Hydrol. Earth Syst. Sci. Discuss., 10, 8875–8944, 2013
www.hydrol-earth-syst-sci-discuss.net/10/8875/2013/
doi:10.5194/hessd-10-8875-2013

Hydrology and Earth System Sciences Discussions
Open Access

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer[1,2], H. Bormann[3], T. Blume[4], W. Buytaert[5], G. B. Chirico[6], J.-F. Exbrayat[7,8], D. Gustafsson[9], H. Hölzel[1], T. Krauße[10], P. Kraft[6], S. Stoll[11], G. Blöschl[12], and H. Flühler[13]

[1]Chair of Hydrology and Water Resources Management, Brandenburg University of Technology, Cottbus, Germany
[2]Department of Civil Engineering, University of Manitoba, Winnipeg, Canada
[3]Department of Civil Engineering, University of Siegen, Siegen, Germany
[4]GFZ German Research Centre for Geosciences, Potsdam, Germany
[5]Department of Civil and Environmental Engineering and Grantham Institute for Climate Change, Imperial College London, London, UK
[6]Dipartimento di ingegneria agraria e agronomia del territorio, Università di Napoli Federico II, Naples, Italy
[7]Institute for Landscape Ecology and Resources Management, University of Giessen, Giessen, Germany
[8]Climate Change Research Centre and ARC Centre of Excellence for Climate System Science, University of New South Wales, Sydney, New South Wales, Australia
[9]Department of Land and Water Resources Engineering, Royal Institute of Technology KTH, Stockholm, Sweden

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

[10]Institute of Hydrology and Meteorology, University of Technology Dresden, Dresden, Germany
[11]Institute of Environmental Engineering, ETH Zurich, Zurich, Switzerland
[12]Institute of Hydraulic Engineering and Water Resources Management, TU Vienna, Vienna, Austria
[13]Department of Environmental System Sciences, ETH Zurich, Zurich, Switzerland

Received: 12 June 2013 – Accepted: 18 June 2013 – Published: 10 July 2013

Correspondence to: H. M. Holländer (hartmut.hollaender@umanitoba.ca)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The purpose of this paper is to stimulate a re-thinking of how we, the catchment hydrologists, could become reliable forecasters.

A group of catchment modellers predicted the hydrological response of a man-made 6 ha catchment in its initial phase (Chicken Creek) without having access to the observed records. They used conceptually different model families. Their modelling experience differed largely. The prediction exercise was organized in three steps: (1) for the 1st prediction modellers received a basic data set describing the internal structure of the catchment (somewhat more complete than usually available to a priori predictions in ungauged catchments). They did not obtain time series of stream flow, soil moisture or groundwater response. (2) Before the 2nd improved prediction they inspected the catchment on-site and attended a workshop where the modellers presented and discussed their first attempts. (3) For their improved 3rd prediction they were offered additional data by charging them pro forma with the costs for obtaining this additional information.

Holländer et al. (2009) discussed the range of predictions obtained in step 1. Here, we detail the modeller's decisions in accounting for the various processes based on what they learned during the field visit (step 2) and add the final outcome of step 3 when the modellers made use of additional data. We document the prediction progress as well as the learning process resulting from the availability of added information. For the 2nd and 3rd step, the progress in prediction quality could be evaluated in relation to individual modelling experience and costs of added information.

We learned (i) that soft information such as the modeller's system understanding is as important as the model itself (hard information), (ii) that the sequence of modelling steps matters (field visit, interactions between differently experienced experts, choice of model, selection of available data, and methods for parameter guessing), and (iii) that added process understanding can be as efficient as adding data for improving parameters needed to satisfy model requirements.

HESSD
10, 8875–8944, 2013

1 Introduction

Predicting hydrological variables in ungauged catchments is one of the major challenges in hydrological sciences (Sivapalan et al., 2003). The success – or equivalently the uncertainty – of predicting the hydrological response of an ungauged catchment to external driving forces depends (i) on the quality and abundance of catchment data, (ii) on the availability of suitable models, (iii) on the existence of comparable catchments, and (iv) on the modeller her- or himself. In science one often tends to believe that prediction is an objective projection based on "hard" author-independent information. Here, we look at how a group of modellers with the same system knowledge but with different modelling experience and philosophy addresses the problem of predicting a catchment response, while the observed response was made unavailable to them.

To estimate discharge from an ungauged catchment one has to address three major uncertainty sources: (i) the model and its structure, (ii) model parameters, and (iii) model inputs (initial and boundary conditions) (Blöschl, 2006). Most previous model comparisons focused on identifying the relative merits of alternative approaches to these three issues. Intercomparison studies in gauged catchments generally tested whether a particular model structure is superior relative to others (e.g. Naef, 1981; Goodrich, 1990; Reed et al., 2004; Breuer et al., 2009).

In case of ungauged catchments the model comparisons often aimed at optimizing the methods of parameter estimation (Parajka et al., 2005; Oudin et al., 2008) as for instance using pedotransfer functions to guess the hydraulic parameters (Wösten et al., 2001) or regionalisation of parameter values (Seibert, 1999). The role of the modeller was neither systematically investigated in these studies nor even intentionally disregarded making the intercomparison as objective as possible assuming the modellers are interchangeable. The results by Holländer et al. (2009) and Bormann et al. (2011) indicate that the modeller per se is an intrinsic part of a modelling study and has a major bearing on the modelled results, even more so in ungauged catchments because of the more numerous degrees of freedom in making modelling decisions. Here we

HESSD
10, 8875–8944, 2013
Impact of modellers' decisions on hydrological a priori predictions
H. M. Holländer et al.

analyse the role of the modeller in repeatedly predicting the response of a particular catchment based on a stepwise improved database.

This is done by pretending that the artificial catchment "Chicken Creek" (Gerwin et al., 2009b) is ungauged. The discharge record was only known to the organizers of this model intercomparison. Ten modellers (Table 1) were invited to predict the discharge from the 6 ha man-made catchment. In a 1st step, all of them received the same sparse data set. They submitted their 1st stage discharge prediction without having had the possibility to visit the catchment (Holländer et al., 2009). After presenting their 1st stage predictions to each other at a 1st workshop and visiting the catchment, they re-did their prediction using the same data set (2nd stage prediction). For the 3rd stage prediction the modellers were offered additional data. They were then charged, in a virtual sense, with the costs of the parameters they actually selected and used for their 3rd stage prediction. The expected improvement of the "a priori" prediction could then be related to the required additional investment for a more detailed parameterisation.

The objective of this paper is to investigate how modellers address the problem of an a priori prediction in an ungauged catchment in terms of (i) making assumptions about the dominant processes, (ii) choosing the model and its structure, (iii) identifying the model parameters, and (iv) defining the initial and boundary conditions. Furthermore, we discuss how the modeller's attitude changes with enhanced system understanding, and as a side benefit, we analyse the cost-benefit aspect of using models with increased parameter requirements.

2 1st stage prediction: impact of modeller's decisions

2.1 Framework and first instalment of data

The following "sparse" data base has been provided to the modellers for the 1st modelling stage: gridded information on elevation of surface and the underlying clay layer, soil texture and soil depth, mean annual vegetation cover (20 m grid), hourly climate

HESSD
10, 8875–8944, 2013

data (precipitation, temperature, relative humidity, global radiation, wind speed and direction), initial groundwater heads, aerial photo, and a shape file of the gully-network. Hence, in many real world applications the database would be even smaller than in case of this experimental catchment (Fig. 1).

The modellers had been asked to provide their 1st stage predictions on short notice within a period of only two months. None of the modellers had an extra budget or time allocated for this modelling task. The time available for model selection and testing was therefore short. The time invested for setting up the model was in the order of one day up to one week, except one modeller who spent one month for this task (Holländer et al., 2009).

2.2 Prior modelling experience

Ten modellers were invited to predict the discharge (Table 1). A detailed description of the models can be found in Holländer et al. (2009). Prior modelling experiences had a major impact on the modeller's choice of the model and its implementation and on the parameterisation (Holländer et al., 2009). Prior to this study all modellers except the CMF user (**C**atchment **M**odelling **F**ramework) had experience with three to five different models (Table 2). CMF, which is a multi-model toolkit, was developed by the user himself and the Hill-Vi user was a member of the developer's group. None of the groups had experience neither with artificial catchments nor with applying hydrological models in the (semi-) continental climate of Lusatia (Table 2). Their experience ranged from cold (CoupModel, SWAT) to moderate (Hill-Vi, SIMULAT, SWAT, both WaSiM-ETH users[1]), to temperate (NetThales), and tropic climates (SIMULAT, SWAT, WaSiM-ETH (Richards)). The Catflow and Topmodel users worked mainly in the Alps and Andes (Tab. 2). The Catflow, CoupModel, NetThales, SIMULAT, and Topmodel user had an

[1]Two modelers used the WaSiM-ETH model. Here, they are distinguished on the basis of their internal model WaSiM-ETH (Richards) and WaSiM-ETH (Topmodel).

in-depth modelling experience while other just started their PhD (Hill-Vi, SWAT, WaSiM-ETH (Richards)).

The SIMULAT and the Topmodel user worked in a wide range of model applications: model validity, uncertainty analysis, development of regionalisation concepts, quantifying the effect of input data availability (only SIMULAT) and the impact of land use and climate change (Bormann et al., 1999, 2007; Elfert and Bormann, 2010; Buytaert and Beven, 2011). The users of Catflow, Hill-Vi, and WaSiM-ETH (Topmodel) focused on rainfall-runoff modelling. The SWAT and WaSiM-ETH (Richards) users worked also in the field of rainfall-runoff modelling, and additionally, with ensemble approaches in ecological modelling (Exbrayat et al., 2010, 2013) and soil erosion (Hölzel and Diekkrüger, 2011), respectively.

All modellers did preliminary evaluations based on their own experience regarding the dominant processes within the catchment. They neglected infiltration excess knowing that the soils in this catchment are predominantly sandy. Some of the modellers included and others explicitly excluded certain processes, e.g. the existence of a snow pack or soil freezing when snow-free (Holländer et al., 2009), and all of them did not make use of the aerial picture which clearly showed the network of gullies.

2.3 Modelling philosophy

For all modellers it appeared to be a straightforward exercise assuming that an artificial catchment dominated by sandy soils is a rather uniform and homogeneous system although they had no analogue for expert similarity analyses. Therefore, all models except the Topmodel were physically based "bottom-up" approaches assuming that the relevant processes can be realistically represented based on the provided data. Another reason for using this model category is the fact that conceptual models need to be calibrated against observed data or need to be adapted according to hydrological similarities of other catchments. All models except SWAT and Topmodel have a spatially distributed structure to make use of the provided spatially distributed characteristics.

HESSD
10, 8875–8944, 2013

The assumption that an artificial catchment built with sandy soil material is homogeneous justified the use of average soil properties derived from texture data (Catflow, CMF, CoupModel, NetThales, Topmodel, and WaSiM-ETH (Richards)). The state-of-the art experience that sandy soils usually do not produce much direct run-off, led to a significant overestimation of infiltration in the 1st prediction run although deeply eroded gullies were documented by the provided aerial photos (Holländer et al., 2009). Therefore, some modellers concluded that surface runoff could be triggered by saturation excess. In addition, the NetThales user understood the small scale physical processes as the elementary basis for the modelling task primarily based on his experience from prior studies with dominant precipitation events during the winter period with typical discharge coefficients of 50 %. This led to a wrong definition of the dominant water transport processes.

Most of the modellers (CoupModel, CMF, Hill-Vi, SIMULAT, SWAT, Topmodel, WaSiM-ETH (Richards)) chose their model before they defined their modelling philosophy mostly based on their earlier modelling experience whereas others did it afterwards (Catflow, NetThales). The modellers who chose their model beforehand felt that this modelling exercise offers an opportunity to answer the following questions: (a) how does their previously used or developed model behave in a prediction context (CMF, SIMULAT, Topmodel, WaSiM-ETH (both modellers)[2]), (b) how does the modeller's model perform in a region where the model was not used before (CMF, Hill-Vi, SWAT), or (c) the leader of their group simply suggested to use the group's model (Hill-Vi). In contrast to the other modellers, the Catflow and the NetThales modellers chose the model they were familiar with and felt that it is suited for this case. Additional and probably most common motives for selecting a particular model were: immediate availability, familiarity with the code, lack of time, and unavailable funding for extra-work.

The SIMULAT and the Catflow user wanted to minimize the influence of the modeller's choice and did not decide a priori on the dominant process(es). This was the

[2] Although the WaSiM-ETH (Topmodel) user joined for stage II, the initial modeller philosophy is stated in this chapter.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |
HESSD
10, 8875–8944, 2013
Impact of modellers' decisions on hydrological a priori predictions
H. M. Holländer et al.

lesson the SIMULAT user learned from previous studies where SIMULAT produced reliable results without prior calibration (Bormann et al., 1999). The Catflow, SWAT, Topmodel, and WaSiM-ETH (Richards) users had similar philosophies: SWAT is a model designed for larger catchments. Although the model is process based, parameters are more of a conceptual nature, e.g. the usage of the Curve–Number approach to simulate infiltration and surface runoff. Therefore, the user minimized the influence of the modeller's decision and used default values with the objective to investigate how SWAT behaves in case of small catchments. The Topmodel user used the "standard" version of the model as a 1st approximation, knowing that some assumptions are invalid (e.g., exponential decrease of transmissivity with depth). The WaSiM-ETH (Richards) modeller also relied on the description of the physical processes in his model and thereby minimized the influence of his own decisions. For the Catflow user the model is primarily a platform for hypothesis testing. The underlying hypothesis was that the best possible representation of the catchment structure and physical properties would yield the best (possible) prediction. Although suspecting that soil surface crusts might be a dominant feature they were not included in the setup because the Catflow user understood that the modellers were supposed to use exclusively the provided data.

2.4 Model parameterisation

All modellers who used spatially distributed models used either a rectangular/curvilinear grid with up to 20 m spacing or irregular grids (CMF). All except the NetThales and SIMULAT user modelled at least a saturated and an unsaturated layer. SIMULAT started with unsaturated conditions but allowed for partial saturation in case of storage based lower boundary condition of the soil columns.

Soil: The modeller's prior modelling experience had a major influence on the parameter estimates, e.g. all of them employed pedotransfer functions (PTF) for assessing soil hydraulic properties: Catflow, Hill-Vi, NetThales, SIMULAT, SWAT, and Topmodel used published PTFs, but national soil data-bases were used for the

HESSD
10, 8875–8944, 2013

CoupModel, CMF, SIMULAT (only bulk densities), and WaSiM-ETH (Richards). Most modellers did not consider the influence of soil freezing on the hydraulic conductivity (K_{sat}) because their model has no routine for snow and frost effects (Catflow, Hill-Vi, Topmodel) or the modeller did not have sufficient experience in this context (CMF, NetThales).

Evapotranspiration: The evapotranspiration parameters were defined primarily based on prior experiences in areas with quite different climatic regimes. None of the modellers had worked with similar climatic conditions before. For instance, the NetThales user tried to match the annual water balance with that of catchments having a similar rainfall regime. The CoupModel user pointed out the importance of soil and snow evaporation based on a surface energy balance.

Initial state: All modellers missed the fact that the catchment and its hydrological behaviour was far from steady state conditions (e.g. the dry initial soil conditions with almost no groundwater). In order to determine the initial conditions, the Catflow, CMF, Hill-Vi, and the WaSiM-ETH (Richards) modeller used several warm-up runs to achieve steady state conditions. The other modellers assumed uniformly distributed soil moisture and groundwater levels. Only the SIMULAT and the SWAT user implemented the dynamics of the lake at the catchment outlet. The parameterisation for and the results of the 1st stage prediction are discussed in detail in Holländer et al. (2009).

The importance of data and parameterisation by the modellers of distributed models can be ranked as (1) terrain information with its lower (clay base below soil) and upper boundary (surface), (2) soil depth and texture, and (3) vegetation coverage.

HESSD
10, 8875–8944, 2013

3 2nd stage prediction: impact of system understanding

3.1 Re-definition of the dominant processes

All modellers except the Topmodel user presented their 1st stage predictions at the 1st workshop. The WaSiM-ETH (Topmodel) user attended this workshop as an observer and subsequently joined the project. The discussions about the major system controls during the workshop, field visit, and in the course of the manuscript preparation (Holländer et al., 2009) changed and homogenised the system understanding of all project partners. Some modellers visited the catchment later in the following spring period (SIMULAT and Topmodel). The workshop revealed that the modellers exploited the initial data set quite differently. All of them neglected the low initial water content of the 2 to 4 m thick soil layer dumped onto the clay base.

Subsurface structures: The V-shaped subterranean clay dam, which was meant to funnel the downslope subsurface flow at the catchment outlet into the lake was handled very differently (Table 3). The CoupModel, Topmodel and WaSiM-ETH (Richards) users did not consider the dam at all and the SIMULAT user assumed a shallow soil layer above the dam. Therefore, the modelled subsurface water storage and flow immediately uphill of the dam differed strongly.

Surface processes: Most of the modellers did not make use of the documented gully network as an indication for the dominant runoff generation process. Possible causes for the observed soil erosion (soil crusts) as assumed but not yet implemented by the Catflow, SIMULAT, and the WaSiM-ETH (Richards) modellers were discussed during the workshop. Based on the fast filling of the lake in early 2006 (Gerwin et al., 2009b) snow melt and soil freezing were postulated as additional explanations for the observed erosive surface runoff. Due to the large variation in the predicted water budget and runoff, the WaSiM-ETH (Topmodel) modeller concluded from the discussions that a completely physically based model for the unsaturated zone is not needed in this case.

Afterwards, during their visit of the catchment the modellers revised their system understanding for the 2nd stage prediction without receiving any additional data. Most important for their system understanding was the on-site inspection of the catchment when the modellers became aware of the deep gullies (Catflow, CMF, Hill-Vi, NetThales, SWAT, Topmodel, WaSiM-ETH (both modellers)). The fact that the models did not predict infiltration excess suggested the necessity to implement soil crusts as observed by Fischer et al. (2010). The biological soil crusts developed within the 1st year after the construction of the catchment.

Vegetation: Those who visited the catchment in spring noticed the fast development of the vegetation (SIMULAT and Topmodel). The CoupModel, the Hill-Vi, the NetThales, the Topmodel, and the two WaSiM-ETH users noticed the different vegetation development in the western and eastern half of the catchment. The WaSiM-ETH (Topmodel) modeller noticed the rapid changes of the surface terrain apparently depending on soil texture.

All modellers except the NetThales user integrated the findings reported in the joint publication (Holländer et al., 2009) for their next prediction. They analysed the scientific arguments and the parameterisation of the other modellers to justify their process and parameter choices (Table 3). The Hill-Vi modeller made use of the information from Gerwin et al. (2009a) about the catchment and the discharge magnitude. The importance of the processes in the catchment and the modeller's decisions were ranked in the following order:

1. Soil crusts enhancing surface runoff (workshop and field visit).
2. Low initial soil water contents and transient conditions during the filling of the soil storage compartments (workshop and field visit).
3. Influence of the V-shaped subsurface clay dam (workshop).

HESSD
10, 8875–8944, 2013

4. Snow melt and soil freezing (workshop and manuscript).

5. Vegetation development (field visit).

6. Analysis of parameterisation (manuscript).

7. Fast change of topography and soil surface structure (field visit).

3.2 Implementation of gained system understanding

The Hill-Vi user implemented most rigorously what he learned in the process. Recognizing that the original model could not adequately describe the actual evapotranspiration (AET), soil crusting, and the effect of the clay dam on the groundwater dynamics without major coding work, he switched to using MIKE SHE[3], redefined the initial conditions, and made small changes to reduce AET and potential evapotranspiration (PET) (Table 4).

Surface condition: Soil crusting was included in six of the ten models by adding a thin layer with a reduced saturated hydraulic conductivity (K_{sat}). It was not included in the CoupModel, CMF, SWAT, and Topmodel. The Catflow modeller modified the van Genuchten parameters to reduce K_{sat} of the crust layer according to Simunek et al. (1998) and the Topmodel user generated a surface runoff of 10 % of the precipitation by changing K_{sat} at the surface (5 mm h^{-1}) and the capillary drive CD (1 mm) (Morel-Seytoux and Khanji, 1974). The SIMULAT modeller used K_{sat} = 2.1 mm h^{-1} for the crust layer taken from Hölzel et al. (2011). The WaSiM-ETH (Richards) modeller used 20 mm h^{-1} based on NAW (2008) (Table 5). All modellers except MIKE SHE, SIMULAT, and SWAT reduced K_{sat} and the WaSiM-ETH (Topmodel) user estimated the hydraulic soil parameters by PTFs according to Wösten and Nemes (2004) and Saxton et al. (1986) and used a set up based on the DEM data. To achieve this the

[3]The Hill-Vi user is referred to as MIKE SHE user hereafter

SIMULAT modeller used information on spatial variability in K_{sat} (standard deviation: $\sigma_{K_{sat}} = 62.5\,\mathrm{mm\,h^{-1}}$) given in Cosby et al. (1984). Similarly, the CoupModel, SIMULAT, Topmodel, and the WaSiM-ETH (Topmodel) modellers changed their parameterisation of the vegetation. Only the MIKE-SHE user accounted additionally for snowmelt by implementing a simple snowmelt routine (degree-day method).

Evapotranspiration: Most of the modellers reduced the influence of the vegetation whereas in SIMULAT the leaf area index was increased to LAI > 1 (**l**eaf **a**rea **i**ndex) (for the year 2008 at selected grid points). The WaSiM-ETH (Richards) modeller split the model into a vegetation-free (before 2008) and a vegetation period with a low density grass canopy. The detailed description of parameterisation of the 2nd stage prediction of WaSiM-ETH (Richards) and SIMULAT can be found in Hölzel et al. (2011) and Bormann (2011). The Topmodel user increased the influence of the vegetation. The MIKE SHE modeller substituted the PET calculation based on Turc with the FAO Penman Monteith method with a dynamic crop coefficient.

Initial state: The CMF, CoupModel, MIKE SHE, NetThales, WaSiM-ETH (Richards), and SWAT user changed the initial state from pre-wetted to quasi-dry. The WaSiM-ETH (Richards) modeller used one pre-run with zero precipitation instead of the three runs with the original precipitation data (Table 5). The CoupModel and the NetThales users implemented dry conditions in a similar way. The SWAT modeller removed the spin-up period completely whereas the WaSiM-ETH (Topmodel) modeller used the initial conditions generated by a one-year pre-run using the 1st year data.

Boundaries: The SIMULAT user changed the lower boundary condition of the individual soil columns to a linear storage based boundary condition to account for the steadily rising groundwater level and enhanced the damming effect of the subsurface V-shaped clay dam by modifying the lower boundary condition (storage based boundary condition), a solution possible due to the one-dimensional nature of this model. In total, more than half of the modellers (Catflow, MIKE SHE, NetThales, SIMULAT, Topmodel, WaSiM-ETH (Richards)) increased or refined the influence of the clay dam. For instance, the WaSiM-ETH (Richards) user defined the formerly constant layer thickness according to the Digital Elevation Model (DEM). The Topmodel user increased the subsurface response slightly (log of the areal average of the local transmissivities at saturation of the soil: lnTe) from -2.5 to $-2\,\mathrm{m}^2\,\mathrm{h}^{-1}$ equal to transmissivity Te from 0.082 to $0.135\,\mathrm{m}^2\,\mathrm{h}^{-1}$ since the clay dam delays the subsurface response and the modeller tried to mimic this by changing lnTe. This is a merely intuitive approach since there was no guidance on how to adapt the parameter.

Various modifications: In addition, many small changes were made by the modellers in the model structure and parameterisation. The WaSiM-ETH (Richards) user included the lake and the SIMULAT and SWAT users changed the volume of the already implemented lake to match with the volume at the spillway. The CMF user accounted for the deeply eroded gully structure. He expected that water exfiltrates through preferential flow pathways into the gullies even when the topsoil is still unsaturated. This substantially increased surface runoff through the gullies. Similarly, the SWAT user allowed reinfiltration from gullies (Table 4).

4 3rd stage prediction: impact of an extended data set

4.1 Revisions to cope with earlier modelling results

A 2nd modelling workshop was held in October 2009. Modellers had a better view of the dominant processes controlling the catchment response. Several modellers focussed on the following key issues: (i) is it necessary to adapt the model structure to represent the dominant processes? (ii) What type of data is required for improving the model parameterisation? (iii) How can the observed heterogeneity be accommodated?

HESSD
10, 8875–8944, 2013

The MIKE SHE modeller improved the implementation of the soil crust by adopting the parameterisation used by the other modellers. Additionally, he found that the change from Turc to Penman Monteith based PET calculations was a step forward but still needed revision again. Similarly, the SIMULAT modeller noticed that the modelled results still varied significantly among the various models. The simulated water balance was consistently wrong, changes in evapotranspiration parameterisation did not seem to be appropriate, and the subsurface storage needed to be better adapted. As a result, the SIMULAT modeller updated the variability in surface K_{sat} and the lower boundary condition of the soil columns once more in order to better describe the infiltration as well as subsurface storage (Bormann, 2011). Additionally, the initial condition was re-defined as dry soil.

The CoupModel user reconsidered the usefulness of the semi-distributed version of the CoupModel and used a 1-D version because he assumed that the erosion gullies are acting as drainage pipes. He felt that a 1-D modelling approach will be more efficient than a 2-D approach and allows incorporating different data, such as soil temperature and soil moisture. The WaSiM-ETH (Richards) user emphasized the relevance of the groundwater dynamic by a more realistic representation of the clay dam (Holländer et al., 2009). He replaced the conceptual 1-D by a process-based 2-D groundwater approach (Hölzel et al., 2013).

The Catflow user changed the spatial resolution of the model grid in order to reduce the numerical problems caused by the implementation of the surface crust on the one hand and computation time on the other hand: the horizontal resolution was reduced to 5 m for the entire slope (before: upper slope 10 m, middle slope 5 m and lower slope 1 m resolution). Furthermore the top 5 cm of the soil were now resolved with vertical increments of 1 cm (before: top 20 cm with 4 cm resolution). This resulted in an 18 % reduction of the total number of nodes.

The NetThales user realised that the model is not suitably structured for an efficient description of the groundwater dynamics, particularly in the initial stage of the emerging groundwater table. The clay dam required a spatial structure of the subsur-

HESSD
10, 8875–8944, 2013

face dynamics. NetThales was designed for catchments with shallow soils as well as for morphology dominated controls in subsurface flow. However, changing the model structure caused substantial costs without a significant reduction of the prediction uncertainty due to the uncertainty in the model parameterisation based on the sparse data set.

The SWAT model user kept the model setup used during the 2nd prediction, but corrected some parameter values based on available data.

4.2 The additional data set

The modelling period for the 3rd stage prediction was extended and covered the period from 29 September 2005 to 4 August 2009. The database contained more and better data because most of the newly installed measuring devices had been installed in 2008. Detailed information on the field equipment and methods can be found in Gerwin et al. (2011) and Mazur et al. (2011). The record of weather station I included hourly data of precipitation, air temperature, wind speed and direction, humidity, global radiation, and the vegetation coverage of 2009. The new data set included all information, apart from discharge:

- K_{sat} measured by slug tests (at 15 grid points);
- K_{sat}, porosity, and bulk density measured in the laboratory on undisturbed samples taken at the four soil pits;
- Soil water retention curves from two soil pits (four depths);
- Carbon content of all observation points at several depths;
- Infiltration rates measured in situ (19 measurements at 10 grid points);
- Daily soil moisture data in the four soil pits at 10, 30, 50, and 80 cm depth;

HESSD
10, 8875–8944, 2013

– Ten minutes, hourly, daily, and monthly weather data monitored at weather station II: air temperature, wind speed and direction, humidity, net and global radiation, each measured at 2, 5, and 10 m elevation above surface, and precipitation, soil temperature and soil heat flux;

– Detailed data about plant species and their distribution at all observation squares;

– DEMs of soil surface elevation determined in November 2005, May 2006, November 2007, and August 2008.

The costs for instrument acquisition, installation, measurement campaigns, and maintenance were estimated according to LAWA (2005) starting at the beginning of the project in 2005. Costs for data inspection and storage are not included. The pro forma costs of data acquisition are documented in Table 6. The cost of the modeller's time was not accounted for.

4.3 Modellers' rationale of data selection

The CMF and the WaSiM-ETH (Topmodel) modeller left the modelling group after completion of their PhD program. Although the WaSiM-ETH (Topmodel) user left the group, he requested additional data. This will be discussed below. All remaining modellers selected K_{sat} derived from slug tests in the field and K_{sat} determined in the laboratory (except CoupModel), porosity (except CoupModel and WaSiM-ETH (Richards)), water retention (except SWAT, Topmodel, and WaSiM-ETH (Topmodel)), and soil moisture data (except Catflow, and MIKE SHE). The carbon content of the soil, infiltration rates, more detailed weather data, DEMs and vegetation data were considered of lesser importance and were selected only by few modellers (Table 7).

Generally, two strategies were followed to request additional data: (i) asking for all data which could be used in the specific model to aim for the best possible prediction and (ii) considering the pro forma costs to achieve a good benefit-cost ratio. Surprisingly, none of the modeller opted for the entire data set. Only the NetThales modeller

HESSD
10, 8875–8944, 2013

tried to obtain the best fit with the best set of the available data. At this stage he was still aiming at improving his perception of the catchment dynamics and was particularly interested in understanding the vertical dynamics (infiltration and soil water redistribution). He felt that he could infer a better parameterisation of the model although he was already aware of the fact that the model could not mimic the actual catchment behaviour, in particular for what concerns the groundwater dynamics in the early stage of development. Similarly, the WaSiM-ETH (Richards) user still relied on the description of the physical processes and therefore, he requested data which allowed improving the model. Similarly, the SIMULAT modeller chose the data based on their usefulness for complementing or revising the model set-up. After a review of the data he used less data than he opted for.

All other modellers selected the data, which were a must or an optional input for the model. For instance, the Topmodel and the WaSiM-ETH (Topmodel) modellers were limited by the conceptual nature of the model, which made it difficult to integrate additional data. Therefore, both modellers chose the same soil and soil moisture data sets. Similarly, the SWAT user did not select the soil moisture data because the HRUs (**H**ydrologic **R**esponse **U**nits) were separated from each other and the model lacked groundwater flow through the soil profile. The MIKE SHE, Catflow, and Topmodel user rigorously requested only the information that appeared to be most important for their model.

The soil moisture data could be used as input or as calibration data for defining soil parameters. Soil moisture data were not used by the Catflow and MIKE SHE modellers. Neither the NetThales nor SWAT modeller calibrated their models against soil moisture. The NetThales user argued that the limited point observations cannot be exploited for calibrating his distributed model at the element scale. In fact, there is a mismatch between the scale of the computation (plot scale) and the scale of the measurements, corresponding to that of soil cores. The WaSiM-ETH (Richards) modeller used these data to have a control on the soil moisture dynamic, but not for calibration as done by the CoupModel, SIMULAT and Topmodel modeller. The Topmodel user averaged all TDR values measured at 10 cm depth and the groundwater levels of observation well L4 (Holländer et al., 2009) as proxies for the storage deficit. He expected that the latter two observations are inversely related and performed a Monte Carlo sensitivity test based on the correlation coefficient as performance measure. He deduced from this that only the amount of water (expressed as a depth) which the soil can hold within the root zone (Sr_{max}) is a sensitive parameter. Finally, he chose Sr_{max} to 0.02 m. Subsequently, the initial root zone storage deficit (Sr_0) and the initial subsurface flow per unit area (q_{s0}) were updated to be compatible with Sr_{max}. The SIMULAT modeller used soil moisture in two steps. First, he evaluated his model with these data and, in a 2nd step, he used them to calibrate the model by adjusting the lower boundary conditions of the soil columns (Bormann, 2011). The CoupModel user employed soil moisture and soil temperature data to calibrate evapotranspiration. This was done by using only the less costly temperature data in a 1st trial and then adding the more expensive soil moisture later on.

4.4 Implementation of data

Soil and surface crust: The SWAT modeller used the extended dataset (e.g. the organic carbon content) for updating soil parameter with observations and the MIKE SHE user modified the soil and crust layer properties (thickness and K_{sat}). The Catflow modeller determined the van Genuchten parameters from the water retention curve data (with $K_{sat} = 110\,\mathrm{mm\,h^{-1}}$). She used a combination of these parameters and the infiltration rate ($9.6\,\mathrm{mm\,h^{-1}}$) to model the surface crust. Similar values were determined by the NetThales modeller. He questioned the K_{sat} values. The lab measured K_{sat} of the top layer cores were similar to those estimated with PTF from textural data ($\sim 100\,\mathrm{mm\,h^{-1}}$) whereas results from slug tests were similar to laboratory measurements on soil cores from the 200–300 cm depths ($\sim 10\,\mathrm{mm\,h^{-1}}$). However, the NetThales modeller stated that the observed values are likely not representative for the hydraulic properties of the soil crust. He concluded from his field observations that the main trigger was in-

HESSD
10, 8875–8944, 2013

filtration excess runoff. Assuming a lognormal distribution of the observed values, he used the 2.5 % value of the cumulative distribution. The Catflow modeller arrived at the same conclusion. She parameterized the soil crust using data by Simunek et al. (1998) ($K_{sat} = 0.06\,\mathrm{mm\,h^{-1}}$). The surface crust parameters were applied to the top soil layer of 2 cm thickness. However, introducing the surface crust caused numerical problems, increased the overall numerical error and computation time due to the extremely large hydraulic conductivity gradient.

Using the additional soil physical data the magnitude of the soil hydraulic parameters used in the preceding SIMULAT simulations was confirmed. The data from the infiltration experiments were used to improve the description of the infiltration behaviour through the surface layer. The field data were complemented by literature values to parameterise the spatial variability of K_{sat} (Cosby et al., 1984).

Vegetation: The NetThales modeller observed that the soil water retention data were consistent with the PTF predictions based on textural data. The time series of measured soil water contents confirmed that the top 30 cm is the portion of the soil column, which exhibits rapid daily and seasonal changes in soil water content. Therefore, the NetThales user kept the value of 20 cm as reference for the root zone depth. The prevalent plant species (*Trifolium arvense*) was parameterized (max. LAI, rooting depth, and stomatal conductance) based on literature values from the PlaPaDa database (Breuer et al., 2003) by the SWAT modeller. The MIKE SHE modeller re-evaluated the evapotranspiration because the change from the Turc to the Penman–Monteith equation did not produce realistic values.

Snow: The CoupModel user wanted to better understand why there is so little snow in his simulations, and whether this had some impact on the water balance simulations. Therefore, he tested a 1-D approach to analyse the flow system. He intended to implement the following flow pathway: (i) infiltration into the soil between

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

the gullies, (ii) lateral drainage into the gullies, and (iii) runoff along the gullies into the lake.

5 Results

5.1 Main results from the 1st prediction (minimal data set)

The main result of the 1st prediction was the huge variability of variables among the models, notably regarding the simulated water budgets (Fig. 2) and daily discharges (Fig. 3). This has been discussed in detail by Holländer et al. (2009). Expressed as percentage of the observed annual discharge, the predicted discharge ranged from 20 to 300 %. This means that the frequency distribution of simulated discharge varied drastically as well. It is remarkable that the variability in simulated actual evapotranspiration was high although most models were based on the Penman–Monteith approach used to predict the potential evapotranspiration. The predicted actual evapotranspiration ranged from 88 to 579 $mm\,yr^{-1}$.

Most models simulated too much discharge due to continuous subsurface flow while the catchment showed an obvious gully network suggesting massive surface runoff (Fig. 3). Due to the continuous lateral subsurface flow into the gullies the subsurface storage changed too slowly and the change in catchment storage over time was not well predicted by any of the models. Note that the initial soil-water content was not defined by the provided data.

The differences among the model predictions could be mainly attributed to different modeller decisions regarding conceptualisation and parameterisation of their models. We learned that the modellers per se are part of the modelling process and have a major impact on the modelled results. Their decisions during the phase of model implementation and parameterisation are strongly influenced by their experience from previous modelling studies. The detailed description of the modelling results and its discussion of this phase can be found in Holländer et al. (2009).

HESSD
10, 8875–8944, 2013

5.2 Main results from the 2nd prediction (after field visit)

The 2nd prediction resulted in the average of the ten predictions in larger PET (+103 mm yr^{-1}), AET (+56 mm yr^{-1}), ΔS (+49 mm yr^{-1}), and lower Q (–60 mm yr^{-1}) (Fig. 2) than calculated in the 1st prediction. The variability among the predictions of PET, AET, and Q decreased considerably but the variation of ΔS increased strongly. This can also be seen in the frequency–discharge relationship (Fig. 5). All models except SIMULAT and WaSiM-ETH (Richards) show discharges, which are more similar compared to each other than in the 1st prediction. However, the model predictions still differ for the maximum discharge Q_{max} (e.g. Q_{max}: CoupModel 65 m^3 d^{-1} and Topmodel 949 m^3 d^{-1}) and in the form of the frequency–discharge relationship.

Most predictions for PET in the hydrological year 2005/2006 were in the range from 600 to 800 mm yr^{-1}, just CMF (443 mm yr^{-1}) and Topmodel (1014 mm yr^{-1}) predicted more extreme values. The corresponding values of the two other years were similar (Table 8). AET, which varied from 157 (SIMULAT) to 364 mm yr^{-1} (WaSiM-ETH (Topmodel)) for the 1st year, was increased in the two following years. In the 2nd year the maximum AET was 465 mm yr^{-1} (Topmodel) and in the last year 381 mm yr^{-1} (Topmodel). Note that the observation period of the last year was shorter (until 8 September 2008). Most models predicted AET in the order of 300, 410 and 330 mm yr^{-1} for the 1st, 2nd, and 3rd year, respectively. Only SIMULAT (157, 266 and 260 mm yr^{-1}) and WaSiM-ETH (Richards) (234, 273 and 285 mm yr^{-1}) predicted significantly lower AET.

The range of the of discharge during the three years was 36 to 174 mm yr^{-1}, 99 to 327 mm yr^{-1} and 125 to 273 mm yr^{-1} which results to 32 to 154 %, 94 to 311, and 111 to 271 % of the measured discharge. Most modellers except the MIKE SHE (–13 mm yr^{-1}) and the SWAT (–24 mm yr^{-1}) user started their model runs with the initially dry state of the catchment and predicted positive storage changes throughout the entire simulation period (Table 4). All other models predicted storage changes in the 1st year of about 50 to 85 mm yr^{-1}. In the 2nd year CoupModel, MIKE SHE, SIMULAT, and SWAT predicted about 20 to 40 mm yr^{-1} of storage change, CMF (108 mm yr^{-1}), NetThales (88 mm yr^{-1}), and WaSiM-ETH (Richards) (53 mm yr^{-1}) predicted larger values. Similar values were obtained for the 3rd year. Catflow could not present discharge data in the 2nd prediction due to numerical problems.

The errors in the mass balance were less than 5 % in most models except in SWAT (8 %) and in the two WaSiM-ETH models (Topmodel: 17 % and Richards: 17 %). A similar mass balance was observed in the following years with SWAT and WaSiM-ETH (Topmodel) producing smaller errors of about 5 %.

The discharge of the 2nd prediction is illustrated in Fig. 4 for the 2nd year. SIMULAT was the only model, which had a zero flow period during 40 % of the simulation period (see Supplement A). CoupModel (2 %) and NetThales (8 %) predicted also periods with zero discharge. Times with base flow below 1 m^3 d^{-1} were also found by CMF. All other models consistently predicted base flow larger than 2 m^3 d^{-1} (Fig. 5). The largest base flow was predicted by WaSiM-ETH (Richards) and WaSiM-ETH (Topmodel) always at least 15 and 10 m^3 d^{-1}, respectively. Although SIMULAT predicted zero discharge for many days, it rose rapidly and produced the largest peak discharge of all models (1433 m^3 d^{-1} in the 2nd year). Also CoupModel and SWAT predicted large discharges with Q_{25} equal to 11 and 9 m^3 d^{-1}, respectively. All other models predicted lower total discharges and hence a low base flow. The CMF, SWAT, and CoupModel predicted peak discharges of less than 100 m^3 d^{-1}, followed by NetThales (287 m^3 d^{-1}), WaSiM-ETH (Topmodel) (701 m^3 d^{-1}), MIKE SHE (742 m^3 d^{-1}), Topmodel (958 m^3 d^{-1}), WaSiM-ETH (Richards) (1028 m^3 d^{-1}), and finally SIMULAT (Fig. 5). All models with a peak discharge larger than 100 m^3 d^{-1} have in common that there is a strong discharge reduction between Q_{100} and Q_{95} of nearly two orders of magnitude.

SIMULAT predicted primarily surface runoff (~ 80 %) and negligible interflow (Table 10). The CoupModel, MIKE SHE, SWAT, and Topmodel computed a surface runoff of about 50 % and cumulated interflow and base flow. Both WaSiM-ETH modellers succeeded to differentiate surface runoff, interflow and base flow. The WaSiM-ETH

(Richards) modeller obtained 5 % surface runoff, 56 % interflow and 39 % base flow, where the WaSiM-ETH (Topmodel) modeller obtained in the average 21 % surface runoff, 19 % interflow and 60 % base flow. CMF and NetThales could not distinguish the discharge components.

Only the CoupModel, CMF, and MIKE SHE generated groundwater levels (Supplement C). The CoupModel used the drainage equation by Hooghoudt (1940) and MIKE SHE Darcy's law. MIKE SHE used a large K_{sat} because the water table draw down was too slow in dry periods. The CoupModel used a too small storage coefficient and predicted too small groundwater table fluctuations. CMF was not responding to drying periods. The groundwater table simulated by CMF was since spring 2008 in a quasi-steady state at a depth of ~ 30 cm. WaSiM-ETH (Richards) mentioned that due to the 1-D groundwater approach used in the 2nd prediction, the role of the clay dam could not be taken into account.

5.3 Main results from the 3rd prediction (extended data set)

Figure 2 shows that for the first two hydrological years the average PET (3rd prediction) was slightly reduced when the modellers had access to the larger data sets. Due to a longer simulation period (until 3 August 2009), the predictions for hydrological year 2007/2008 were not directly comparable with those of the earlier predictions. Despite the longer simulation period, PET was smaller than in the 2nd prediction. The reduction in PET resulted in a lowered AET. The reduction of PET was nearly equal to that of AET. Since the storage changes were in average also smaller than in the 2nd prediction, the resulting discharge was larger.

The results for the additional simulation period (2008/2009) produced similar results as obtained for the preceding years. AET was in average in the order of 300 mm yr^{-1} and discharge 90 mm yr^{-1}. Only PET and the storage changes showed considerably deviations from previous estimates (PET ~ 500 mm yr^{-1} and a negative $\Delta S \sim -50$ mm yr^{-1}).

HESSD
10, 8875–8944, 2013

PET simulated for 2006/2007 did not change from the 2nd to the 3rd prediction in case of NetThales, SIMULAT, and Topmodel because they did not make use of the additional weather data. MIKE SHE reduced PET slightly (−13 mm yr^{-1}) although a denser vegetation was assumed (Table 5). Only the SWAT and WaSiM-ETH (Richards) user requested the data of weather station II. The results of the PET changes were opposite: SWAT predicted about 30 mm yr^{-1} more PET (847 mm yr^{-1}) whereas WaSiM-ETH (Richards) reduced PET by about 30 to 682 mm yr^{-1}. These minor changes were due to the use of the additional weather data provided by the 2nd weather station.

The changes in AET corresponded to the changes in PET. The models, which did not use new weather data, calculated AET values, which differed by < 10 mm yr^{-1} from AET in the 2nd prediction. Only the Topmodel calculated considerably smaller AET (420 mm yr^{-1}, previously 465 mm yr^{-1}). SWAT predicted the largest changes in AET. Although PET only increased by 33 mm yr^{-1}, AET increased by 118 mm yr^{-1} during the 2nd hydrological year. The large AET changes predicted by SWAT are probably due to the parameter changes of the vegetation. AET calculated by MIKE SHE decreased by about 50 mm yr^{-1} despite a larger PET.

The largest changes in the water budget are those of discharge and storage since all modellers made use of additional soil property data: MIKE SHE (+53 mm yr^{-1}), NetThales (+88 mm yr^{-1}), SIMULAT (+21 mm yr^{-1}), and Topmodel (+48 mm yr^{-1}) predicted larger discharge, whereas SWAT (−101 mm yr^{-1}) and WaSiM-ETH (Richards) (−65 mm yr^{-1}) calculated less discharge, the latter showing an increase in AET. Catflow simulated 255 mm yr^{-1} discharge and therefore the 2nd largest discharge in 2006/2007 (SIMULAT 291 mm yr^{-1}). The changes in discharge were due to changes in K_{sat}, e.g. Catflow used a larger K_{sat} of 110 mm h^{-1} and NetThales increased K_{sat} to 100 mm h^{-1} (Table 5). Additionally, K_{sat} of the soil crust was a new input: the NetThales modeller implemented soil freezing and the soil crust into his model using a K_{sat} of 3 mm h^{-1} and thereby reducing the infiltration. Similarly, the MIKE SHE user reduced infiltration by changing K_{sat} of the soil crust.

HESSD
10, 8875–8944, 2013

The storage changes from the 2nd to 3rd prediction were less than from the 1st to 2nd prediction in case of WaSiM-ETH (Richards) (−171 mm yr^{-1}), NetThales (−49 mm yr^{-1}), and MIKE SHE (−6 mm yr^{-1}) while they were larger in case of SWAT (+21 mm yr^{-1}) and SIMULAT (+20 mm yr^{-1}). Topmodel and WaSiM-ETH (Topmodel) did not account for storage changes. The large storage changes of WaSiM-ETH (Richards) were mainly caused by correcting the mass balance error as described in the preceding chapter.

Figure 6 shows the discharge of the 3rd prediction for the hydrological year 2006/2007. The measured peak discharge on the 27 May 2007 was 897 m^3 d^{-1}. The range of predictions was large, from 106 m^3 d^{-1} (SIMULAT) to 1481 m^3 d^{-1} (NetThales) (Supplement B), but somewhat smaller than during the 2nd prediction (Supplement A; 24 m^3 d^{-1} (CMF) to 1433 m^3 d^{-1} (SIMULAT)). A similar behaviour is observed for other events during 2006/2007.

The discharge–frequency relationship of MIKE SHE, NetThales, Topmodel, and WaSiM-ETH (Richards) shown in Fig. 7 are quite similar for high discharge (> 40 m^3 d^{-1} at 90 % of all events). Only SIMULAT and SWAT predicted considerably smaller discharges and Catflow larger ones. Low flow was more frequent among models compared to those of the 2nd prediction: NetThales predicted for about 75 % of all events a discharge of less than 1 m^3 d^{-1}, whereas Catflow predicts at 5 % of all events a discharge of more than 20 m^3 d^{-1}. The discharge characteristic of the Topmodel predictions showed the slightest change.

Only the Topmodel and SWAT had similar discharge components comparing the 3rd and 2nd prediction, producing about 50 % of surface runoff and base flow (Table 9). Catflow shows the lowest runoff of all models although the surface runoff gets larger from less than 1 % in the 1st year to 25 % in the 3rd year. Also SIMULAT generated little surface runoff (22 %) and mainly interflow and therefore the opposite of what it simulated in the 2nd prediction. The discharge components of WaSiM-ETH (Richards) changed significantly over the period of all three years: in the 1st year surface runoff, interflow, and base flow were similar. Surface runoff dominated in the 2nd year (∼ 45 %)

HESSD
10, 8875–8944, 2013

with base flow the lowest (~ 20 %) and interflow dominated in the 3rd year (~ 45). In the 3rd year the surface runoff component was the lowest (~ 25 %). MIKE SHE simulated the largest surface runoff off all models, having runoff of about 87 %. The accumulated interflow and base flow contributed about 13 %. NetThales could not distinguish the discharge components.

Only Catflow and MIKE SHE presented groundwater predictions, which were as in the 2nd prediction and close to the actual measurements (Supplement D). On the other hand, Catflow started with too much groundwater (water table 1 m below ground instead of 2.5 m) and reaching complete water saturation of the whole aquifer early in 2008 (Supplement D).

6 Discussion

The successive predictions changed mainly due to modified process descriptions (1st to 2nd prediction) and due to the availability of additional data (2nd to 3rd prediction), which affected the parameterisation.

6.1 Impact of changing process assumptions and descriptions

During the 1st prediction half of the modellers checked the effect of assuming different dominant processes, but it was difficult to identify them. In this phase of the 1st prediction runs the modeller's experience was crucial (Holländer et al., 2009). Defining the major controls was the modellers' main interest during the 1st workshop. The discussion about the role of soil crusts, initial conditions, and the role of the V-shaped subsurface clay dam (Sect. 3.1) and the field visit after the 1st prediction resulted in more consistent predictions of the water budgets predicted in the 2nd prediction (Fig. 2). The standard deviation of the simulated AET decreased from 100, 137, 118 mm yr^{-1} to 66, 78, 56 mm yr^{-1} for the 1st, 2nd, and 3rd year, respectively. The mean of all items in the predicted water budgets changed in the same direction throughout all three years,

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

e.g. AET became larger and Q smaller. The exception was water storage with a larger variance of the simulated ΔS, which is due to two reasons: (i) several modellers (Catflow, SIMULAT, Topmodel, and WaSiM-ETH (Topmodel)) did not account for the initially dry soil conditions and (ii) such transient conditions with a rising water table have not been experienced by any of the modellers since all of them dealt with natural "mature" catchments so far (Table 2). Since the other modellers accounted for the dry initial conditions there was an overall trend to larger subsurface storage in the 2nd compared to the 1st prediction.

During the 1st workshop the soil crust was recognized as the most crucial property in the early phase of the developing catchment (Sect. 3.1). This concept superseded the views prevailing for the 1st prediction attempt when most modellers tried to reduce infiltration by modifying the van Genuchten parameters. Catflow implemented a soil crust with K_{sat} of 0.06 mm h^{-1} whereas WaSiM-ETH (Richards) used 20 mm h^{-1} (Table 5). Since WaSiM-ETH (Richards) used the largest K_{sat} for the soil crust, the predicted infiltration and the discharge were highest of all models (Table 8) causing a base flow of 51 to 67 % of the discharge. Other modellers used higher K_{sat} of the soil crust (e.g. Topmodel: 5 mm h^{-1}) and predicted a very fast subsurface flow (up to 74 %). It is obvious from the discharge–frequency relationship (Fig. 5) that the base flow has been drastically reduced in the 2nd prediction. Since the maximum discharges still reached a similar magnitude, the implementation of the soil crust mainly influenced the base flow.

Accounting for the clay wall increased the amount of water stored in the subsurface since the accumulated groundwater could not seep away fast enough. The base flow was overestimated by nearly all modellers in the 1st prediction, so that they felt pushed to reduce K_{sat}. The impact of introducing the clay wall cannot be determined. The only model having a low base flow component in the 1st prediction (CoupModel, 18 to 33 %) increased the base flow in the 2nd prediction. It was between 39 to 54 % of total discharge although the modeller chose a smaller K_{sat} for the soil compared to the

HESSD
10, 8875–8944, 2013

choice in the 1st prediction. Since the base flow component increased an impact of the clay wall does not seem to be the prime reason.

Although soil freezing and snow melt was discussed during the workshop as an important process only the NetThales and MIKE SHE users added it into their model. This enabled the NetThales user to account for the significant snowmelt event in January 2006 (Gerwin et al., 2011) quite effectively. CoupModel where these processes were already included in the 1st prediction predicted a minor discharge caused by this snow melt event on the 20 and 21 January 2006 of less than 1 % of the observed discharge. MIKE SHE did not predict any additional discharge during that event although the process was included in the model. All other models missed the consequences of this event.

Finally, most modellers adapted the parameterization of their model. Reducing K_{sat} obviously reduced the base flow (e.g. SIMULAT: from nearly 100 to 20 %) and raised the groundwater table (Supplement C). CMF, using the highest K_{sat} in the 1st prediction reduced K_{sat} by nearly one order of magnitude to a mean K_{sat} of 60 mm h^{-1}. Although this was still a rather high K_{sat} compared to the other models, it decreased the discharge by about one third.

Several modellers (Table 5) accounted for the larger vegetation by using a larger LAI. This had only a minor impact on PET. Only the MIKE SHE user calculated a significantly different PET, since he switched from using the Turc to the Penman–Monteith model. Similarly, modifications related to the gullies (CMF and SWAT) had a minor influence on the results.

Generally, the impact of the modeller's experience was much less pronounced in the 2nd prediction, because the discussions during the workshop and field visit concerning the dominant system controls harmonized the modellers' views. Therefore, the results are more similar with smaller standard deviations in the water budget components (Fig. 2) and the smaller span in the discharge–frequency relationship is smaller as well (Fig. 5). However, the differences among the model simulations remained considerably.

HESSD
10, 8875–8944, 2013

6.2 Impact of additional data

The process assumptions for the 3rd prediction remained largely the same as for the preceding prediction. Only the NetThales user added soil crusting to his model (Sect. 4.1). All parameter changes were based on the additional data.

The harmonised ensemble of modelled water budgets, as seen during the 2nd prediction, was continued for discharge and storage. In contrast, the range of PET and AET increased in the 3rd prediction (Fig. 2). The reduced variation of discharge and storage is clearly due to the soil parameterisation based on the data formally purchased by all modellers. In particular, the field data of K_{sat} were used, so that only the spatial regionalisation differed between the models.

Only the CoupModel, SIMULAT, and SWAT user opted for additional vegetation data (Table 7). While the plant parameterisation of SIMULAT was not changed, the SWAT user changed them. Therefore, AET did not change for SIMULAT while for SWAT AET in the 2nd year increased significantly from 410 mm yr^{-1} (2nd prediction) to 528 mm yr^{-1} (3rd prediction). However, since the PET did not significantly change for SWAT the changes of AET were not related to the vegetation data but probably to other changes like the introduction of re-infiltration, which increased the soil water storage. WaSiM-ETH (Richards) also used additional climate data, which lead to an opposite trend in PET compared to SWAT. All other models except MIKE SHE used the same plant parameterisation resulting in smaller AET, primarily due to the soil parameterisation. MIKE SHE increased the vegetation density resulting in a minor increase of AET (+14 mm yr^{-1} in the 2nd year, equal to 2 % of total AET).

Changes in simulated discharge into the lake are mostly in opposite direction compared to changes in AET. Therefore, since most models used information on soil data to adjust soil parameterisation, runoff generation and subsurface storage seem to be main drivers for changes in discharge since changes in subsurface storage are small in the 2nd year (Table 9).

Mike SHE predicted only small changes in catchment storage, but most models produced larger changes between the 2nd and the 3rd prediction. NetThales and WaSiM-ETH (Richards) significantly reduced the storage changes in all three year. While NetThales did so by 10 to 80 mm yr^{-1}, WaSiM-ETH (Richards) reduced the storage changes by 115 to 181 mm yr^{-1} between the 2nd and 3rd prediction. The changes of NetThales are related to the introduction of the soil crust in the 3rd prediction and additional changes in K_{sat}. This has already been seen with those models which introduced the soil crust in the 2nd prediction. The huge changes in ΔS of WaSiM-ETH (Richards) are due to updated soil hydraulic parameters. The user evaluated the simulated soil moisture dynamic using the additional data set. In the 1st year SIMULAT started the model run with an initially dry soil resulting in a ΔS of 28 mm yr^{-1}. In the 2nd year the changes were smaller but increased again in the 3rd year. These changes may be due to modified lower boundary conditions. Similarly, SWAT predicted about the same changes as SIMULAT, however these changes were due to the re-infiltration of surface water directly affecting ΔS (Sect. 4.4).

Depending on model philosophy and parameterisation strategy, the various models respond differently in terms of changes in runoff generation mechanisms (surface runoff, interflow, base flow; Sects. 5.2, 5.3, and Table 10). While for Topmodel runoff generation did not change from 2nd to 3rd prediction, SIMULAT and SWAT show a remarkable decrease in surface runoff generation. For SIMULAT this is due to the change in K_{sat} of the surface crust (infiltration tests, Table 5). For WaSiM-ETH (Richards) the contribution of base flow increased at the cost of interflow, which may be caused by the choice of K_{sat} of the soil columns. For all other models no comparison can be made because either the modeller could no longer participate in this study or because some of them did not provide information on the discharge components for the 2nd prediction.

The parameterisation, which changed the runoff components, also affected peak flows and the duration of zero-flow in some of the models. For example, the peak flows (Supplement A and B) simulated by SIMULAT decreased significantly due to changes of the surface crust K_{sat}, which diminished the contribution of surface runoff.

HESSD
10, 8875–8944, 2013

In contrast, SWAT generated higher peak flows in the 3rd than in the 2nd prediction due to a significant increase in interflow mainly at the cost of base flow. Similar changes due to parameterisation of the surface crust are seen for NetThales predictions where the hydrographs cannot be attributed to the various discharge components.

Those models showing increased fast discharge components due to the parameterisation showing zero- or low flow periods ($< 1\,m^3\,d^{-1}$) (CMF, MIKE SHE, NetThales, SIMULAT and SWAT) while Catflow, Topmodel, WaSiM-ETH (Richards), and WaSiM-ETH (Topmodel) still simulate continuous flow. Due to the increased surface crust K_{sat} the zero-flow periods predicted by SIMULAT were shortened.

Examining this process of iteratively "improving" predictions raises the question about the significance of the modeller's experience versus the modelling strategy per se. Since none of the modellers had any experience with artificial catchments and the particular climatic conditions of this area, there is no difference in this respect. Differences among the modellers were found in terms of experience with (i) conceptually different models, which did not matter in this case since because all modellers chose process/physically based models, (ii) the number of different models applied (Table 2), and (iii) the modelling experience in terms of their career phase or of the number of catchments modelled. An interesting detail in this data is that the most experienced modellers chose the simplest models, either in terms of dimensionality (CoupModel and SIMULAT), or in terms of physical process representation, (Topmodel) in order to represent the hydrological behaviour of the Chicken Creek catchment.

6.3 Relation between additional data chosen (cost) and the improvement of the model performance (benefit)

We used the root mean square error (RMSE) and the Nash–Sutcliffe index (NSE) (Nash and Sutcliffe, 1970) to compare the discharge predictions (Figs. 8 and 9). The prediction improvements for the 1st year were relatively poor throughout all three prediction stages as shown by the RSME (Fig. 8) and Nash–Sutcliffe index (Fig. 9) mainly because of the large error related to the intensive snowmelt event on the 20 and 21

HESSD
10, 8875–8944, 2013

January 2006 (Holländer et al., 2009; Gerwin et al., 2009b). Only NetThales predicted a significant amount of snowmelt in the 3rd prediction although the predicted discharge was consistently less than measured.

Excluding the snowmelt event results in smaller RMSE (Fig. 8). However, the NSE shows a different picture. Only discharge few predictions were rated better. Therefore, starting with quasi-dry or dry conditions as most modellers did in the 2nd prediction had no positive impact as rated by the NSE. RSME shows similar results. However, the impact on the water budget predictions was positive. Only SIMULAT, SWAT, and Topmodel reduced the RMSE from the 1st to the 2nd prediction (Fig. 8).

The RMSE shows an improvement of all the modelled results in the 2nd year (Fig. 8). This can also be seen in the NSE for most model but later predictions of some models got worse. The WaSiM-ETH (Richards) predictions improved strongly from the 1st to the 3rd prediction but the 2nd prediction of the 2nd year was the worst. The RMSE improved strongly for all models from the 1st to the 2nd prediction, which left little room for further improvements for the last prediction (Fig. 8). The NSE shows the best results for the 1st prediction in the 3rd year while the 3rd prediction is the worst.

Looking for a measure, which enables us to value the model predictions of each prediction stage against the preceding prediction, we calculated the relative change of the RMSE-coefficient $\alpha_{i,j}$ [–] for each hydrological year i between two predictions $j-1$ and j as defined by:

$$\alpha_{i,j} = \frac{\mathrm{RMSE}_{i,j-1} - \mathrm{RMSE}_{i,j}}{\mathrm{RMSE}_{i,j-1}}. \quad (1)$$

Therefore, the larger $\alpha_{i,j}$ the larger is the relative prediction improvement. Similarly, the relative change of the NSE $\beta_{i,j}$ [–] is defined by:

$$\beta_{i,j} = \frac{-(\mathrm{NSE}_{i,j-1} - \mathrm{NSE}_{i,j})}{1 - \mathrm{NSE}_{i,j-1}}. \quad (2)$$

HESSD
10, 8875–8944, 2013

Figure 10a and b shows the relative improvement from the 2nd to the 3rd prediction related to the costs of using additional data (Table 7) using the relative RMSE and NSE, respectively. The graphs show for all three years that improving predictions by investing more to obtain additional data becomes less efficient. Obviously, this trend is statistically weak since the number of modellers and prediction stages were small, a limitation, which cannot easily be overcome.

Most of the differences can be attributed to the additional soil data, some of them less costly in total such as K_{sat} (640 Euro), bulk densities (10 Euro), and infiltration rates (410 Euro) and some very expensive (soil moisture: 9300 Euro) (Table 7). Both types of data seem to be equally valuable for improving the model parameterisation and for an adequate description of the initial conditions.

The prediction improvements from the 1st to the 2nd prediction stage were larger than those of the following step. The costs for visiting the field site and exchanging ideas during the workshop were definitely lower than data costs, but they are too arbitrary for a similar comparison because they depend on the modeller's travel costs. These results suggest that the sequence of modelling steps could or should follow cost efficiency criteria. The large improvement of the predictions from 1st and the 2nd stage can be explained by a more detailed view on the particular features of the site, by collective learning about the dominant controls as discussed by Holländer et al. (2009), and the modeller's experience in grasping the important features and assimilating convincing arguments brought up by colleagues. In order to look at the potential role of the modellers' experience we use an index for rating their experience taking into account five attributes (Table 2):

1. number of different models they used before,
2. amount of modelling years of each modeller,
3. amount of different regions where the modellers were active,
4. number of years they worked with the model used in this comparison, and finally

5. closeness of contact of the modeller to the developer team of their model.

All attributes were rated on a scale from 1 to 3 where 1 is little and 3 is top experience. Only the attribute (5) could be rated with a zero in case of no or a very minor connection to the model developers. The overall experience is the sum of the five ratings (Table 11).

We compare the indexed modeller's experience against the relative change of the RMSE and the NSE between the 1st and 2nd prediction (Fig. 11a and b). The comparison against the relative change of the RMSE (Fig. 11a) shows that there is no significant impact of the modeller's experience on the improvement of the predictions. The relative change of the NSE (Fig. 11b) gives a diffusive picture: almost no impact in the 1st year, a strong positive impact in the 2nd year, and a strong negative impact in the last year. The more experienced modellers exchanged their ideas with all other modellers during the workshop so that all modellers agreed on the dominant processes and implemented those (Tables 3 and 4).

Although not all measures document an improvement of the prediction there is a clear improvement with respect to process implementation throughout the modelling steps. The models did improve in so far as they became more realistic, even though the NSE did not get better.

In Fig. 12 we compare the deviation of the ensemble mean of predicted daily discharge $Q_{\text{pred, mean}}$ from the actually observed discharge Q_{obs}. In this context we use the predicted mean of the daily mean $Q_{\text{pred, mean}}$ ($24.2\,\text{m}^3\,\text{d}^{-1}$) instead of the mean of the daily median $Q_{\text{pred, median}}$ ($14.3\,\text{m}^3\,\text{d}^{-1}$) because the former corresponds well with the observed mean of daily discharge Q_{obs} ($21.8\,\text{m}^3\,\text{d}^{-1}$). Hence the mean of the ensemble mean is the better predictor than the mean of the median. This is consistent with the conclusions about the mean being the best predictor drawn by Surowiecki (2004) based on many statistical exercises.

The discharge is systematically over-predicted for discharge rates below the average daily Q_{obs} (Fig. 12). The larger the discharge above this rate is, the more pronounced is the under-prediction by the ensemble mean $Q_{\text{pred, mean}}$ down to about two fifth of the

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

actual Q_{obs}. In the specific case of Chicken Creek in its initial development phase the models therefore tended, in the average, to over-estimate the non-event discharge and to massively under-predict the event-discharge. In Fig. 12 we excluded the extreme influence of the singular snowmelt event (January 2006) and the discharge in the first months after leaving the completed catchment surface to be shaped by nature.

7 Conclusions

Anticipating the hydrological response of a catchment to external forces is the ultimate goal of catchment hydrology. Here we show how expertise and added information affects the quality of such predictions. The accuracy of scenarios modelled under given or assumed initial and boundary conditions depends (i) on the availability of pertinent data, (ii) on the suitability of available models, which should include the major system controls, and (iii) on the modeller's expertise to choose and adapt a suitable model, which must rely on sufficient modelling experience as well as on a profound system understanding. All of the above requirements consume a fraction of the almost always limited resources. Therefore, it is essential to know more about the gain of prediction quality relative to the needed investment of time and resources.

Each catchment is unique. A predictive model must be tailored to the case-specific features. The man-made Chicken Creek catchment challenged the modellers because of the unusual initial conditions (dry soil material) and the dynamic transition from the state of completion ($t = 0$) to that of converging toward a quasi-equilibrium (3rd year).

Table 12 summarizes the relative discharge prediction "success" for the first three years of this catchment. The 1st prediction was a difficult task because the modellers were confronted with three special features of the newly constructed catchment, (i) the initially dry soil, (ii) the impact of an unusually intensive snowmelt event, (iii) which enhanced the gully formation on the not yet stabilized bare surface. The gully network was, however, known to the modellers beforehand (aerial picture in the initial data set). The 1st predictions for the 2nd year were somewhat better. The real progress was

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

made with the 2nd prediction after the field visit and the discussions among the modellers during the 1st workshop. Adding additional data for the 3rd prediction improved those made for the 2nd but not those for the 3rd year and even decreased the predictive accuracy of several models in the last step. The modellers chose additional data based on two philosophies, using low investment data to optimize cost efficiency or to perfect parameter guessing by maximizing the data base. The former had a better effect for the model performance than using expensive data such as detailed information on vegetation, weather data and newer digital elevation models. The latter statement might be case-specific.

On a regional scale a priori predictions are quite realistic (e.g. Bormann et al., 1999) while on small spatial scales they often fail due to missing data (e.g. Holländer et al., 2009; Bormann et al., 2011). Despite predominantly similar process descriptions the variability of variables predicted by the various models was surprisingly large possibly because of the catchment's artificial features and the transient conditions in this early phase of its development (Bormann et al., 2011). The modellers did not systematically identify all possible controls of this system (Holländer et al., 2009).

The differences between the 1st and 2nd prediction were definitely larger compared to those between the 2nd and 3rd prediction. This underpins the value of soft information (field visit, workshop discussions, and experience). However, local measurements such as infiltration tests – provided for the 3rd prediction – certainly contributed to the improved predictions as well. They were apparently better suited to define the soil parameters than those estimated based on pedotransfer functions (Holländer et al., 2009; Bormann et al., 2011).

Most modellers struggled with estimating the initial soil moisture conditions since is not something that needs to be done in natural "mature" catchments. This corrupted the catchment's storage behaviour during the 1st years of development. Hard information on soil moisture was therefore essential to define the initial conditions (see also Bormann, 2011; Bormann et al., 2011).

For some models, the use of soft and hard data leads to contrasting effects. E.g., SIMULAT introduced a soil crust for the 2nd prediction based on literature and reduced the impact of crusting in the 3rd prediction based on infiltration measurements.

From this modelling exercise we conclude (i) that soft information such as the modeller's system understanding is as important as the model itself, (ii) that the sequence of different modelling steps impacts the relative improvement attributed to the different steps (e.g. field visit, expert discussion, choice of model, selection of available data, parameter estimation), and (iii) that additional process understanding gained during the modelling process can be as efficient as improving data availability for optimising parameters needed to satisfy model requirements.

8 Conjectures and implications

The number of model screws one can adjust for making such a prediction or for improving it is very large and it differs from model to model. Adjusting the model screws is where experience comes in. The modellers achieved similar effects by turning different screws. Knowing this dilemma in advance gives the modeller a nudge for a lead.

Being faced with the request for a real world prediction we have different options to start with: on-site inspection which is important as we showed in this study, getting a better handle on available data, using local knowledge of residents, or asking differently experienced colleagues to join the team for a 1st guessing phase, some of them being the real expert and others being recently educated, scientifically up-to-date and not blocked by their previous experiences. Such a team might be a better investment (and likely to come at smaller financial cost) than to demand additional, possibly useless data to satisfy parametric needs of the chosen model.

Hence the sequence of modelling steps for making a forecast – a real prediction and not a re-prediction – has to be carefully planned. There is no universal recipe for the "right" strategy. It depends on the case, which might be similar or differ from already encountered cases.

HESSD
10, 8875–8944, 2013

The most important lesson is what became routine in recent years in climate modelling. It is the ensemble of reasonable and well founded predictions, which yields the envelope of the possible outcomes. Such an ensemble is not solely a matter of how good the model is, but how well the steps of making a prediction are being sequenced and being based on solid knowledge.

Supplementary material related to this article is available online at: http://www.hydrol-earth-syst-sci-discuss.net/10/8875/2013/hessd-10-8875-2013-supplement.zip.

Acknowledgements. These investigations were supported by the German Research Foundation (DFG) and the Brandenburg Ministry of Science, Research and Culture (MWFK, Potsdam) in the framework of the Transregional Collaborative Research Centre 38 (SFB/TRR 38). Werner Gerwin gave tremendous help on supporting this work. The authors thank Vattenfall Europe Mining AG for providing the research site.

References

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: Large area hydrologic modelling and assessment part I: model development, J. Am. Water Resour. Assoc., 34, 73–89, 1998.

Beven, K. J., Lamb, R., Quinn, P., Romanowicz, R., and Freer, J. E.: Topmodel, in: Computer Models of Watershed Hydrology, Water Resources Publications, Colorado, 627–668, 1995.

Blöschl, G.: Rainfall–runoff modeling of ungauged catchments, in: Encyclopedia of Hydrological Sciences, edited by: Anderson, M. G., John Wiley & Sons, Ltd., Chichester, 2061–2080, 2006.

Bormann, H.: Sensitivity of a regionally applied soil vegetation atmosphere scheme to input data resolution and data classification, J. Hydrol., 351, 154–169, 2008.

Bormann, H.: Treating an artificial catchment as ungauged: increasing the plausibility of an uncalibrated, process-based SVAT scheme by using additional soft and hard data, Phys. Chem. Earth, 36, 615–629, doi:10.1016/j.pce.2011.04.006, 2011.

Bormann, H., Diekkrüger, B., and Renschler, C.: Regionalisation concept for hydrological modelling on different scales using a physically based model: results and evaluation, Phys. Chem. Earth B, 24, 799–804, doi:10.1016/s1464-1909(99)00083-0, 1999.
Bormann, H., Breuer, L., Gräff, T., and Huisman, J. A.: Analysing the effects of soil properties changes associated with land use changes on the simulated water balance: a comparison of three hydrological catchment models for scenario analysis, Ecol. Model., 209, 29–40, doi:10.1016/j.ecolmodel.2007.07.004, 2007.
Bormann, H., Holländer, H. M., Blume, T., Buytaert, W., Chirico, G. B., Exbrayat, J.-F., Gustafsson, D., Hölzel, H., Kraft, P., Krauße, T., Nazemi, A., Stamm, C., Stoll, S., Blöschl, G., and Flühler, H.: Comparative discharge prediction from a small artificial catchment without model calibration: representation of initial hydrological catchment development, Bodenkultur, 62, 23–29, 2011.
Breuer, L., Eckhardt, K., and Frede, H.-G.: Plant parameter values for models in temperate climates, Ecol. Model., 169, 237–293, doi:10.1016/s0304-3800(03)00274-6, 2003.
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 Resour., 32, 129–146, doi:10.1016/j.advwatres.2008.10.003, 2009.
Buytaert, W. and Beven, K.: Models as multiple working hypotheses: hydrological simulation of tropical alpine wetlands, Hydrol. Process., 25, 1784–1799, doi:10.1002/hyp.7936, 2011.
Chirico, G. B., Grayson, R. B., and Western, A. W.: On the computation of the quasi-dynamic wetness index with multiple-flow-direction algorithms, Water Resour. Res., 39, 1115, doi:10.1029/2002wr001754, 2003.
Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R.: A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils, Water Resour. Res., 20, 682–690, doi:10.1029/WR020i006p00682, 1984.
Diekkrüger, B. and Arning, M.: Simulation of water fluxes using different methods for estimating soil parameters, Ecol. Model., 81, 83–95, doi:10.1016/0304-3800(94)00162-B, 1995.
Elfert, S. and Bormann, H.: Simulated impact of past and possible future land use changes on the hydrological response of the Northern German lowland "Hunte" catchment, J. Hydrol., 383, 245–255, doi:10.1016/j.jhydrol.2009.12.040, 2010.

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

Exbrayat, J.-F., Viney, N. R., Seibert, J., Wrede, S., Frede, H.-G., and Breuer, L.: Ensemble modelling of nitrogen fluxes: data fusion for a Swedish meso-scale catchment, Hydrol. Earth Syst. Sci., 14, 2383–2397, doi:10.5194/hess-14-2383-2010, 2010.

Exbrayat, J.-F., Viney, N. R., Frede, H.-G., and Breuer, L.: Using multi-model averaging to improve the reliability of catchment scale nitrogen predictions, Geosci. Model Dev., 6, 117–125, doi:10.5194/gmd-6-117-2013, 2013.

Fischer, T., Veste, M., Schaaf, W., Dümig, A., Kögel-Knabner, I., Wiehe, W., Bens, O., and Hüttl, R. F.: Initial pedogenesis in a topsoil crust 3 years after construction of an artificial catchment in Brandenburg, NE Germany, Biogeochemistry, 101, 165–176, doi:10.1007/s10533-010-9464-z, 2010.

Gerwin, W., Raab, T., Biemelt, D., Bens, O., and Hüttl, R. F.: The artificial water catchment "Chicken Creek" as an observatory for critical zone processes and structures, Hydrol. Earth Syst. Sci. Discuss., 6, 1769–1795, doi:10.5194/hessd-6-1769-2009, 2009a.

Gerwin, W., Schaaf, W., Biemelt, D., Fischer, A., Winter, S., and Hüttl, R. F.: The artificial catchment "Chicken Creek" (Lusatia, Germany) – a landscape laboratory for interdisciplinary studies of initial ecosystem development, Ecol. Eng., 35, 1786–1796, doi:10.1016/j.ecoleng.2009.09.003, 2009b.

Gerwin, W., Schaaf, W., Biemelt, D., Winter, S., Fischer, A., Veste, M., and Hüttl, R. F.: Overview and first results of ecological monitoring at the artificial watershed Chicken Creek (Germany), Phys. Chem. Earth, 36, 61–73, doi:10.1016/j.pce.2010.11.003, 2011.

Goodrich, D. C.: Geometric simplification of a distributed rainfall-runoff model over a range of basin scales, Technical Reports No. HWR 91-010, Hydrology Department, University of Arizona, 361 pp., 1990.

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.

Hölzel, H. and Diekkrüger, B.: Predicting the impact of linear landscape elements on surface runoff, soil erosion, and sedimentation in the Wahnbach catchment, Germany, Hydrol. Process., 26, 1642–1654, doi:10.1002/hyp.8282, 2011.

Hölzel, H., Rössler, O., and Diekkrüger, B.: Grope in the dark – hydrological modelling of the artificial Chicken Creek catchment without validation possibilities, Phys. Chem. Earth, 36, 113–122, doi:10.1016/j.pce.2010.04.017, 2011.

Hölzel, H., Diekkrüger, B., Biemelt, D., and Gädeke, A.: Impact of dumped sediment structures on hydrological modelling in the artificial Chicken Creek catchment, Germany, J. Hydrol., 477, 189–202, doi:10.1016/j.jhydrol.2012.11.029, 2013.

Hooghoudt, S. B.: Bijdragen tot de kennis van enige natuurkundige grootheden van de ground, Versl. Landb. Onderz, 42, 449–541, 1940.

Jansson, P.-E. and Moon, D. S.: A coupled model of water, heat and mass transfer using object orientation to improve flexibility and functionality, Environ. Modell. Softw., 16, 37–46, doi:10.1016/s1364-8152(00)00062-1, 2001.

Kraft, P., Vaché, K. B., Breuer, L., and Frede, H.-G.: A solute and water flux library for catchment models, in: Proceedings of the iEMSs Fourth Biennial Meeting: International Congress on Environmental Modelling and Software, Barcelona, 2008.

LAWA: Leitlinien zur Durchführung dynamischer Kostenvergleichsrechnungen, 7th Edn., Deutsche Vereinigung für Wasserwirtschaft, A.u.A. e.V., Berlin, 186 pp., 2005.

Maurer, T.: Physikalisch begründetete, zeitkontinuierliche Modellierung des Wassertransports in kleinen ländlichenen Einzugsgebieten, Inst. für Hydrologie und Wasserwirtschaft (IHW), Universität Karlsruhe, Karlsruhe, 1997.

Mazur, K., Schoenheinz, D., Biemelt, D., Schaaf, W., and Grünewald, U.: Observation of Hydrol. Process. and structures in the artificial Chicken Creek catchment, Phys. Chem. Earth, 36, 74–86, doi:10.1016/j.pce.2010.10.001, 2011.

Morel-Seytoux, H. J. and Khanji, J.: Derivation of an equation of infiltration, Water Resour. Res., 10, 795–800, doi:10.1029/WR010i004p00795, 1974.

Naef, F.: Can we model the rainfall–runoff process today?/Peut-on actuellement mettre en modèle le processus pluie–écoulement?, Hydrolog. Sci. J., 26, 281–289, doi:10.1080/02626668109490887, 1981.

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

NAW: DIN 4220 Pedologic site assessment – designation, classification and deduction of soil parameters (normative and nominal scaling), Deutsches Institut für Normung (DIN) e.V., Berlin, 50 pp., 2008.

Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged catchments: a comparison of regionalization approaches based on 913 French catchments, Water Resour. Res., 44, W03413, doi:10.1029/2007wr006240, 2008.

HESSD

10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, Hydrol. Earth Syst. Sci., 9, 157–171, doi:10.5194/hess-9-157-2005, 2005.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., and DMIP Participants: Overall distributed model intercomparison project results, J. Hydrol., 298, 27–60, doi:10.1016/j.jhydrol.2004.03.031, 2004.

Saxton, K. E., Rawls, W. J., Romberger, J. S., and Papendick, R. I.: Estimating generalized soil-water characteristics from texture, Soil Sci. Soc. Am. J., 50, 1031–1036, doi:10.2136/sssaj1986.03615995005000040039x, 1986.

Schulla, J. and Jasper, K.: Model Description WaSiM, ETH Zurich, Zurich, 181, 2007.

Seibert, J.: Regionalisation of parameters for a conceptual rainfall-runoff model, Agr. Forest Meteorol., 98–99, 279–293, doi:10.1016/S0168-1923(99)00105-7, 1999.

Simunek, J., Angulo-Jaramillo, R., Schaap, M. G., Vandervaere, J.-P., and van Genuchten, M. T.: Using an inverse method to estimate the hydraulic properties of crusted soils from tension-disc infiltrometer data, Geoderma, 86, 61–81, doi:10.1016/S0016-7061(98)00035-4, 1998.

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

Surowiecki, J.: The Wisdom of Crowds, Anchor Books, New York, 2004.

Weiler, M. and McDonnell, J. J.: Virtual experiments: a new approach for improving process conceptualisation in hillslope hydrology, J. Hydrol., 285, 3–18, 2004.

Wösten, J. H. M., Pachepsky, Y. A., and Rawls, W. J.: Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics, J. Hydrol., 251, 123–150, doi:10.1016/s0022-1694(01)00464-4, 2001.

Wösten, J. H. M. and Nemes, A.: Pedotransfer functions for Europe, in: Developments in Soil Science, edited by: Pachepsky, Y. and Rawls, W. J., Elsevier, 431–435, 2004.

HESSD

10, 8875–8944, 2013

Table 1. Catchment models.

model	full name of acronym	institution	reference
Catflow (hillslope version)[1]		GFZ German Research Centre for Geosciences	Maurer (1997)
CMF	**C**atchment **M**odelling **F**ramework	University of Giessen	Kraft et al. (2008)
CoupModel	**Coupl**ed Heat and Mass Transfer **Model** for Soil–Plant–Atmosphere System	Royal Institute of Technology KTH Stockholm	Jansson and Moon (2001)
Hill-Vi/MIKE-SHE		University of Freiburg; now: ETH Zurich	Weiler and McDonnell (2004)
NetThales		University of Naples	Chirico et al. (2003)
SIMULAT[2]		University of Oldenburg; now: University of Siegen	Bormann (2008); Diekkrüger and Arning (1995)
SWAT	**S**oil and **W**ater **A**ssessment **T**ool	University of Giessen; now: University of New South Wales	Arnold et al. (1998)
Topmodel	**Top**ography-based **model**	University of Bristol: now: Imperial College London	Beven et al. (1995)
WaSiM-ETH (Richards)[2]	**Wa**ter Balance **Sim**ulation Model-ETH	University of Bonn; now: Vattenfall Europe Mining AG	Schulla and Jasper (2007)
WaSiM-ETH (Topmodel)[3]	**Wa**ter Balance **Sim**ulation Model-ETH	University of Technology Dresden	Schulla and Jasper (2007)

[1] The catchment was simplified as a single hillslope.
[2] This is not a catchment model in the proper sense, but was adapted to be used as such.
[3] Two modellers used the WaSiM-ETH model. Here, they are distinguished based on their internal structure of the soil model (physical based versus conceptual: WaSiM-ETH (Richards) and WaSiM-ETH (Topmodel).

HESSD
10, 8875–8944, 2013

Table 2. Prior modelling experience (SD = fully or partially) spatially distributed, L = lumped, PB = physically based, and C = conceptual).

model	modelling experience	other models	regions, climates	SD	L	PB	C
Catflow	Ph.D., Postdoc	2	Alps, Andes	×		×	
CMF	teaching	2		×		×	
CoupModel	12 yr	10	Scandinavia	×		×	
Hill-Vi/MIKE SHE	Diploma thesis	3	Western Germany	×		×	
NetThales	temperate climates, no snow		Temperate climates (Australia, Southern Europe)		×	×	
SIMULAT	10 catchments 1-D to quasi 3-D	4	Western Germany, Western Africa	×	×	×	×
SWAT	Master thesis	6	Scandinavia, Central Europe			×	×
Topmodel	Ph.D., Postdoc	> 3	Ethiopia, Andes	×	×	×	×
WaSiM-ETH (Richards)	Ph.D. thesis	3	Western Germany, Cuba	×		×	
WaSiM-ETH (Topmodel)	Ph.D. thesis	2	Germany	×		×	

Table 3. Process understanding gained and lessons learnt during the three steps of the project work.

model	1st workshop	field trip	manuscript preparation
Catflow	– extremely wide range of predictions – initial condition	– dominant processes (role of soil crusts and surface runoff) – uncertainties of field measured data due to monitoring set up	– large variations of soil hydraulic parameters obtained from various pedo transfer functions and their impact on the results
CMF	– initial condition	– gullies considered to be important	– reasons for different results with different models (assumptions and catchment perception
CoupModel	– initial condition – construction of catchment – different assumptions and model structure	– vegetation characteristics – uncertainties with interpretation of field measurements	– reflections regarding own modelling approach – impact of soil hydraulic properties
Hill–Vi/MIKE SHE	– initial condition – important processes (soil freezing, clay wall)	– dominant process (surface runoff, gullies) – catchment size and characteristics – structures and spatial distribution of the vegetation	– development of the catchment – information about other modeller's decision and results – wrong assumption on PET
NetThales	– initial condition – dominant processes (soil freezing, clay wall)	– dominant processes (soil crusts and surface runoff dominated by infiltration excess) – spatial variability of soil ⇒ impact on soil moisture and vegetation	
SIMULAT	– impact of clay wall on groundwater – dominant process (surface runoff)	– soil crusts – impact of clay wall – vegetation coverage [3]	– justifications of other modellers for their decisions – importance of modeller's decisions during parameterisation process
SWAT	– initial condition, lake volume – identical data are interpreted differently by modellers	– dominant processes (soil crusts and surface runoff)	– information about other models and their results
Topmodel	[2]	– clay wall – dominant process (surface runoff by infiltration excess) – vegetation coverage [3]	– information about other models and their results – good water balance ⇒ weak influence of model implementation
WaSiM–ETH (Richards)	– dominant processes (soil crusts and surface runoff) ⇒ K_{sat} most sensitive parameter – initial condition – model weakness (constant layer thickness, no clay wall, no lake)	– catchment size and characteristic and of structures (e.g. gullies) – spatial distribution of the vegetation, soil crust	– qualitative information (e.g. water budget)
WaSiM–ETH (Topmodel)	– physically–based model for the unsaturated zone not needed [1]	– dominant process (surface runoff) – rapid catchment, land use, and land form changes	[1]

[1] Participated only for the 2nd prediction.
[2] 1st workshop not attended.
[3] Field visit in Jun 2009.

HESSD
10, 8875–8944, 2013

Table 4. Major modifications in the conceptualization of catchment processes and components between the 1st, 2nd and 3rd prediction.

model	initial condition		soil crust		clay wall		other
prediction stage	2nd	3rd	2nd	3rd	2nd	3rd	
Catflow	×		×		×[2]		
CMF	×	[3]		[3]		[3]	discharge into gullies at unsaturated conditions
CoupModel	×	[3]		[3]		[3]	
Hill-Vi/MIKE SHE[1]	×		×		×		Penman–Monteith, snow melt
NetThales	×			×	×		soil freezing
SIMULAT		×	×		×	×	plant parameterisation
SWAT	×						re-infiltration
Topmodel			×		[4]		
WaSiM-ETH (Richards)	×		×		×		soil thickness, soil cluster, lake
WaSiM-ETH (Topmodel)	[5]	[3]	[5]	[3]	[5]	[3]	[5]

[1] Using another model.
[2] Clay wall as no flow boundary instead of clay.
[3] No prediction made in that prediction stage.
[4] Use of lumped model did not allow implementation.
[5] Only part of the 2nd prediction.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

Table 5. Major modifications in the parameterisation between the 1st to 2nd and 2nd to 3rd prediction.

model	initial condition 2nd	initial condition 3rd	K_{sat} of soil crust 2nd	K_{sat} of soil crust 3rd	other 2nd	other 3rd
Catflow			$0.06\,mm\,h^{-1}$	$0.06\,mm\,h^{-1}$	smaller K_{sat}	K_{sat} $110\,mm\,h^{-1}$
CMF	dry state (pF 2.5)	[2]		[2]	$K_{sat} 60 \pm 30\,mm\,h^{-1}$, LAI = 1	[2]
CoupModel	dry state	[2]		[2]	smaller K_{sat}	[2]
Hill-Vi/MIKE SHE[3]	dry state		smaller	K_{sat}^{1}		K_{sat}^{1}, larger vegetation
NetThales	dry state			$3\,mm\,h^{-1}$	smaller K_{sat}	K_{sat} $100\,mm\,h^{-1}$
SIMULAT		dry state	$2.1\,mm\,h^{-1}$	$11.6\,mm\,h^{-1}$	LAI > 1 (2008)	σK_{sat}, min. water level for lower boundary condition
SWAT	dry state					vegetation[1], soil carbon content
Topmodel			$5\,mm\,h^{-1}$		smaller K_{sat}, larger vegetation, clay wall (Te $0.135\,m^2\,h^{-1}$)	K_{sat}
WaSiM-ETH (Richards)	quasi dry state		$20\,mm\,h^{-1}$		smaller K_{sat}	
WaSiM-ETH (Topmodel)	wet conditions	[2]		[2]		[2]

[1] Properties derived directly from data set (Gerwin et al., 2011; Mazur et al., 2011).
[2] Not part of the modelling group.

HESSD
10, 8875–8944, 2013

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

Table 6. Virtual costs of data (in Euro) provided in the 3rd prediction.

measured property	amount of observation locations	available since	costs [Euro yr^{-1}]	total costs [Euro]
soil hydraulic conductivity (field K_{sat})	15	[4]		590
soil hydraulic conductivity (laboratory K_{sat})	2	[4]		50
porosity, bulk density	2	[4]		10
water retention curves	2	[4]		510
carbon content	129	[1]		660
infiltration rates	10	[2,4]		410
soil moisture (TDR)	4	2008	6200	9300
weather station II	1	2008	4200	6300
Digital elevation model (DEM)	4			770
vegetation	120	2006	5210[4]	15 630[4]

[1] Data taken in 2005.
[2] Data taken in 2006.
[3] Data taken in 2009.
[4] Costs are average values.

Table 7. Data chosen by the modeller for the 3rd prediction.

model	K_{sat} (field)	K_{sat} (lab.)	porosity	water retention	carbon content	infiltration rates	soil moisture	weather station II	DEM	vegetation	total costs [Euro]
Catflow	×	×	×	×		×					1570
CoupModel	×			×	×		×[2]			×	26 690
Hill-Vi/MIKE SHE	×	×	×	×							1160
NetThales	×	×	×	×	×	×	×				11 530
SIMULAT	×	×	×	×[1]	×[1]	×	×[2]			×	27 930
SWAT	×	×	×		×		×	×		×	32 540
Topmodel	×	×	×				×[2]				9950
WaSiM-ETH (Richards)	×	×		×			×	×			16 700
WaSiM-ETH (Topmodel)	×	×	×				×				9950

[1] Review of data without using them later in the modelling process.
[2] Soil moisture data were used for model calibration.

HESSD
10, 8875–8944, 2013

Table 8. Water budget components of the 2nd year (2nd prediction) (NA: not available).

model	P (mm yr^{-1})	PET (mm yr^{-1})	AET (mm yr^{-1})	discharge (mm yr^{-1})	storage (mm yr^{-1})	balance (mm yr^{-1})
Catflow	*	*	*	*	*	*
CMF	565	433	334	109	108	14
CoupModel	635	NA	417	179	39	−0
MIKE SHE	565	621	452	103	20	−10
NetThales	566	NA	374	104	88	0
SIMULAT	565	688	266	269	17	13
SWAT	565	815	410	148	23	−16
Topmodel	565	1021	465	99	NA	1
WaSiM-ETH (Richards)	565	705	280	327	42	−84
WaSiM-ETH (Topmodel)	716	801	459	221	NA	36

* Catflow did not predict a complete water budget due numerical problems (see Sect. 4.4).

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

Table 9. Water budget components of the 2nd year (3rd prediction) (NA: not available).

model	P (mm yr^{-1})	PET (mm yr^{-1})	AET (mm yr^{-1})	discharge (mm yr^{-1})	storage (mm yr^{-1})	balance (mm yr^{-1})
Catflow	565	NA	197	255	75	38
CMF	*	*	*	*	*	*
CoupModel	*	*	*	*	*	*
MIKE SHE	565	635	403	155	9	−2
NetThales	565	NA	262	192	14	−1
SIMULAT	565	688	268	291	−5	11
SWAT	565	847	528	47	7	−17
Topmodel	565	1021	420	146	NA	−1
WaSiM-ETH (Richards)	565	682	248	244	72	1
WaSiM-ETH (Topmodel)	*	*	*	*	*	*

* CMF, CoupModel, and WaSiM-ETH (Topmodel) did not take part of the 3rd prediction.

HESSD
10, 8875–8944, 2013

Table 10. Model specific runoff components of the 2nd year (3rd prediction) (all numbers in % of total runoff).

	2nd prediction			3rd prediction		
	surface runoff	interflow	base flow	surface runoff	interflow	base flow
Catflow				11	89	
CMF						
CoupModel	54		46			
MIKE SHE	55		45	87	13	
NetThales						
SIMULAT	79	~ 0	21	22	4	74
SWAT	39	61		40	60	
Topmodel	39	61		40	60	
WaSiM-ETH (Richards)	5	56	39	36	38	26
WaSiM-ETH (Topmodel)	21	19	60			

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

Table 11. Indexed experience of the modeller.

model	amount of models[1]	different regions[2]	modelling years[3]	years with the used model[4]	model development[5]	total (max. 15)
Catflow	1	2	2	3	1	9
CMF	1	1	2	2	3	9
CoupModel	3	1	3	3	2	12
MIKE SHE	2	1	1	1	2	7
NetThales	1	2	3	3	2	11
SIMULAT	2	2	3	3	1	11
SWAT	3	2	2	2	0	9
Topmodel	3	2	2	2	2	11
WaSiM-ETH (Richards)	2	2	1	2	0	7
WaSiM-ETH (Topmodel)	1	1	1	1	0	4

[1] Rating of amount of models: 1: < 3 models, 2: 3–5 models, 3: > 5 models.
[2] Rating of different regions: 1: 1 region, 2: 2–5 regions, 3: > 5 regions.
[3] Rating of modelling years: 1: during Ph.D., 2: < 5 yr, 3: > 5 yr.
[4] Rating of years with the used model: 1: < 2 yr, 2: 2–3 yr, > 3 yr.
[5] Rating of model development: 0: no or email contact, 1: developers within the range of the modeller, 2: being a developer, 3: being the main developer.

HESSD
10, 8875–8944, 2013

Table 12. Rating of prediction progress compared to measurements in the course of consecutive model improvements (0 = poor to 3 = good).

Prediction stage	Development stage of the catchment 1st year	 2nd year	 3rd year
	highly dynamic surface processes, gully formation and snowmelt event	increasing role of feedbacks, crust formation, emerging plant cover	partly stabilized surface, approaching quasi-steady state
1st prediction	0	1	1
2nd prediction	0	2	3
3rd prediction	0	2	2

HESSD
10, 8875–8944, 2013

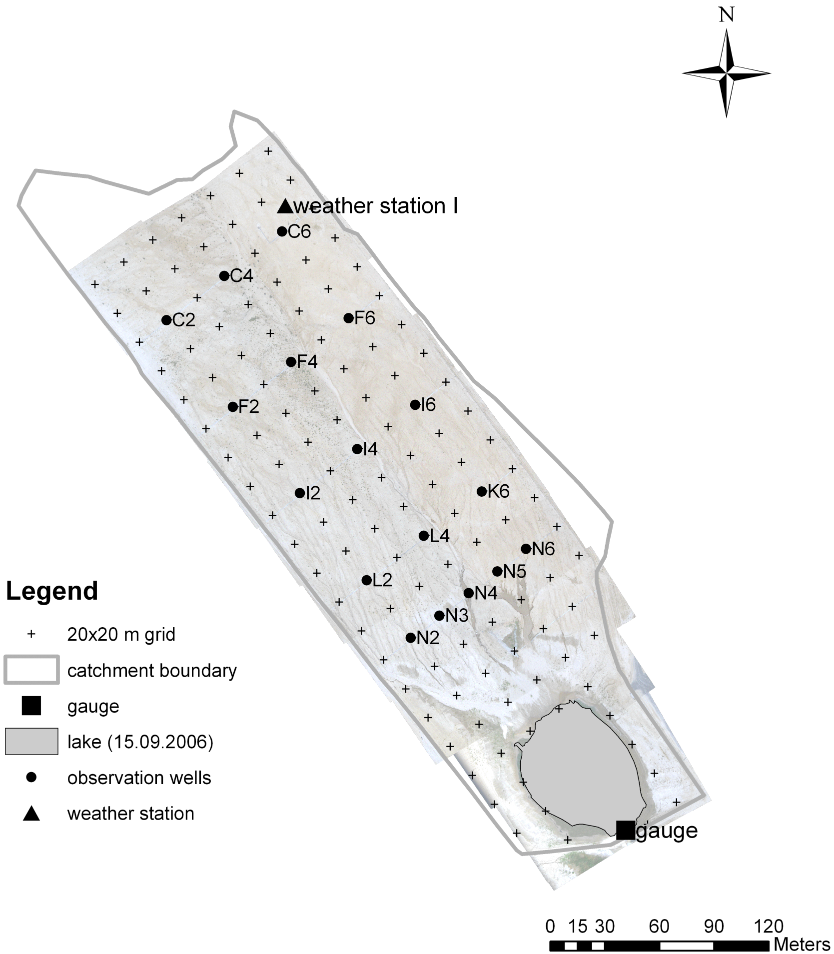

Fig. 1. GIS framework of the Chicken Creek catchment.

HESSD
10, 8875–8944, 2013

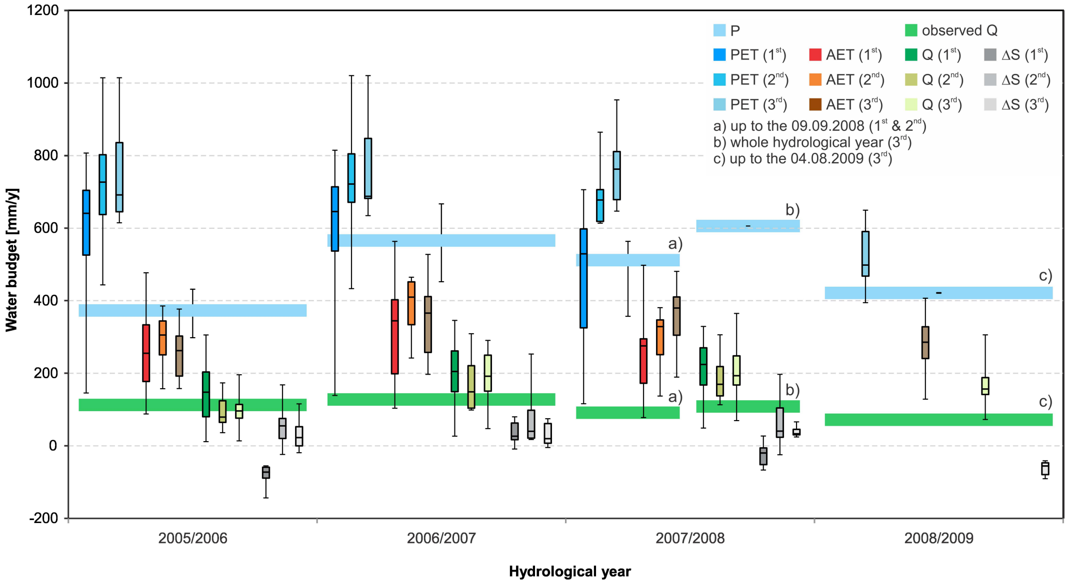

Fig. 2. Mean and standard deviation of the water budget components simulated in the 1st, 2nd and 3rd prediction.

HESSD
10, 8875–8944, 2013

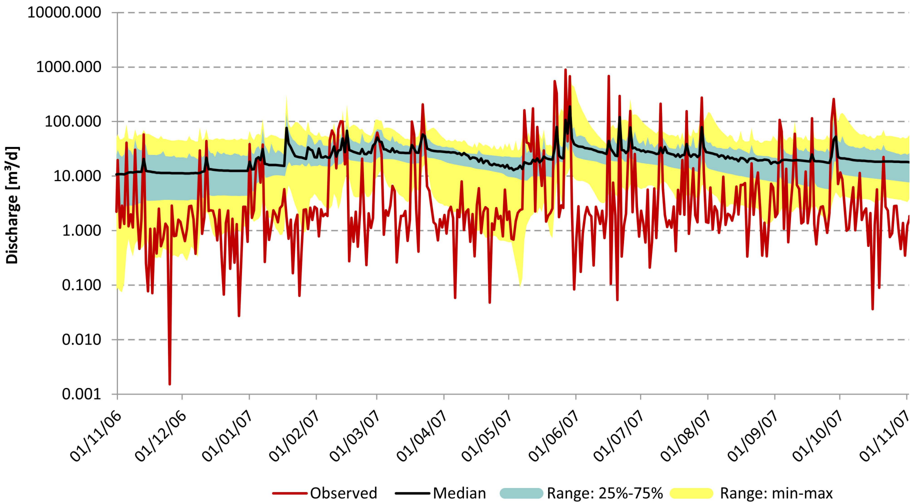

Fig. 3. Discharge predicted for the hydrological year 2006/2007 (1st prediction).

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

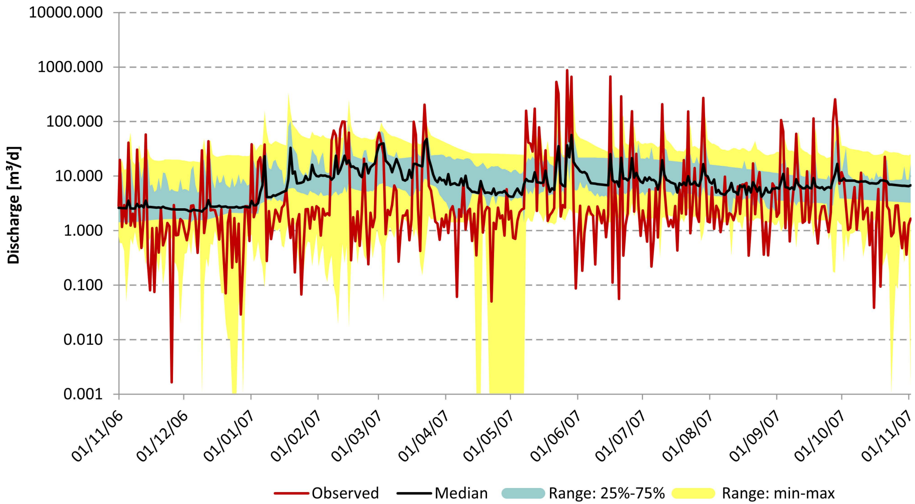

Fig. 4. Discharge predicted for the hydrological year 2006/2007 (2nd prediction).

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

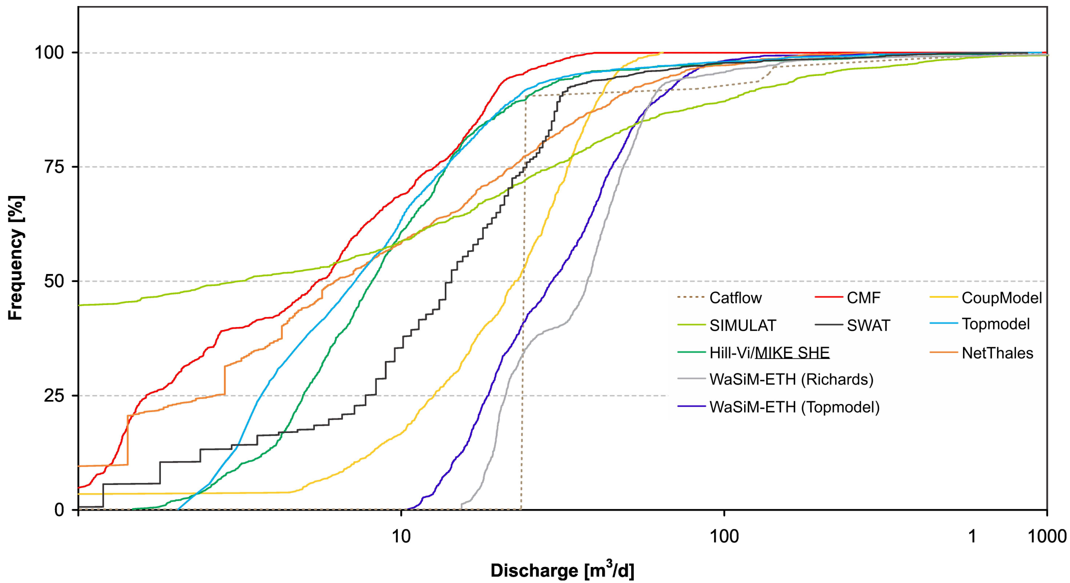

Fig. 5. Discharge–frequency relationship of the ten remaining predictions for the 2nd prediction.

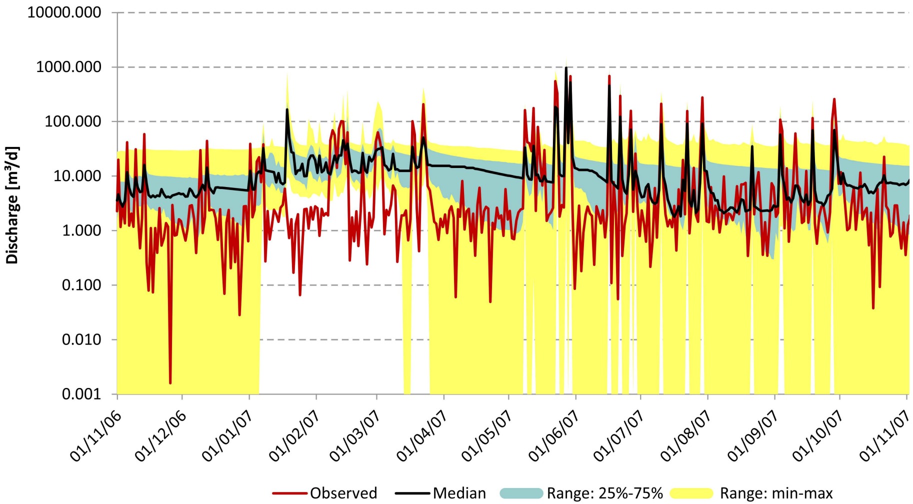

Fig. 6. Discharge predicted for the hydrological year 2006/2007 (3rd prediction).

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

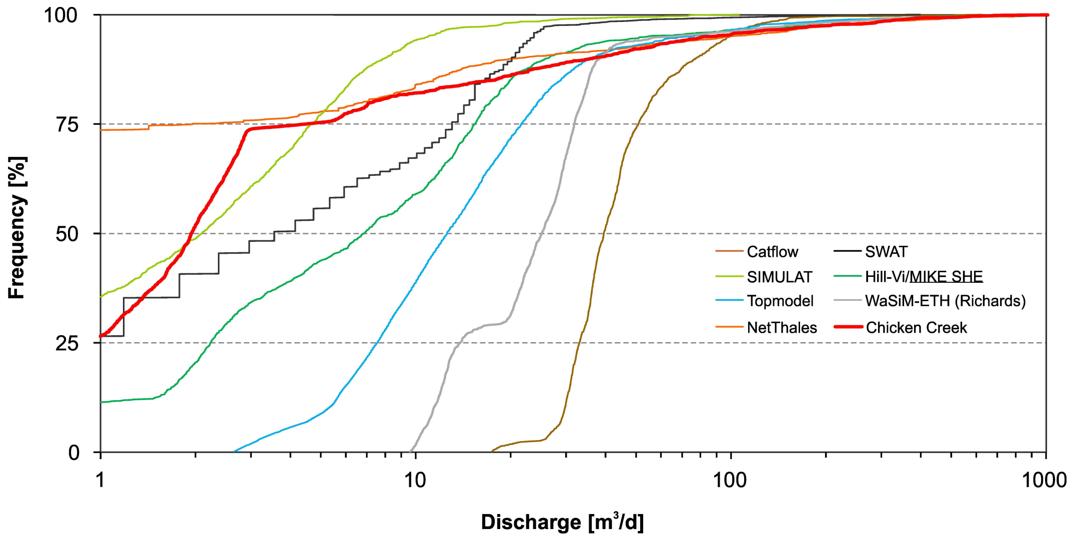

Fig. 7. Discharge–frequency relationship of observed and simulated daily discharge (3rd prediction 2005–2008).

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

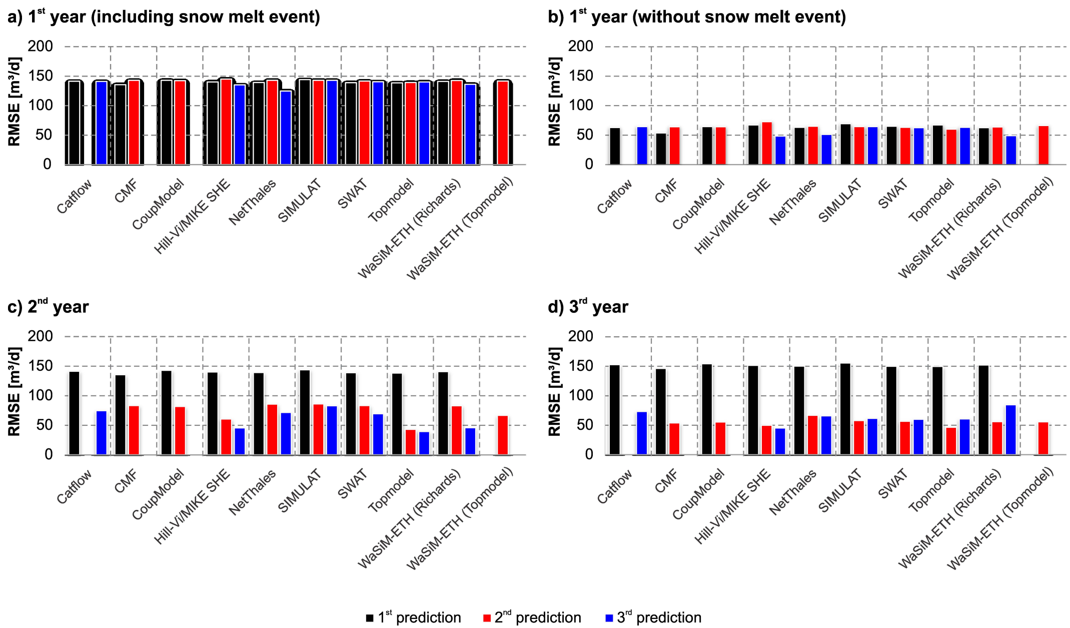

Fig. 8. RMSE of each simulated discharge prediction against the observed discharge for **(a)** the 1st year including the snow melt event, **(b)** the 1st year without the snow melt event, **(c)** the 2nd year, and **(d)** the 3rd year.

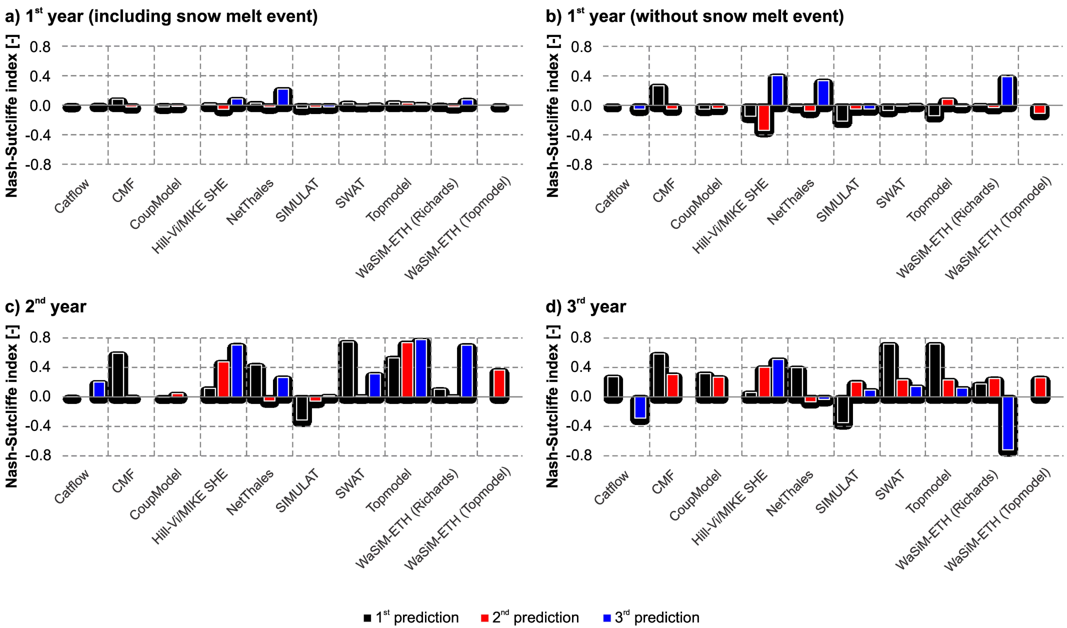

Fig. 9. Nash–Sutcliffe index of each simulated discharge prediction against the observed discharge for **(a)** the 1st year including the snow melt event, **(b)** the 1st year without the snow melt event, **(c)** the 2nd year, and **(d)** the 3rd year.

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

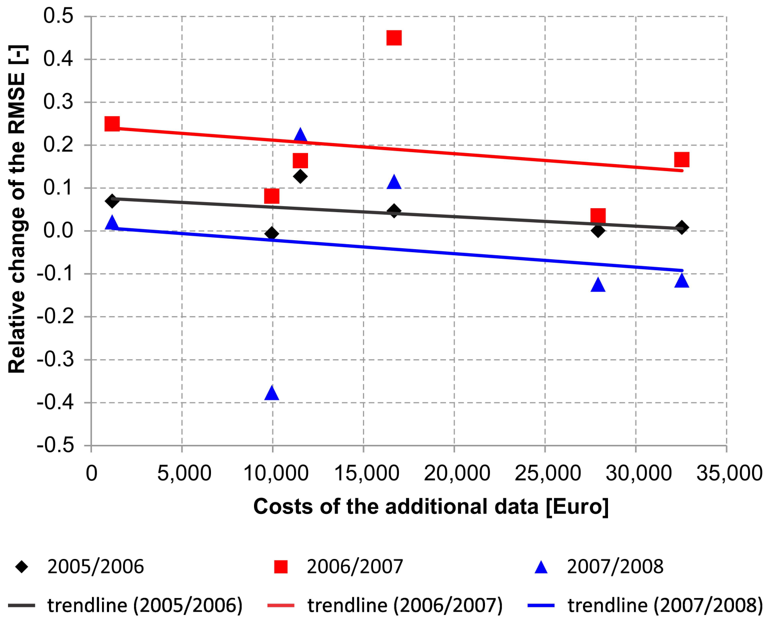

Fig. 10a. Relationship between the costs of the additional data and the relative change of the RMSE (comparison 2nd and 3rd prediction).

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

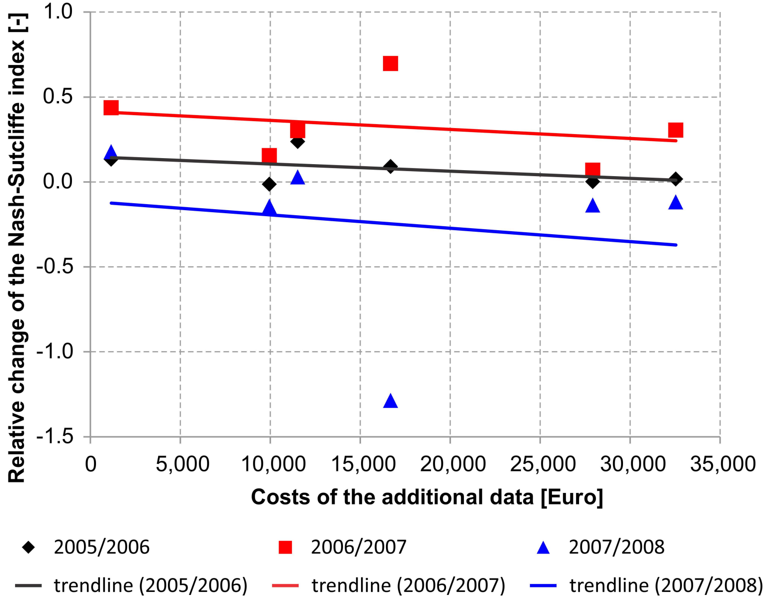

Fig. 10b. Relationship between the costs of the additional data and the relative change of the Nash–Sutcliffe index (comparison 2nd and 3rd prediction).

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

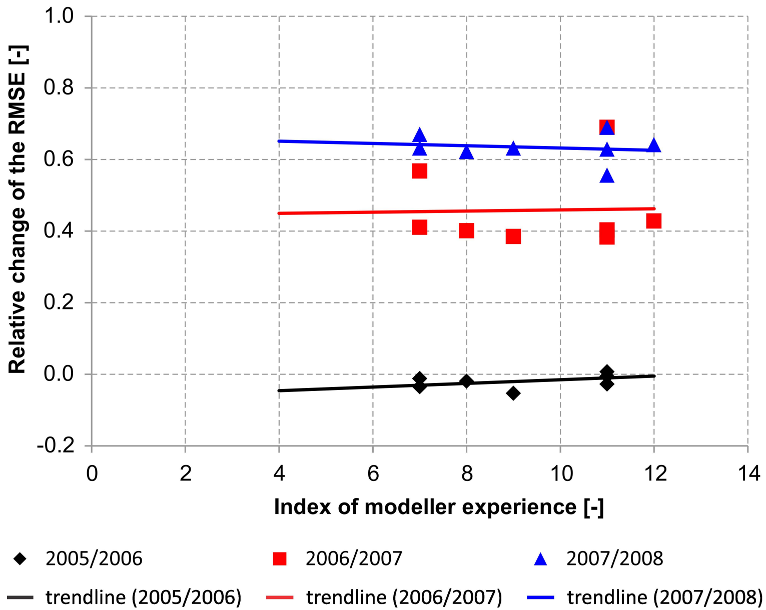

Fig. 11a. Relationship between the indexed modeller experience and the relative change of the RMSE (comparison 1st and 2nd prediction).

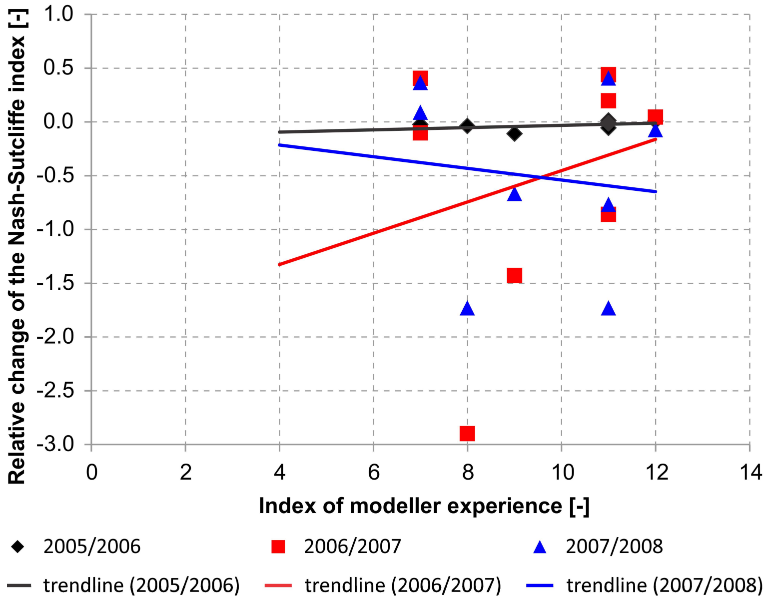

Fig. 11b. Relationship between the indexed modeller experience and the relative change of the Nash–Sutcliffe index (comparison 1st and 2nd prediction).

HESSD
10, 8875–8944, 2013

Impact of modellers' decisions on hydrological a priori predictions

H. M. Holländer et al.

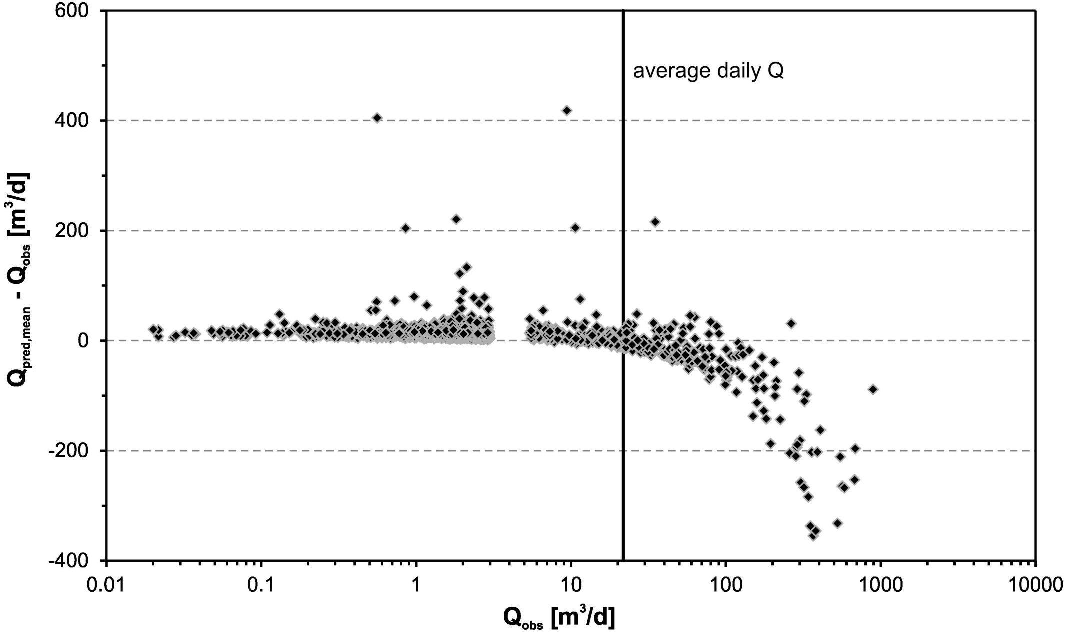

Fig. 12. Deviation of the predicted ensemble means $Q_{\text{pred, mean}}$ of daily discharge from the observed discharge Q_{obs}. To plot the log Q_{obs} we added $5 \times 10^{-5}\,\text{m}^3\,\text{d}^{-1}$ to the average daily discharge.