Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration

S. Gharari\textsuperscript{1,2}, M. Hrachowitz\textsuperscript{1}, F. Fenicia\textsuperscript{3,4}, H. Gao\textsuperscript{1}, H. H. G. Savenije\textsuperscript{1}

\textsuperscript{1}Delft University of Technology, Faculty of Civil Engineering and Geosciences, Water Resources Section, Delft, the Netherlands
\textsuperscript{2}Public Research Center-Gabriel Lippmann, Belvaux, Luxembourg
\textsuperscript{3}Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland

Abstract

Conceptual environmental systems models, such as rainfall runoff models, generally rely on calibration for parameter identification. Increasing complexity of this type of models for better representation of hydrological process heterogeneity typically makes parameter identification more difficult. Although various, potentially valuable, approaches for better parameter identification were developed in the past, strategies to impose general conceptual understanding of how a catchment works into the process of parameter identification of a conceptual model has still not been fully explored. In this study we assess the effect of imposing semi-quantitative, relational expert knowledge for model development and parameter selection, efficiently exploiting the complexity of a semi-distributed model formulation. Making use of a topography driven rainfall-runoff modeling (FLEX-TOPO) approach, a catchment was delineated into three functional units, i.e. wetland, hillslope and plateau. Ranging from simplicity to complexity, three model set-ups, FLEX\textsuperscript{A}, FLEX\textsuperscript{B} and FLEX\textsuperscript{C} have been developed based on these functional units. While FLEX\textsuperscript{A} is a lumped representation of the study catchment, the semi-distributed formulations FLEX\textsuperscript{B} and FLEX\textsuperscript{C} introduce increasingly more complexity by distinguishing 2 and 3 functional units, respectively. In spite of increased complexity, FLEX\textsuperscript{B} and FLEX\textsuperscript{C} allow modelers to compare parameters as well as states and fluxes of their different functional units to each other. Parameter estimation was performed using semi-quantitative, relational constraints imposed onto three models structures. Increased model complexity allowed the identification of additional constraints. It was shown that a constrained but uncalibrated semi-distributed model, FLEX\textsuperscript{C}, can predict runoff with similar performance to a calibrated lumped model, FLEX\textsuperscript{A}. In addition, when constrained and calibrated, the semi-distributed model FLEX\textsuperscript{C} exhibits not only higher performance but also lower predictive uncertainty than the calibrated, lumped FLEX\textsuperscript{A} model.

1- Introduction:

Lumped conceptual and distributed physically based models are the two endpoints of the modeling spectrum, ranging from simplicity to complexity. These two approaches are characterized by their very own advantages and limitations. In hydrology, physically based models are typically applied under the assumptions that (a) the spatial resolution and the complexity of the model are warranted by the available data, and (b) the catchment response is a mere aggregation of small scale processes. However, these two fundamental assumptions are violated in many cases. As a result, not only the predictive power but also the hydrological insights that these models provide is limited (e.g. Beven, 1989, 2001; Grayason et al., 1992, Blöschl, 2001; Pomeroy et al., 2007; Sivapalan, 2006; McDonnell et al., 2007; Hrachowitz et al., 2013b).
In contrast, lumped conceptual models require less data for identifying model parameters. This advantage comes at the expense of considerable limitations. Representing system integrated processes, model structures and parameters are not directly linked to observable quantities. Their estimation therefore strongly relies on calibration. To limit parameter identifiability issues arising from calibration, these models are often oversimplified abstractions of the system. If inadequately tested they may act as “mathematical marionettes” (Kirchner, 2006), frequently resulting in good calibration performance. They may outperform more complex distributed models (e.g. Refsgaard and Knudsen 1996; Ajami et al., 2004; Reed et al., 2004), but they often fail to provide realistic representations of the underlying processes, leading to limited predictive power (e.g. Freer et al., 2003; Seibert, 2003; Kirchner, 2006; Beven, 2006; Kling and Gupta, 2009; Andréassian et al., 2012; Euser et al., 2013; Gharari et al., 2013).

Various strategies have been suggested in the past to allow for increased model complexity and to thereby improve the physical realism of conceptual models. These strategies included the attempt to incorporate different data sources in the parameter estimation process, such as ground- and soil water dynamics (e.g. Seibert and McDonnell, 2002; Freer et al., 2004; Fenicia et al., 2008b; Matgen et al., 2012; Sutanudjaja et al., 2013), remotely sensed evaporation (e.g. Winsemius et al., 2008), snow dynamics (e.g. Parajka and Blöschl, 2008) or tracer data (e.g. Vache and McDonnell, 2006; Dunn et al., 2008; Son and Sivapalan, 2007; Birkel et al., 2011; Hrachowitz et al., 2013a). Alternatively, it was tried to extract more information from available data, for example through the development of signatures representing different aspects of the data (e.g. Gupta et al. 1998, 2008; Boyle et al., 2000, 2001; Madsen 2000; Fenicia et al., 2006; Rouhani et al., 2007; Khu et al., 2008; Winsemius et al., 2009; Bulygina and Gupta, 2010; McMillan et al., 2011, Clark et al., 2011; Euser et al., 2013; Hrachowitz et al, 2013b).

Traditionally, parameter estimation of conceptual models relied on the availability of calibration data, which, however, are frequently not available for the time period or the spatio-temporal resolution of interest. A wide range of regionalization techniques for model parameters and hydrological signatures were thus developed to avoid calibration in such data scarce environments (e.g. Bardossy, 2007; Yadav et al., 2007; Perrin et al., 2008; Zhang et al., 2008; Kling and Gupta, 2009; Samaniego et al., 2010; Kumar et al., 2010; Wagener and Montanari, 2011; Kapangaziwiri et al., 2012, Viglione et al., 2013). However, it was for a long time considered to be challenging to identify suitable functional relationships between catchment characteristics and model parameters (e.g. Merz and Blöschl, 2004; Kling and Gupta, 2009). Only recently, Kumar et al. (2010, 2013a) showed that making use of multi-scale parameter regionalization (MPR) can yield global parameters which perform consistently over different catchment scales. In a further study they successfully transferred parameters obtained by the MPR technique to ungauged catchments in Germany and the USA (Kumar et al., 2013b). Without any further calibration the transferred global parameters were capable to adequately reproduce runoff as well as other hydrological responses of the catchments.

Related to the above discussed difficulties with parameterization, the frequent lack of sufficient processes heterogeneity, i.e. complexity, in conceptual models introduces further limitations to the degree of realism in these models. The concept of hydrological response unit (HRU) can be exploited as a strategy for an efficient tradeoff between model simplicity, required for adequate parameter identifiability, and a more realistic representation of hydrological processes. HRUs are units within a catchment, characterized by a different hydrological function. Individual HRUs can be represented by different model structures to account for hydrologically heterogeneous behavior based on data availability and desired resolution of process representation. This helps to enhance model realism while keeping the
necessary complexity and related identifiability issues comparatively low. In most cases HRUs are defined based on soil types, land cover and similar physical catchment characteristics (e.g. Knudsen et al., 1986; Flügel, 1995; Grayson and Blöschl, 2000; Krcho, 2001; Winter, 2001; Scherrer and Naef, 2003; Uhlenbrook et al., 2004; Wolock et al., 2004; Pomeroy et al., 2007; Scherrer et al., 2007; Schmocker-Fackel et al., 2007; Efstratiadis et al., 2008; Lindström et al., 2010; Nalbantis et al., 2011; Kumar et al., 2010).

A wide range of studies also points towards the potential value of using topographical indices, which are readily available from digital elevation models (DEM) to account for process heterogeneity (e.g. McGlynn and McDonnell, 2003; Seibert et al., 2003; McGuire at al., 2005; Hrachowitz et al., 2009; Jensco et al., 2009; Detty and McGuire, 2010; Gascuel-Odoux et al., 2010). As standard metrics of landscape organization, such as absolute elevation, slope or curvature, as used in the catena concept (Milne, 1935; Park and Van de Giesen, 2004), are often not strong enough descriptors to infer hydrological function, alternative concepts were sought. The development of derived metrics such as the Topographic Wetness Index (Beven and Kirkby, 1979) facilitated an important step forward, being at the core of TOPMODEL (e.g. Beven and Kirkby, 1979; Beven and Freer, 2001b), which has proven to be a valuable approach in specific environmental settings meeting the assumptions of the model. A different descriptor allowing a potentially more generally applicable and hydrologically meaningful landscape classification has recently been suggested by Rennó et al. (2008): the Height Above the Nearest Drainage (HAND). Nobre et al. (2011) showed the hydrological relevance of HAND by investigating long term groundwater behavior and land use. In a further study, this metric facilitated the identification of hydrologically similar landscape units, such as wetlands, hillslopes and plateaus in a Luxembourgish catchment (Gharari et al., 2011).

Explicitly invoking the co-evolution of topography, vegetation and hydrology, Savenije (2010) argued that catchments, as self-organizing systems, need to fulfill the contrasting hydrological functions of efficient drainage and sufficient water storage in order to allow, in a feedback process, topography and vegetation to develop the way they did. These distinct hydrological functions can then be associated with different landscape elements or HRUs as defined by HAND and slope, such that each HRU is represented by a model structure best representing its function in the ecosystem (cf. Savenije, 2010).

While HAND-based landscape classification can potentially show a way forward, it does not solve the problem arising when moving from lumped to HRU-guided, semi-distributed model formulations: multiple parallel model structures typically result in an increased number of parameters, which, when not adequately constrained, may increase equifinality and thereby predictive uncertainty (e.g. Gupta and Sorooshian, 1983; Beven, 2006; Gupta et al., 2008). In order to better satisfy the contrasting priorities of model complexity and predictive power, new strategies are sought to more efficiently utilize the modelers’ understanding of the system and the frequently scarce available data for constraining the feasible model- and parameter space (e.g. Gupta et al., 2008; Wagener and Montanari, 2011; Singh and Bárdossy, 2012; Andréassian et al., 2012; Gharari et al., 2013; Hrachowitz et al., 2013b; Razavi and Tolson, 2013). In contrast to earlier attempts to constrain models using multiple evaluation criteria or a priori information on catchment properties such as land use or soil type (e.g. Koren et al., 2008), the utility of a different and so far underexploited type of constraints, based on a priori understanding of the system, has been tested in this study. The concept of topography-driven conceptual modeling introduced by Savenije (2010) involves the identification of HRUs that operate in parallel. Linked to the technique of regularization (e.g. Tikhonov, 1963; Engl et al., 1996), this opens the possibility to impose semi-quantitative, expert knowledge based, relational constraints of catchment behavior on model parameters, similar to what was suggested by Pokhrel et al. (2008) and Yilmaz et al. (2008). To restrict the posterior parameter distributions, hydrologically meaningful relations between parallel HRUs are
introduced. Based on expert knowledge and expressed as relational constraints they ensure that similar processes between parallel model structures in the semi-distributed model are represented in an internally consistent way, thereby reducing the parameters’ potential for compensating for errors. The advantage of this method is that there is only limited need to precisely quantify the constraints or the prior parameter distributions (e.g. Koren et al., 2000, 2003; Kuzmin et al., 2008; Duan et al., 2006). This could allow for a meaningful and potentially more realistic representation of the system in which each model component is, within certain limits, forced to do what it is designed to do, rather than allowing it to compensate for data and model structural errors.

The objectives of this paper are thus to test the hypothesis if the use of semi-distributed, conceptual models, representing HRUs defined by hydrologically meaningful, topography-based landscape classification combined with model constraints (1) can increase model internal consistency and thus the level of process realism as compared to lumped model set-ups, (2) can increase the predictive power compared to lumped model set-ups and (3) can reduce the need for model calibration by the use of expert knowledge based on relations between parameters, fluxes and states.

2- Study area and data:

The outlined methodology will be illustrated and tested with a case study using data of the Wark catchment in the Grand Duchy of Luxembourg. The catchment has an area of 82 km² with the catchment outlet located downstream of the town of Ettelbrück at the confluence with the Alzette River (49.85° N, 6.10° E, Figure 1). With an annual mean precipitation of 850 mm yr⁻¹ and an annual mean potential evaporation of 650 mm yr⁻¹ the annual mean runoff is approximately 250 mm yr⁻¹. The geology in the northern part is dominated by schist while the southern part of the catchment is mostly underlain by sandstone and conglomerate. Hillslopes are generally characterized by forest, while plateaux and valley bottoms are mostly used as crop land and pastures, respectively. Drouge et al. (2002) quantified land use in the catchment as 4.3% urban areas, 52.7% agricultural land and 42.9% forest. In addition they reported that 61% of catchment is covered by permeable soils while the remainder is characterized by lower permeability substrate. The elevation varies between 195 to 532 m, with a mean value of 380 m. The slope of the catchment varies between 0-200%, with a mean value of 17 % (Gharari et al., 2011).

The hydrological data used in this study include discharge measured at the outlet of the Wark catchment, potential evaporation estimated by the Hamon equation (Hamon, 1961) with temperature data measured at Luxembourg airport (Fenicia et al., 2008a); and precipitation measured by three tipping bucket rain gauges located at Reichlange. The temporal resolution used in this study is 3 h.

3- FLEX-TOPO framework:

Realizing the potential of “reading the landscape” in a systems approach (cf. Sivapalan, 2003), Savenije (2010) argued that due to the co-evolution of topography, soil and vegetation, all of which define the hydrological function of a given location, an efficient, hydrologically meaningful descriptor of topography together with land use could be used to distinguish different HRUs. HAND, which can be loosely interpreted as the hydraulic head at a given location in a catchment, may be such a descriptor as it potentially allows for meaningful landscape classification (e.g. Rennó et al., 2008; Gharari et al., 2011). It was argued previously (Gharari et al., 2011) that, in Central European landscapes, HAND can efficiently distinguish between wetlands, hillslopes and plateaus. These are landscape elements that may
also be assumed to fulfill distinct hydrological functions (HRUs) in the study catchment (Savenije, 2010). Wetlands, located at low elevations above streams, are characterized by shallow ground water tables with limited fluctuations. Due to reduced storage capacity between ground water table and soil surface, potentially exacerbated by the relative importance of the capillary fringe, wetlands tend to be saturated, and thus connected, earlier during a rainfall event than the two other landscape elements with arguably higher storage capacity, thus frequently becoming the dominant source of storm flow during comparably dry periods (e.g. Seibert et al., 2003; McGlynn et al., 2004; Molenat et al., 2005; Blume et al., 2008; Anderson et al., 2010; Kavetski et al., 2011). The dominant runoff process in wetlands can therefore be assumed to be saturation overland flow. In contrast, forested hillslopes, landscape elements with steeper slopes than the wetlands or plateaus, require a balance between sufficient storage capacity and efficient drainage to develop and maintain the ecosystem (Savenije, 2010). A dual system combining sufficient water storage in the root zone and efficient lateral drainage through preferential flow networks, controlled by a suite of activation thresholds as frequently observed on hillslopes (e.g. Hewlett, 1961; Beven and Germann, 1982; Sidle et al., 2001; Freer et al., 2002; Weiler et al., 2003; McNamara et al., 2005; Tromp van Meerveld and McDonnell, 2006a, 2006b; Zehe and Sivapalan, 2009; Spence, 2010) can be seen as the dominant mechanism. Finally, plateaus are landforms with low to moderate slopes and comparably deep ground water tables. In absence of significant topographic gradients and due to the potentially increased unsaturated storage capacity, it can be hypothesized that the primary functions of plateaus are sub-surface storage and groundwater recharge (Savenije, 2010). Although plateaus may experience infiltration excess overland flow in specific locations, the topographical gradients may not be sufficient to generate surface runoff connected to the stream network. In the FLEX-TOPO approach the proportions of the hydrologically distinct landscape units, i.e. HRUs, in a given catchment need to be determined on the basis of topographical and land cover information. Subsequently, suitable model structures and parameterizations (read constitutive functions) will be assigned to the different HRUs (Fenicia et al. 2011, Kavetski et al., 2011, Clark et al., 2009). The integrated catchment output, i.e. runoff and evaporative fluxes, can then be obtained by combining the computed proportional outputs from the individual HRUs. Note that the three landscape classes tested for suitability in this study, i.e. wetland, hillslope and plateau together with their assumed dominant runoff process are designed for the Wark catchment and are likely to be different for other environmental settings (e.g. Gao et al., 2014).

3.1- Landscape classification:

As the objective of FLEX-TOPO is to efficiently extract and use hydrologically relevant information from worldwide readily available topographic data, i.e. DEMs, the Height Above the Nearest Drainage (HAND; Rennó et al., 2008; Nobre et al., 2011; Vannametee et al., 2014) is a potentially powerful metric to classify landscapes into HRUs with distinct hydrological function, as discussed above. Testing a suite of HAND-based classification methods Gharari et al. (2011) found that results best matching observed landscape types could be obtained by using HAND together with the local slope. Based on a probabilistic framework to map the desired HRUs which were then compared with in-situ observations they obtained a threshold for HAND and slope of approximately 5 m and 11 % for the Wark catchment. Following that, wetlands were defined to be areas with HAND ≤ 5 m. Areas with HAND > 5 m and local slopes > 11 % were classified as hillslopes, while areas with HAND > 5 m and slope ≤ 11 % were defined as plateaus. The HAND and slope map of the study catchment together with the classified landscape entities (wetland, hillslope and plateau) are
presented in Fig. 1. The proportion of the individual HRUs wetland, hillslope and plateau are 15%, 45% and 40% respectively.

3-2- Model setup:

In this study a lumped conceptual model of the Wark catchment, hereafter referred to as FLEXA, is used as a benchmark since similar lumped conceptual models are frequently used in catchment hydrology, particularly in small- to mesoscale catchments (e.g. Merz and Blöschl, 2004; Clark et al., 2008; Perrin et al., 2008; Seibert and Beven, 2009; Fenicia et al., 2013). The above discussed concept of FLEX-TOPO (Savenije, 2010) is thereafter tested with a stepwise increased number of parallel landscape units (FLEX B, FLEX C), thereby increasing the conceptualized process heterogeneity and thus the model complexity. The core of the three model set-ups is loosely based on the FLEX model (Fenicia et al., 2006).

3-2-1-FLEXA: This model set-up represents the catchment in a lumped way. The FLEXA model structure consists of four storage elements representing interception, unsaturated, slow
(i.e. groundwater) and fast responding reservoirs (i.e. preferential flow and saturation overland flow). A schematic illustration of FLEXA is shown in Fig. 2a. The water balance and constitutive equations used are given in Table 2.

3-2-1-1-Interception reservoir ($S_L$): The interception reservoir is characterized by its maximum storage capacity ($I_{\text{max}}$ [L]). After precipitation ($P$ [LT$^{-1}$]) enters this reservoir the excess precipitation, hereafter referred to as effective precipitation ($P_e$ [LT$^{-1}$]), is distributed between the unsaturated ($S_U$), slow ($S_S$) and fast reservoir ($S_F$).

3-2-1-2-Unsaturated reservoir ($S_U$): The unsaturated reservoir is characterized by a parameter that loosely reflects the maximum soil moisture capacity in the root zone ($S_{U,\text{max}}$ [L]). Part of the effective precipitation ($P_e$) enters the unsaturated zone according to the coefficient $C_r$, which here is defined by a power function with exponent $\beta$ [-], reflecting the spatial heterogeneity of thresholds for activating fast lateral flows from $S_F$. This coefficient $C_r$ will be 1 when soil moisture ($S_U$) is lower than a specific percentage of maximum soil moisture capacity ($S_{U,\text{max}}$) defined by relative soil moisture at field capacity ($F_C$[-]), meaning that the entire incoming effective precipitation ($P_e$) at a given time step is stored in the unsaturated reservoir ($S_U$). The soil moisture reservoir feeds the slow reservoir through matrix percolation ($R_p$ [LT$^{-1}$]), expressed as a linear relation of the available moisture in the unsaturated zone ($S_U$) and the maximum percolation capacity ($P_{\text{Per}}$ [LT$^{-1}$]). The reverse process, capillary rise ($R_c$), feeds the unsaturated reservoir from the saturated zone. Capillary rise ($R_c$ [LT$^{-1}$]) has an inverse linear relation with the moisture content in the unsaturated zone and is characterized by the maximum capillary rise capacity ($C$ [LT$^{-1}$]). Soil moisture is depleted by plant transpiration. Transpiration is assumed to be moisture constrained when the soil moisture content is lower than a fraction $L_p$ [-] of the maximum unsaturated capacity ($S_{U,\text{max}}$). When the soil moisture content in the unsaturated reservoir is higher than this fraction ($L_p$) transpiration is assumed to be equal to the potential evaporation ($E_{\text{pot}}$ [LT$^{-1}$]).

3-2-1-3- Splitter and transfer functions: The proportion of effective rainfall which is not stored in the unsaturated zone, i.e. $1-C_r$, is further regulated by the partitioning coefficient ($D$ [-]), distributing flows between preferential groundwater recharge ($R_S$ [LT$^{-1}$]) to $S_S$ and water that is routed to the stream by fast lateral processes from $S_F$ (e.g. preferential flow or saturation overland flow, $R_F$). Both fluxes are lagged by rising linear lag functions with parameters $N_{\text{lagf}}$ and $N_{\text{lags}}$, respectively (e.g. Fenicia et al., 2008b).

3-2-1-4-Fast reservoir ($S_F$): The fast reservoir is a linear reservoir characterized by reservoir coefficient $K_F$.

3-2-1-5-Slow reservoir ($S_S$): The slow reservoir is a linear reservoir characterized by a reservoir coefficient $K_S$.

3-2-2-FLEXB: As discussed above, a range of process studies suggested that wetlands can frequently exhibit storage-discharge dynamics that are decoupled from other parts of a catchment, in particular due to their typically reduced storage capacity and closeness to the stream. FLEXB explicitly distinguishes wetlands from the rest of the catchment, the “remainder” (i.e. hillslopes and plateaus), which is represented in a lumped way, to account for this difference. The FLEXB model set-up therefore consists of two parallel model structures which are connected with a common groundwater reservoir (Figure 2b), similar to what has been suggested by Knudsen et al. (1986). One major difference between the two parallel structures is that capillary rise is assumed to be a relevant process only in the wetland,
while it is considered negligible in the remainder of the catchment due to the deeper groundwater. Further, since the wetlands are predominantly exfiltration zones of potentially low permeability, preferential recharge is considered negligible in wetlands. The areal proportions of wetland and the remainder (i.e. hillslope and plateau) of the catchment are 15% and 85%, respectively (Gharari et al., 2011).

3-2-3-FLEX\textsuperscript{C}: This model set-up offers a complete representation of the three HRUs in the study catchment: wetland, hillslope and plateau (Figure 2c). The formulation of the wetland module in FLEX\textsuperscript{C} is identical to the one suggested above for FLEX\textsuperscript{B}. The hillslope HRU is represented by a model structure resembling the FLEX\textsuperscript{A} set-up. Plateaus are assumed to be dominated by vertical fluxes, while direct lateral movement in the form of Hortonian overland flow is considered negligible compared to those generated from hillslope and wetland HRUs. Therefore the plateau model structure does not account for these fast fluxes. Analogous to FLEX\textsuperscript{B}, the FLEX\textsuperscript{C} set-up is characterized by one single groundwater reservoir linking the three dominant HRUs in this catchment. The individual proportions of wetland, hillslope and plateau are 15%, 45% and 40%, respectively (Gharari et al., 2011). The proportions of these HRUs are used to compute the total discharge based on the contribution of each landscape unit.

The connection between the parallel structures of FLEX\textsuperscript{B} and FLEX\textsuperscript{C} is through the surface drainage network (the stream network) and through the slow (groundwater) reservoir.

![Figure 2](image)

**Figure 2-** The model structures for (a) FLEX\textsuperscript{A}, (b) FLEX\textsuperscript{B} and (c) FLEX\textsuperscript{C}.

**Table 1-** Uniform prior parameter distributions for the three model set-ups
Table 2- Water balance and constitutive equations used in FLEX A

### Interception reservoir

\[
\frac{dS_I}{dt} = P - I - P_e \quad (1)
\]

\[
I = \begin{cases} 
E_{pot} & S_I > 0 \\
0 & S_I = 0 
\end{cases} \quad (2)
\]

\[
P_e = \begin{cases} 
0 & S_I < I_{\text{max}} \\
P & S_I = I_{\text{max}} 
\end{cases} \quad (3)
\]

### Unsaturated reservoir

\[
\frac{dS_U}{dt} = R_u - T - R_p + R_C \quad (4)
\]

\[
R_u = C_r P_e \quad (5)
\]

\[
C_r = \begin{cases} 
1 & \left( \frac{S_u - S_{u,max} F_C}{S_{u,max} - S_{u,max} F_C} \right) > 1 \\
0 & \left( \frac{S_u - S_{u,max} F_C}{S_{u,max} - S_{u,max} F_C} \right) \leq 1 
\end{cases} \quad (6)
\]

\[
T = K_T (E_{pot}) \quad (7)
\]

\[
K_T = \begin{cases} 
\frac{S_u}{S_{u,max} L_p} & S_u < S_{u,max} L_p \\
1 & S_u \geq S_{u,max} L_p 
\end{cases} \quad (8)
\]

\[
R_p = \left[ S_u / S_{u,max} \right] P_{per} \quad (9)
\]

\[
R_C = \left[ 1 - \left( S_u / S_{u,max} \right) \right] C \quad (10)
\]

### Fast reservoir

\[
\frac{dS_F}{dt} = R_{F,lag} - Q_F \quad (11)
\]

\[
R_p = (1 - D)(1 - C_r) P_e \quad (12)
\]

\[
R_{F,lag} = R_F * N_{lagf} \quad (13)
\]

\[
Q_F = K_F S_F \quad (14)
\]

### Slow reservoir

\[
\frac{dS_S}{dt} = R_{S,lag} - Q_S + R_p - R_C \quad (16)
\]

\[
R_s = D(1 - C_r) P_e \quad (17)
\]

\[
R_{S,lag} = R_S * N_{lags} \quad (18)
\]

\[
Q_S = K_S S_S \quad (19)
\]

* is the convolution operator.

3-3- Introducing realism constraints in selecting behavioral parameter sets:

With increasing process heterogeneity from FLEX A over FLEX B to FLEX C, the respective model complexities and therefore the number of calibration parameters also increase. This, in
the frequent absence of sufficient suitable data to efficiently constrain a model, typically leads
to a situation where parameters have increased freedom to compensate for errors in data and
model structures, as recently reiterated by Gupta et al. (2008). As a consequence, the resulting
higher level of equifinality can substantially reduce a model’s predictive power. In this study,
two fundamentally different types of model constraints were applied to test their value for
reducing equifinality in complex model set-ups. Firstly, conditions between parameters of
parallel model units, hereafter referred to as parameter constraints, were imposed before each
model evaluation run. These a priori constraints ensure that the individual parameter values
for the same process in the parallel units, reflect the modeler’s perception of the system. For
example, it can be argued that the maximum interception capacity ($I_{max}$) of a forested HRU
needs to be higher than the one in a not forested HRU. In the absence of more detailed
information this does not only allow overlapping prior distributions but it also avoids the need
for quantification of the constraints themselves. The second type of constraints is process
constraints, which can only be applied after each evaluation run during the calibration phase.
These a posteriori constraints compare the modeled output of the individual HRUs and ensure
that these outputs follow the modeler’s perception of the system’s internal dynamics. For
example, it can be argued that the modeled evaporation in forested HRUs needs to be higher
than in not forested HRUs. The parameter and process constraints imposed on the models in
this study are described in detail below. Note that the choice of constraints to impose is the
modeler’s choice and that with increasing number of different HRUs an increasing number of
constraints can be applied. While here FLEX$A$ only allows for two constraints, i.e. one
parameter- and one process constraint, all constraints suggested below can be applied to
FLEX$C$.

**3-3-1-Parameter constraints:**

A number of a priori constraints is imposed on different model parameters in order to exclude
unrealistic parameter combinations. The constraints are guided by considerations on what the
model components are designed to reproduce. The number of constraints that can be imposed
increases with increasing model complexity. The full set of parameter constraints detailed
below were applied to FLEX$C$ and when applicable also to FLEX$A$. In contrast, only one
parameter constraint could be used for FLEX$A$, as for this model no more obvious
relationships between parameters could be identified. In the following, the subscripts $w$, $h$ and
$p$ indicate parameters for wetland, hillslope and plateau, respectively.

**3-3-1-1-Interception:**

The different land cover proportions of each landscape unit, here wetlands, hillslopes and
plateaus, can be used to define the relation between interception thresholds ($I_{max}$) of these
individual units. The land uses are defined as two general classes for this case study, forested
areas and grass or pasture-land areas. The maximum interception capacity ($I_{max}$) for each
landscape entity can be estimated from the proportion of land-use classes and their maximum
interception capacities, selected from their respective prior distributions as given in Table 1:

$$I_{max,w} = a_w I_{max,forest} + b_w I_{max,cropland}$$ (20)

$$I_{max,h} = a_h I_{max,forest} + b_h I_{max,cropland}$$ (21)

$$I_{max,p} = a_p I_{max,forest} + b_p I_{max,cropland}$$ (22)

The proportions of forested area are indicated with $a_w$, $a_h$ and $a_p$ for wetland, hillslope and
plateau and are fixed at 42%, 60% and 29%, respectively. The proportions of cropland and
green land areas are indicated by $b_w$, $b_h$ and $b_p$ for wetland, hillslope and plateau and are fixed
at 58%, 40% and 71%, respectively. Moreover the parameter sets which are selected for
maximum interception capacity of forest are expected to be higher than crop- or grassland:
\[ I_{\text{max,cropland}} < I_{\text{max,forest}} \] (23)

3-3-1-2- Lag functions:

Preferential recharge (Rs) is routed to the slow reservoir by a lag function. Due to a deeper
groundwater table on plateaus it can be assumed that the lag time for Rs is longer for plateaus
than for hillslopes. It can also be assumed that the lag function used for fast reservoir for
hillslopes is longer than for wetlands due to the on average higher distance of and therefore
longer travel times from hillslopes to the stream.
\[ N_{\text{lags,w}} \leq N_{\text{lags,h}} \leq N_{\text{lags,p}} \] (24)

3-3-1-3-Soil moisture capacity:

Many experimental studies support the assumption that wetlands have shallower groundwater
tables than the other two landscape entities in this study. Therefore the unsaturated zone of
wetland should be shallower, i.e. the maximum soil moisture capacity \( S_{U,\text{max}} \) of hillslopes
and plateaus can be assumed to be higher. Moreover, as hillslopes in the study catchment are
predominantly covered with forest, it can, due to the deeper root zone of forests, be expected
that the maximum unsaturated soil moisture capacity \( S_{U,\text{max}} \) in the root zone of hillslopes is
deeper than the other two landscape entities.
\[ S_{U,\text{max,w}} < S_{U,\text{max,p}} < S_{U,\text{max,h}} \] (25)

3-3-1-4-Reservoir coefficients:

The reservoir coefficient of the wetland fast reservoir \( K_F \) is assumed to be lower than
reservoir coefficient of the hillslope fast reservoir as, once connectivity is established, the
flow velocities of saturation overland flow in wetlands are assumed to exceed the integrated
flow velocities of preferential flow networks (cf. Anderson et al., 2009). Likewise, the
retention time of the slow reservoir should be higher than both wetland and hillslope fast
reservoirs.
\[ K_S < K_{F,h} < K_{F,w} \] (26)

The reservoir constraints can be applied to all models while the other constraints can only be
applied to FLEX\(^B\) and FLEX\(^C\).

3-3-2-Process constraints:

In contrast to the parameter constraints discussed above, which are set \textit{a priori}, process
constraints are applied \textit{a posteriori}. Only parameters which generate model internal flux
dynamics in agreement with the modeler’s perception of these dynamics are retained as
feasible. Hence, while with the use of parameter constraints there is no need to run the model
for rejected parameter sets, here it is necessary to run the model to evaluate it with respect to the process constraints.

Process constraints are defined for dry and wet periods as well as for peak-, high- and low flows. Here wet periods were defined to be the months from December to March, while the dry periods in the study catchment occur between April and November. The thresholds for distinguishing between high and low flow were chosen to be 0.05 and 0.2 mm(3h)$^{-1}$ respectively for dry and wet periods. Furthermore, events during which discharge increases with a rate of more than 0.2 mm(3h)$^{-2}$ are defined as peak flows. Note that in the following the subscripts peak, high and low indicate peak-, high- and low flows.

3-3-2-1 Transpiration:

Transpiration typically exhibits a clear relationship with the normalized difference vegetation index (NDVI, Szilagyi et al., 1998). Therefore the ratios between NDVI values of different landscape units can serve as constraints on modeled transpiration obtained from the individual parallel model components. A rough estimation of the ratio between transpiration from plateau and hillslope can be derived from LANDSAT 7 images. For this ratio seven cloud free images were selected (acquisition dates of 20/4/2000, 6/3/2000, 11/9/2000, 18/2/2001, 6/3/2001, 26/3/2001 and 29/8/2001). The ratio of transpiration between hillslope and plateau ($R_{\text{trans}}$) can be estimated by assuming a linear relation (Szilagyi et al., 1998) with slope of $\alpha$ and intercept zero between transpiration and mean NDVI for each landscape unit ($\mu_{\text{NDVI}}$).

$$R_{\text{trans}} = \frac{\alpha_{\mu_{\text{NDVI},h}}}{\alpha_{\mu_{\text{NDVI},p}}} = \frac{\mu_{\text{NDVI},h}}{\mu_{\text{NDVI},p}}$$ (27)

Mean ($\mu_{R_{\text{trans}}}$) and standard deviation ($\sigma_{R_{\text{trans}}}$) of $R_{\text{trans}}$ can be used to estimate acceptable limits of the transpiration ratios for hillslope and plateau.

Therefore the annual transpiration can be confined between two values as follows:

$$\mu_{R_{\text{trans}}} - \sigma_{R_{\text{trans}}} < \int \frac{T_{h} \, dt}{T_{p} \, dt} < \mu_{R_{\text{trans}}} + \sigma_{R_{\text{trans}}}$$ (28)

Based on the mean ($\mu_{R_{\text{trans}}} = 1.2$) and standard deviation ($\sigma_{R_{\text{trans}}} = 0.2$) of the seven LANDSAT 7 images used the following process constraint on transpiration from hillslope ($T_{h}$) and plateau ($T_{p}$) was imposed:

$$1.0 < \frac{\int T_{h} \, dt}{T_{p} \, dt} < 1.4$$ (29)

Similar constraints can be imposed between transpiration fluxes from wetland, hillslope or plateau; however, the spatial resolution of LANDSAT 7 data with resolution of 30 meters is coarser than the required 20-meter DEM resolution for distinguishing wetlands from other landscape entities (Gharari et al., 2011).

3-3-3-2 Runoff coefficient:

The runoff coefficient is a frequently used catchment signature (e.g. Sawicz et al., 2011; Euser et al., 2013) and can be used as a behavioral constraint (e.g. Duan et al., 2006; Winsemius et al., 2009). In this study the runoff coefficients of dry and wet periods as well as the annual runoff coefficient were used. Parameters that result in modeled runoff coefficients that substantially deviate from the observed ones are therefore discarded. In case of absence of suitable runoff data, the mean annual runoff coefficient can be estimated from the regional
Budyko curve using for example the Turc-Pike relationship (Turc, 1954; Pike, 1964; Arora, 2002). However in this study the runoff coefficients of each individual year, and their respective dry and wet periods was used and determined the mean and standard deviation of the runoff coefficients for these periods. Here, as a conservative assumption, the limits are set to three times the standard deviation around the mean runoff coefficient. Note that the runoff coefficient is the only process constraint that is not related to model structure in this study and can therefore also be applied to the lumped FLEXA set-up.

\[
\frac{\int Q_m \, dt}{P \, dt} < 0.43 \quad (30)
\]

\[
\frac{\int Q_w \, dt}{P \, dt} > 0.16 \quad (31)
\]

\[
\frac{\int Q_{m,\text{dry}} \, dt}{P_{\text{dry}} \, dt} < 0.36 \quad (32)
\]

\[
\frac{\int Q_{m,\text{dry}} \, dt}{P_{\text{dry}} \, dt} > 0 \quad (33)
\]

\[
\frac{\int Q_{m,\text{wet}} \, dt}{P_{\text{wet}} \, dt} < 0.71 \quad (34)
\]

\[
\frac{\int Q_{m,\text{wet}} \, dt}{P_{\text{wet}} \, dt} > 0.40 \quad (35)
\]

3-3-3-3-Preferential recharge:

The slow reservoir can be recharged by both preferential and matrix percolation from the unsaturated reservoirs. Here, hillslopes and plateaus contribute to the slow reservoir by preferential recharge. It can be assumed that in a realistic model setup the long term contribution volume of preferential recharge ratio between hillslope and plateau should not be unrealistically high or low. For example, it can be assumed unrealistic that the ratio is zero or infinity, meaning that one landscape unit is constantly feeding the slow reservoir while another one is not contributing at all. To avoid such a problem, a loose and very conservative constraint was imposed on the ratio of contribution of the two fluxes.

\[
0.2 < \frac{\int R_{s,\text{h}} \, dt}{R_{s,p} \, dt} < 5 \quad (36)
\]

3-3-3-4-Fast component discharge:

During dry periods, hillslopes and plateaus can exhibit significant soil moisture deficits, limiting the amount of fast runoff generated from these landscape elements. In contrast, due to their reduced storage capacity, wetlands are likely to generate fast flows at lower moisture levels, thus dominating event response during dry periods (cf. Beven and Freer, 2001a; Seibert et al., 2003; Molenat et al., 2005; Anderson et al., 2010; Birkel et al., 2010). It can thus be assumed that during both, the entire dry periods as well as peak flows in dry periods
the fast component of wetlands \( Q_{f,w,dry}; Q_{f,w,dry,peaks} \) contributes to runoff more than the fast component of hillslopes \( Q_{f,h,dry}; Q_{f,h,dry,peaks} \). In contrast, high flows during wet periods are predominantly generated by hillslopes \( Q_{f,h,wet}; Q_{f,h,wet,high} \). This process constraint is also applied to Flexb:

\[
\frac{\int Q_{f,h,dry,peaks} \, dt}{\int Q_{f,w,dry,peaks} \, dt} < 1 \quad (37)
\]

\[
\frac{\int Q_{f,h,wet,high} \, dt}{\int Q_{f,w,wet,high} \, dt} > 1 \quad (38)
\]

\[
\frac{\int Q_{f,h,wet} \, dt}{\int Q_{f,w,wet} \, dt} > 1 \quad (39)
\]

### 3-3-4- Calibration algorithm and objective functions:

Based on uniform prior parameter distributions as well as on the parameter- and process constraints the model was calibrated using MOSCEM-UA (Vrugt et al., 2003). As a brief description, MOSCEM-UA uses a Latin Hypercube sampling strategy for the random sampling of the entire parameter space. Penalizing the objective function(s) based on the number of unsatisfied constraints, however, may lead to non-smooth objective functions which potentially may cause instabilities in the search algorithm or create invalid results. A recently developed stepwise search algorithm was therefore used for finding parameter sets which satisfy both parameter- and process constraints (Gharari et al., this issue). These parameter sets were then used as initial sampling parameter sets for MOSCEM-UA instead of the traditionally used Latin Hypercube sampling strategy.

The models were evaluated on the basis of three different objective functions to emphasize different characteristics of the system response: (i) the Nash-Sutcliffe efficiency of the flows \( E_{NS} \), (ii) the Nash-Sutcliffe efficiency of the logarithm of the flows \( E_{NS,\log} \) and (iii) the Nash-Sutcliffe efficiency of the flow duration curve \( E_{NS,FDC} \). These criteria evaluate the models’ ability to simultaneously reproduce high flows, low flows and flow duration curves respectively. The model set-ups have been constrained and calibrated for the year 2002-2005 and validated for year 2006-2009. The year 2001 was used as warm up period.

### 3-4- Model validation and parameter evaluation:

To assess the value of incorporating parameter and process constraints in increasingly complex models a four-step procedure as outlined below was followed. Note that for each step the respective model (parameters) was evaluated against the constrained and calibrated lumped FlexA benchmark model.

#### 3-4-1- Evaluating models with “constrained but uncalibrated” parameter sets:

Firstly, all parameter sets which satisfy all the applied constraints were evaluated based on their ability to reproduce the observed hydrograph. Hereafter these parameter sets are referred to as constrained but uncalibrated parameter sets because they were obtained without any calibration on the observed hydrographs. Based on the retained, feasible parameter sets, the
mean performance of the three constrained but uncalibrated models FLEX\(^A\), FLEX\(^B\) and FLEX\(^C\), for the three objective functions \((E_{NS}, E_{NS,log}, E_{NS,FDC})\) together with their uncertainty ranges for both the calibration and the validation periods are compared. FLEX\(^A\), FLEX\(^B\) and FLEX\(^C\) have an increasing number of constraints. It was thus tested whether the higher complexity models also result in better model performance and how the predictive uncertainty is affected by increased complexity and model realism.

To investigate how well the hydrographs generated with parameters satisfying all constraints match the observed hydrograph, the 95% uncertainty intervals of simulated hydrographs based on these parameter sets were generated for the three models. The uncertainty was estimated on the basis of the area indicated by 95% uncertainty intervals based on simulated hydrographs.

### 3-4-2- Evaluating models with “constrained and calibrated” parameter sets:

In the second step, the three models FLEX\(^A\), FLEX\(^B\) and FLEX\(^C\) have been calibrated within the parameter space which satisfied all the imposed parameter and process constraints. The models were calibrated using a multi-objective strategy \((E_{NS}, E_{NS,log}, E_{NS,FDC})\). The obtained Pareto optimal model parameters are in the following referred to as constrained and calibrated.

Analogous to the previous step uncertainty intervals based on the constrained and calibrated Pareto optimal parameters, were generated. The uncertainty was estimated on the basis of the area of the uncertainty bands.

### 3-4-3- Comparison of model performance and uncertainty for “constrained but un-calibrated” and “constrained and calibrated” parameter sets:

To assess the added value of incorporating constraints in higher complexity models, the performance and uncertainties of the three models FLEX\(^A\), FLEX\(^B\) and FLEX\(^C\) were compared for both the “constrained but un-calibrated” and the “constrained and calibrated” case during calibration (2002-2005) and validation (2006-2009) periods.

### 3-4-4- Comparison of modeled hydrograph components for different model structures:

One of the main reasons for imposing constraints on model parameters is to ensure the realistic internal dynamic of a model. Comparing different fluxes contributing to the modeled hydrograph can give an insight into the performance of imposed constrained on the model. The effect of imposing behavioral constraints on fast and slow components of the three models structures, FLEX\(^A\), FLEX\(^B\) and FLEX\(^C\) is compared visually. The fast component of lumped model, FLEX\(^A\), is compared with fast components of FLEX\(^B\) which are wetland and remainder of catchment and fast components of FLEX\(^C\) which are wetland and hillslope. This visual comparison is based on normalized average contribution of each component for Pareto optimal parameter sets in every time step.

### 4- Results and discussion

#### 4-1- Evaluating the performance of constrained but uncalibrated parameter sets:

The median and the 95% uncertainty intervals of the performance of modeled hydrographs for constrained but uncalibrated parameter sets is presented in Table 3 for the calibration period.
The lumped FLEX\textsuperscript{A} model has only one parameter and one process constraint, i.e. the runoff coefficient. Hence, this model is free within the limits of this apparently relatively weak condition, resulting in a wide range of possible parameters, many of which cannot adequately reproduce the system response. As a consequence, the overall performance is poor ($E_{NS,\text{median}}=0.19$, $E_{NS,\text{log,median}}=0.13$, $E_{NS,FDC,\text{median}}=0.38$) (Table 3, Figure 3).

FLEX\textsuperscript{B}, run with the set of constrained but uncalibrated parameters shows a substantial improvement in overall performance ($E_{NS,\text{median}}=0.59$, $E_{NS,\text{log,median}}=0.44$, $E_{NS,FDC,\text{median}}=0.92$) compared to FLEX\textsuperscript{A}, as FLEX\textsuperscript{B} not only allows for more process heterogeneity but, more importantly, it is conditioned with an increased number of constraints.

The additional process heterogeneity and constraints allowed by FLEX\textsuperscript{C}, results in the highest overall performance for all three objective functions ($E_{NS,\text{median}}=0.68$, $E_{NS,\text{log,median}}=0.54$, $E_{NS,FDC,\text{median}}=0.95$) (Table 3, Figure 3).

These results clearly illustrate that the imposed relational constraints force the model and its parameters towards a more realistic behavior, which significantly improves model performance.

The 95% uncertainty intervals indicate that FLEX\textsuperscript{C}, which might be expected to produce the highest uncertainty interval due to its complexity, is providing a lower uncertainty compared to FLEX\textsuperscript{B}. Although FLEX\textsuperscript{C} cannot outperform FLEX\textsuperscript{A} in terms of a lower uncertainty interval, overall performance of this model is better than FLEX\textsuperscript{A} as discussed earlier.

<table>
<thead>
<tr>
<th></th>
<th>$E_{NS}$</th>
<th>$E_{NS,\text{log}}$</th>
<th>$E_{NS,FDC}$</th>
<th>95% uncertainty area [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLEX\textsuperscript{A}</td>
<td>Calibration</td>
<td>0.19 [0.12 0.28]</td>
<td>0.13 [-0.13 0.41]</td>
<td>0.38 [0.29 0.56]</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>0.20 [0.10 0.33]</td>
<td>0.37 [0.18 0.56]</td>
<td>0.40 [0.27 0.61]</td>
</tr>
<tr>
<td>FLEX\textsuperscript{B}</td>
<td>Calibration</td>
<td>0.59 [0.25 0.75]</td>
<td>0.44 [0.16 0.61]</td>
<td>0.92 [0.81 0.95]</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>0.54 [0.23 0.76]</td>
<td>0.59 [0.31 0.73]</td>
<td>0.92 [0.76 0.98]</td>
</tr>
<tr>
<td>FLEX\textsuperscript{C}</td>
<td>Calibration</td>
<td>0.68 [0.46 0.77]</td>
<td>0.54 [0.07 0.65]</td>
<td>0.95 [0.91 0.96]</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>0.66 [0.42 0.77]</td>
<td>0.63 [0.05 0.77]</td>
<td>0.97 [0.94 0.99]</td>
</tr>
</tbody>
</table>
Figure 3- The observed hydrograph and the 95% uncertainty interval of the modeled hydrograph derived from the complete set of constrained but un-calibrated parameter sets for the three different model set-ups (a) FLEX\textsuperscript{A}, (b) FLEX\textsuperscript{B} and (c) FLEX\textsuperscript{C} for two years (2002-2003) of calibration.

4-2- Evaluating the performance of constrained and calibrated parameter sets:

The comparison of the constrained and calibrated model set-ups shows that all three models set-ups can reproduce the hydrograph similarly well (Table 4, Figure 4). FLEX\textsuperscript{A} exhibits a slightly better calibration performance compared to the other two model set-ups. This can partly be attributed to the lower number of parameters which leads, with the same number of samples, to a more exhaustive sampling of the parameter space and a smoother identification of Pareto optimal solutions. In addition, FLEX\textsuperscript{A} has the lowest number of imposed constraints, i.e. only the runoff coefficient and one parameter constraints, compared to FLEX\textsuperscript{B} and FLEX\textsuperscript{C}. This model set-up therefore allows more freedom in exploiting the parameter space to produce mathematically good fits between observed and modeled system response in the calibration period.

For the validation period, arguably more important for model evaluation, as in contrast to the calibration period, it gives information on model consistency (cf. Klemes, 1986; Andréassian et al., 2009; Euser et al., 2013) and predictive uncertainty, the performances of the three model set-ups exhibit quite different patterns (Table 4). The simplest model, the lumped FLEX\textsuperscript{A}, is characterized by the highest performance deterioration from calibration to validation. FLEX\textsuperscript{B} shows a better validation/calibration performance ratio than FLEX\textsuperscript{A}. Despite the expectation that increasingly complex models will have increasingly poor validation/calibration performance ratios, due to higher degrees of freedom, FLEX\textsuperscript{C} exhibited a more stable performance between calibration and validation.

In addition, the absolute performance of FLEX\textsuperscript{C} in the validation period is in general higher than the performances of FLEX\textsuperscript{A} and FLEX\textsuperscript{B} (Table 4). Although, strictly speaking, no meaningful comparison between Nash-Sutcliffe efficiencies from different periods can be made, these results nevertheless indicate that the most complex model set-up, i.e. FLEX\textsuperscript{C}, is the most consistent model-set-up with the lowest predictive uncertainty, which has important implications that will be discussed below. The explanation is that in spite of the high degree of process heterogeneity, the high number of constraints in FLEX\textsuperscript{C} prevents the calibration
algorithm to over-fit this complex model set-up, thus reducing the probability of seriously misrepresenting reality.

Table 4- The median model performances (in brackets their corresponding Pareto uncertainty intervals) and the area spanned by the uncertainty interval of the hydrograph derived from the Pareto optimal solutions of the constrained and calibrated model set-ups FLEX\(^A\), FLEX\(^B\) and FLEX\(^C\) for the three modeling objectives \(E_{NS}\), \(E_{NS,\text{log}}\), \(E_{NS,FDC}\) in the calibration and validation periods.

<table>
<thead>
<tr>
<th></th>
<th>(E_{NS})</th>
<th>(E_{NS,\text{log}})</th>
<th>(E_{NS,FDC})</th>
<th>95% uncertainty area [mm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLEX(^A)</td>
<td>Calibration</td>
<td>0.74 [0.51, 0.82]</td>
<td>0.74 [0.66, 0.80]</td>
<td>0.96 [0.95, 0.98]</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>0.66 [0.48, 0.78]</td>
<td>0.75 [0.70, 0.81]</td>
<td>0.97 [0.94, 0.98]</td>
</tr>
<tr>
<td>FLEX(^B)</td>
<td>Calibration</td>
<td>0.74 [0.60, 0.80]</td>
<td>0.73 [0.59, 0.79]</td>
<td>0.96 [0.94, 0.98]</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>0.72 [0.57, 0.79]</td>
<td>0.78 [0.64, 0.82]</td>
<td>0.96 [0.94, 0.98]</td>
</tr>
<tr>
<td>FLEX(^C)</td>
<td>Calibration</td>
<td>0.69 [0.60, 0.79]</td>
<td>0.66 [0.60, 0.67]</td>
<td>0.96 [0.94, 0.96]</td>
</tr>
<tr>
<td></td>
<td>Validation</td>
<td>0.68 [0.57, 0.78]</td>
<td>0.74 [0.66, 0.79]</td>
<td>0.98 [0.95, 0.98]</td>
</tr>
</tbody>
</table>

Figure 4- The observed hydrograph and the 95% Pareto uncertainty interval of the modeled hydrograph for constrained and calibrated parameter sets for the three different model set-ups (a) FLEX\(^A\), (b) FLEX\(^B\) and (c) FLEX\(^C\) for the two years (2008-2009) of validation period.

4-3- Comparison of “constrained but uncalibrated” and “constrained and calibrated” models:

The following comparison of the performances of FLEX\(^A\), FLEX\(^B\) and FLEX\(^C\) for “constrained but uncalibrated” and “constrained and calibrated” parameter sets focused on
\(E_{NS}\) only, for the reason of brevity (Figure 5). In Figures 5a and 5b the model performances based on the “constrained but uncalibrated” parameter sets, that satisfy the full set of constraints, are shown for the calibration and validation periods. As discussed in detail above, although uncalibrated, increasing the number of constraints from FLEX\(^A\) to FLEX\(^C\) increases the overall performance of the models while reducing uncertainty (Figures 5c and 5d; note that these are zoom-ins).

Figure 5e compares model performance based on constrained and calibrated parameter sets for the calibration period. As discussed earlier, it can be clearly seen that the simple lumped model, FLEX\(^A\), shows the best calibration performance with lowest uncertainty. However, when comparing the individual model performances of the constrained and calibrated models during the validation period (Figure 5f), it can be seen that FLEX\(^A\) not only shows the strongest performance deterioration compared to the calibration period but also that FLEX\(^A\) is also the model with the poorest performance in the validation period. This implies that although FLEX\(^C\) is the most complex model, the realism constraints imposed on this model generate the most reliable outputs when used for prediction, i.e. in the validation period. This strongly underlines that the widely accepted notion of complex models necessarily being subject to higher predictive uncertainty is not generally valid when the feasible parameter space can be well constrained based on assumptions of realistic functionality of a catchment.

In addition a second crucial aspect was revealed by comparing “constrained but uncalibrated” and “constrained and calibrated” models. It can be seen that, for the study catchment, a calibrated lumped model, FLEX\(^A\) (Figure 5f, left plot) can on average not outperform a more complex constrained but uncalibrated model, i.e. FLEX\(^C\) (Figure 5d, right plot). This has potentially important implications for selecting suitable parameter values for models applied in ungauged basins as it highlights the value of semi- and non-quantitative hydrological expert knowledge, even in the absence of reliable model regionalization tools and detailed soil or geological information, as discussed in detail below.

Figure 5- Model performance (\(E_{NS}\)) based on constrained but uncalibrated (a-d) and constrained and calibrated (e-f) parameter sets for calibration (2002-2005) and validation (2006-2009) periods for the three different model set-ups FLEX\(^A\), FLEX\(^B\) and FLEX\(^C\). Note that (c) and (d) are zoom-ins of (a) and (b).
4-4- Comparison of flow contributions from different model components:

The comparison of the fluxes generated from the individual model components in the three model set-ups helps to assess to which degree the model internal dynamics reflect the modeler’s perception of the system and thus to a certain degree the realism of the models.

Fast and slow responses of each tested model set-up have been visually illustrated in Fig. 6. Predominance of slow responses of all the three models are indicated by green color; predominance of fast responses of FLEX^A, fast responses of the remainder of the catchment of FLEX^B and fast responses of hillslope of FLEX^C is indicated by red color; wetland fast responses of FLEX^B and FLEX^C are indicated by predominance of blue color.

The colors in Fig. 6 are an illustration using three colors (red, green and blue) for the models’ responses based on their weight of contribution to the modeled runoff. As it can be seen in Fig. 6a the fast component of FLEX^A is dominant just during peak flows and even the recession shortly after peak flows are accounted for mainly by ground water. Analysis of the individual model components computed by Pareto optimal parameter sets (not shown here for brevity), indicates that some Pareto optimal parameters can generate peak flows by predominant contributions from slow responses while fast reaction is tend to be inactive during these events.

In accordance with the perception of the system that wetlands are predominantly responsible for peak flows during dry conditions, Fig. 6b and c show that wetland fast responses in FLEX^B and FLEX^C control the rapid response during wetting up periods (dry to wet transition), before hillslope fast processes become more important at higher moisture levels. When the system is saturated the hillslope contribution to modeled runoff becomes significantly higher compared to the wetland response. Note that the response of the wetland may not correspond well to individual events, as a consequence of the fact that the corresponding constraint was set for an aggregated period.
Figure 6- Comparison between mean proportions of Pareto members for model components of the three model set-ups in part of the calibration periods (August 2002- June 2003) (a) FLEX$^A$, (b) FLEX$^B$, and (c) FLEX$^C$. The green color indicates the relative contribution of the slow reservoir for the three different models. Red indicates relative contribution from the fast components, i.e. fast reservoir in FLEX$^A$, fast reservoir of the remainder of the catchment in FLEX$^B$ and fast reservoir of hillslope of FLEX$^C$. The blue color indicates the relative contribution of fast wetland component of FLEX$^B$ and FLEX$^C$.

4-5- General discussion:

The results of this study quite clearly indicate that discretizing the catchment into hydrological response units (HRUs) and incorporating expert knowledge in model development and testing is a potentially powerful strategy for runoff prediction, even where insufficient data for model calibration (e.g. Koren et al., 2003; Duan et al., 2006; Winsemius et al., 2009) or only comparatively unreliable regionalization tools are available (e.g. Wagener and Wheater, 2006; Bárdossy, 2007; Parajka et al., 2007; Oudin et al., 2008; Laaha et al., 2013). It was found that the performance and the predictive power of a comparatively complex uncalibrated conceptual model, based on posterior parameter distributions obtained merely from relational, semi- and non-quantitative realism constraints inferred from expert knowledge, can be as efficient as the calibration of a lumped conceptual model (Fig. 5). Typically it is expected that, if not warranted by data, models with higher complexity suffer from higher predictive uncertainty. As stated by Beven (2001): “More complexity means more parameters, more parameters mean more calibration problems, more calibration problems will often mean more uncertainty in the predictions, particularly outside the range of the calibration data”. Thus, more parameters would allow better fits of the hydrograph but would not necessarily imply a better and more robust understanding of catchment behavior or more reliable predictions.

A complex model may include many processes, i.e. hypotheses, which can usually not be rigorously tested with the available data. However, a wide range of previous studies has demonstrated that hydrologically meaningful constraints can help to limit the increased uncertainty caused by incorporating additional processes, i.e. parameters (e.g. Yadav et al., 2007; Zhang et al., 2008; Kapangaziwiri et al., 2012). These studies generally include a large set of catchments and try to relate model parameters to catchment characteristics. Although regional constraints are important, the importance of expert knowledge on the catchment scale, which leads to better understanding of hydrological behavior is highlighted in this study.

In a similar attempt, Pokhrel et al. (2008, 2012) demonstrated use of regularization for model parameters and reduction of model parameter space dimensionality by linking model parameters using super-parameters to catchment characteristics. However, no explicit hydrological reasoning is typically applied for such “regularization rules” (e.g. Pokhrel et al., 2012). On the other hand, Kumar et al. (2010, 2013) parameterize and successfully regionalize their models using empirical transfer functions with global parameters, developed from extensive literature study and iterative testing in a large sample of catchments. In contrast, the use of relational parameter- and process constraints, as presented in this study, is based on semi-quantitative, hydrologically explicit and meaningful reasoning avoiding the need for empirical transfer functions to link catchments characteristics and model parameters.

Including prior knowledge for parameters of physically-based models for estimating runoff in ungauged basins was quite successfully investigated in the past (e.g. Otte and Uhlenbrook...
2004, Vinegradov et al., 2011, Fang et al., 2013, Semenova et al., 2013). These studies specifically indicate that calibration can be replaced by prior information which is a significant contribution to Predictions in Ungauged Basins (PUB). While physically-based models need detailed information of catchment behavior for model parameters, the here proposed semi-distributed conceptual modeling framework, exploiting relational constraints, can be more efficiently set up using the least prior information necessary. In this study, the performances and uncertainties of the three tested model set-ups for constrained but uncalibrated parameters indicate the potential of the presented FLEX-TOPO framework for Predictions in Ungauged Basins (PUB). Hence, this framework can efficiently use expert knowledge for improving model parameter value selection in complex conceptual hydrological models, not only to increase model performance but also to reduce model predictive uncertainty even in the absence of calibration.

It should be noted that the model set-ups suggested within the FLEX-TOPO framework are hypotheses that still need to undergo further tests, ideally confronting them with additional, system internal information, such as groundwater dynamics (e.g. Seibert and McDonnell 2003, 2013; Fenicia et al., 2008) or tracer data (e.g. Madsen 2006, Campbell et al., 2012; Birkel et al., 2011; Hrachowitz et al., 2013a). To make more efficient use of relational constraints, model sensitivities to these constraints need to be evaluated in the future. It is also emphasized that the constraints introduced in this study are based on the authors’ subjective understanding of catchment behavior and can and should be discussed further. However, we would like to stress the notion that reaching an agreement on the relations between parameters and fluxes in different landscape units is potentially much easier than finding the most adequate parameter values together with associated uncertainties for a conceptual model based on field observations or available data on geology or soil types.

5- Conclusion:

In this study it was tested if a topography-driven semi-distributed formulation of a catchment-scale conceptual model, conditioned by expert knowledge based relational parameter- and process constraints, can increase the level of process realism and predictive power while reducing the need for calibration compared to a lumped model set-up.

It was found that:

(1) A constrained but uncalibrated semi-distributed model exhibited an equivalent performance compared to a constrained and calibrated lumped model when used for prediction. This illustrates the potential value of the combined use of higher complexity models and relational constraints for predictions in ungauged basins, where no calibration data are available.

(2) The use of relational parameter- and process constraints in model calibration ensured a high degree of process realism. Thus, in spite of the comparatively high complexity, the overall model performance and uncertainty showed better prediction results than for a lumped model. It was shown that higher model complexity therefore does not necessarily entail reduced predictive power.

(3) Semi-distributing a model on the basis of HRUs derived from topographic data can increase model internal consistency as it better accounts for fundamentally different runoff generating processes active at different wetness conditions.

(4) In contrast to constraints based on more detailed and frequently unavailable regionalization relationships or catchment data, such as geology and soils, hydrologically meaningful relational constraints can be applied with a minimum amount of information.
Acknowledgment: Anke Luijben’s assistance is acknowledged for the data analysis of the study area. Shervan Gharari is funded during his PhD program by the Fonds National de la Recherche (FNR) of Luxembourg with Aides à la Formation-Recherche (AFR) project number of 1383201.

6- References:


Kumar, R.; Livneh, B. & Samaniego, L. Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme Water Resources Research, 49(9), 2013b.


