Hydrology and Earth System Sciences

Open Access

Discussions

The influence of conceptual model structure on model performance: a comparative study for 237 French catchments

W. R. van Esse[1,*], C. Perrin[2], M. J. Booij[1], D. C. M. Augustijn[1], F. Fenicia[3,4], and F. Lobligeois[2]

[1]University of Twente, Faculty of Engineering Technology, Department of Water Engineering and Management, Enschede, the Netherlands
[2]Irstea, Hydrosystems and Bioprocesses Research Unit (HBAN), Antony, France
[3]Delft University of Technology, Faculty of Civil Engineering and Geosciences, Water Resources Section, Delft, the Netherlands
[4]Public Research Centre – Gabriel Lippmann, Belvaux, Luxembourg
[*]now at: Nelen & Schuurmans, Utrecht, the Netherlands

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

HESSD

10, 5457–5490, 2013

The influence of conceptual model structure on model performance

W. R. van Esse et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀◀ | ▶▶

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Abstract

In hydrological studies models with a fixed structure are commonly used. For various reasons, these models do not always perform well. As an alternative, a flexible modelling approach could be followed, where the identification of the model structure is part of the model set-up procedure. In this study, the performance of twelve different conceptual model structures from the SUPERFLEX framework with varying complexity and the fixed model structure of GR4H were compared on a large set of 237 French catchments. The results showed that in general the flexible approach performs better than the fixed approach. However, the flexible approach has a higher chance of inconsistent results when implemented on two different periods. The same holds for more complex model structures. When for practical reasons a fixed model structure is preferred, this study shows that models with parallel reservoirs and a power function to describe the reservoir outflow perform best. In general, conceptual hydrological models perform better on large or wet catchments than on small or dry catchments. The model structures performed poorly when there was a climatic difference between the calibration and validation period, for catchments with flashy flows or disturbances in low flow measurements.

1 Introduction

Building efficient hydrological models remains a challenging issue, despite the huge efforts by the community to develop and improve models since the pioneering work of Linsley and Crawford (1960). This has resulted in a large variety of models and modelling approaches, ranging from lumped conceptual models to distributed physically-based models. Making an exhaustive literature review of modelling approaches is out of the scope of this paper, but the reader may refer to the reviews by Singh and Frevert (2002a,b) or Pechlivanidis et al. (2011). Here we will only focus on two types

of modelling approaches, namely the *fixed* and *flexible* modelling approaches. In this paper, these two approaches are limited to the case of lumped conceptual models.

The fixed modelling approach assumes that a given model structure can be applied in various contexts and conditions. As examples, one may mention TOPMODEL
5 (Beven et al., 1995), HBV (Lindström et al., 1997), Xinanjiang (Zhao and Liu, 1995), NAM (DHI, 2008) or GR4J (Perrin et al., 2003) to name a few. Although these models have been continuously improved and adapted over the years, the core of their structure remained more or less similar and it was assumed to be general enough to be applicable in a variety of basins. For example, Bergström (1995) provides a detailed
10 review of the applications of the HBV model on catchments over the five continents. Hence, an end-user would take the model as it is proposed by the developers and apply it on his case study. This approach has several advantages. It saves time to the end-user who can take the model from the shelf, without bothering about the tedious process of developing a model. This is particularly useful when one wishes to
15 apply the model on many catchments, as it may be the case in operational conditions. This is also interesting on catchments where data may be too limited to develop a full model from scratch. Despite its attractiveness, the fixed modelling approach is based on the assumption that a given model structure can be applied in many different catchments. However, this may not be the case. Processes included in the model may not
20 correspond to the dominant ones occurring in the studied catchment. This may cause structural errors and strongly limit model efficiency. Hence, the model could prove to be unable to actually reproduce catchment hydrological behaviour. Several examples of fixed models' failures have been described in the literature (e.g. Le Moine et al., 2008; Goswami and O'Connor, 2010; Refsgaard and Hansen, 2010).
25 The flexible modelling approach assumes that no one-size-fits-all model structure exists and that for each catchment, a set of alternatives should be considered instead. Given that model structural errors are often a first-order source of uncertainty, the flexible approach proposes to search among a set of predefined structures or model components for the one that fits catchment's behaviour best. The main advantage

of this approach is to give room to reduce structural uncertainty when applying the model, which should result in more robust model applications. Models such as RRMT (Wagener et al., 2001), MMS (Leavesley et al., 2002), FLEX (Fenicia et al., 2008), FUSE (Clark et al., 2008) and SUPERFLEX (Fenicia et al., 2011; Kavetski and Fenicia, 2011) are aimed at increasing model flexibility in the model construction phase of a modelling study. However this approach may be more time consuming for the end-user. It may also end up with several model structures performing equivalently.

Several researchers have already compared the performance of different model structures. The results of these studies have provided different insights. For instance, Chiew et al. (1993) indicated that relatively simple model structures can be used for larger time scales (months, years), where Refsgaard and Knudsen (1996) concluded that models of different complexity performed equally well. Perrin et al. (2001) showed that complex models outperform simple ones in calibration but not in validation. Reed et al. (2004) found that a lumped model, used as benchmark, generally showed equivalent or better overall performance than distributed models. More recent studies of Breuer et al. (2009) and Seiller et al. (2012) showed slight differences between models of differing complexity. In most of these intercomparisons and in hydrological modelling studies in general, there is a common aim to find an appropriate model for a particular purpose or condition, considering hydrological signature, catchment type and spatial and temporal scale (Rogers, 1978; Moussa and Bocquillon, 1996; Booij, 2003).

The objective of this study is to evaluate the influence of model structure on model performance for different catchment properties. It extends the work of Fenicia et al. (2011), Kavetski and Fenicia (2011) and Fenicia et al. (2013) carried out on only a few catchments. Based on a large set of 237 French catchments, this study tries to find more general conclusions on connections between model structure, catchment characteristics and model performance and ultimately help reducing errors or subjectivity in model conceptualization.

The questions we wish to answer are:

- What is the influence of different model structures on average model performance?

- What is the influence of different catchment characteristics on the relationship between model structure and performance?

- What are the differences in performance between a fixed and flexible model structure?

To address these questions, we will use SUPERFLEX as a modelling framework to construct and compare different model structures. SUPERFLEX is an example of a flexible modelling approach and is based on 12 individual model structures. The GR4H model is tested as an example of a fixed modelling approach. This hourly model and its daily version, GR4J, are a widely applied model that has shown good average performance on many catchments in different parts of the world (Le Moine et al., 2007; Coron et al., 2012; Perrin et al., 2003; Valéry et al., 2010), but for which structural inadequacy or the lack of flexibility in model structure is suggested as one of the main reasons for its failures (Kavetski and Fenicia, 2011; Perrin et al., 2003; Le Moine, 2008; Wagener, 2003; Andréassian et al., 2010).

In Sect. 2, we will present the data and models used, as well as the evaluation methodology. Section 3 discusses the general results, followed by a discussion and conclusions of this work in Sects. 4 and 5 respectively.

2 Material and methods

2.1 Data

This study is based on a large data set of 237 catchments spread throughout France (Fig. 1). They represent a large variety of conditions in terms of physical characteristics (size, geology, etc.) and climate (Table 1). Precipitation, potential evapotranspiration (PE) and discharge data were available at hourly time steps from 1997 to 2006 for

all catchments. Precipitation originates from the reanalysis produced by Metéo-France (Tabary et al., 2012) and discharge data were taken from the French data base Banque Hydro (MEDD, 2007). Physical characteristics of the 237 catchments were also available (Bourgin et al., 2011; European Environment Agency, 2006).

2.2 Catchment classification

To investigate the performance of the model structures on different types of catchments, the 237 catchments were classified based on four catchment characteristics: area, Wetness Index, permeability and the ratio of runoff coefficients in summer and winter ($RC_{S/W}$). The latter is considered a more integral property of climate and catchment characteristics and distinguishes groundwater dominated catchments from catchments where runoff occurs more directly. $RC_{S/W}$ was calculated using:

$$RC_{S/W} = \frac{RC_S}{RC_W} = \frac{\overline{P}_S/\overline{Q}_S}{\overline{P}_W/\overline{Q}_W} \tag{1}$$

where $\overline{P}_S$ and $\overline{Q}_S$ (resp. $\overline{P}_W$ and $\overline{Q}_W$) are the mean precipitation and discharge during three summer (resp. winter) months (July–September, resp. January–March) over ten-year time series (1997–2006). This ratio is best explained by two extreme cases. (1) High $RC_{S/W}$: when runoff compared to rainfall is high in summer, it is likely that the river is fed by additional groundwater that was stored in winter. This means that some rainfall in winter did not go to runoff and thus lowering the runoff coefficient in winter. This case is classified as "groundwater dominated runoff". (2) Low $RC_{S/W}$: when runoff compared to rainfall in summer is low, water might be lost to potential evapotranspiration and little water from storage will flow in the river. Rainfall in winter is then likely to flow in the river more directly as there is little storage. This case is classified as "direct runoff".The Wetness Index (also referred to as Aridity Index, Middleton and Thomas, 1992) was calculated as the ratio between mean precipitation and PE over the ten-year time series.

For catchment area, Wetness Index and RC$_{S/W}$, the catchments were divided into three classes with approximately equal number of catchments. Table 2 shows the ranges of each class and the number of catchments per class. The qualitative scales (e.g. small, medium and large for catchment area) chosen here may not correspond to the scales often found in the literature for these four characteristics, but were chosen here to distinguish between catchments with below-median, median or above-median characteristics. Permeability was determined following the information provided by the European Environment Agency (2006) based on the type of bedrock underlying the catchments.

2.3 Models

The SUPERFLEX modelling approach that includes 12 different model structures, and the GR4H model were used to investigate the influence of model structure on model performance. All thirteen structures are lumped and use the same rainfall and PE over the whole catchment as inputs and generate discharge as output.

2.3.1 SUPERFLEX

Twelve structures (SF01–SF12) as proposed by Fenicia et al. (2013) are used in the SUPERFLEX framework. They cover a broad range of model complexities (Fig. 2). Starting from a very simple structure (SF01), the complexity is gradually increased by adding reservoirs, lag-functions and junction elements. In this way, the influence of individual components can be assessed.

In the SUPERFLEX structures, rainfall (Pt) and potential evaporation are used as inputs. Potential evapotranspiration is systematically corrected with a calibrated ratio Ce to fulfil the water balance. Actual evapotranspiration (noted Ei, Eu or Ef for the interception, unsaturated zone and fast reservoirs, respectively) empties one or two reservoirs in each structure.

SF01 consists of a single fast reservoir (FR) with a residence time Kf and a power function denoted by α. SF02 consists of a single reservoir as well, but it is an unsaturated zone reservoir (UR) that uses both a linear and a threshold type power function (β) to describe the outflow from the reservoir.

In structures SF03 to SF05, an unsaturated zone reservoir (UR) is connected in series to the fast reservoir. These three structures vary in the functions by which the flow to and from the reservoirs is described and the number of calibrated parameters. SF05 uses power functions to describe flow from both reservoirs, denoted by α and β, and a lag-function that distributes flow over multiple time steps. Structures SF06, SF07, SF09, SF10 and SF11 all use three reservoirs, but of different types and with the introduction of more complex connections and functions. SF07 is the first structure with a riparian zone reservoir (RR) connected in parallel allowing for more independent flow along two flow paths. SF08 is a simple structure with only two parallel reservoirs: a slow reservoir (SF) is introduced with a residence time Ks independent to that of the fast reservoir. From this structure onward, complexity is again increased up to the introduction of an interception reservoir (IR) with a threshold type function in structure SF12. Parameters M and D are ratios that divide water between different reservoirs. Table 3 shows an overview of the model structures and the type of functions used. A more detailed description of these structures can be found in Fenicia et al. (2013).

2.3.2 GR4H

The GR4H model (Le Moine, 2008) is used as an example of a fixed model structure. In GR4H (see diagram in the bottom right corner of Fig. 2), rainfall and PE are subtracted to find the net precipitation P_n or net evaporation E_n. P_n is partitioned between storage into a soil moisture accounting reservoir $S(P_s)$, and effective rainfall ($P_t = P_n - P_s$). This reservoir is depleted by a percolation function Perc that is added to effective rainfall. Effective rainfall is then routed to the outlet via a two-branch routing module. The first branch (10 % of effective rainfall) is routed by a single unit hydrograph. The other 90 % are routed by a unit hydrograph and a non-linear reservoir named routing store R.

A water exchange function F is applied to the two flow components, to simulate import or export of groundwater with the underlying aquifer or neighbouring catchments.

2.4 Model evaluation

2.4.1 Evaluation procedure

All model structures were calibrated and validated using the split sample test (Klemes, 1986), in which ten years of the available data were split into two independent sub-periods (1997–2001 and 2002–2006). Calibration was performed on each sub-period followed by validation on the other sub-period. To reduce initialisation problems, an initial value was given to the level in the reservoirs (as a fraction of reservoir capacity) and three years of data (1994–1996 and 1999–2001 resp.) preceding the test sub-periods, were used for model warm-up. Note that the data used for warm-up prior to 1997 originate from the SAFRAN reanalysis (Vidal et al., 2010).

Parameter optimization was done according to the methods described in Kavetski and Fenicia (2011) using the Bayesian Total Error Analysis (BATEA) framework (Kavetski et al., 2006; Kavetski and Evin, 2011). Here, the objective function is a weighted least square (WLS) scheme. It accounts for the knowledge that prediction errors are commonly larger for high flows and intends to give more balance between high and low flows. A quasi-Newton optimization was applied using twenty different initial values across the parameter space. This local optimization method was shown to be effective and efficient using a limited number of multistarts (i.e. twenty) on a smoothed parameter space (Kavetski and Kuczera, 2007).

Model performance was evaluated based on the validation results using the four criteria described in the next section. The thirteen model structures were compared using the average score of the evaluation criteria for all catchments. For the comparison of the flexible approach with the fixed structure of GR4H, consistency rules were applied to ensure that each catchment is represented by a structure and parameter set consistently over the calibration and validation periods.

2.4.2 Evaluation criteria

The four evaluation criteria (CR1 to CR4, by Eqs. (2)–(5)) focus on different aspects of model performance (high flow, low flow, volume error and variability of predictions):

$$CR1 = \frac{CR1^*}{2 - CR1^*} \quad \text{with } CR1^* = 1 - \frac{\sum_{i=1}^{N}(Q_{\text{obs},i} - Q_{\text{sim},i})^2}{\sum_{i=1}^{N}(Q_{\text{obs},i} - \overline{Q_{\text{obs}}})^2} \tag{2}$$

$$CR2 = \frac{CR2^*}{2 - CR2^*} \quad \text{with } CR2^* = 1 - \frac{\sum_{i=1}^{N}\left(\frac{1}{Q_{\text{obs},i}+\varepsilon} - \frac{1}{Q_{\text{sim},i}+\varepsilon}\right)^2}{\sum_{i=1}^{N}\left(\frac{1}{Q_{\text{obs},i}+\varepsilon} - \overline{\frac{1}{Q_{\text{obs}}+\varepsilon}}\right)^2} \tag{3}$$

$$CR3 = \frac{CR3^*}{2 - CR3^*} \quad \text{with } CR3^* = 1 - \left| \sqrt{\frac{\sum_{i=1}^{N} Q_{\text{sim},i}}{\sum_{i=1}^{N} Q_{\text{obs},i}}} - \sqrt{\frac{\sum_{i=1}^{N} Q_{\text{obs},i}}{\sum_{i=1}^{N} Q_{\text{sim},i}}} \right| \tag{4}$$

$$CR4 = \begin{cases} -1 + 2\frac{\sigma_{\text{sim}}}{\sigma_{\text{obs}}}, & \sigma_{\text{obs}} > \sigma_{\text{sim}} \\ -1 + 2\frac{\sigma_{\text{obs}}}{\sigma_{\text{sim}}}, & \sigma_{\text{obs}} < \sigma_{\text{sim}} \end{cases} \tag{5}$$

where Q_{obs} and Q_{sim} represent the observed and simulated discharge respectively, at time step i, N is the number of time steps, the overbar represents an average over the selected period, ε is a small constant (one-hundredth of mean flow, see Pushpalatha et al., 2012) and σ is the standard deviation over the selected period.

CR1* is the well-known Nash–Sutcliffe efficiency (Nash and Sutcliffe, 1970) which is most sensitive to peaks in discharge (Perrin et al., 2003). CR2* is the Nash–Sutcliffe efficiency based on the inversed discharge emphasizing low flow errors (Pushpalatha et al., 2012). CR3* is based on the Relative Volume error and thus emphasizes any error in the water balance between observed and simulated discharge (Perrin et al., 2003). CR1* to CR3* have values between 1 (perfect fit) and $-\infty$ and are transformed

to a value between 1 and −1 to avoid the influence of very low negative values on the calculation of mean performance (Mathevet et al., 2006; Pushpalatha et al., 2012).

The fourth criterion (CR4) is the ratio between standard deviations of observed and simulated discharges with a maximum value of 1 meaning the simulated discharge was able to reproduce the variability in the observed discharge and a minimum value of −1 when the difference in standard deviation becomes very large (Gupta et al., 2009). Since all the selected criteria are non-dimensional, the average of CR1–CR4 over the validation periods was used as the measure of overall performance to compare model structures.

2.4.3 Consistency

The stability of the model between the two calibration periods for a given catchment was evaluated by checking for parameter and structural consistency. Here, a model structure was considered parametrically inconsistent when at least one of the parameters departed by more than 50 % of the average between the two periods. In that case, this structure was left out of the final comparison.

In the case of SUPERFLEX, structural inconsistency was considered to occur when the best model structure identified on the two calibration periods was not the same. Therefore, only if a SUPERFLEX structure scores within the best 10 % in both calibration periods (of one catchment), it is considered consistent and eligible for comparison. The best SUPERFLEX structure for a catchment is the simplest consistent structure within 10 % of the highest scoring structure (see "Rank" in Table 3). When two equivalently performing structures have the same number of parameters, then the one that has the lower number of functions is preferred. Here complexity is measured by the number of parameters (see Table 3).

Thanks to the double model calibration and validation procedure and the consistency rules, models that give efficient and consistent results on a catchment could be identified. Despite being somewhat arbitrary, the consistency rules have made the model evaluation more severe.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

3 Results

3.1 Average performance of the thirteen individual structures

Figure 3 shows the distribution of performance for all model structures on the 237 catchments for the two validation sub-periods. The figure shows that the seven best performing model structures (GR4H, SF04–SF07, SF11 and SF12) show very similar average performance despite their structural differences. Six of the model structures (SF01, SF02, SF03, SF08, SF09 and SF10) perform poorly compared to the best ones and do not seem to be good candidate structures for a fixed-structure approach. By comparing the performance and complexity, one can see that structures with an unsaturated zone reservoir or a power function perform on average considerably better than structures without these components.

The calibrated values of the power function parameter α (describing the outflow of the fast reservoir FR) show a wide range for the different catchments and a high correlation between the two periods (not shown here; see Van Esse, 2012, for more details). This indicates that this is an effective parameter to calibrate. The results are similar for power function parameter β (describing the outflow of the unsaturated zone reservoir UR). However, for structures with both power functions, the values of power β are clustered around 1. This indicates that adding a second power function to the structure is far less effective than the first one and that the power function at the outflow of a reservoir at the end of a structure (like FR) is more effective.

Comparing several pairs of structures showed that the lag-function and interception reservoir (IR) do not increase performance for most catchments. For the riparian zone reservoirs (RR), the analysis shows larger differences: for some catchments this reservoir does increase performance while for others, the effects are negative. These results show that, at least within the SUPERFLEX framework, the lag-function and the interception and riparian zone reservoirs do not increase model performance on average, which questions their usefulness.

The average performance and performance range of the GR4H model are close to those of models SF04–SF07. The fixed power functions describing reservoir outflow in GR4H are expected to be important components, just as the power functions in the SUPERFLEX structures. The more complex models SF11 and SF12 are more robust by limiting the number of strong model failures, which is very valuable. However, on average, they are not able to outperform the simpler models (including GR4H).

3.2 Performance across catchment classes

To investigate the effect of catchment characteristics on model performance, average performance is analysed for four catchment characteristics: catchment area, Wetness Index, permeability and the ratio of runoff coefficients in summer and winter.

Figure 4a shows that model performance on large catchments is generally better than that on small catchments. Apparently, it is easier to model the rainfall–runoff relationship in catchments where hydrological processes are mixed and have a smoother response. This corroborates previous findings by Merz et al. (2009) on a large set of Austrian catchments.

Figure 4b shows that model performance on wet catchments is generally better than that on dry catchments. The difference between dry catchments and the catchments that are classified as wet or moist is large for all structures. This is in agreement with literature showing that dry catchments are generally more difficult to model due to the higher non-linearity in processes (e.g. Atkinson et al., 2002). Only SF11 and SF12 show little difference in performance over the three classes (however their performance is still worse for dry catchments). The parallel fast and slow reservoirs plus the unsaturated zone reservoir with a power function seem to be better able to simulate both wet and dry catchments.

Figure 4c shows that model structures with two reservoirs in series (SF03–SF07) perform better on impermeable catchments than on semi-permeable or permeable catchments. Adding a parallel slow reservoir in structures SF08 to SF12 decreases performance on impermeable catchments. The performance is again somewhat better

for SF11 and SF12, as these structures likely benefit from a power function that was introduced in these structures. GR4H shows the same order in efficiencies over the permeability classes as the structures with reservoirs in series (SF04–SF06). The GR4H structure does include a parallel flow path, but there are some distinct differences: one path is without a reservoir and it is not possible to calibrate them independently since the division over the flow paths is fixed and both paths use routing unit hydrographs based on the same parameter x_4.

The $RC_{S/W}$ classification in Fig. 4d separates the structures in an interesting way. Structures SF04 to SF07 perform much worse on the groundwater driven catchments than on the other catchments, while SF09 and SF10 simulate these catchments best of the three classes. This reversed order of performance can be explained by the parallel slow reservoir component in SF09 and SF10. This component allows independent fast and slow flow, hence enabling a slow groundwater component while maintaining the ability to produce high flow in case of a storm event. The parallel riparian zone reservoir in SF07 does not have this effect because it is not connected to the unsaturated zone reservoir and only a maximum of 20 % of rainfall is routed through this reservoir.

The difference in performance between structures with reservoirs in series or parallel was already established by e.g. Jakeman and Hornberger (1993), and more recently by Kavetski and Fenicia (2011) and Fenicia et al. (2013). Jakeman and Hornberger found that the most commonly identified configuration for a rainfall-runoff model is two storages in parallel. This difference is clearly observed when comparing the performance per catchment of SF05 and SF11 using the $RC_{S/W}$ classification. Figure 5 shows the average performance of SF05 against that of SF11, two identical models apart from the use of a parallel slow reservoir in SF11. The figure confirms that the more complex SF11 does not perform better overall than SF05 (mean performance of 0.57 and 0.56, respectively), but that it significantly improves the performance for most groundwater dominated catchments. This shows the value of the flexible distribution of flow (ratio D) and the independent residence times (Kf and Ks) of the parallel reservoirs for these type of catchments.

3.3 Comparison of fixed and flexible modelling approaches

Figure 6 shows the performance distribution of the flexible and fixed approaches with and without accounting for the cases where inconsistent results between calibration periods occurred. Without the consistency rules, the flexible approach performs better on the catchment set, which is a clear advantage of the flexible approach. However, when the inconsistent catchments are removed from the set, both approaches perform very similar. The flexible approach is inconsistent more often (90 times) than the fixed approach (67), possibly caused by the multiple model structures in the flexible approach resulting in structural inconsistency besides parametric inconsistency (in both approaches). Note that the comparison of performance when accounting for consistency rules is made on two different sub-sets, hence results are not completely comparable. However, the mean results on the overlapping (116 catchments) are very similar, for GR4H 0.62 and SUPERFLEX 0.61.

It must be noted that the choice for the 50 % threshold for parameter inconsistency and the 10 % range for structural inconsistency considerably influence these results: e.g. a 5 % structural inconsistency range would render 134 catchments with inconsistent results for the flexible approach while a 25 % range results in only 26 inconsistent catchments. Also, many of the SUPERFLEX structures are more prone to parameter inconsistency as they have more parameters.

To find out more about which of the flexible model structures perform well, we look at the number of times each structure is selected as best. Table 4 shows that SF04, SF09 and SF11 are selected most often. These structures are successful because of their use of only the most effective components. Very simple structures and structures with only small differences from these best three are selected much less often. This shows a disadvantage of hypothesising a model structure a priori: many structures give similar results which make choosing the best one difficult. Combining all 13 structures, GR4H is selected the largest number of times (138) showing the good generality of the model. For 63 other catchments, however, one of the twelve SUPERFLEX structures

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

is favoured. For these catchments, choosing a customized structure gives a better performance.

4   Discussion

Fenicia et al. (2011), Kavetski and Fenicia (2011) and Fenicia et al. (2013) performed
similar studies, but only on a few catchments. The results of this study on a large set
of 237 French catchments extend the previous works in two ways. First, the results
show similar correspondences between catchment characteristics and model structure, especially in the case of groundwater dominated catchments. Results obtained
here are also consistent with results shown by Perrin et al. (2001), who showed that
choosing a specific structure for a given catchment significantly improved the performance compared to using a fixed structure. Second, this research has shown some
of the limitations of using the SUPERFLEX approach on a large scale: large similarity in performance of different models makes selection of the best model difficult and
thus, reduces the amount of knowledge on connections between model structure and
hydrological processes that can be gained. This becomes even more difficult given
the large differences in results between the two calibration periods and the demand
that the best model structure should work well for that specific catchment under all
possible conditions (see also best-compromise model, Seiller et al., 2012). Furthermore, the SUPERFLEX approach poses a practical limitation in that modellers have to
deal with a different model for every catchment, and thus becoming experienced with
the approach takes more time. The general advantages and disadvantages of model
structures and components and their relevance for different catchment types can only
be studied based on a large set of catchments (Andréassian et al., 2006).

For the calibration, a weighted least square (WSL) scheme was used putting more
emphasis on low flow than the usual least square approach. However, in spite of this
improved objective function, all model structures scored poorly on the low flow criterion (CR2[*]). As noted by Thyer et al. (2009), the WSL is expected to only be a small

improvement compared to a standard least square method. These limited results on low flows are a clear indication of the difficulty for models to produce satisfactory low-flow simulations even with a variety of modelling options, which corroborates results by Pushpalatha et al. (2011).

Further analysis (not detailed here for the sake of brevity) of the results indicated three possible reasons for inconsistent or poor overall performance: (i) severe differences in climate between calibration and validation periods in catchments with a large groundwater component in the discharge, (ii) flashy flow in response to extreme rainfall events predominantly occurring in the Mediterranean area, and (iii) disturbed flow measurements, particularly during low-flow conditions. Strong relations between poor model performance and catchment characteristics available in this research could not be established and are left for further research. Understanding the reasons for these model failures in more detail can prove to be vital for knowledge and model development. For example, in depth diagnosis of model parameter transposability in time was investigated by several authors recently (Vaze et al., 2010; Merz et al., 2011; Coron et al., 2012).

Considering the components of the SUPERFLEX structures, new systematic tests could be additionally carried out to further investigate the sensitivity of model results to the various modelling options identified as relevant here (power functions and parallel flow paths). Interestingly, in the case of the GR4H model and its daily version GR4J, recent investigations independent of the present study corroborate to some extent the results of this research: the addition of a free parameter in the formulation of the water exchange function significantly improves the simulation of low flows (Le Moine, 2008), and the addition of a second routing store (with an additional parameter) in parallel to the existing ones also significantly improves results (Pushpalatha et al., 2011).

5 Conclusions

This study found that relatively simple model structures with some key components can lead to a good simulation of the discharge from a catchment, which corroborates previous findings (Jakeman and Hornberger, 1993; Perrin et al., 2001). The analysis of thirteen individual lumped conceptual model structures on a large catchment set with different characteristics showed that:

- increasing model complexity does not always lead to higher performance however, complex structures perform poorly for fewer catchments;

- conceptual hydrological models generally perform better on large or wet catchments than on small or dry catchments;

- the use of a power function to describe reservoir outflow significantly increases mean model performance compared to models that use linear functions;

- independently calibrated parallel reservoirs increase model performance in permeable, groundwater dominated catchments; and

- a lag-function between reservoirs or an interception reservoir is unlikely to lead to a significant increase in model performance.

On the whole catchment set, the flexible modelling approach provides better average results than the fixed modelling approach. Generally, selecting the best model structure for each catchment gives the best results. However, the results of the two approaches are equivalent when applying consistency rules on parameters and structures for two calibration periods. The risk of equifinality between structures and differences in model performance between periods mean that care should be taken when selecting the best model structure and shows a disadvantage of hypothesizing a model structure a priori. Therefore, appropriate model structure evaluation schemes should be designed to enhance the robustness of the selected model structures and avoid unexpected model failures (Andréassian et al., 2009, 2010).

References

5 Andréassian, V., Hall, A., Chahinian, N., and Schaake, J.: Introduction and synthesis: why should hydrologists work on a large number of basin data sets?, IAHS Publ., 307, 1–5, 2006.

Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M.-H., and Valéry, A.: HESS Opinions "Crash tests for a standardized evaluation of hydrological models", Hydrol. Earth Syst. Sci., 13, 1757–1764, doi:10.5194/hess-13-1757-2009, 2009.

Andréassian, V., Perrin, C., Parent, E., and Bárdossy, A.: The court of miracles of hydrology: can failure stories contribute to hydrological science?, Hydrolog. Sci. J., 55, 849–856, doi:10.1080/02626667.2010.506050, 2010.

15 Atkinson, S. E., Woods, R. A., and Sivapalan, M.: Climate and landscape controls on water balance model complexity over changing timescales, Water Resour. Res., 38, 1314, doi:10.1029/2002WR001487, 2002.

Beven, K. J., Lamb, R., Quinn, P., Romanowicz, R., and Freer, J.: TOPMODEL, in: Computer Models of Watershed Hydrology, edited by: Singh, V. P., Water Resources Publications, Highlands Ranch, Colorado, 627–668, 1995.

20 Booij, M. J.: Determination and integration of appropriate spatial scales for river basin modelling, Hydrol. Process., 17, 2581–2598, doi:10.1002/hyp.1268, 2003.

Bourgin, P. Y., Lobligeois, F., Peschard, J., Andréassian, V., Le Moine, N., Coron, L., Perrin, C., Ramos, H., and Khalifa, A.: Description des caractéristiques morphologiques, climatiques et hydrologiques de 4436 bassins versants français – guide d'utilisation de la base de données hydro-climatique, Internal document, Unité HBAN, Equipe Hydrologie, Irstea, Antony, 2011.

Breuer, L., Huisman, J. A., Willems, P., Bormann, H., Bronstert, A., Croke, B. F. W., Frede, H.-G., Gräff, T., Hubrechts, L., Jakeman, A. J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D. P., Lindström, G., Seibert, J., Sivapalan, M., and Viney, N. R.: Assessing the impact of land

use change on hydrology by ensemble modeling (LUCHEM), 1: Model intercomparison with current land use, Adv. Water Res., 32, 129–146, 2009.

Chiew, F., Stewardson, M., and McMahon, T.: Comparison of six rainfall–runoff modelling approaches, J. Hydrol., 147, 1–36, 1993.

5 Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): a modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, W00B02, doi:10.1029/2007wr006735, 2008.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., and Hendrickx, F.: Crash
10 testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments, Water Resour. Res., 48, W05552, doi:10.1029/2011wr011721, 2012.

DHI: MIKE11 – Reference and User's Manual, DHI Water and Environment, Hørsholm, Denmark, 2008.

European Environment Agency: Corine Land Cover 2006 raster data from EEA, available at:
15 http://www.eea.europa.eu//themes/landuse/dc (last access: 14 September 2012), 2006.

Fenicia, F., McDonnell, J. J., and Savenije, H. H. G.: Learning from model improvement: on the contribution of complementary data to process understanding, Water Resour. Res., 44, W06419, doi:10.1029/2007wr006386, 2008.

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

Fenicia, F., Kavetski, D., Savenije, H. H. G., Schoups, G., Pfister, L., Clark, M. P., and Freer, J.: Catchment properties and conceptual model structure: is there a correspondence?, Hydrol. Process., online first, doi:10.1002/hyp.9726, 2013.

25 Goswami, M. and O'Connor, K. M.: A "monster" that made the SMAR conceptual model "right for the wrong reasons", Hydrolog. Sci. J., 55, 913–927, 2010.

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

30 Jakeman, A. J. and Hornberger, G. M.: How much complexity is warranted in a rainfall-runoff model?, Water Resour. Res., 29, 2637–2649, 1993.

Kavetski, D. and Evin, G.: BATEA The User Guide (v 7.020.005), University of Newcastle, Australia, 2011.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Kavetski, D. and Fenicia, F.: Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights, Water Resour. Res., 47, W11511, doi:10.1029/2011wr010748, 2011.

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.

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, Water Resour. Res., 42, W03407, doi:10.1029/2005wr004368, 2006.

Klemes, V.: Operational testing of hydrological simulation-models, Hydrolog. Sci. J., 31, 13–24, 1986.

Leavesley, G. H., Markstrom, S. L., Restrepo, P. J., and Viger, R. J.: A modular approach to addressing model design, scale, and parameter estimation issues in distributed hydrological modelling, Hydrol. Process., 16, 173–187, doi:10.1002/hyp.344, 2002.

Le Moine, N.: Le bassin versant de surface vu par le souterrain: une voie d'amélioration des performances et du réalisme des modèles pluie-débit?, Ph.D. thesis, Université Pierre et Marie Curie, Paris, 324 pp., 2008.

Le Moine, N., Andréassian, V., Perrin, C., and Michel, C.: How can rainfall–runoff models handle intercatchment groundwater flows? Theoretical study based on 1040 French catchments, Water Resour. Res., 43, W06428, doi:10.1029/2006WR005608, 2007.

Le Moine, N., Andréassian, V., and Mathevet, T.: Confronting surface- and groundwater balances on the La Rochefoucauld-Touvre karstic system (Charente, France), Water Resour. Res., 44, W03403, doi:10.1029/2007wr005984, 2008.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, J. Hydrol., 201, 272–288, 1997.

Linsley, R. K. and Crawford, N. H.: Computation of a synthetic streamflow record on a digital computer, IAHS Publ., 51, 526–538, 1960.

Mathevet, T., Michel, C., Andréassian, V., and Perrin, C.: A bounded version of the Nash–Sutcliffe criterion for better model assessment on large sets of basins, large sample basin experiments for hydrological model parameterization: results of the Model Parameter Experiment – MOPEX, IAHS Publ., 307, 211–219, 2006.

MEDD: Banque Hydro, Retrieved 2012, from Le Ministère de l'Ecologie et du Développement Durable (MEDD), Direction de l'Eau et le SCHAPI, available at: www.hydro.eaufrance.fr/ (last access: 27 August 2012), 2007.

Merz, R., Parajka, J., and Blöschl, G.: Scale effects in conceptual hydrological modeling, Water Resour. Res., 45, doi:10.1029/2009wr007872, 2009.

Merz, R., Parajka, J., and Blöschl, G.: Time stability of catchment model parameters: implications for climate impact analyses, Water Resour. Res., 47, W02531, doi:10.1029/2010WR009505, 2011.

Middleton, N. and Thomas, D. S. G.: World Atlas of Desertification, United Nations Environment Programme (UNEP), London, 1992.

Moussa, R. and Bocquillon, C.: Criteria for the choice of flood-routing methods in natural channels, J. Hydrol., 186, 1–30, 1996.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part 1 – a discussion of principles, J. Hydrol., 10, 282–290, 1970.

Pechlivanidis, I. G., Jackson, B. M., McIntyre, N. R., and Wheater, H. S.: Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications, Global Nest J., 13, 193–214, 2011.

Perrin, C., Michel, C., and Andreassian, V.: Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, J. Hydrol., 242, 275–301, 2001.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, J. Hydrol., 279, 275–289, doi:10.1016/s0022-1694(03)00225-7, 2003.

Pushpalatha, R., Perrin, C., Le Moine, N., Mathevet, T., and Andréassian, V.: A downward structural sensitivity analysis of hydrological models to improve low-flow simulation, J. Hydrol., 411, 66–76, 2011.

Pushpalatha, R., Perrin, C., Le Moine, N., and Andréassian, V.: A review of efficiency criteria suitable for evaluating low-flow simulations, J. Hydrol., 420, 171–182, doi:10.1016/j.jhydrol.2011.11.055, 2012.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D., and Dmip participants: Overall distributed model intercomparison project results, J. Hydrol., 298, 27–60, 2004.

Refsgaard, J. C. and Hansen, J. R.: A good-looking catchment can turn into a modeller's nightmare, Hydrolog. Sci. J., 55, 899–912, doi:10.1080/02626667.2010.505571, 2010.

The influence of conceptual model structure on model performance

W. R. van Esse et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►|

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Refsgaard, J. C. and Knudsen, J.: Operational validation and intercomparison of different types of hydrological models, Water Resour. Res., 32, 2189, doi:10.1029/96WR00896, 1996.

Rogers, P.: On the choice of "appropriate model" for water resources planning and management, Water Resour. Res., 14, 1003–1010, 1978.

Seiller, G., Anctil, F., and Perrin, C.: Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions, Hydrol. Earth Syst. Sci., 16, 1171–1189, doi:10.5194/hess-16-1171-2012, 2012.

Singh, V. P. and Frevert, D. K. (Eds.): Mathematical Models of Small Watershed Hydrology and Applications, Water Resources Publications, Highlands Ranch, Colorado, 2002a.

Singh, V. P. and Frevert, D. K. (Eds.): Mathematical Models of Large Watershed Hydrology, Water Resources Publications, Highlands Ranch, Colorado, 2002b.

Tabary, P., Dupuy, P., L'Henaff, G., Gueguen, C., Moulin, L., Laurantin, O., Merlier, C., and Soubeyroux, J.-M.: A 10-yr (1997–2006) reanalysis of quantitative precipitation estimation over France: methodology and first results, IAHS Publ., 351, 255–260, 2012.

Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S. W., and Srikanthan, S.: Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: a case study using Bayesian total error analysis, Water Recour. Res., 45, W00B14, doi:10.1029/2008wr006825, 2009.

Valéry, A., Andréassian, V., and Perrin, C.: Regionalization of preciptation and air temperature over high altitude catchments – learning from outliers, Hydrolog. Sci. J., 55, 928–940, doi:10.1080/02626667.2010.504676, 2010.

Van Esse, W. R.: Demystifying hydrological monsters – can flexibility in model structure help explain monster catchments?, M.Sc. thesis, University of Twente, Enschede, the Netherlands, 111 pp., 2012.

Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J. M., Viney, N. R., and Teng, J.: Climate non-stationarity – validity of calibrated rainfall-runoff models for use in climate change studies, J. Hydrol., 394, 447–457, 2010.

Vidal, J.-P., Martin, E., Franchisteguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-yr high-resolution atmospheric reanalysis over France with the Safran system, Int. J. Climatol., 30, 1627–1644, 2010.

Wagener, T.: Evaluation of catchment models, Hydrol. Process., 17, 3375–3378, doi:10.1002/hyp.5158, 2003.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

|◀ | ▶|

◀ | ▶

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Wagener, T., Lees, M. J., and Wheater, H. S.: A toolkit for the development and application of parsimonious hydrological models, in: Mathematical Models of Small Watershed Hydrology, edited by: Singh, V. P., Frevert, R., and Meyers, D., Water Resources Publications, LLC, USA, 2001.

5 Zhao, R. J. and Liu, X. R.: The Xinanjiang model, in: Computer Models of Watershed Hydrology, edited by: Singh, V. P., Water Resources Publications, Highlands Ranch, Colorado, 215–232, 1995.

The influence of conceptual model structure on model performance

W. R. van Esse et al.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Table 1. Characteristics of the 237 catchments.

	Area [km^2]	Mean annual rainfall [mm yr^{-1}]	Mean annual PE [mm yr^{-1}]	Mean annual discharge [mm yr^{-1}]
Minimum	16	61	605	84
Average	567	988	732	383
Maximum	6836	1961	1182	1329

Table 2. Classification ranges of catchment characteristics and number of catchments in each range.

Characteristic	Classification	Range	# Catchments
Area (km^2)		16–6836	(237)
	Small	< 200	85
	Medium	200–600	79
	Large	> 600	73
Wetness Index (–)		0.41–2.03	(237)
	Dry	< 1.2	78
	Moist	1.2–1.5	83
	Wet	> 1.5	76
Permeability			(237)
	Impermeable	Classified by the European	89
	Semi-permeable	Environment Agency (2006)	85
	Permeable		63
RC$_{S/W}$ (–)		0.03–1.00	(237)
	Direct runoff	< 0.15	87
	Mixed	0.15–0.24	74
	Groundwater dominated	> 0.24	76

Table 3. Main characteristics of the twelve SUPERFLEX structures and GR4H where N_{res} and N_θ are the number of reservoirs and calibrated parameters respectively. "Rank" denotes the complexity rank of each structure based on the number of parameters, and number and type of function(s), where 1 is the most simple and 13 is de most complex.

Structure	N_{res}	Connection	Type of function(s)	N_θ	Rank
SF01	1	–	Power	3	1
SF02	1	–	Power + Linear	4	2
SF03	2	Series only	Threshold + Power	4	3
SF04	2	Series only	Power	5	6
SF05	2	Series only	Lag + Power	6	8
SF06	3	Series only	Threshold + Lag + Power	7	10
SF07	3	Series and parallel	Lag + Power + Linear	8	12
SF08	2	Parallel only	Linear	4	4
SF09	3	Series and parallel	Linear	5	7
SF10	3	Series and parallel	Lag + Linear	6	9
SF11	3	Series and parallel	Lag + Power + Linear	7	11
SF12	4	Series and parallel	Threshold + Lag + Power + Linear	8	13
GR4H	2	Series and parallel	Lag + Power + Linear	4	5

Table 4. Number of times each structure is selected for the flexible approach alone and when GR4H is considered as one of thirteen structures (combined).

Model (Rank)	Flexible	Combined
SF01 (1)	0	0
SF02 (2)	0	0
SF03 (3)	3	3
SF04 (6)	50	22
SF05 (8)	1	1
SF06 (10)	0	0
SF07 (12)	13	7
SF08 (4)	2	2
SF09 (7)	38	13
SF10 (9)	7	0
SF11 (11)	32	15
SF12 (13)	1	0
Total SUPERFLEX	147	63
GR4H (5)	–	138
Inconsistent	90	36
Total	237	237

Full Screen / Esc
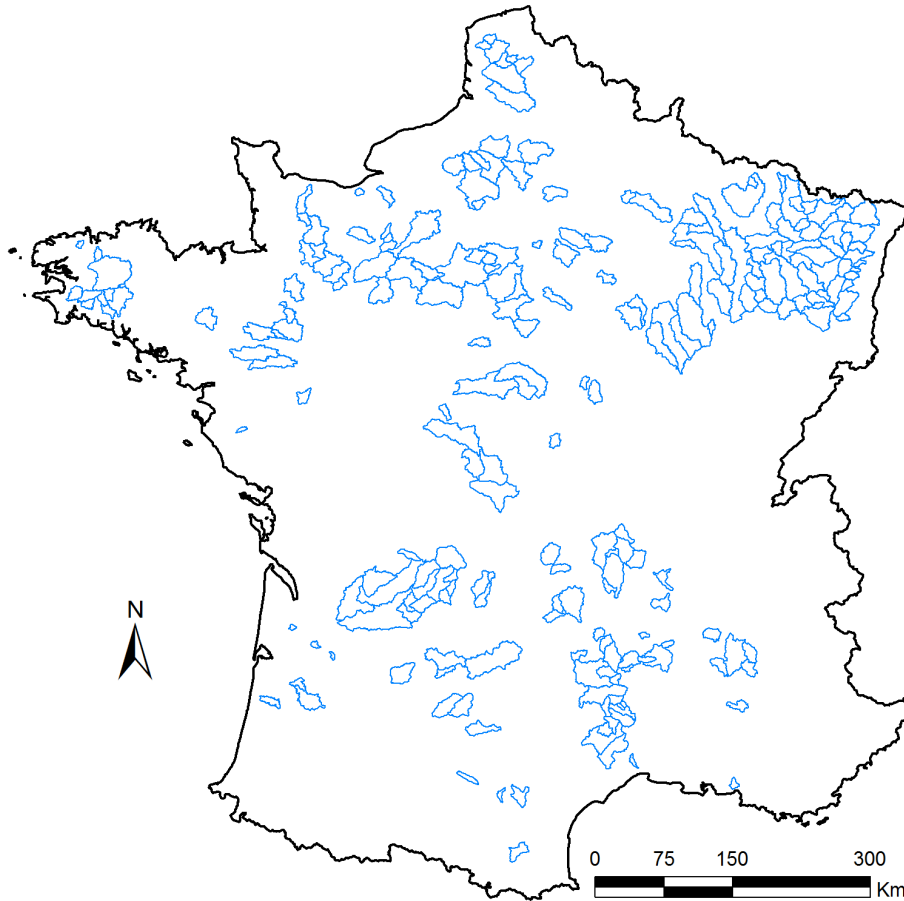
Fig. 1. Location and boundaries of the 237 catchments in France (after Le Moine, 2008).
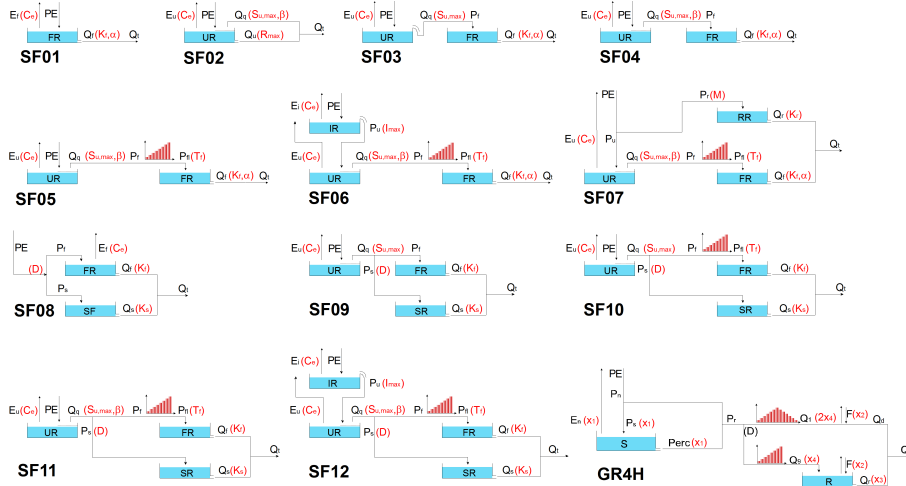
Fig. 2. Schemes of the twelve model structures in the SUPERFLEX framework (SF01–SF12 by Fenicia et al., 2012) and the GR4H model (Perrin et al., 2003). A flux of water is denoted by a letter in black and parameters are in parentheses.

Fig. 3. Boxplots (maximum, 75th percentile, median, 25th percentile and minimum) of averaged CR1–CR4 values of all model structures on the 237 catchments in the validation periods, including notes on differences in SUPERFLEX structures. Where "+β" means adding a power function β, "+lag" means adding a lag-function, UR = unsaturated zone reservoir, FR = fast reservoir, SR = slow reservoir, IR = interception reservoir and RR = riparian zone reservoir. The x-axis shows the twelve SUPERFLEX structures plus GR4H. The value between parentheses denotes the complexity rank. At the top of the figure the mean values for model performance are given.
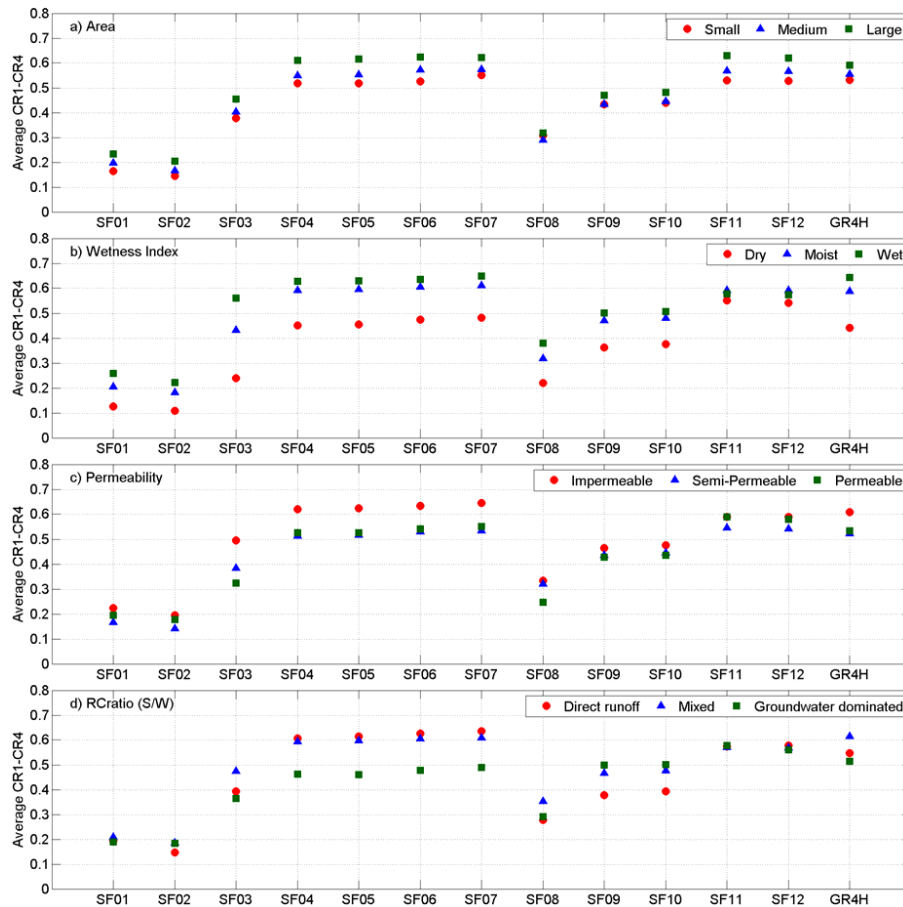
Fig. 4. Average of CR1–CR4 in validation in three classes of catchment area **(a)**, wetness index **(b)**, permeability **(c)** and RC$_{S/W}$ **(d)**.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◀◀ | ▶▶

◀ | ▶

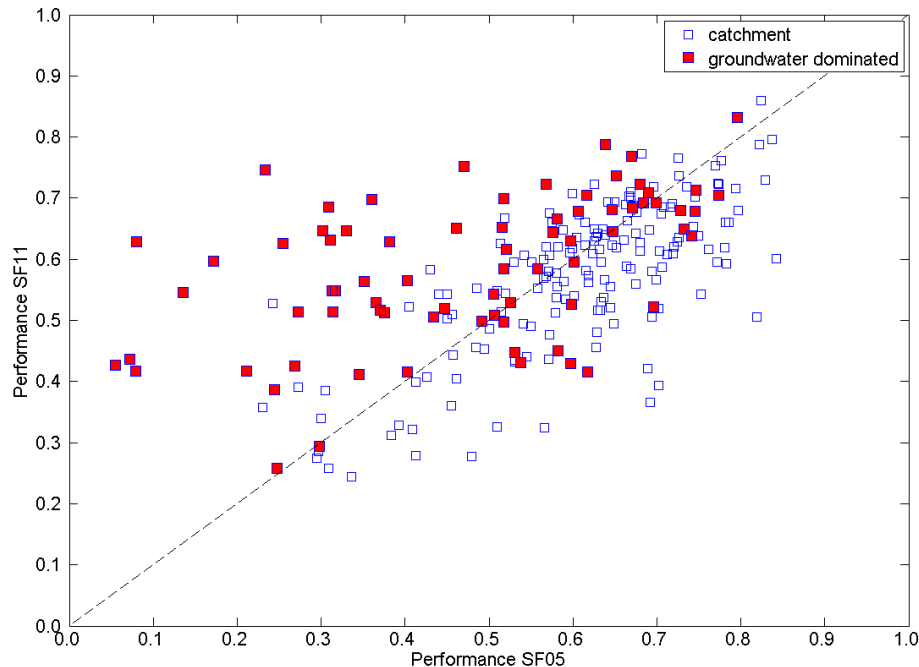Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Fig. 5. Average performance of model structure SF05 (without parallel slow reservoir) against SF11 (with parallel slow reservoir). Groundwater dominated catchments are marked to show the specific performance of these catchments compared to the others.
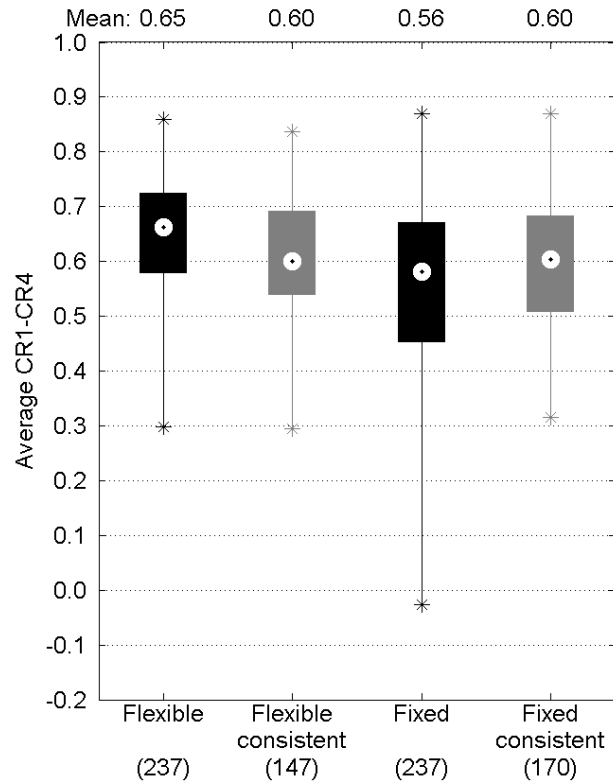
Fig. 6. Distribution in average performance in validation for the flexible, fixed and combined approach. These distributions represent the minimum and maximum (whiskers), the area between the 25th and 75th percentile (box) and the median. The dark boxes represent the distributions including inconsistent results, the grey boxes represent only consistent results. Number of catchments in each group is noted between parentheses.

Title Page

Abstract | Introduction

Conclusions | References

Tables | Figures

◄ | ►

◄ | ►

Back | Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion