Hydrol. Earth Syst. Sci., 15, 3447–3459, 2011
www.hydrol-earth-syst-sci.net/15/3447/2011/
doi:10.5194/hess-15-3447-2011

Hydrology and Earth System Sciences

Comparison of hydrological model structures based on recession and low flow simulations

M. Staudinger[1], **K. Stahl**[2], **J. Seibert**[1], **M. P. Clark**[3], **and L. M. Tallaksen**[4]

[1]Department of Geography, University of Zurich, Zurich, Switzerland
[2]Institute of Hydrology Freiburg, Albert-Ludwigs University, Freiburg, Germany
[3]University Cooperation for Atmospheric Research, Boulder, Colorado, USA
[4]Department of Geosciences, University of Oslo, Oslo, Norway

Received: 8 July 2011 – Published in Hydrol. Earth Syst. Sci. Discuss.: 13 July 2011
Revised: 27 October 2011 – Accepted: 1 November 2011 – Published: 17 November 2011

Abstract. Low flows are often poorly reproduced by commonly used hydrological models, which are traditionally designed to meet peak flow situations. Hence, there is a need to improve hydrological models for low flow prediction. This study assessed the impact of model structure on low flow simulations and recession behaviour using the Framework for Understanding Structural Errors (FUSE). FUSE identifies the set of subjective decisions made when building a hydrological model and provides multiple options for each modeling decision. Altogether 79 models were created and applied to simulate stream flows in the snow dominated headwater catchment Narsjø in Norway (119 km^2). All models were calibrated using an automatic optimisation method. The results showed that simulations of summer low flows were poorer than simulations of winter low flows, reflecting the importance of different hydrological processes. The model structure influencing winter low flow simulations is the lower layer architecture, whereas various model structures were identified to influence model performance during summer.

1 Motivation

Hydrological low flow periods and droughts affect water supply for drinking water, irrigation, industrial needs, hydropower production and ecosystems. Their occurrence is also of importance regarding environmental flow and water quality requirements, which are strongly connected to critical low flows (Vogel and Fennessey, 1995). Low flow and droughts affect many sectors and occur in every country albeit in different perceived severity. There is a wide range of consequences related to low flow and drought and monitoring and modelling of low flow are crucial for their analysis and prediction. However, low flows are poorly reproduced by many hydrological models since these are traditionally designed to simulate the runoff response to rainfall.

A revision of model concepts regarding low flows requires a clear understanding of the model's structural deficits; in other words "when does it go wrong and which part of the model is the origin?" (Reusser et al., 2009). A common approach to investigate the impact of the differences in model structure is to perform model intercomparison experiments (e.g. Henderson-Sellers et al., 1993; Reed et al., 2004; Duan et al., 2006; Breuer et al., 2009 and Holländer et al., 2009). Such experiments have been helpful to explore model simulation performance of lumped (Duan et al., 2006; Breuer et al., 2009), semi-distributed (Duan et al., 2006; Holländer et al., 2009) and distributed (Henderson-Sellers et al., 1993; Reed et al., 2004; Holländer et al., 2009) models in a consistent way using the same input data. The reasons for the differences, however, remain unclear since each model uses different interacting parametrisations to simulate the hydrological processes (Clark et al., 2008). Perrin et al. (2001) studied the relation between the number of optimized parameters and model performance in a multi-model, multi-catchment experiment, and discussed the problem of over-parametrisation and parameter uncertainty.

Discrepancies between observed and simulated streamflow can arise from errors in the input data rather than weaknesses in model structure. This complicates the investigation

Correspondence to: M. Staudinger
(maria.staudinger@geo.uzh.ch)

of the impact of the differences in model structure. Clark et al. (2008) created a computational framework that enables a separate evaluation of each model component. The Framework for Understanding Structural Errors (FUSE) differs from others as it modularises individual flux equations instead of linking available submodels. FUSE identifies the set of subjective decisions while creating a hydrological model and offers multiple options for each model decision. This approach can thus help to get a better understanding of the hydrological processes occurring. Clark et al. (2008) first introduced FUSE, as a diagnostic tool to evaluate the performance of hydrological model structures using the Nash-Sutcliffe efficiency for two climatically different catchments. Clark and Kavetski (2010) evaluated several classes of numerical time stepping schemes in order to find appropriate numerical methods used to solve the governing model equations of hydrological models. The experimental setup included beside different distinct time stepping algorithms, eight conceptual rainfall runoff models derived from the parent models. Another recent application of FUSE is documented in the two-part series of McMillan et al. (2011) and Clark et al. (2011b). First, they used precipitation, soil moisture and streamflow data to estimate the dominant hydrological processes of a catchment. Then, plausible representations of these processes in conceptual models were formulated (McMillan et al., 2011). In the second part, they evaluated FUSE models regarding their capability to simulate those processes (Clark et al., 2011b).

Commonly, streamflow recession is modelled as the outflow from a, or a set of, linear or non-linear reservoirs. In periods with no input, i.e. precipitation or snow melt, outflow from the reservoirs control the streamflow and thus, the model behaviour during low flow. Real hydrological processes can be more complex. Therefore, it is of interest to have a closer look at the hydrograph recession, and carefully evaluate model simulations of recession behaviour. The shape of the observed recession curve reflects the gradual depletion of water stored in a catchment during periods with little or no precipitation. Initially, the recession curve is steep as quick flow components like overland flow and subsurface flow contribute to streamflow. The recession curve flattens with time as e.g. delayed water from deeper subsurface storages contributes, and may become nearly constant if sustained by outflow from the groundwater storage or from a glacier (Smakhtin, 2001). The recession curve describes in an integrated manner how different factors in a catchment influence the generation of streamflow in dry weather periods (Tallaksen, 1995). Hydrogeology, relief and climate have been found to be the most important catchment properties affecting the recession rate (Tallaksen, 1995). Catchments with a slow recession rate are typically groundwater dominated, while impermeable catchments with little storage show faster recession rates. Moreover, summer recessions are usually faster than autumn or winter recessions (e.g. Federer, 1973; Tallaksen, 1995).

Several studies exist that link recession analysis with the structure of hydrological models (e.g. Ambroise et al., 1996; Wittenberg, 1999; Clark et al., 2009; Harman et al., 2009). In this study the model structures are systematically analysed using FUSE. The associated model performance is evaluated with respect to the ability to simulate low flows and recession behaviour. This is done for one catchment only to allow a more detailed insight in the model structures. The main objective is to investigate the relative influence of a single model structure on the model performance. As there are distinct differences in the recession rates found for summer and winter, one task is to study how model structure is connected to the seasonal performance for low flow simulation. This paper aims to contribute to the improvement of hydrological models for low flow prediction.

2 Data and study area

The data are from the 119 km^2 headwater catchment Narsjø, located in the South-East of Norway (Fig. 1) with an altitude range between 737 and 1595 m a.s.l. (Engeland, 2002). Narsjø is a subcatchment of the Upper Glomma basin, which is characterised by a continental climate with cold winters and relatively warm summers (Engeland, 2002). The annual snow melt flood dominates the hydrological regime. The most pronounced low flow period occurs in winter, caused by precipitation being stored in snow and ice. A second low flow period occurs in summer, caused by a lack of precipitation and losses due to evapotranspiration (Engeland, 2002).

The geology can be divided into two main areas: one area consists of schists and phyllites that occur in combination with fine grained till soil, the other area consists of igneous rocks (granite, gneiss and gabbro) usually in combination with coarser till (Engeland, 2002). This geological characteristic influences the properties of soil and vegetation. The quaternary remains, consisting of several types of till and fluvial deposits as well as bogs and lakes, form a wide, open mountain landscape with gentle slopes. The land cover is barely influenced by humans (0.3 % agricultural land) and is composed of 23.7 % forest, 60.9 % open land, 12.0 % bogs and 3.0 % lakes (Engeland, 2002).

The streamflow data used are daily time series of observed discharge measured at the outlet of the Narsjø catchment (provided by the Norwegian Water Resources and Energy directorate, NVE). In addition, daily time series of precipitation interpolated from 12 surrounding meteorological stations and potential evaporation (Beldring et al., 2003) were available. The time series cover the period from 6 May 1981 to 31 December 1995.

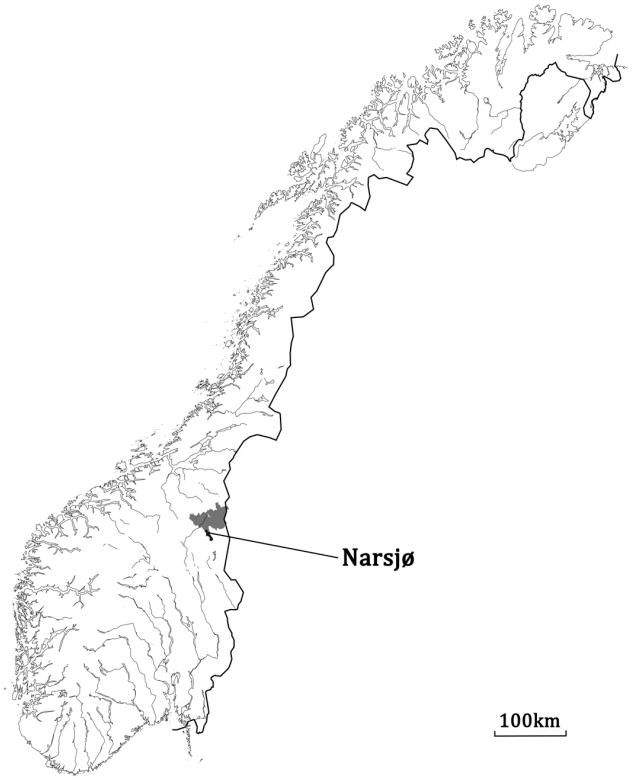

Fig. 1. Location of the Narsjø catchment (modified after Beldring et al., 2003).

3 Methods

3.1 Snow accumulation and melt

Narsjø is a snow dominated catchment, however, there was no snow routine implemented in the version of FUSE used for this study. Hence, the input data was pre-processed with a snow accumulation and melt model. This corresponds to an implemented snow routine. Here, a simple degree day method was applied. The daily change in snow water equivalent ΔSWE [mm day^{-1}] is equal to the difference in the daily snow accumulation a_s [mm day^{-1}] and the daily snow melt m_s [mm day^{-1}] (Eq. 1).

$$\Delta\text{SWE} = a_s - m_s. \quad (1)$$

The snow model separates the precipitation P [mm day^{-1}] into rain and snow using a temperature threshold. Hence, there is only snow accumulation a_s in the catchment when the measured temperature T [°C] is below the threshold temperature T_{acc} (Eq. 2).

$$a_s = \begin{cases} 0, & T \geq T_{acc}, \\ P, & T < T_{acc}. \end{cases} \quad (2)$$

In this study T_{acc} was set to 1.0 °C. The daily snow melt m_s was computed (Eq. 3) with a melt factor M_f of 3.0 °C^{-1} day^{-1} and a melt threshold temperature T_{melt} of 0 °C.

$$m_s = \begin{cases} M_f\,(T - T_{melt}), & T \geq T_{melt} \text{ and SWE} > 0, \\ 0, & T < T_{melt} \text{ and SWE} = 0. \end{cases} \quad (3)$$

The chosen melt factor was based on Seibert (1999) who found melt factors in Sweden to vary between 1.5 and 4 °C^{-1} day^{-1}, where the first value is suited for open and the latter for forested sites. The degree day method was extended with a refreeze factor r_f [−] which accounts for rain that does not directly contribute to runoff due to the water holding capacity of an existing snow cover (Eq. 4).

$$P = \begin{cases} 0, & T \geq T_{acc}, \\ P, & T \geq T_{acc} \text{ and } P \geq r_f \text{ SWE}, \\ (1 - r_f)\, m_s, & T \geq T_{acc} \text{ and } P < r_f \text{ SWE}. \end{cases} \quad (4)$$

3.2 FUSE framework

The use of FUSE as a diagnostic tool to detect the impact of model structure involved the following three steps: (1) prescription of the type of model (2) definition of the major model-building decisions and (3) preparation of multiple options for each model building decision (Clark et al., 2008). In this study, the type of model was limited to lumped hydrological, that were run at a daily time step (although the models are not limited to a daily time step). Four conceptual parent models were selected to be recombined to new FUSE-models: ARNO-VIC (Zhao, 1977), TOPMODEL (Beven and Kirkby, 1979), PRMS (Leavesley et al., 1983) and SACRAMENTO (Burnash, 1995). Simplified wiring diagrams of the generating parent models are shown in Fig. 2. The selection of the parent FUSE models was here limited to four well known models, covering common principles used in conceptual hydrological models.

All parent models consist of equally plausible structures and the important processes could be broken down into fluxes occurring in the upper layer and lower layer, evaporation, percolation, subsurface flow and surface runoff (model building options).

Some processes were not explicitly modelled, including interception by the vegetation canopy as well as specific surface energy balance calculations. Routing was calculated by a two parameter Gamma distribution (Press et al., 1992). Thus, all models represent the subsurface with a similar level of detail and thus differences that emerged from different plausible model structures were emphasised rather than differences due to the set of processes represented. The model decision options that were made separately for each of the FUSE models are described next (more details to the decision options e.g. equations can be found in Clark et al., 2011b). A summary of those decisions that were permuted for this study can be found in Table 1 and the abbreviations from Table 1 will be referred to later in the text.

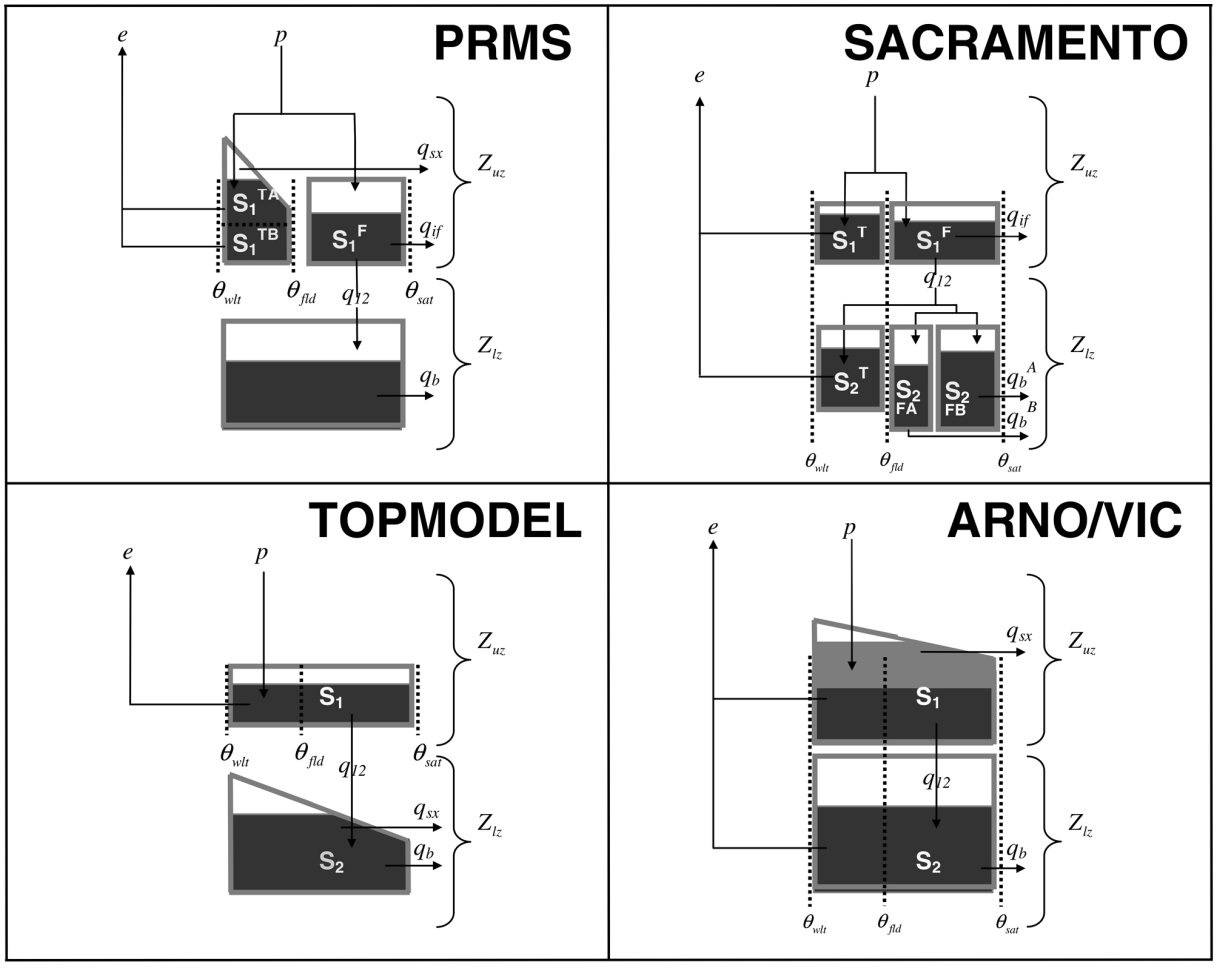

Fig. 2. Simplified wiring diagrams of the parent models (modified after Clark et al., 2008).

3.2.1 Upper layer

The water content of the upper soil layer was either defined as a single state variable or split into tension storage and free storage, with an additional option to further subdivide the free storage into below and above field capacity (Table 1).

3.2.2 Lower layer

The lower soil layer was either defined by a single state variable with unlimited storage and no lower layer evapotranspiration, by a single state variable with fixed storage and no lower layer evapotranspiration or as a tension storage combined with two parallel tanks (Table 1). All subsurface flow options (see below) are closely connected to the lower layer, this is why the choice of subsurface flow and lower-layer option is realised as a single model decision within FUSE (Clark et al., 2008).

3.2.3 Evaporation

Evaporation was parameterised by the sequential evaporation scheme (Clark et al., 2008): first potential evaporative demand is supplied by evaporation from the upper layer and then any residual demand by water from the lower layer.

3.2.4 Percolation

In FUSE there are three percolation options each having two parameters (Table 1). The architecture of the parent model VIC is equivalent to the gravity drainage term in the Richard's equation (e.g. Boone and Wetzel, 1996), often resulting in a large exponent to limit drainage below field capacity (water can percolate from the wilting point to saturation). The equation used in PRMS does not allow drainage below field capacity (water can percolate from the field capacity to saturation). Non-linearities in the SACRAMENTO parametrisation are controlled by the moisture content in the lower layer, meaning percolation will be fastest when the

Table 1. FUSE model decision options.

Model structure	Model option	Abbreviation
Upper layer architecture U	Upper layer divided into tension and free storage	U_{tension1}
	Free storage plus tension storage sub-divided into recharge and excess	U_{tension2}
	Upper layer defined by a single state variable	U_{onestate}
Lower layer architecture and subsurface flow L	Tension storage combined with two parallel tanks	L_{tens2pll}
	Storage of unlimited size combined with linear fraction rate	L_{unlimfrc}
	Storage of unlimited size combined with power recession	L_{unlimpow}
	Storage of fixed size with non-linear storage function	L_{fixedsiz}
Surface runoff S	ARNO/Xzang/VIC parametrisation	$S_{\text{arno/vic}}$
	PRMS variant; fraction of upper tension storage	S_{prms}
	TOPMODEL parametrisation	S_{tmdl}
Percolation P	Water from field capacity to saturation available for percolation	P_{f2sat}
	Water from wilting point to saturation available for percolation	P_{w2sat}
	Percolation defined by moisture content in lower layer architecture	P_{lower}

lower layer is dry (Clark et al., 2008). All three options were used as model decision options.

3.2.5 Subsurface flow

There are four subsurface flow options (Table 1). Subsurface flow was modelled either by a single linear storage, by two parallel connected linear reservoirs or by nonlinear storage functions like in ARNO/VIC or TOPMODEL (Clark et al., 2008). TOPMODEL requires a distribution of topographic index values for each catchment (Beven and Kirkby, 1979). For the Narsjø catchment the distribution was derived using a three-parameter Gamma distribution following Sivapalan et al. (1987).

3.2.6 Surface runoff

Surface runoff was generated using a saturation-excess mechanism, when it rains on saturated areas of the basin. The surface runoff is distributed according to the topographic index distribution (defined in Clark et al., 2008).

3.2.7 Bucket overflow

Additional fluxes of water may occur when one of the storages reaches its capacity. In the upper layer, the bucket overflow from the primary tension storage carries over precipitation that falls into the second tension storage. The bucket overflow from a tension storage carries precipitation into a free storage and from the free storage it adds to surface runoff. In the lower soil layer, the bucket overflow from tension storage forms additional percolation into free storage and from free storage again additional subsurface flow. Following Kavetski and Kuczera (2007), logistic functions were used to smooth the thresholds associated with a fixed capacity of model storages.

3.2.8 Routing

The time delay in runoff was modelled using a two-parameter Gamma distribution (Press et al., 1992), with an adjustable mean of the Gamma distribution. The shape of the time delay histogram, however, was fixed by setting the shape parameter to 3.0 to keep the number of adjustable parameters small.

3.3 Model calibration

All FUSE models were calibrated using the Shuffled Complex Evolution algorithm (SCE) which was parameterised based on the recommendations of Duan et al. (1994). A maximum of 10 000 trials was allowed before the optimisation was terminated. Within five shuffling loops the value had to change by 10 % or the optimisation was terminated. The number of complexes in the initial population was set to 10. Each complex contained $2N_{\text{opt}} + 1$ points, each sub-complex $N_{\text{opt}} + 1$ points and $2N_{\text{opt}} + 1$ evolution steps were allowed for each complex before shuffling, where N_{opt} was the number of parameters to be optimised in the calibration procedure, respectively. The algorithm was used to minimise the mean absolute relative error (F_{MARE}) (Eq. 5). F_{MARE} ranges between zero and infinity with the optimum at zero.

$$F_{\text{MARE}} = \frac{1}{n} \sum_{i=1}^{n} \frac{|Q_{\text{obs}}(i) - Q_{\text{sim}}(i)|}{Q_{\text{obs}}(i)} \quad (5)$$

The calibration was performed for 15 yr using a three years spin up period. As recommended by Clark and Kavetski (2010) for conceptual hydrological models, the fixed step implicit Euler method was used as numerical time stepping scheme.

3.4 Low flow and recession analysis

The performance of the model was then evaluated using the logarithmic Nash-Sutcliffe efficiency. F_{logNSE} was based on log-transformed streamflow series from observation Q_{obs} and simulation Q_{sim} (Eq. 6). This metric ranges between minus infinity and one and a perfect model would result in 1.

$$F_{\text{logNSE}} = 1 - \frac{\sum_{i=1}^{n} \left(\ln\left(Q_{\text{obs}}(i)\right) - \ln\left(Q_{\text{sim}}(i)\right)\right)^2}{\sum_{i=1}^{n} \left(\ln\left(Q_{\text{obs}}(i)\right) - \ln\left(\bar{Q}_{\text{obs}}\right)\right)^2} \quad (6)$$

As a good model should be able to produce reasonable results for a range of objective functions the performance was evaluated using F_{logNSE}, whereas the models were calibrated using F_{MARE}. Calibration and validation by the two objective functions is done on the entire series, but both objective functions chosen emphasize the lower flow ranges of the hydrograph.

Several studies use recession analysis to infer the exponent in a non-linear storage (Ambroise et al., 1996; Wittenberg, 1999; Clark et al., 2009; Kirchner, 2009), or, more generally, provide guidance on the structure of a hydrological model (Clark et al., 2009; Harman et al., 2009). Recession analysis is also useful as a diagnostic tool for model evaluation (McMillan et al., 2011; Clark et al., 2011b). In this study the relationship between the negative change in streamflow over time $-\frac{\mathrm{d}Q}{\mathrm{d}t}$ [mm day^{-2}] and the corresponding streamflow Q [mm] was analysed using the method of Brutsaert and Nieber (1977). For the evaluation of the model performance of recessions both modelled and observed data were used. The method was modified by using flexible (instead of fixed) time steps scaled to the observed streamflow ΔQ between time steps as recommended by Rupp and Selker (2006). Our study was based on daily observations and similar to Palmroth et al. (2010), the lower and upper limits of the time step were set to 1 and 5 days, respectively. The time step was then found by setting the maximum difference in ΔQ (threshold) between to time steps equal to 0.1 % of the mean observed streamflow at that point. As both $-\frac{\mathrm{d}Q}{\mathrm{d}t}$ and Q span several orders of magnitude, their relation is plotted in log-log-space. The data points in the plots including all recessions of the hydrograph and might thus be composed of both subsurface and overland flow. Overland flow would mainly affect the upper range of streamflow values. Hence, the upper range in the plots of $-\frac{\mathrm{d}Q}{\mathrm{d}t}$ and Q should be treated with special care if interpreted regarding storage release. In case of an exponential recession (simple linear storage model) the relation can be expressed as in Eq. (7), where p is a constant. However, a power function results in Eq. (8), with the additional coefficient q.

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = -p\,Q \quad (7)$$

$$\frac{\mathrm{d}Q}{\mathrm{d}t} = -p\,Q^q \quad (8)$$

The $-\frac{\mathrm{d}Q}{\mathrm{d}t}$ versus Q plots can become noisy. Therefore, points in a certain range of Q were averaged to one value representative for this range (binned). Then, a polynomial function was fitted to the relationship between $-\frac{\mathrm{d}Q}{\mathrm{d}t}$ and Q (Eq. 9) (Kirchner, 2009).

$$\ln\left(\frac{-\mathrm{d}Q/\mathrm{d}t}{Q}\right) \approx a + b\ln(Q) + c\,(\ln(Q))^2 \quad (9)$$

The polynomial coefficients were fitted using a least squares regression model. The significance of the regression model was tested with the Kolmogorov-Smirnov goodness-of-fit test (Massey Jr., 1951). The polynomial fitted to the observed recessions is used as a benchmark model (see Seibert, 2001) similar to the mean streamflow being used as a benchmark model for the Nash-Sutcliffe efficiency (F_{NSE}). Hence, passing the Kolmogorov-Smirnov test, similar to a F_{NSE} above zero, is used as an objective decision for acceptable models (similar or better than the benchmark). The choice of a polynomial follows Kirchner (2009). It was used because of it offers both enough flexibility to adapt to the data and enough smoothness to allow moderate extrapolation beyond the binned relationships. Scatter plots of the coefficients b and c in Eq. (9) were then used to compare observed and simulated recession behaviour for the FUSE models that passed the Kolmogorov-Smirnov goodness-of-fit test. The relationship between $-\frac{\mathrm{d}Q}{\mathrm{d}t}$ and Q is in the following referred to as the "recession relationship".

The recession behaviour was analysed for both the whole year and the individual seasons. The seasonal recessions were derived by splitting the recessions for the whole year into summer and winter recessions. Winter was defined as the time from 15 October, when precipitation generally begins to fall as snow in the catchment, to 15 June, which is usually towards the end of the snowmelt period.

4 Results

4.1 Calibration

For 73 out of 79 FUSE models the F_{logNSE} was greater than zero. In Fig. 3 a scatter plot of the resulting values of the objective functions for both calibration (F_{MARE}) and evaluation (F_{logNSE}) is shown. The axes are ordered from high to low model performance for both measures, which means that the points of best performance group in the lower left corner. It appears that the F_{logNSE} and the F_{MARE} show a similarly good model performance for the F_{logNSE} range from 1 to 0.8. However, for lower F_{logNSE} the two objective functions differ. While the models are considered poorer for F_{logNSE}, F_{MARE} remains at the same level.

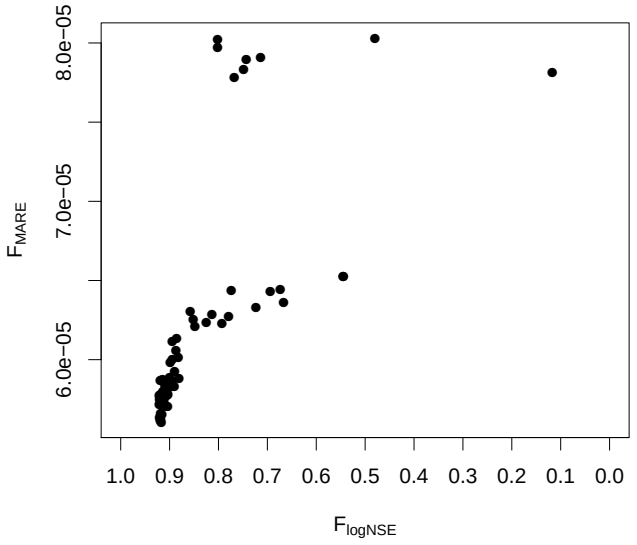

Fig. 3. F_{logNSE} versus F_{MARE} for the 79 FUSE models after calibration with SCE (Shuffled Complex Evolution algorithm).

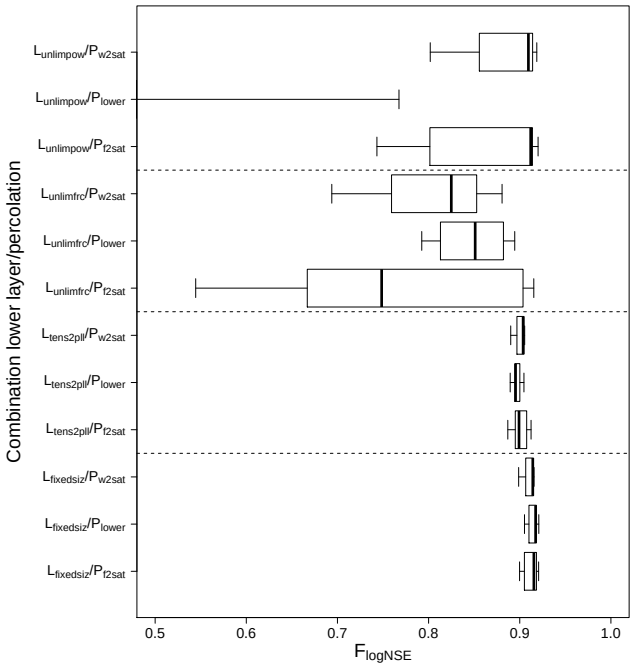

Fig. 4. Boxplots of the performance of models using different lower layer and percolation combinations. The box of models using L_{unlimpow} and P_{lower} includes model performances (F_{logNSE}) below zero.

4.2 Model performance during low flows

All models with $F_{\text{logNSE}} < 0$ used the same combination of lower layer/subsurface flow and percolation options L_{unlimpow} and P_{lower} (see Fig. 4). The best models ($F_{\text{logNSE}} > 0.8$) used varying combinations. The majority of the best models, however, used a lower layer/subsurface flow combination of either L_{tens2pll} or L_{fixedsiz}. Many of the poor models used a combination of L_{unlimfrc} for lower layer/subsurface flow and P_{f2sat} for percolation. The poorest models in the group with $F_{\text{logNSE}} > 0$ primarily used the same combination of lower layer/subsurface flow and percolation options as found for the poorest performing models ($F_{\text{logNSE}} < 0$). All possible upper layer and surface runoff options were found for the poorest performing models.

4.3 Recession behaviour

The observed flow values in the recession periods ranged between 0.2 and 40 mm day^{-1} for Q and between 0.001 and about 15 mm day^{-2} for $-\frac{\mathrm{d}Q}{\mathrm{d}t}$ and in general showed a linear recession relationship with higher $-\frac{\mathrm{d}Q}{\mathrm{d}t}$ for higher Q. Most of the modelled recession relationships were similar in range, their shapes, however, differed: some appeared more convex, others more concave and a third group showed nearly a linear recession relationship. In comparison to the observed range, some of the models produced an unrealistic scatter. For example, low flow values were modelled that were below the observed range (Fig. 5f) and their associated recession slopes were too steep (Fig. 5e and f). The latter behaviour was only found for models containing a combination of the lower layer/subsurface flow L_{unlimpow} and the percolation P_{lower}. The model decision options for the example models in Fig. 5 are listed in Table 2. The combinations including the S_{prms} surface runoff option (Fig. 5b, d and f) show linear relationships, while the combinations including S_{tmdl} (Fig. 5c and e) show convex or concave relationships. Figure 5e includes the lower layer/subsurface flow and percolation options L_{unlimpow} and P_{lower} and shows a large range in $-\frac{\mathrm{d}Q}{\mathrm{d}t}$ for the same flow values.

The coefficients b and c from Eq. (9) are shown in Fig. 6. The b coefficient describes the slope and the c coefficient the curvature of the binned recession relationships. The observation pair can be found at the edge of the group resulting from the simulations having a large b coefficient and a small c coefficient. Most pairs are located in the lower right quarter, i.e. in the area of positive slope and negative curvature. A smaller group can be found for positive b and c coefficients and only few models resulted in negative b and c coefficients. None was fitted with negative slope and positive curvature. The few models that resulted in negative slope and negative curvature used L_{unlimpow} for lower layer and subsurface flow, S_{prms} for surface runoff and P_{lower} for percolation.

The models that resulted in both coefficients being positive predominantly used U_{onestate} for the upper layer architecture, often combined with L_{unlimfrc} for lower layer/subsurface flow. The only differing model decision option for the upper layer architecture within this group was U_{tension2}. All surface runoff structure model options were found in this group. However, the S_{tmdl} parametrisation was found only in

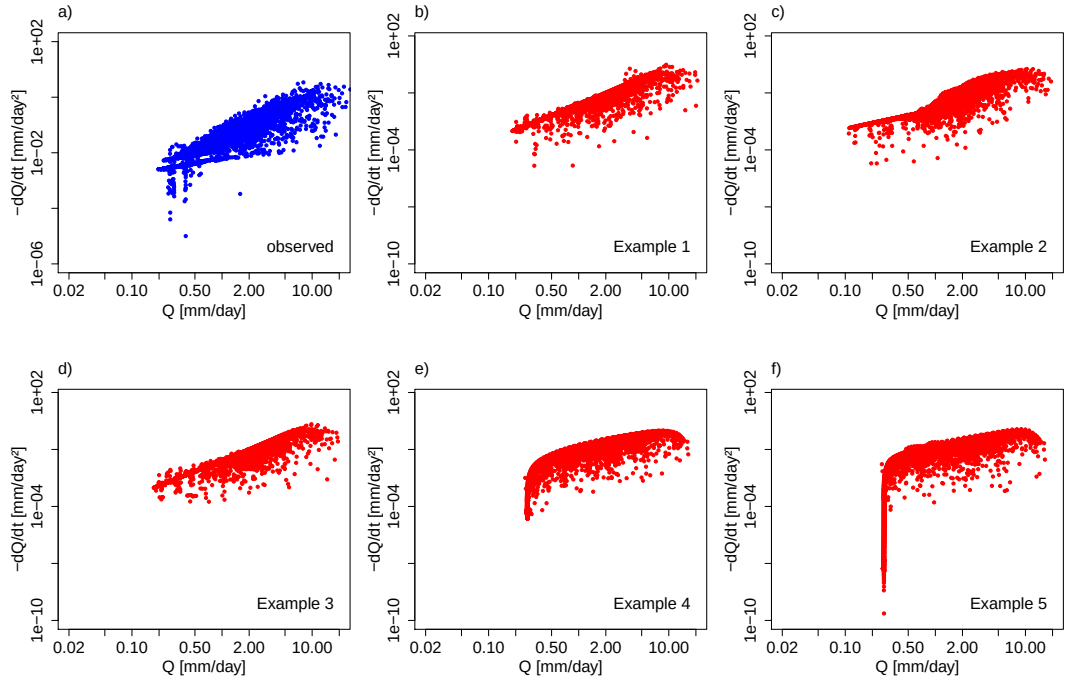

Fig. 5. Plots of recession relationships **(a)** observed recessions in blue, and **(b)–(f)** five examples of simulated recessions in red. The model decision options for the examples can be found in Table 2.

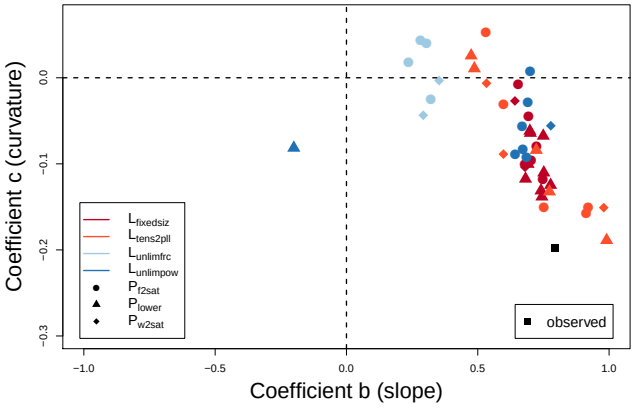

Fig. 6. Relation between b and c coefficients of the polynomial function fitted to binned recession relationships.

the particular combination with U_{onestate} and L_{unlimfrc} for the upper and lower layer architecture, respectively. The steepest slopes (coefficient b) were found for models containing the option L_{tens2pll} for lower layer/subsurface flow.

4.4 Seasonal analysis

F_{logNSE} values separated for summer and winter differed from each other and also from those derived for the whole year (Fig. 7). Model performance was generally lower for the summer season, with $F_{\text{logNSE}} < 0.4$ for all models. Eight models had F_{logNSE} values below zero. They all used the same lower layer, subsurface flow and percolation structure combination as those model that performed poorest for the whole year. The models showing the best performance of summer recessions used all combinations including the TOPMODEL surface runoff structure S_{tmdl} (Fig. 8).

However, in combination with L_{unlimpow} for subsurface flow and lower layer models using S_{tmdl} performed poorer. The direct comparison of the performance for summer and winter resulted in a higher F_{logNSE} value for winter for almost all models. Two models showed the opposite. Both consist of a tension storage in the upper layer (either U_{tension1} or U_{tension2}) and had exactly the same lower layer, subsurface flow/percolation structure (L_{unlimfrc}). All models where summer shows a better performance than winter use the percolation structure P_{f2sat}. All but one of the seven models with a F_{logNSE} less than zero in winter, used the percolation option P_{lower} in combination with L_{unlimpow} for lower layer and subsurface flow. The same subsurface flow and lower layer option in combination with either P_{f2sat} or P_{w2sat} improved the model performance. Models using P_{lower} in combination with L_{fixedsiz} had a high F_{logNSE}, and an even higher F_{logNSE} when S_{tmdl} was the surface runoff modeling option. The L_{tens2pll} combined with any model option for the other structures always performed better than a F_{logNSE} of 0.9 in winter. Most combinations of L_{unlimfrc} with P_{f2sat} were found to range between F_{logNSE} 0.2 and 0.7. Combined

Table 2. Model decision options for the examples in Fig. 5.

Example	Upper Layer	Lower Layer	Percolation	Surface runoff
1	U_{tension2}	L_{unlimpow}	P_{f2sat}	S_{prms}
2	U_{tension2}	L_{unlimfrc}	P_{f2sat}	S_{tmdl}
3	U_{tension2}	L_{fixedsiz}	P_{lower}	S_{prms}
4	U_{tension2}	L_{unlimpow}	P_{lower}	S_{tmdl}
5	U_{tension2}	L_{unlimpow}	P_{lower}	S_{prms}

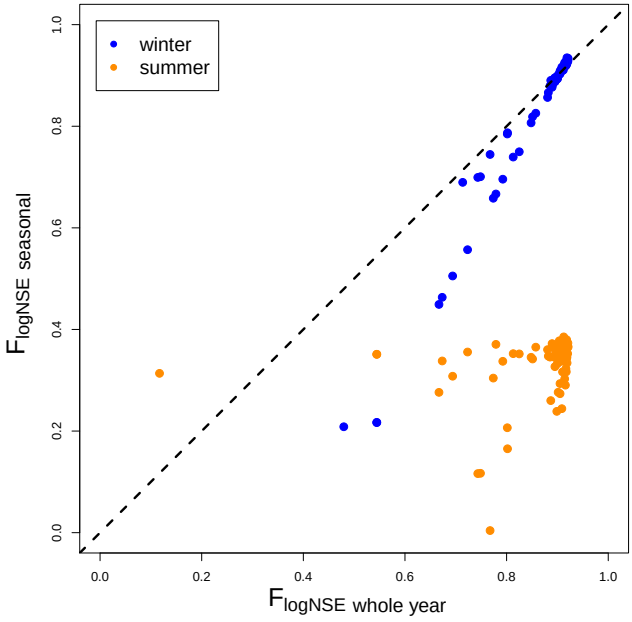

Fig. 7. F_{logNSE} for summer and winter compared to F_{logNSE} for the whole year; the 8 models with $F_{\text{logNSE}} < 0$ are not shown.

with the surface runoff option S_{tmdl} it resulted in F_{logNSE} values of about 0.9.

Generally, in summer observed recession slopes were steeper and flows were higher as compared to winter recessions which were slower with less steep slopes. Sometimes, a distinct non-linearity in recession slope was found with a considerably steeper recession slope from flow values of about 0.001 mm day^{-2} upwards. The recession relationships could be modelled with the polynomial (passed Kolmogorov-Smirnov-test) for 29 models for the winter season, for 44 for the whole year and for 28 models for the summer season. The polynomial described different recession relationships for summer and winter. The winter b and c coefficients of the polynomials are similar to those of the whole year. The structures of the underlying FUSE models were similar to the ones found for the whole year, but the lower layer and subsurface flow parametrisation were dominated by L_{tens2pll_2}. Only some models used L_{unlimfrc}, which was the dominant option for lower layer/subsurface flow for the whole year.

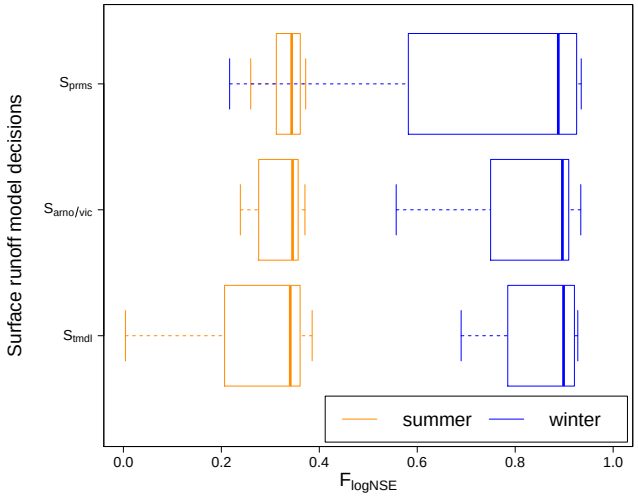

Fig. 8. Boxplots of model performance for summer and winter streamflow simulations for the three surface runoff decision options.

In summer, more models had positive c coefficients and indeed there were cases where both coefficients were negative (Fig. 9).

5 Discussion

5.1 Model structures

The basic assumption in this study was that different model structures are the reason for the differences in model performance. Only four models performed well regarding the F_{logNSE} for both the whole year and for summer and winter. All used a combination of the lower layer/subsurface flow L_{fixedsiz}, upper layer U_{tension2} and the percolation P_{f2sat}, containing at least two of the three components. For all other well performing models a systematic influence of a specific structural decision could not clearly be found. The models performed either better in one of the seasons or for the whole year.

Structural decisions that cause poor performance could be tracked based on the performance criteria F_{logNSE} and the simulation of the recession relationships. Such a structural decision is the lower layer/subsurface flow L_{unlimpow} in combination with the percolation P_{lower}. This combination caused poor low flow simulations for the whole year as well as for the seasonal time series. Most of the binned versions of this combination could not be estimated using the polynomial as they did not pass the Kolomogorov Smirnov test. However, those that did pass, distinguished themselves by steep recession slopes.

The comparison of the slopes of summer and winter recessions reveals no seasonal differences for models with exactly this lower layer/subsurface flow and percolation combination. Clark et al. (2008) explain that here the lower

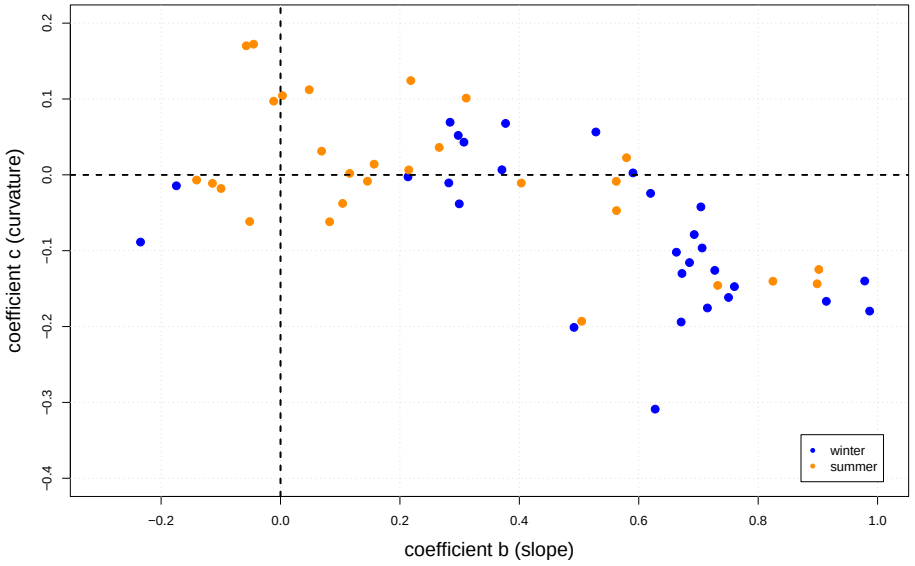

Fig. 9. Coefficients of the polynomial fitted to seasonal $-dQ/dt$ to Q relationships.

layer is defined as a single state variable with no evaporation from this depth. The lower layer corresponds to the subsurface flow which is conceptualised by a power law originating from the parent model TOPMODEL. The main difference between the subsurface flow parametrisation in TOPMODEL and the other parent models is its dependency on the underlying distribution of the topographic index. The storage capacity in TOPMODEL also depends on the topographic index distribution and can hence be smaller or greater depending on the topography. In this study the Gamma distribution was used to define the distribution of the topographic index to keep some flexibility for calibration. Generally, the Gamma distribution is considered to be an appropriate assumption for the topographic index distribution of most catchments (Sivapalan et al., 1987). However, the models that used the TOPMODEL options may not have represented the topography in the Narsjø catchment well enough.

The percolation option P_{lower} is dependent on the lower layer decision. It thus strengthens the assumptions made with the lower layer/subsurface flow decision. The percolation option causes the fastest drainage when the lower layer is dry (Clark et al., 2008). Steep recession slopes were modelled with the combination of P_{lower} and L_{unlimpow}. The calibration with this combination appears to have caused a small water holding capacity of the lower layer resulting recessions that are steeper than recessions in the observed data.

For the winter recessions of the models containing this combination for lower layer/subsurface flow another fact should be kept in mind: in winter a snow storage is included. The precipitation data was pre-processed with the same snow routine for all FUSE models. Models input in winter is precipitation plus snow melt. Towards the end of the winter season (May/June) this process might fill the storages with small amounts of melt water and produce a prolongation of the recessions. The recessions modelled with the combination of lower layer/subsurface flow and percolation options L_{unlimpow} and P_{lower} are too fast and this results in unrealistic shapes of the recession relationships. The percolation option P_{lower} hence seems inappropriate for a combination with the lower layer/subsurface flow L_{unlimpow} as it results in recessions that are too fast in summer and in streamflow that are too low in winter. None of the model decision combinations has such a distinct influence on model performance as the combination of L_{unlimpow} and P_{lower}.

There are further combinations that systematically influence the seasonal performance: models containing the combination of L_{unlimfrc} for lower layer and P_{f2sat} for percolation perform poorly for winter low flows. P_{f2sat} seems to influence the models ability to simulate low flows as it was used by all poorest performing models for winter. This means that the assumption of a percolation based on the field capacity should not be used to simulate winter recessions.

In summer, however, other model decisions cause a poor performance: one example is S_{tmdl} that models poor summer recessions. S_{tmdl} differs from other structures by surface runoff based on the distribution of the topographic index. Many model combinations in summer perform poorer when they contain the S_{tmdl} surface runoff. In summer, surface runoff plays a larger role for recessions than in winter.

Generally, model performance for low flows is easier to analyse for winter than for summer. In summer, there are several fast responding storages that contribute to the streamflow. The longer the recessions last, the less important become quickly draining storages that are prone to evaporation while slowly draining storages gain more influence. In addition, there can be a considerable influence by transpirating

vegetation (Federer, 1973). In winter, the only storages that are important are lower layer storages and snow. Since only one snow storage option was modelled, only the lower layer storages matter. The results point out that the most important features for winter recession are directly connected to the lower storages. Hence, it is rather surprising to find a distinct modeling decision that causes a similar performance for both winter and summer recessions (L_{unlimpow} plus P_{lower}).

In this study the choice of model structures was constrained to the structures of only four parent models. To keep the analysis manageable, in addition some processes were explicitly exempt, similar to the approach used in the original FUSE model (Clark et al., 2008). This includes climate input and hence required the preprocessing of the input data was with a snow accumulation and melt model instead of including several structural decisions of a snow model. Snow is in fact important in the Narsjø catchment, and testing structures describing the processes connected to snow might be worthwhile. This study, however, focused on the impact of model structures used to represent groundwater storage and release behaviour. Future applications should consider testing more structures describing processes of snow melt and accumulation, but also interception and evapotranspiration, all of which were described with a single structural decision in this study. Further, the storage structural decisions included in this study are not the only options. Combinations of linear and non-linear reservoirs in series or parallel as tested in other studies could be appropriate for the Narsjø catchment as well (e.g. Wang, 2011). Generally, it should be considered that an exclusion of alternative process representations using multiple hypothesis methods as FUSE can lead to the realization and evaluation of a model being biased by the modelers view (Clark et al., 2011a).

5.2 Data quality

During the analysis some data issues common to winter streamflow measurements emerged. When ice forms in the river and at the gauging station, backwater effects may result due to ice blocking the channel. This will affect the validity of the rating curve or stop measuring devices alltogether requiring data gaps to be filled later (see e.g. Moore et al., 2002). A few mostly horizontal stripes can be seen in the Narsjø data when plotting flow on a log scale (Fig. 5). However, here no gaps were filled (NVE, personal communication, 2010). Rupp and Selker (2006) also mention that measurement accuracy and changing rating curves in general may be the source of stripe-like patterns as in Fig. 5a. The difficulties of measuring low flows, particularly in winter, are well known and difficult to avoid. More detailed discussions can be found, for example, in Tallaksen and van Lanen (2004).

In general, validation of models with observed data of poor quality may lead to the rejection of models that might in fact be appropriate. A way to avoid the evaluation of model performance by standard metrics, such as the mean squared error, is to use diagnostic signatures (Yilmaz et al., 2008; Gupta et al., 2008). To include additional data on individual processes within a catchment may be necessary to identify scientifically defensible modeling strategies. Examples of application of diagnostic signatures in recession analysis can be found in e.g. McMillan et al. (2011) and Clark et al. (2011b).

6 Conclusions

In this study the impact of model structure on low flow simulations and recession behaviour has been assessed using the Framework for Understanding Structural Errors (FUSE). Using specific model structure combinations of different conceptual models resulted in different model performances for summer and winter low flows. Overall, individual structural decisions never appeared to be an exclusive reason, but rather the combinations of specific structural decisions affected model performance. Evaluating with F_{logNSE} as objective function, led to only a small number of models that performed well. While most well performing models did not allow for the detection of a systematic influence of a model structure combination on the model performance, poor performance was more clearly linked to specific model structures.

A specific structural combination for lower layer, subsurface flow and percolation was found that performed poorly in both seasons. The lower layer and subsurface flow structures influenced the winter low flow simulation, particularly. One main finding of this study was that there is a difference in model performance for summer and winter low flow and recession. In fact, all the structural decision combinations that were salient in this study were season specific – beside one combination that led to the poorest performance, independent on the time period.

An important task would be to test this further for additional catchments with a seasonal flow regime (with snow in winter). In order to elucidate to which extent the influence of the considered model on low flow simulations are catchment specific or can be generalized, it should be replicated in other catchments. Those catchments should ideally be located in different topographical, geological and climatological regions.

The method itself, i.e. a systematic analysis of the structures of hydrological models within the FUSE framework, using objective functions targeting at low flow and recession behaviour, seems promising. For low flow modelling it seems appropriate to use multiple objective functions and not to rely too much on a single function that is based on a comparison between simulated and observed data. Then, using FUSE allows to look at the model structures separately and to investigate the influence of the model structure on the model performance during low flow.

Acknowledgements. The authors thank the German Academic Exchange Service (DAAD) for their financial support and the Norwegian Water Resources and Energy Directorate (NVE) for providing the data on which this study is based.

Edited by: N. Basu

References

Ambroise, B., Beven, K., and Freer, J.: Toward a generalization of the TOPMODEL concepts: Topographic indices of hydrological similarity, Water Resour. Res., 32, 2135–2145, 1996.

Beldring, S., Engeland, K., Roald, L. A., Sælthun, N. R., and Voksø, A.: Estimation of parameters in a distributed precipitation-runoff model for Norway, Hydrol. Earth Syst. Sci., 7, 304–316, doi:10.5194/hess-7-304-2003, 2003.

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

Boone, A. and Wetzel, P.: Issues related to low resolution modeling of soil moisture: Experience with the PLACE model, Global Planet. Change, 13, 161–181, 1996.

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), I: Model intercomparison with current land use, Adv. Water Res., 32, 129–146, 2009.

Brutsaert, W. and Nieber, J.: Regionalized drought flow hydrographs from a mature glaciated plateau, Water Resour. Res., 13, 637–643, 1977.

Burnash, R.: The NWS river forecast system-catchment modeling, Computer Models of Watershed Hydrology, Water Resources Publications Highlands Ranch, CO, 311–366, 1995.

Clark, M. and Kavetski, D.: Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes, Water Resour. Res., 46, 1–27, 2010.

Clark, M., Rupp, D., Woods, R., Tromp-van Meerveld, H., Peters, N., and Freer, J.: Consistency between hydrological models and field observations: linking processes at the hillslope scale to hydrological responses at the watershed scale, Hydrol. Process., 23, 311–319, 2009.

Clark, M., Kavetski, D., and Fenicia, F.: Pursuing the method of multiple working hypotheses for hydrological modeling, Water Resour. Res., 47, 1–16, 2011a.

Clark, M., McMillan, H., Collins, D., Kavetski, D., and Woods, R.: Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses, Hydrol. Process., 25, 523–543, 2011b.

Clark, M. P., Slater, A., Rupp, D., Woods, R., Vrugt, J., Gupta, H., Wagener, T., and Hay, L.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, 12, 2008.

Duan, Q., Sorooshian, S., and Gupta, V.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, J. Hydrol. (Amsterdam), 158, 265–284, 1994.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H. V., Gusev, Y. M., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O. N., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E. F.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, J. Hydrol., 320, 3–17, 2006.

Engeland, K.: ECOMAG – Application to the Upper Glomma catchment (Four separate reports), Department of Geosciences, University of Oslo, Norway, 2002.

Federer, C.: Forest transpiration greatly speeds streamflow recession, Water Resour. Res., 9, 1599–1604, 1973.

Gupta, H., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol. Process., 22, 3802–3813, 2008.

Harman, C., Sivapalan, M., and Kumar, P.: Power law catchment-scale recessions arising from heterogeneous linear small-scale dynamics, Water Resour. Res., 45, W09404, doi:10.1029/2008WR007392, 2009.

Henderson-Sellers, A., Yang, Z., and Dickinson, R.: The project for intercomparison of land-surface parameterization schemes, B. Am. Meteorol. Soc., 74, 1335–1349, 1993.

Holländer, H. M., Blume, T., Bormann, H., Buytaert, W., Chirico, G. B., Exbrayat, J.-F., Gustafsson, D., Hölzel, H., Kraft, P., Stamm, C., Stoll, S., Blöschl, G., and Flühler, H.: Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data, Hydrol. Earth Syst. Sci., 13, 2069–2094, doi:10.5194/hess-13-2069-2009, 2009.

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.

Kirchner, J.: Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, Water Resour. Res., 45, 1–34, doi:10.1029/2008WR006912, 2009.

Leavesley, G. H., Lichty, R. W., Troutman, B. M., and Saindon, L. G.: Precipitation-runoff modeling system users manual, USGS Series Water-Resources Investigations Report, p.207, 1983.

Massey Jr., F.: The Kolmogorov-Smirnov test for goodness of fit, J. Am. Stat. Assoc., 46, 68–78, 1951.

McMillan, H., Clark, M., Bowden, W., Duncan, M., and Woods, R.: Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, Hydrol. Process., 25, 511–522, 2011.

Moore, R., Hamilton, A., and Scibek, J.: Winter streamflow variability, Yukon Territory, Canada, Hydrol. Process., 16, 763–778, 2002.

Palmroth, S., Katul, G., and Oren, R.: Estimation of long-term basin scale evapotranspiration from streamflow time series, Water Resour. Res., 46, 1–13, 2010.

Perrin, C., Michel, C., and Andréassian, 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.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B.: Numerical recipes in FORTRAN: the art of scientific computing, Cambridge Univ Press, 1992.

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

Reusser, D. E., Blume, T., Schaefli, B., and Zehe, E.: Analysing the temporal dynamics of model performance for hydrological models, Hydrol. Earth Syst. Sci., 13, 999–1018, doi:10.5194/hess-13-999-2009, 2009.

Rupp, D. and Selker, J.: Information, artifacts, and noise in $\mathrm{d}Q/\mathrm{d}t - Q$ recession analysis, Adv. Water Res., 29, 154–160, 2006.

Seibert, J.: Regionalisation of parameters for a conceptual rainfall-runoff model, Agr. Forest Meteorol., 98, 279–293, 1999.

Seibert, J.: On the need for benchmarks in hydrological modelling, Hydrol. Process., 15, 1063-1064, 2001.

Sivapalan, M., Beven, K., and Wood, E.: On hydrologic similarity 2. A scaled model of storm runoff production, Water Resour. Res., 23, 2266–2278, 1987.

Smakhtin, V. U.: Low flow hydrology: a review, J. Hydrol., 240, 147–186, 2001.

Tallaksen, L. and van Lanen, H.: Hydrological drought: processes and estimation methods for streamflow and groundwater, Elsevier, 2004.

Tallaksen, L. M.: A review of baseflow recession analysis, J. Hydrol., 165, 349–370, 1995.

Vogel, R. and Fennessey, N.: Flow Duration Curves II: A Review of Applications in Water Resources Planning, J. Am. Water Resour. Assoc., 31, 1029–1039, 1995.

Wang, D.: On the base flow recession at the Panola Mountain Research Watershed, Georgia, United States, Water Resour. Res., 47, 1–10, 2011.

Wittenberg, H.: Baseflow recession and recharge as nonlinear storage processes, Hydrol. Process., 13, 20–33, 1999.

Yilmaz, K., Gupta, H., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resour. Res., 44, 1–18, doi:10.1029/2007WR006716, 2008.

Zhao, R.: Flood forecasting method for humid regions of China, East China Institute of Hydraulic Engineering, Nanjing, China, 1977.