^D$ are somewhat closer to the observed response than the results of $FLEX^L$ (Fig. 9b and d).

However, $FLEX^D$ could not mimic the event in a satisfactory way and hence the failure of $FLEX^D$ to adequately represent the event can be attributed to an oversimplified model structure. This is supported by the results of $FLEX^T$. From Fig. 10 it can be seen that the modelled flow generated by this storm event mostly originated in the wetlands, and a smaller proportion originated in grassland hillslopes. Contribution from forest hillslopes and bare-soil/rock hillslopes is negligible. The catchment average temperature on that day is close to the threshold temperature (Fig. 10). Thus, at lower elevations, which are mainly characterized by wetland/terrace, grassland and forest hillslopes,

the precipitation was in the form of rainfall. The temperature and precipitation records show that the preceding days were dry and warm (Fig. 10), translating into comparatively elevated evaporation and, linked to that, relatively high soil moisture deficits. In addition, the deep root zone on the forest hillslopes provides considerable storage capacity in the soil before discharge is generated. At a higher elevation, which is mainly characterized by bare soil/rock and grassland hillslopes, the precipitation was in a solid state, subsequently stored as snowpack. When the temperature increased again in the following days, the snow melted gradually. However, due to the slow melt rates and the dry antecedent conditions, the snowmelt water was almost completely infiltrated into the groundwater and did not contribute to the storm flow, even when the temperature increased several days later (Fig. 10b–d). Therefore, there is apparently enough information, not only in landscapes but also in the observed discharge, to parameterize the FLEX^T model. In summary, FLEX^T allowed the low and high elevation areas to reduce the storm flow for this specific event by different mechanisms, resulting in a very limited response to the event, in close agreement with the observed response.

6.1.2 Realism testing of FLEX^T model

The FLEX^T model is based on landscape classification, which is an observable prior to enhance model realism. Subsequently, based on our knowledge and understanding of different dominant hydrological processes in different landscapes, we assigned suitable hydrological process representations to these landscapes to highlight landscape heterogeneity. Significant differences in hydrological function, for example between wetland and hillslope, are well documented by a wide range of experimental studies (e.g. McGlynn and McDonnell, 2003; Seibert et al., 2003; Molenat et al., 2008; Jencso et al., 2009; Detty and McGuire, 2010). The consideration of landscape heterogeneity, reflected in runoff generation mechanisms and differences in water budgets, makes the FLEX^T model perform better than FLEX^L and FLEX^D, reproducing hydrographs (Fig. 9) but also FDCs (Fig. 8), as emergent catchment properties. The suggested model hypotheses were not only tested for temporal transferability between calibration and validation periods, but also for spatial transferability to sub-catchments. This successful transferability (Table 8; Figs. 8 and 9) is strong supporting evidence for the hypothesis that landscape carries crucial information on hydrological function and it illustrates that, by allowing for increased process heterogeneity in models, their degree of realism, i.e. their skill to adequately represent critical features in response dynamics, increases. More convincingly, the benchmark model (FLEX^D) with a higher number of parameters (40 free parameters) did not improve transferability, even with nearly twice the number of parameters as in FLEX^T (23 free parameters). However, note that here "realism" is not primarily and necessarily linked to an improved fit of the hydrograph. It is rather linked to the incorporation of more knowledge available from observation and experiments, thereby allowing the development of models whose internal and external dynamics correspond with this information, which in turn increases model realism.

Note that this paper aims to test the realism of models, meaning that it investigates to which degree a more realistic representation has been achieved, but it does not intend to claim that the developed model is realistic in absolute terms. In that respect it emphasizes that model realism should always be seen in relation to uncertainties arising from data error as well as from the choice of constitutive functions and parameters. In this paper, a detailed analysis of the influence of these uncertainties was omitted as this was beyond the scope of this research. Further, some of the hypotheses of the proposed model structure could not be tested individually for lack of available data. They will be further investigated during future research, using additional information. Thirdly, at this point, the interpretation is only valid in the study catchment and subsequent studies will have to test this hypothesis further in other regions in order to evaluate the generality of these findings.

6.2 Translating topography information into hydrological models

It is intriguing to find that the landscape classification based on topography information (HAND, slope, elevation and aspect) closely reflects the patterns and shapes of the land cover map (Fig. 6a and b). In other words, it clearly illustrates that topography has great influence on the energy and water availability and the evolution of vegetation cover. Certain types of vegetation cover evolve under specific topographic conditions. Elevation greatly influences the amount of precipitation and available energy. HAND and slope are two important factors defining water retention and drainage. Aspect influences the energy balance and precipitation. Normally, the south-facing hillslopes receive more solar energy. Thus, the potential evaporation on the south-facing hillslopes is larger than on north-facing ones, while aspect influences the distribution of forest and grassland in arid/semi-arid regions.

Topography does not only directly influence the groundwater level and the occurrence of saturation overland flow, but it also controls the soil and vegetation cover in certain geological and climatic condition, and consequently the dominant hydrological processes (Savenije, 2010). The presented modelling approach can therefore be seen as a step towards making more efficient use of topographic information for use in conceptual hydrological models. The successfully linked topographic information, land cover classification and hydrological model structure supports the hypothesis that topographic information can be used to distinguish landscape elements with different hydrological functions (Wagener et al., 2007; Savenije, 2010).

6.3 The value of soft data in hydrological modelling ($FLEX^{T0}$ vs. $FLEX^{T}$)

Hydrological modelling should also be seen as an art (Savenije, 2009). To ensure that models better reflect our understanding of reality we should make use of our experience and creativity. In addition to available data, hydrologists often have extensive, yet sometimes only semi-quantitative expert knowledge about specific study sites. However, this "soft" knowledge is with some exceptions (e.g. .Seibert and McDonnell, 2002, 2013), regularly underexploited in hydrological modelling. In general four types of soft data can be valuable for hydrological modelling. The first one is our explicit or inferred knowledge of the hydrological processes occurring in reality. For example, in this study, streams in high elevation tributaries, characterized by a dominance of relatively erodible bare soil/rock, exhibit relatively high levels of turbidity (Fig. 2e), thus indicating the importance of soil erosion, which in turn supports the existence of Hortonian surface runoff in these locations. Another type of "soft" data is the expert knowledge on meaningful acceptable prior parameter ranges, such as the maximum storage of the unsaturated reservoir at the catchment scale (S_{uMax}), which is closely linked to rooting depth and soil structure and strongly depends on the ecosystem. The third kind of valuable "soft" data is the understanding of the relative magnitude of specific parameters in different landscapes (Gharari et al., 2013a), providing further constraints on model parameters, and eliminating unrealistic parameter combinations. For example, in this study it was argued that forest canopy, undergrowth and litter on forested hillslopes can intercept more precipitation (I_{maxFH}) than grass-dominated hillslopes (I_{maxGH}) (Table 7). Fourthly, simulation results can be constrained by "soft" data, such as NDVI maps indicating inequalities between forest and grassland transpiration (Table 7). Making use of these four types of soft data, a landscape driven model, $FLEX^{T}$, based on our understanding of the hydrological processes in the Upper Heihe, was developed and constrained. Although the use of these additional constraints resulted in a slightly reduced calibration performance of $FLEX^{T}$ as compared to the $FLEX^{T0}$ set-up, the more successful sub-catchments validation illustrated the value of soft data and clearly indicated that the efficient use of soft data allows for a more realistic representation of catchment heterogeneity, leading to higher predictive power.

6.4 The role of forest in the Upper Heihe

Since forest is an important land cover in the Upper Heihe and many other catchments, the hydrological impact of forest is essential for understanding the catchment water cycle (Andréassian et al., 2004; Lyon et al., 2012), but also for an efficient implementation of water resource management policies. The role of forest on the catchment scale is subject of ongoing discussion in ecohydrology (Moore and Wondzell, 2005; Sriwongsitanon and Taesombat, 2011).

Various earlier studies found very diverse conclusions (Bosch and Hewlett, 1982; Robinson et al., 1991; Sahin and Hall, 1996; Andréassian, 2004; Moore and Wondzell, 2005). In this study, the $FLEX^{T}$ model generated little runoff in forested hillslopes in the Upper Heihe, with most of the rainfall on the forest being intercepted and transpired. In addition, these results are supported by other studies in this catchment based on remote sensing information (Tian et al., 2013), statistical analysis (Wang et al., 2011), paired catchment analysis in this region (Huang et al., 2003; Qin et al., 2011) and the simulation of an ecohydrological model (Yu et al., 2009). Also, field observations and experimental studies in the Upper Heihe (B. Ye, personal communication, 2012) gave evidence of limited runoff from forests. This phenomenon is most likely linked to the climatic conditions in the Upper Heihe. Since the precipitation in this region reaches on average only $430\,\mathrm{mm\,a^{-1}}$, with a maximum observed daily catchment average precipitation of below $45\,\mathrm{mm\,d^{-1}}$ (corrected by elevation), with ample storage available in the root zone, the forest hillslopes in the Upper Heihe remain largely below the moisture content necessary to establish connectivity conditions necessary to significantly contribute to storm flow.

7 Conclusions

We compared four model structures on the Upper Heihe in China: a lumped model ($FLEX^{L}$), a semi-distributed model with different parameter sets for different Thiessen polygons ($FLEX^{D}$), a conceptual model whose model structure is determined based on hypotheses of how topography influences hydrological processes ($FLEX^{T0}$), and $FLEX^{T}$, which is based on $FLEX^{T0}$ but constrained by soft data. $FLEX^{T0}$ and $FLEX^{T}$ perform in almost the same way as the two previous models in calibration and split-sample validation, but much better when validated in two sub-catchments. The increased performance of $FLEX^{T0}$ and $FLEX^{T}$ with respect to $FLEX^{L}$ and $FLEX^{D}$ in the sub-catchments, in particular with respect to the flow duration curves and for some specific events, indicates the following: (1) the topography-driven model, using landscapes classification as prior information and describing different dominant hydrological mechanisms in different landscapes, reflects the catchment heterogeneity in a more realistic way; (2) the natural land cover may be identified by topographic information to some extent, because the topography greatly influences the local energy and water budget; (3) the better performance of $FLEX^{T}$ compared to $FLEX^{T0}$ indicates further that the use of realism constraints guided by soft data reduced unrealistic compensation between fluxes and increased model transferability. In summary, making use of topography-derived data as

prior information to guide model development is a promising avenue.

References

Abbott, M. B. and Refsgaard, J. C.: Distributed hydrological modelling, Springer, Dordrecht, the Netherlands, 1996.

Andréassian, V.: Waters and forests: From historical controversy to scientific debate, J. Hydrol., 291, 1–27, doi:10.1016/j.jhydrol.2003.12.015, 2004.

Andréassian, V., Oddos, A., Michel, C., Anctil, F., Perrin, C., and Loumagne, C.: Impact of spatial aggregation of inputs and parameters on the efficiency of rainfall-runoff models: A theoretical study using chimera watersheds, Water Resour. Res., 40, W05209, doi:10.1029/2003wr002854, 2004.

Beven, K. J. and Germann, P.: Macropores and water flow in soils revisited, Water Resour. Res., 49, 3071–3092, doi:10.1002/wrcr.20156, 2013.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrolog. Sci.-Bull., 24, 43–69, 1979.

Bosch, J. M. and Hewlett, J. D.: A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration, J. Hydrol., 55, 3–23, doi:10.1016/0022-1694(82)90117-2, 1982.

Chen, R. S., Kang, E., Yang, J. P., and Zhang, J. S.: Application of topmodel to simulate runoff from heihe mainstream mountainous basin, J. Desert Res., 23, 428–434, 2003.

Detty, J. and McGuire, K.: Topographic controls on shallow groundwater dynamics: Implications of hydrologic connectivity between hillslopes and riparian zones in a till mantled catchment, Hydrol. Process., 24, 2222–2236, 2010.

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S., and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, Hydrol. Earth Syst. Sci., 17, 1893–1912, doi:10.5194/hess-17-1893-2013, 2013.

Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: Is the groundwater reservoir linear? Learning from data in hydrological modelling, Hydrol. Earth Syst. Sci., 10, 139–150, doi:10.5194/hess-10-139-2006, 2006.

Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: Understanding catchment behavior through stepwise model concept improvement, Water Resour. Res., 44, W01402, doi:10.1029/2006wr005563, 2008.

Fenicia, F., Kavetski, D., and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour. Res., 47, W11510, doi:10.1029/2010wr010174, 2011.

Gao, H., He, X., Ye, B., and Pu, J.: Modeling the runoff and glacier mass balance in a small watershed on the central tibetan plateau, china, from 1955 to 2008, Hydrol. Process., 26, 1593–1603, doi:10.1002/hyp.8256, 2012.

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

Gharari, S., Shafiei, M., Hrachowitz, M., Fenicia, F., Gupta, H. V., and Savenije, H. H. G.: A strategy for "constraint-based" parameter specification for environmental models, Hydrol. Earth Syst. Sci. Discuss., 10, 14857–14871, doi:10.5194/hessd-10-14857-2013, 2013a.

Gharari, S., Hrachowitz, M., Fenicia, F., and Savenije, H. H. G.: An approach to identify time consistent model parameters: sub-period calibration, Hydrol. Earth Syst. Sci., 17, 149–161, doi:10.5194/hess-17-149-2013, 2013b.

Grayson, R. and Blöschl, G.: Spatial patterns in catchment hydrology: Observations and modelling, Cambridge University Press, 2001.

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, Hydrol. Process., 22, 3802–3813, doi:10.1002/hyp.6989, 2008.

Hamon, W. R.: Estimating potential evapotranspiration, J. Hydraul. Div.-ASCE, 87, 107–120, 1961.

Hrachowitz, M., Savenije, H., Bogaard, T. A., Tetzlaff, D., and Soulsby, C.: What can flux tracking teach us about water age distribution patterns and their temporal dynamics?, Hydrol. Earth Syst. Sci., 17, 533–564, doi:10.5194/hess-17-533-2013, 2013a.

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

Huang, M., Zhang, L., and Gallichand, J.: Runoff responses to afforestation in a watershed of the loess plateau, china, Hydrol. Process., 17, 2599–2609, 2003.

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E., and Marshall, L. A.: Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale, Water Resour. Res., 45, W04428, doi:10.1029/2008wr007225, 2009.

Jia, Y., Ding, X., Qin, C., and Wang, H.: Distributed modeling of landsurface water and energy budgets in the inland Heihe river basin of China, Hydrol. Earth Syst. Sci., 13, 1849–1866, doi:10.5194/hess-13-1849-2009, 2009.

Jin, X., Wan, L., and Hu, G.: Distribution characteristics of mountain vegetation and the influence factors in upstream of heihe river basin, J. Arid Land Resour. Environ., 22, 140–144, 2008.

Kang, E., Cheng, G., Lan, Y., Chen, R. S., and Zhang, J.: Application of a conceptual hydrological model in the runoff forecast of a mountainous watershed, Adv. Earth Sci., 17, 18–26, 2002.

Kavetski, D. and Kuczera, G.: Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration, Water Resour. Res., 43, W03411, doi:10.1029/2006wr005195, 2007.

Khu, S.-T., Madsen, H., and di Pierro, F.: Incorporating multiple observations for distributed hydrologic model calibration: An approach using a multi-objective evolutionary algorithm and clustering, Adv. Water Resour., 31, 1387–1398, 2008.

Klemeš, V.: Dilettantism in hydrology: Transition or destiny?, Water Resour. Res., 22, 177S–188S, doi:10.1029/WR022i09Sp0177S, 1986.

Klemeš, V.: The modelling of mountain hydrology: The ultimate challenge, IAHS report, IAHS, Wallingford, Oxfordshire, 1990.

Knudsen, J., Thomsen, A., and Refsgaard, J. C.: Watbal a semi-distributed, physically based hydrological modelling system, Nord. Hydrol., 17, 347–362, 1986.

Li, X., Li, X., Li, Z., Ma, M., Wang, J., Xiao, Q., Liu, Q., Che, T., Chen, E., Yan, G., Hu, Z., Zhang, L., Chu, R., Su, P., Liu, Q., Liu, S., Wang, J., Niu, Z., Chen, Y., Jin, R., Wang, W., Ran, Y., Xin, X., and Ren, H.: Watershed allied telemetry experimental research, J. Geophys. Res., 114, D22103, doi:10.1029/2008jd011590, 2009.

Li, Z., Xu, Z., and Li, Z.: Performance of wasmod and swat on hydrological simulation in yingluoxia watershed in northwest of china, Hydrol. Process., 25, 2001–2008, doi:10.1002/hyp.7944, 2011.

Lyon, S. W., Nathanson, M., Spans, A., Grabs, T., Laudon, H., Temnerud, J., Bishop, K. H., and Seibert, J.: Specific discharge variability in a boreal landscape, Water Resour. Res., 48, W08506, doi:10.1029/2011wr011073, 2012.

McGlynn, B. L. and McDonnell, J. J.: Quantifying the relative contributions of riparian and hillslope zones to catchment runoff, Water Resour. Res., 39, 1310, doi:10.1029/2003wr002091, 2003.

Molenat, J., Gascuel-Odoux, C., Ruiz, L., and Gruau, G.: Role of water table dynamics on stream nitrate export and concentration in agricultural headwater catchment (france), J. Hydrol., 348, 363–378, doi:10.1016/j.jhydrol.2007.10.005, 2008.

Moore, R. D. and Wondzell, S. M.: Physical hydrology and the effects of forest harvesting in the pacific northwest: A review, J. Am. Water Resour. Assoc., 41, 763–784, 2005.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part i – a discussion of principles, J. Hydrol., 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Nobre, A., Cuartas, L., Hodnett, M., Rennó, C., Rodrigues, G., Silveira, A., Waterloo, M., and Saleska, S.: Height above the nearest drainage – a hydrologically relevant new terrain model, J. Hydrol., 404, 13–29, 2011.

Pokhrel, P. and Gupta, H. V.: On the ability to infer spatial catchment variability using streamflow hydrographs, Water Resour. Res., 47, W08534, doi:10.1029/2010wr009873, 2011.

Qin, J., Ding, Y., Ye, B., Zhou, Z., and Xie, Z.: Regulating effect of mountain landscapes on river runoff in northwest china, J. Glaciol. Geocryol., 33, 397–404, 2011.

Refsgaard, J. C. and Knudsen, J.: Operational validation and inter-comparison of different types of hydrological models, Water Resour. Res., 32, 2189–2202, doi:10.1029/96wr00896, 1996.

Rennó, C., Nobre, A., Cuartas, L., Soares, J., Hodnett, M., Tomasella, J., and Waterloo, M.: Hand, a new terrain descriptor using srtm-dem: Mapping terra-firme rainforest environments in amazonia, Remote Sens. Environ., 112, 3469–3481, doi:10.1016/j.rse.2008.03.018, 2008.

Robinson, M., Gannon, B., and Schuch, M.: A comparison of the hydrology of moorland under natural conditions, agricultural use and forestry, Hydrolog. Sci. J., 36, 565–577, 1991.

Sahin, V. and Hall, M. J.: The effects of afforestation and deforestation on water yields, J. Hydrol., 178, 293–309, doi:10.1016/0022-1694(95)02825-0, 1996.

Savenije, H. H. G.: Determination of evaporation from a catchment water balance at a monthly time scale, Hydrol. Earth Syst. Sci., 1, 93–100, doi:10.5194/hess-1-93-1997, 1997.

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

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.

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

Seibert, J.: Estimation of parameter uncertainty in the hbv model, Nord. Hydrol., 28, 247–262, 1997.

Seibert, J. and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, Water Resour. Res., 38, 1241, doi:10.1029/2001wr000978, 2002.

Seibert, J. and McDonnell, J.: Gauging the ungauged basin: The relative value of soft and hard data, J. Hydrol. Eng., doi:10.1061/(ASCE)HE.1943-5584.0000861, in press, 2013.

Seibert, J., Bishop, K., Rodhe, A., and McDonnell, J. J.: Groundwater dynamics along a hillslope: A test of the steady state hypothesis, Water Resour. Res., 39, 1014, doi:10.1029/2002wr001404, 2003.

Sivapalan, M.: Pattern, process and function: Elements of a unified theory of hydrology at the catchment scale, in: Encyclopedia of hydrological sciences, John Wiley & Sons, Ltd, Hoboken, New Jersey, 2006.

Sivapalan, M.: The secret to 'doing better hydrological science': Change the question!, Hydrol.l Process., 23, 1391–1396, doi:10.1002/hyp.7242, 2009.

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., and Zehe, E.: Iahs decade on predictions in ungauged basins (pub), 2003–2012: Shaping an exciting future for the hydrological sciences, Hydrological Sciences Journal – des Sciences Hydrologiques, 48, 857–880, 2003.

Spence, C. and Woo, M.-k.: Hydrology of subarctic canadian shield: Soil-filled valleys, J. Hydrol., 279, 151–166, doi:10.1016/S0022-1694(03)00175-6, 2003.

Sriwongsitanon, N. and Taesombat, W.: Effects of land cover on runoff coefficient, J. Hydrol., 410, 226–238, doi:10.1016/j.jhydrol.2011.09.021, 2011.

Tian, F., Qiu, G., Yang, Y., Lü, Y., and Xiong, Y.: Estimation of evapotranspiration and its partition based on an extended three-temperature model and modis products, J. Hydrol., 498, 210–220, doi:10.1016/j.jhydrol.2013.06.038, 2013.

Tromp-van Meerveld, H. J. and McDonnell, J. J.: Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis, Water Resour. Res., 42, W02411, doi:10.1029/2004wr003800, 2006.

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

Uhlenbrook, S., Roser, S., and Tilch, N.: Hydrological process representation at the meso-scale: The potential of a distributed, conceptual catchment model, J. Hydrol., 291, 278–296, doi:10.1016/j.jhydrol.2003.12.038, 2004.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resour. Res., 39, 1214, doi:10.1029/2002wr001746, 2003.

Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment classification and hydrologic similarity, Geography Compass, 1, 901–931, doi:10.1111/j.1749-8198.2007.00039.x, 2007.

Wang, J., Li, H., and Hao, X.: Responses of snowmelt runoff to climatic change in an inland river basin, Northwestern China, over the past 50 years, Hydrol. Earth. Sci., 14, 1979–1987, doi:10.5194/hess-14-1979-2010, 2010.

Wang, N.: Isotope hydrology in heihe, Chinese Sci. Bull., 54, 2148–2152, 2009.

Wang, Y., Yu, P., Feger, K. H., Wei, X., Sun, G., Bonell, M., Xiong, W., Zhang, S., and Xu, L.: Annual runoff and evapotranspiration of forestlands and non – in selected basins of the loess plateau of china, Ecohydrology, 4, 277–287, 2011.

Westerberg, I. K., Guerrero, J.-L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., Freer, J. E., and Xu, C.-Y.: Calibration of hydrological models using flow-duration curves, Hydrol. Earth Syst. Sci., 15, 2205–2227, doi:10.5194/hess-15-2205-2011, 2011.

Western, A. W., Grayson, R. B., Blöschl, G., Willgoose, G. R., and McMahon, T. A.: Observed spatial organization of soil moisture and its relation to terrain indices, Water Resour. Res., 35, 797–810, doi:10.1029/1998wr900065, 1999.

Xia, J., Wang, G. S., Lv, A. F., and Tan, G.: A research on distributed time variant gain modeling, Acta Geogr. Sin., 58, 789–796, 2003.

Yu, P., Krysanova, V., Wang, Y., Xiong, W., Mo, F., Shi, Z., Liu, H., Vetter, T., and Huang, S.: Quantitative estimate of water yield reduction caused by forestation in a water-limited area in northwest china, Geophys. Res. Lett., 36, L02406, doi:10.1029/2008gl036744, 2009.

Zang, C. F., Liu, J., van der Velde, M., and Kraxner, F.: Assessment of spatial and temporal patterns of green and blue water flows under natural conditions in inland river basins in Northwest China, Hydrol. Earth Syst. Sci., 16, 2859–2870, doi:10.5194/hess-16-2859-2012, 2012.

Zhao, R.-J.: The xinanjiang model applied in china, Journal of Hydrology, 135, 371–381, doi:10.1016/0022-1694(92)90096-E, 1992.

Zhou, J., Li, X., Wang, G., Hu, H. C., and Chao, Z. H.: An improved precipitation-runoff model based on mms and its application in the upstream basin of the heihe river, J. Nat. Resour., 23, 724–736, 2008.