



Originally published as:

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., Flüher, H. (2009): Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data. - Hydrology and Earth System Sciences, 13, 11, 2069-2094

<http://www.hydrol-earth-syst-sci.net/13/2069/2009/hess-13-2069-2009.pdf>

Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data

H. M. Holländer¹, T. Blume², H. Bormann³, W. Buytaert^{4,*}, G.B. Chirico⁵, J.-F. Exbrayat⁶, D. Gustafsson⁷, H. Hölzel⁸, P. Kraft⁶, C. Stamm⁹, S. Stoll¹⁰, G. Blöschl¹¹, and H. Flühler¹²

¹Chair of Hydrology and Water Resources Management, Brandenburg University of Technology Cottbus, 03046 Cottbus, Germany

²Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Telegrafenberg, C4 2.25, 14473 Potsdam, Germany

³Department of Biology and Environmental Sciences, Carl von Ossietzky University of Oldenburg, 26129 Oldenburg, Germany

⁴School of Geographical Sciences, University of Bristol, BS8 1SS, UK

⁵Dipartimento di ingegneria agraria e agronomia del territorio, Università di Napoli Federico II, 80055 Naples, Italy

⁶Institute for Landscape Ecology and Resources Management, University of Giessen, 35392 Giessen, Germany

⁷Department of Land and Water Resources Engineering, Royal Institute of Technology KTH, 10044 Stockholm, Sweden

⁸Department of Geography, University of Bonn, 53113 Bonn, Germany

⁹Department Environmental Chemistry, Eawag, 8600 Dübendorf, Switzerland

¹⁰Institute of Environmental Engineering, ETH Zurich 8093 Zürich, Switzerland

¹¹Institute of Hydraulic Engineering and Water Resources Management, TU Vienna, 1040 Vienna, Austria

¹²Department of Environmental Sciences, ETH Zurich, 8092 Zürich, Switzerland

* now at: Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, UK

Received: 1 April 2009 – Published in Hydrol. Earth Syst. Sci. Discuss.: 15 April 2009

Revised: 25 September 2009 – Accepted: 30 September 2009 – Published: 4 November 2009

Abstract. Ten conceptually different models in predicting discharge from the artificial Chicken Creek catchment in North-East Germany were used for this study. Soil texture and topography data were given to the modellers, but discharge data was withheld. We compare the predictions with the measurements from the 6 ha catchment and discuss the conceptualization and parameterization of the models. The predictions vary in a wide range, e.g. with the predicted actual evapotranspiration ranging from 88 to 579 mm/y and the discharge from 19 to 346 mm/y. The predicted components of the hydrological cycle deviated systematically from the observations, which were not known to the modellers. Discharge was mainly predicted as subsurface discharge with little direct runoff. In reality, surface runoff was a major flow component despite the fairly coarse soil texture. The actual evapotranspiration (AET) and the ratio between actual and potential ET was systematically overestimated by nine of the

ten models. None of the model simulations came even close to the observed water balance for the entire 3-year study period. The comparison indicates that the personal judgement of the modellers was a major source of the differences between the model results. The most important parameters to be presumed were the soil parameters and the initial soil-water content while plant parameterization had, in this particular case of sparse vegetation, only a minor influence on the results.

1 Rationale and scientific concept

Hydrological catchment modelling is a tool for testing the assumptions and the conceptualization of the dominant system properties. It advances our process understanding of discharge formation. Often, the discharge record is known to the modeller when setting up the model, but in the case of ungauged catchments, this is not the case. The PUB research initiative (*Predictions in Ungauged Basins*) addresses



Correspondence to: H. M. Holländer
(hartmut.hollaender@tu-cottbus.de)

the problem of a priori predicting an unknown system response (Sivapalan et al., 2003). Such endeavours are typical for real world applications when the dominant processes are unknown and the data are too sparse to meet the model requirements. An important question is how to improve the predictive model performance by acquiring additional information on process understanding and catchment characteristics and/or by reducing the parametric requirements.

In this study, we make use of data obtained in an artificial catchment for a comparative prediction of discharge. Artificial catchments are per se the opposite of ungauged catchments because they are supposed to provide a well documented case (e.g. a clear definition of catchment geometry and boundary conditions). We use conceptually different models to predict the discharge – yet unknown to the modellers – based on minimum information. The purpose of this collective exercise is neither a rating of model suitability nor success, but the question about the crucial elements of discharge modelling for an “a priori prediction” of the catchment response. This prediction exercise is the first of three steps. In a second step, more detailed information on the catchment characteristics will be provided to the modellers. In a third step, the entire database including the discharge records will be made available to the modellers, which enables them to calibrate the model. The process of stepwise in satisfying the model needs will allow us to relate the gain of predictive performance to the efforts and costs of providing the information needed for the model parameterization. This paper documents the first step of the exercise and focuses on the comparison of the underlying model assumptions and the role of the modeller’s experience.

2 Artificial catchments and predictions in ungauged basins

Artificial catchments are an approximation to hydrological systems in their initial phase, because of the short time span since construction. Hydrological processes have been studied in artificial catchments, e.g. in China (Gu and Freer, 1995), Canada (Barbour et al., 2001), Spain (Nicolau, 2002), and Germany (Gerwin et al., 2009). The main objective of most of these studies was to determine the water and element budgets of catchments under well-defined boundary conditions to identify the flow paths through and the storage behaviour of the various catchment compartments by characterizing the processes of runoff formation (Hansen et al., 1997; Kendall et al., 2001). There is a general agreement that a good correspondence of observed and calculated discharge at a catchment outlet is a weak and insufficient criterion for the validity of a hydrological model (Grayson and Blöschl, 2000a). Additional knowledge on internal variables is required for calibration (e.g. Beven, 1989). Both local boundary conditions (e.g. catchment surface and subsurface size) and internal structures (e.g. discharge points and strat-

ification) can be controlled and more precisely documented in artificially constructed systems. Detailed observations of discharge, soil-water status and groundwater dynamics, both in terms of quantity and quality, allow for verifying the hypotheses about the causes of the system’s multi-responses provided the catchment properties do not change too rapidly during the very initial phase of catchment formation. Such data sets reduce the uncertainties by using part of them for an “a posteriori” calibration. In our case, we will use the artificial catchment data set only after having predicted the system response based on information that is usually available in catchments at the regional scale.

The “a priori” attempt – when target variables such as discharge are yet unknown – is an important step in any model application if the system, including its boundary conditions, changes or if a calibrated model is used for analogous but ungauged catchment. This can only work if the dominant and system-relevant processes are known and can be adequately described. Here, we use the artificial catchment “Chicken Creek” in Lusatia, Germany (Gerwin et al., 2009, this issue) to test the “a priori” attempt of discharge prediction.

Predicting state variables within and fluxes between compartments, as well as across catchment boundaries, is often hampered due to the considerable uncertainties which may be due to catchment heterogeneity and poorly defined boundary and initial conditions. The PUB initiative aims to develop and improve methods for such cases. Sivapalan et al. (2003) propose several approaches to addressing this problem either by conceptually simplifying process-based models and/or by using more comprehensive data including proxy data. Pretending that the Chicken Creek catchment is a data-poor, ungauged catchment allows us to investigate the dependence of the predictive performance on the amount of data available to the modellers.

3 Experiment and models

3.1 Chicken Creek catchment

The Chicken Creek catchment (Fig. 1) is 6 ha in size and currently the largest artificial catchment worldwide. It was built in 2005 by Vattenfall Europe Mining in scientific cooperation with the Brandenburg University of Technology (Gerwin et al., 2009). It is located in an open mining pit area in Lusatia, Germany. The catchment bottom is a 2 m thick tertiary clay layer placed on top of the reclaimed mining land. The clay layer forms a 450 m long and 150 m wide catchment, which drains into a depression at the bottom outlet. This depression is now a small lake which collects the outflow from the catchment. The longitudinal slope is 1 to 5% and 0.5 to 2% in transverse direction (Fig. 2a and b). A 2 to 3 m thick sand layer has been put onto the clay basement. It consists mainly of quaternary sand with variable fractions of 2 to 25% silt and 2 to 16% of clay. The slope of the surface is roughly

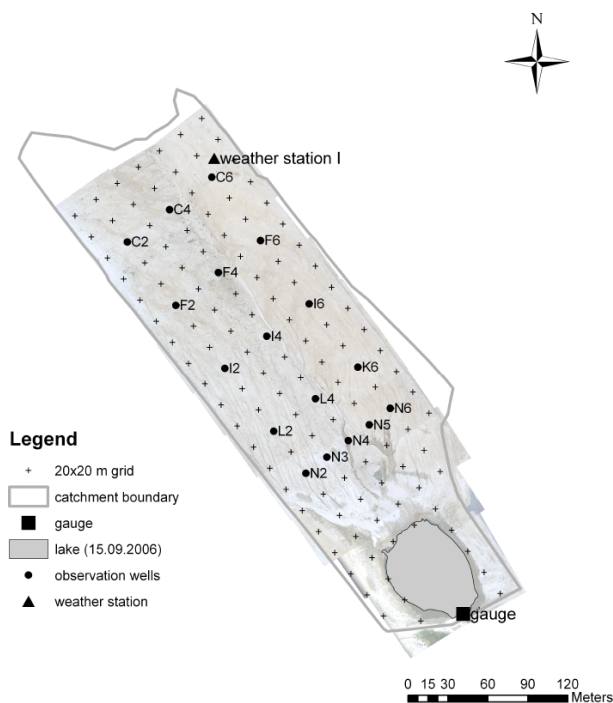


Fig. 1. GIS framework of the Chicken Creek catchment.

given by the slope of the clay base but the thickness of the sand layer tapers off towards the lake. Hence, the clay layer forms the lake bottom. The catchment boundary is defined by the high edges of the clay layer. The catchment and the depression are separated by a V-shaped clay dam to funnel the deep seepage through a narrow outlet into the depression (Fig. 2b). The climate is temperate and humid. Annual precipitation in the past decades has varied from 335 mm (1976) to 865 mm (1974), and the mean annual temperature is about 9.3°C (1971–2000). The catchment remained unplanted after the construction, and the establishment of the natural vegetation is being closely monitored (Gerwin et al., 2009).

3.2 Hydrological models

In this section, we describe the conceptual differences of the ten models, which were independently used by ten groups for predicting the discharge. The models are listed and followed by a brief description and pertinent model references (Table 1). We discuss the underlying assumptions and the basic concepts such as the dimensionality of the various approaches from 1-D to 3-D, and the different handling of surface processes, e.g. the links to the channel network. Furthermore, we highlight the similarities of the models, e.g. the description of evapotranspiration. We use the term “physically-based” according to the wording where the model is being discussed in the literature, not inferring that the process description is based on “ab initio” physical laws.

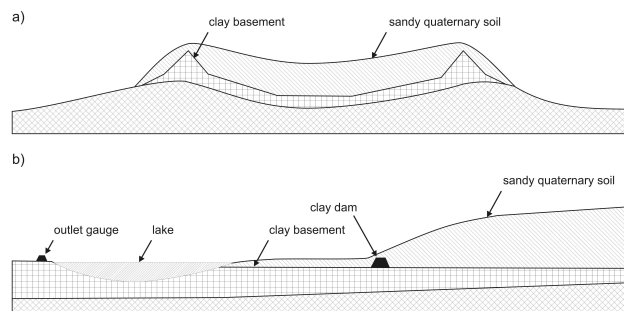


Fig. 2. Schematic of the transverse (a) and longitudinal (b) transect of the Chicken Creek catchment.

3.2.1 Catflow

Catflow (Maurer, 1997; Zehe and Flüher, 2001a; Zehe and Bloeschl, 2004; Zehe et al., 2005) is a physically-based model. It relies on a detailed process representation: the soil-water dynamic is described with the Richards equation (mixed form), evapotranspiration by the Penman-Monteith equation, surface runoff by the convection-diffusion equation, which is an approximation to the 1-D Saint Venant equation. Surface saturation, infiltration excess runoff, re-infiltration of surface runoff, lateral subsurface flow and return flow can be simulated by Catflow. It has been used as a virtual landscape generator to investigate the role of initial soil moisture and precipitation in runoff processes (Zehe et al., 2005), for simulating water flow and bromide transport in a loess catchment (Zehe and Flüher, 2001b), and for process analysis within a slowly moving landslide terrain (Lindenmaier et al., 2005), among other applications. Here, we used the quasi-3-D hillslope module of the model.

3.2.2 CMF

The *Catchment Modelling Framework* (CMF) is a multi-model toolkit. The work on it is still in progress (Kraft et al., 2008). The main objective of the model framework is to connect local scale transport models with lateral transport processes between neighbouring sites. So far, a model similar to DHSVM (Distributed Hydrology Soil Vegetation Model) (Wigmosta et al., 1994) has been implemented in CMF, based on previous work by Vaché and McDonnell (2006). The model represents subsurface transport and water flow by the 3-D solution of the Richards equation. Infiltration and unsaturated percolation is calculated with the Richards equation, and the lateral saturated flow with Darcy’s law. Infiltration excess and ponded water is directly routed to the stream network using a mass balance approach and re-infiltration is neglected. We used the two layer approach with an unsaturated and a saturated zone per cell, where the depth of the boundary between the two layers changes according to the saturation of the soil column.

Table 1. Catchment models.

model	full name of acronym	modeller	institution
Catflow		T. Blume	GFZ Potsdam
CMF	C atchment M odelling F ramework	P. Kraft	Univ. of Giessen
CoupModel	C oupled Heat and Mass T ransfer M odel for Soil-Plant-Atmosphere System	D. Gustafsson	Royal Institute of Technology KTH Stockholm
Hill-Vi		S. Stoll	ETH Zürich
HYDRUS-2D ^a		C. Stamm	Eawag
NetThales		G. B. Chirico	Univ. of Naples
SIMULAT ^a		H. Bormann	Univ. of Oldenburg
SWAT	S oil and W ater A ssessment T ool	J.-F. Exbrayat	Univ. of Giessen
Topmodel	T opography-based m odel	W. Buytaert	Univ. of Bristol
WaSiM-ETH	W ater B alance S imulation M odel- E TH	H. Hölzel	Univ. of Bonn

^a Although HYDRUS-2D and SIMULAT are not catchment models in its proper sense, they are adapted to be used as such.

3.2.3 CoupModel

The CoupModel is a physically-based model for coupled heat and mass transfer in soil-plant-atmosphere systems (Jansson and Moon, 2001). Vertical movement of water in a 1-D soil profile is described with the Richards equation using a water retention function (Brooks and Corey, 1964) and an unsaturated hydraulic conductivity function (Mualem, 1976) for each soil layer. Lateral water fluxes are considered as a drainage system, with horizontal outflow from saturated soil layers to a hypothetical drainage pipe following the Hooghoudt drainage equation (Hooghoudt, 1940). Semi-2-D and semi-3-D representation is achieved by taking the outflow from one or several 1-D soil column as lateral inputs to a downstream column. The model accounts for soil freezing, including effects on the thermal and hydraulic conductivity (Stähli et al., 1996). Water and heat exchange between soil and atmosphere are calculated separately for different surface compartments including bare soil, snow, vegetation, and interception, with individual energy balance sub-models.

3.2.4 Hill-Vi

The physically-based hillslope model Hill-Vi was developed by Weiler and McDonnell (2004) to test the benefit of virtual experiments to hillslope hydrology. Subsequently, it has been modified to simulate nutrient flushing (Weiler and McDonnell, 2006) and the effects of preferential flow networks (Weiler and McDonnell, 2007).

At each grid cell there are two storage compartments: the unsaturated zone from the soil surface to the water table and the saturated zone from the water table to the impermeable soil-bedrock interface. The water balance of the unsaturated zone is calculated based on precipitation input, actual evapo-

transpiration, and vertical recharge into the saturated zone, described by gravity flow and using the equations by van Genuchten (1980). The lateral water exchange in the saturated zone are controlled by the Dupuit-Forchheimer assumption (Freeze and Cherry, 1979), based on an explicit grid cell approach, as presented by Wigmosta and Lettenmaier (1999).

3.2.5 HYDRUS-2D

HYDRUS-2D simulates the movement of water, heat and solutes in 2-D variably saturated porous media. The Richards equation is numerically solved for the saturated-unsaturated flow region considering vertical and horizontal flow under variable boundary conditions such as atmospheric conditions, free drainage or seepage faces. A detailed manual describes the relevant technical details (Simunek et al., 1999). Lateral groundwater and unsaturated flow is represented by Richards' equation. All precipitation infiltrates into the soil except in some scenarios during frozen soil conditions. Evapotranspiration is determined by the Penman-Monteith method. Here, we use HYDRUS-2D in a catchment context and simulate the water flow through the longitudinal transect of the catchment.

3.2.6 NetThales

NetThales (Chirico et al., 2003) is a distributed, continuous, terrain-based hydrological model, simulating the hydrological processes distributed on a spatial network of elements. The properties are defined by terrain analysis of DEMs, which provides the spatial dimensions of the elements, the flow directions within the elements and the connectivity between the elements.

The water fluxes are calculated at the element scale with a computational time-step of one hour, accounting for the following processes: evapotranspiration, surface and subsurface lateral flow. Rainfall is assumed to infiltrate completely into the soil unless the soil column is entirely saturated. Overland flow occurs by exfiltration when the element soil column is saturated by lateral subsurface flow. The vertical distribution of the water within the soil column is not modelled. The lateral surface and subsurface flow are modelled as one-dimensional within each element. The processes controlling the subsurface lateral movement are vertically lumped in a nonlinear kinematic subsurface module.

3.2.7 SIMULAT

SIMULAT (Diekkrüger and Arning, 1995; Bormann, 2001, 2008) is a physically-based and time-continuous hydrological SVAT model (Soil Vegetation Atmosphere Transfer), which has been developed to simulate local-scale (vertical 1-D) hydrological processes and nutrient fluxes. It solves the Richards equation to estimate infiltration and soil-water fluxes and uses the approach by Feddes et al. (1978) to estimate root water uptake and the approach by Ritchie (1972) for evaporation as a function of surface soil moisture. Lateral groundwater flow is represented by concentration time. Surface runoff is estimated by a semi-analytical solution of the Richards equation and the interflow based on Darcy's law. In this study, a quasi 2-D slope version of SIMULAT (Giertz et al., 2006) represented by a 1-D soil column is used where the slope is represented by the number of soil columns (e.g. three to four).

3.2.8 SWAT 2005

The Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998) has been developed to simulate the long-term water and nutrient balance in mesoscale catchments. It is a physically-based semi-distributed model (Gassmann et al., 2007). The surface of each sub-catchment is divided into *Hydrological Response Units* (HRU) corresponding to single combinations of land use classes and soil types. Each HRU is an idealized hillslope and there are no interactions between them. Each HRU has a double groundwater system. Infiltration is estimated by SCS (Soil Conservation Service) curve number method. The soil-water fluxes are represented as a bucket model depending on the soil-water content and other soil properties. Lateral flow is calculated by the Hooghoudt drainage equation (Hooghoudt, 1940).

Although SWAT was developed to simulate mesoscale catchments, we used the model version SWAT 2005 (<http://www.brc.tamus.edu/swat/>) to examine the predictive power in comparison to other models for small catchments.

3.2.9 Topmodel

Topmodel is a semi-distributed hydrological model built around the concept of the topographic index, which is the ratio between the surface area that drains through a given location and the local slope (Beven and Kirkby, 1979; Beven et al., 1995; Beven, 2001). The topographic index represents the tendency of a location in the catchment to develop saturated soil conditions, and thus to generate saturated overland flow. Pixels with a similar topographic index are expected to behave hydrologically in a similar way and are, therefore, lumped in 16 classes.

Topmodel assigns a combination of storage compartments to each topographic index class such as the root zone, unsaturated and saturated zone. Water enters the root zone calculated by the Green-Ampt equation, which is affected by evapotranspiration and overflows into the unsaturated zone. A time delay function controls vertical flow from the unsaturated into the saturated zone. Finally, saturated subsurface flow is calculated by an exponential transmissivity function.

3.2.10 WaSiM-ETH

The *Water Balance Simulation Model* (WaSiM-ETH) is a physically-based and spatially distributed hydrological model. It is capable in calculating climate change effects in heterogeneous catchments and includes the major water cycle processes (Schulla and Jasper, 2007). WaSiM-ETH focuses on spatially-variable atmospheric boundary conditions and has been widely used (Niehoff et al., 2002; Bronstert et al., 2007; Jasper, 2005).

All algorithms, except the saturated soil zone routine configuration, are physically-based. The infiltration is represented by the Green-Ampt equation and the unsaturated zone by Richards equation. Flow in the aquifer was described by a linear storage approach. Here, we use the version 7.9.11.

3.3 The data set

The data set provided to the modellers represents the information which is usually available or easily accessible in case of an ungauged catchment. It contained the following:

- Coordinates of instrument locations and observation 20×20 m squares (Fig. 1).
- Digital Elevation Models (DEM) of soil surface and clay layer surface.
- Soil texture (mean value and standard deviation) of samples from all observation squares.
- Gully network imaged on an aerial photo (summer 2007) (Fig. 1).
- Hourly, daily, and monthly record of weather data monitored at the Chicken Creek weather station during the

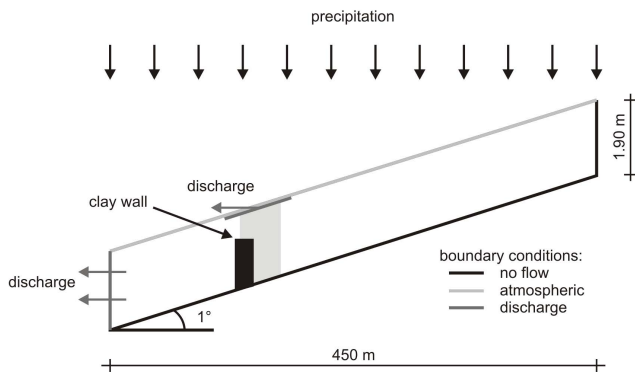


Fig. 3. Geometric representation and spatial arrangement of boundary conditions used for the HYDRUS-2D simulations.

period 29 September 2005 to 9 September 2008 (precipitation, air temperature, wind speed and direction, humidity, global radiation).

- Yearly vegetation coverage in the observation squares (once per year).
- Initial hydraulic head in the groundwater observation wells (installed from the surface down to the impervious clay base at 15 locations in the catchment) observed on 19 September 2005.

The wind direction, wind speed, air temperature, and humidity are measured by instruments from THEIS (wind transmitter “compact” $\pm 3\%$; temperature and humidity sensor “compact” (± 2 K and $\pm 2\%$). Precipitation is measured by a tipping bucket. The maximal error is 2% up and the measurement range is 0–7 mm/min. Radiation data are measured by Hukseflux instruments. The error range is $\pm 5\%$. The discharge at the outlet from the lake is measured by a combination of a V-notch weir and a tipping bucket system for small discharge. The instruments were tested in the experimental flow channel at the Brandenburg University of Technology Cottbus. The influence of small waves on the lake is accounted for by installing scumboards and a triplet of water level logger (diver accuracy ± 2 mm). The same divers are used for groundwater measurements along the No. 4 column of the observation grid (positions C4, F4, I4, L 4 and N4). The hydraulic heads at all observation wells are manually determined every 2 weeks with a hand-held pressure transducer.

The comparison of precipitation data with a second weather station in the catchment (distance 300 m) showed a maximal difference of 5%. The DEMs are based on analogue aerial photos. The GIS technique, which was used, has an error of ± 30 cm.

None of the modelling groups visited the field site before they presented their predictions during the 1st workshop (Cottbus, 1/2 December 2008). During this workshop, the

catchment was visited by all participants except by the SIMULAT and Topmodel modellers.

The data set is accessible at <https://www-fs.tu-cottbus.de/SFB38/PUBLIC>. Password requests should be addressed to the corresponding author.

3.4 Conceptualization of catchment features

The basic features of the ten models are listed in Table 2. Here we discuss these features and the underlying arguments for their choice.

- (a) *Deep seepage*: since the shape of the catchment’s soil surface as well as that of the clay base are well defined in the provided data set, all modelling groups assumed zero flow through the clay layer and across the lateral catchment boundary.
- (b) *Sensitivity analysis and scenarios*: all groups but one (HYDRUS-2D) presented, as suggested, the results for one single run. This exercise simulates the situation of a modeller being confronted with the request for a first prediction guess. HYDRUS-2D computed six scenarios. Two of them were carried out with the empirical pore tortuosity/connectivity parameter $L=0.5$ (Muallem, 1976) and four of them with $L=-0.78$ because recent studies reported considerable deviations from $L=0.5$ (Schaap et al., 2001). The precipitation events were grouped into two categories: (i) precipitation as an immediate infiltration during the day of occurrence and (ii) precipitation onto frozen soil being directly routed to discharge. This was done for an L of 0.5 and -0.78 . For the other two scenarios with $L=-0.78$, the hydraulic parameters were modified to decrease the unsaturated hydraulic conductivity and, hence, to generate more discharge.
- (c) *Dimensionality and catchment feature*: the Catflow modeller used the single hillslope module, which is only part of the full catchment model because the catchment is small. The runoff routing judged to have little effect on the overall response, and most of the gullies oriented in parallel. The two 2-D models, Catflow and HYDRUS-2D, modelled the catchment as a single slope (Fig. 3) and did, therefore, not include the gully network. All other modellers used 3-D or semi-3-D (Coup-Model) models (Table 2). CMF used an irregular grid of about 3.000 Thiessen polygons. CoupModel, Hill-Vi, and WaSiM-ETH used regular grids. The SIMULAT user used a 1-D model to represent the hydrological dynamics because it was assumed that overland flow as well as interflow, and therefore neighbourhood relations, do not play a major role in the catchment. The Topmodel user generated a 2 m resolution digital elevation map (DEM) from the available elevation

Table 2. Conceptualization of catchment features.

model	dimension	discretization horizontal Δx	vertical Δz	pre-calculation/ pre-consideration	scenarios	data estimation
Catflow	2-D	uniform along the elevation contour lines upper slope: $\Delta x=10$ m, else: $\Delta x=1$ m	$0 < z < 20$ cm: $\Delta z=4$ cm $z > 20$ cm: $\Delta z=20$ cm	runoff routing judged to have little effect on the overall response	one	
CoupModel	semi-3-D	20×20 m grid	elevation difference between soil and clay base surface averaged over the grid cell; $\Delta z \geq 0.5$ m		one	
CMF	3-D	irregular digital elevation network	unsaturated and saturated zone with each with a time-variant layer thickness: $\Delta z \geq 0.5$ m		one	
Hill-Vi	3-D	10×10 m grid			one	
HYDRUS-2D	2-D	uniform along the elevation contour lines		assuming that surface runoff hardly ever occurs based on comparison of rainfall intensities and soil hydraulic properties	$L=0.5$ (Mualem, 1976) and four runs with $L=-0.78$ because recent studies reported that $L \gg 0.5$ (Schaap et al., 2001).	
NetThales	3-D		no unsaturated zone	assuming that infiltration hardly ever occurs based on comparison of rainfall intensities and soil hydraulic properties	one	Control of evapotranspiration: initial root-zone depth $\Delta z_{\text{root}}=5$ cm yields a runoff-rainfall ratio of 70%. Ratio was considered as being too high based on the modeller's knowledge. Thus, Δz_{root} was increased to 30 cm, which reduced the runoff-rainfall ratio to about 50% at the annual scale
SIMULAT	1-D	20×20 m grid	soil layer thickness directly taken from soil data set		one	soil considered to be compacted and used the highest bulk density class according to .Adhoc AG Boden (2005)
SWAT	3-D		unsaturated zone and shallow aquifer, no deep aquifer		one	
Topmodel	3-D	topographic index with 16 classes based on a 2 m resolution DEM		maximum root zone storage deficit and flow velocity estimated from available catchment data	one	transmissivity, maximum root zone storage deficit and flow velocity estimated from data set.; recession curve parameter m estimated from literature values
WaSiM-ETH	3-D	5×5 m grid		sparse vegetation was neglected no macropores	one	available soil depths averaged; no macropores because the soil has been recently dumped effective parameters are upscaled measurement-derived parameters

measurements and used it to calculate the topographic index map. The index values were sorted into 16 classes (Table 2).

- (d) *Discretization*: all models except NetThales modelled at least a saturated and an unsaturated layer. In Catflow the top soil is described with a five times higher resolution (Table 2) because the near-surface processes were

assumed to be important. CMF divided each soil column into a saturated and unsaturated zone with time-variant layer thickness to shorten the computing time. The SWAT modeller described an unsaturated zone and a shallow groundwater compartment. In the CoupModel the elevation difference between soil and clay base surfaces is averaged over each grid cell. The resulting grid cell value was, for numerical reasons, kept at least

0.5 m. WaSiM-ETH reduced the calculation effort by aggregating the DEM to a 5×5 m raster. The aggregated DEM does not resolve the gully structures nor the clay dam.

- (e) *Surface runoff*: the aerial photo of summer 2007 showed evidence of surface runoff across the entire catchment. However, the modellers, except CoupModel, neglected it due to the soil texture data. The HYDRUS-2D group compared rainfall intensities and texture-derived estimates of soil hydraulic properties and concluded that surface runoff (not accounted for by HYDRUS-2D) would hardly ever occur. Similarly, the NetThales modellers argued that infiltration excess runoff cannot be generated using a 1-D Richard equation based infiltration model because the soil hydraulic conductivity (estimated with pedotransfer functions from soil texture) was definitely larger than the maximum hourly rainfall intensity. The only dominant runoff generation mechanism was, therefore, saturation excess runoff (Table 2). HYDRUS-2D generated runoff by modifying the porosities and hydraulic conductivities upslope of the clay dam (Fig. 3). The soil parameters were estimated according to Schaap et al. (2001) using the routine implemented in the HYDRUS-2D program. The CMF modeller did not make use of the provided gully network, because the shape and depth of the gullies were lacking. However, the mere existence of gullies was included as infiltration excess. The Hill-Vi group assumed that surface runoff is important because of the distinctive gully network but they had difficulties in accounting for large hydraulic conductivities on one hand, and large amounts of surface runoff on the other. Hill-Vi recalculated the drainage network for every time step so that the information of the gullies was not incorporated in the model. Preliminary Hill-Vi test runs with a snowmelt routine did not yield notable effects. Snow was, therefore, disregarded in the model. The CoupModel group did not use the information on the initial ground water levels assuming that the catchment already existed long enough to be “initialized”. The role of the gullies was incorporated in the parameterization of the surface runoff by reducing the surface pool threshold to get a faster surface runoff response. The SIMULAT user neglected the information on existing gullies. The NetThales modeller considered evapotranspiration and the “root-zone depth” Δz_{root} to be critical features. Initially, they assumed that $\Delta z_{\text{root}} = 5$ cm. This led to an annual runoff-rainfall ratio of 70%. Based on the modeller’s knowledge of relatively dry Austrian and German catchments, the NetThales modellers argued that in Brandenburg this ratio is less than 30%. Since the plant cover was almost non-existent, a larger runoff ratio was expected, but certainly not 70%. Also the baseflow contribution of the initial simulations was

considered too high in this climate. Thus, the Δz_{root} was increased to 30 cm, which reduced the runoff-rainfall ratio to about 50% at the annual scale.

- (f) *Soil parameters*: catflow treated the soil as a homogeneous loamy sand, parameterized after Carsel and Parrish (1988), because soil texture of the soil layer shows little variability across the catchment and with depth. The Hill-Vi modeller applied the Rosetta data base (Schaap et al., 2001) to estimate soil hydraulic parameters with hierarchical pedotransfer functions. For the CoupModel the hydraulic properties of the soil layer were estimated from the numerous soil-water retention data of Swedish sandy soils (Lundmark and Jansson, 2009). In SIMULAT the thickness of the soil layer was directly taken from the soil data set. The SIMULAT modeller treated the soil to be compacted because it was dumped and shaped with large machines and used the highest bulk density class according to Adhoc AG Boden (2005). Based on the soil and the soil layer information, it was concluded that subsurface runoff exceeds surface runoff with a minor contribution of interflow, making baseflow the dominant runoff component. The main principle of the soil parameterisation was “as simple as possible”. Therefore, the data from each soil depth were aggregated to a single average value. This was parameterised with literature values (AdHoc-AG Boden, 1999). The WaSiM-ETH user did not consider macropores because the soil material had been recently dumped and repacked and also because of the initial state of the vegetation. In WaSiM-ETH the effective parameters are upscaled measurement-derived parameters, which are gathered “normally” during the calibration by measured outputs. Therefore, they were taken from another headwater catchment in Germany (Hölzel and Diekkrüger, in press, 2008).
- (g) *Process assumptions*: topmodel does not account for several processes that do occur in this particular catchment, such as snowmelt, gully erosion. Its semi-distributed nature does not allow for describing the clay dam. Although Topmodel could be customised to indirectly include such processes, the modeller decided not to do so at this stage of the modelling process, in order to provide a reference performance. Transmissivity, maximum root zone storage deficit, and flow velocity were estimated from the available catchment data. Only one parameter, the shape of the recession curve, was estimated from literature values.

3.5 Process concepts and implementation

3.5.1 Infiltration, saturated and unsaturated flow

The saturated and unsaturated flow was simulated either as 1-D linear storage (CoupModel, Topmodel,

Table 3. Methods for calculating infiltration, saturated and unsaturated flow.

model	infiltration	saturated flow	unsaturated flow
Catflow	Richards equation (mixed form)	Richards equation (mixed form)	Richards equation (mixed form)
CMF	Richards equation with an assumed transition zone of 5 cm thickness	Darcy's law	Richards equation using Brooks-Corey retention curve
CoupModel	modified Darcy's law infiltration (Jansson and Halldin, 1979) infiltration capacity depend on saturated hydraulic conductivity in both matrix and macro pores, with correction for frozen soil conditions (Stähli et al., 1996)	drainage equation by Hooghoudt (1940)	Richards equation, matrix and macro pore flow
Hill-Vi	infiltration capacity=saturated hydraulic conductivity Mualem-van Genuchten equation	Dupuit-Forchheimer assumption (Freeze and Cherry, 1979; Wigmosta and Lettenmaier, 1999)	simplified Richards equation (gravity flow)
HYDRUS-2D	Richards equation	Richards equation	Richards equation (matrix flow; macropore flow mimicked as described under 3.3.5)
NetThales	no infiltration excess is simulated rainfall is assumed to infiltrate totally into the soil. Exfiltration occurs when the soil column saturates.	lateral non-linear kinematic flow	no unsaturated flow is simulated. The timing of the vertical redistribution of the water into the soil column is neglected lateral flow occurs when average soil moisture is above the field capacity
SIMULAT	semi-analytical solution of the Richards equation for separation of surface runoff and infiltration (Smith and Parlange, 1978) interflow (based on Darcy's law), groundwater recharge (flow across the lower boundary of a soil column)	concentration time	Richards equation
SWAT	SCS (Soil Conservation Service) curve number method	drainage equation by Hooghoudt (1940)	soil properties and water content
Topmodel	Green-Ampt infiltration	time delay function	exponential transmissivity function
WaSiM-ETH	Green-Ampt approach modified by Peschke (1987)	linear storage approach	Richards equation parameterized based on van Genuchten (1980)

WaSiM-ETH), 1-D Richards equation (SIMULAT), 2-D (Catflow, HYDRUS-2D) or complete 3-D (CMF, Hill-Vi). Unsaturated flow was calculated with the Richards equations, except in the case of Topmodel, which used an exponential transmissivity function. NetThales did not calculate the flow in the unsaturated zone. Richards equation was used to calculate saturated flow (Catflow, HYDRUS-2D), the Dupuit-Forchheimer assumption (Hill-Vi), or Darcy's law (CMF). Detailed information is provided in Table 3.

In seven models, except SWAT, Topmodel, and WaSiM-ETH, infiltration was handled as unsaturated flow described by the Richards equation, with the latter representing the

infiltration excess mechanism. SWAT used the SCS curve number method and Topmodel and WaSiM-ETH used the Green-Ampt approach.

In some scenarios, HYDRUS-2D routed 10% of the precipitation directly to the bottom layer above the clay base bypassing the entire soil (preferential flow), due to hydrophobic conditions in summer. This was achieved by introducing a flux boundary at the soil bottom. In a similar way, precipitation in frost periods was directly routed downstream as surface runoff due to frozen top soil and was not accumulated as snow.

Table 4. Methods for calculating snow melt and interception.

model	snow melt	interception
Catflow	not represented	LAI dependent bucket approach method (seasonal cycle)
CMF	no snow accumulation modelled	20% of total rainfall
CoupModel	snow melt/refreeze based on energy balance, including surface heat exchange, radiation, and near surface soil heat flux. Precipitation is assumed to be snow below $T < 0^{\circ}\text{C}$, and a mixture of rain and snow in a temperature range $0 < T < +2^{\circ}\text{C}$	LAI dependent bucket model with specific interception capacities for snow and rain (higher for snow) (Ståhli and Gustafsson, 2006) sky-view fraction and direct throughfall exponential function of LAI LAI was assumed a seasonal cycle (0 to maximum), and an inter-annual increase (see supporting material)
Hill-Vi	no snow routine implemented	no interception
HYDRUS-2D	cumulative precipitation during periods of snowfall periods is directly converted into discharge upon soil thawing	no vegetation cover assumed
NetThales	no snow fall and snow accumulation is simulated snow has been considered negligible after a preliminary analysis	no interception is simulated
SIMULAT	degree day approach	LAI dependent bucket approach
SWAT	snowfall at $T < 1^{\circ}\text{C}$ snowmelt above 0.5°C based on degree-day approach	LAI function daily updated as function of a maximum value
Topmodel	no snow routine implemented	no interception
WaSiM-ETH	temperature-index method	LAI depended bucket approach method

3.5.2 Stream flow routing

The catchment is relatively small and has a maximal extension of 450 m. Therefore, some modelling groups assumed that stream flow is of minor importance (CoupModel, Hill-Vi, and HYDRUS-2D). Catflow and WaSiM-ETH approximated the stream flow as a kinematic wave using either the 1-D Saint-Venant or the Manning-Strickler equation. Simple mass balance approaches were used by CMF and NetThales. SIMULAT assumed a concentration time based approach and Topmodel a simple time delay function, both neglecting the gully network. SWAT used the gully network map to define the stream-network. They neglected the existence of the lake, allowing ArcView to define a stream network routing the water directly to the lake outlet.

3.5.3 Snow accumulation, snowmelt and interception

Snow accumulation and snow melt had a strong influence during the winter 2005/2006 with a period of 42 days below 0°C with 15.6 mm precipitation, but it was not important for the other winter periods. The two processes were taken care of by CoupModel, SIMULAT, SWAT, and WaSiM-ETH. These models are using the energy balance and temperature

index or degree day method to accumulate and melt the snow (Table 4). The other models include neither snow nor soil frost, but some HYDRUS-2D scenarios included the frozen soil by routing the precipitation directly to surface runoff.

Interception was mostly neglected because vegetation was very sparse in the initial phase after catchment construction. However, the vegetation developed rapidly and will probably affect future predictions. Catflow, CMF, CoupModel, SIMULAT, and WaSiM-ETH explicitly describe the interception losses from plant surfaces. CMF used a constant 20% loss of all precipitation events whereas the other four models used a leaf-area-index (LAI) dependent approach (Table 4).

3.5.4 Evapotranspiration

Potential evapotranspiration (PET) was calculated by most models using the Penman-Monteith equation. Hill-Vi used the Turc equation and SWAT relied on the Hargreaves equation. Additionally, the CoupModel calculated soil and snow evaporation based on a surface energy balance. For all models the actual evapotranspiration (AET) was determined on the basis of PET and the available soil-water. The CoupModel also includes the root zone soil temperature as a

Table 5. Methods for calculating the potential and actual evapotranspiration (PET and AET, respectively).

model	PET	AET
Catflow	Penman-Monteith equation but not returned as output	Plate and Zehe (2008)
CMF	Penman-Monteith equation (Allen et al., 1998)	piecewise linear function of the soil-water content within the “root-zone”
CoupModel	potential transpiration and potential interception evaporation using Penman-Monteith equation (Monteith, 1965), with radiative and vapour pressure deficit regulation of stomatal resistance (Lohammar et al., 1980) soil (and snow) evaporation by surface energy balance, i.e. bulk transfer equations (Alvenäs and Jansson, 1997; Gustafsson et al., 2001)	soil moisture and temperature regulation of actual root water (Jansson and Halldin, 1979) soil surface vapor pressure function of surface temperature and water content of upper soil layer; snow surface vapor pressure correspond to saturation over ice (dry snow) or water (melting snow)
Hill-Vi	Turc (1961)	linear function of soil-water content in the unsaturated zone
HYDRUS-2D	Penman-Monteith	
NetThales	Penman-Monteith equation (Allen et al., 1998; Kroes et al., 2008)	linear function of the soil-water content within the “root-zone”
SIMULAT	Penman-Monteith equation	reduction of PET depends on actual soil matric potential, root distribution (Feddes et al., 1978) for transpiration and a soil factor as well as the number of days after the last rainfall in case of evaporation (Ritchie, 1972)
SWAT	Hargreaves empirical method (Hargreaves et al., 1985)	evaporates canopy storage until PET is reached if $PET > \text{canopy storage}$, remaining evaporative demand is partitioned between vegetation and snow/soil
Topmodel	Penman-Monteith equation (Allen et al., 1998)	function of root zone storage deficit
WaSiM-ETH	Penman-Monteith (Monteith and Unsworth, 1990)	suction depended reduction approach

parameter in this calculation (Table 5). WaSiM-ETH neglected the sparse vegetation and included only evaporation.

3.5.5 Clay dam

The clay dam is supposed to funnel the saturated subsurface flow towards the narrow outlet into the alluvial region next to the lake. All 3-D models, except CoupModel, Topmodel, and WaSiM-ETH, incorporated the subsurface clay dam using the two DEMs describing the elevation of the surfaces of the soil and the clay base. This reduced the depth of the sandy soil layer immediately above the clay dam to a few centimetres. In SIMULAT the clay dam was considered as a locally shallow soil layer, but this did not affect the concentration time of subsurface flow. Lateral transport processes were considered by a concentration-time based approach neglecting neighbourhood relations. In WaSiM-ETH, the clay dam was neglected by using a constant soil layer thickness of 181 cm. Topmodel implemented the subsurface dam by calculating the topographic index based on the subsurface topography rather than on the surface topography. The topographic index distribution function did not show large dif-

ferences. The soil thickness was constant for the whole catchment (300 cm). CoupModel calculated the sand layer thickness from the elevation difference between the sand surface and the clay base surface averaged over the observation squares. The sand layer thickness was, for numerical reasons, not allowed to be smaller than 0.5 m. Thus, the clay dam was only represented as a shallow sand layer. To represent the clay dam, the 2-D models (Catflow and HYDRUS-2D) used a constant sand layer thickness with a reduced hydraulic conductivity (Fig. 3). HYDRUS-2D simulations were run with a low porosity soil material being placed uphill of the dam to mimic the funnelling effect of the subsurface dam. Its porosity and hydraulic conductivity was about one fifth of the sand layer. This forced the streamlines towards the soil surface above the clay layer producing a seepage face, which allowed runoff generation (Fig. 3).

3.6 Parameterization of physical soil properties

For describing the physical properties of the saturated and the unsaturated zone, all modeller groups received only the information on soil texture. This was the basis for estimating

Table 6. Parameterization of hydraulic conductivity, porosity, and of the unsaturated zone.

model	hydraulic conductivity ^a	unsaturated zone ^a	porosity ^a
Catflow	Carsel and Parrish (1988)	after Carsel and Parrish (1988) Mualem-van Genuchten (Mualem, 1976; van Genuchten, 1980)	after Carsel and Parrish (1988)
CoupModel	Swedish sand (Lundmark and Jansson, 2009)	hydraulic conductivity function of Mualem (1976) and water retention function of Brooks and Corey (1964)	Input parameter (estimated by analogy)
CMF	estimated	AG Boden (1994)	AG Boden (1994)
Hill-Vi	Schaap et al. (2001)	Mualem-van Genuchten (parameterized according to (Schaap et al., 2001))	Schaap et al. (2001)
HYDRUS-2D		Mualem-van Genuchten (Schaap et al., 2001), for the L factor we used also the data base implemented in HYDRUS yielding different values	
NetThales	Rawls and Brakensiek (1985)	Rawls and Brakensiek (1985) PTFs have been used to estimate the saturated and residual water content according (Romano and Santini, 2002)	FWC has been quantified by analyzing a drainage process (Romano and Santini, 2002), simulated with the SWAP model (van Dam et al., 1997). The FWC value is assumed equal to the average water content in the top 30 cm when the drainage flux at 30 cm depth is equal to 0.10 mm/d.
SIMULAT	Rawls and Brakensiek (1985)	Brooks and Corey (1964)	Adhoc AG Boden (2005)
SWAT	Rawls and Brakensiek (1985)		computed by SWAT as a function of bulk density
Topmodel	Saxton et al. (1986)	unsaturated zone time delay per unit storage deficit from literature values (Gallart et al., 2007; Choi and Beven, 2007)	not used explicitly
WaSiM-ETH	Adhoc AG Boden (2005)	Adhoc AG Boden (2005)	Adhoc AG Boden (2005)

^a The parameter sets are included in the Supplement (<http://www.hydrol-earth-syst-sci.net/13/2069/2009/hess-13-2069-2009-supplement.pdf>).

the porosity and the saturated and unsaturated hydraulic conductivity. Catflow, CMF, HYDRUS-2D, and NetThales considered hydraulic conductivity being constant for the whole catchment. CoupModel, Hill-Vi, SIMULAT, SWAT, Topmodel, and WaSiM-ETH used hydraulic conductivities with a spatial variation based on the soil particle distribution.

In case of NetThales, SIMULAT, and SWAT the parameters were estimated on the basis of the transfer functions of Rawls and Brakensiek (1985) (Table 6). They obtained similar mean saturated hydraulic conductivities K_{sat} (NetThales: 50 mm/h; SIMULAT: 61 mm/h; SWAT: 75 mm/h). Also, the modeller of HYDRUS-2D (54 mm/h) and Topmodel (58 mm/h) obtained a value in that range using the approach of Saxton et al. (1986). Slightly larger K_{sat} were used

in the Hill-Vi (90 mm/h, calculated after Schaap et al., 2001) and CoupModel (84 mm/h, in analogy to Swedish sands – Lundmark and Jansson, 2009). WaSiM-ETH used a German soil definition (Adhoc AG Boden, 2005) and obtained 118 mm/h. Catflow used the approach of Carsel and Parrish (1988) and estimated a value of 146 mm/h for the aquifer. The largest hydraulic conductivity was used by CMF. CMF derived the hydraulic properties using the German soil mapping manual (AG Boden, 1994). Since in-situ saturated conductivity is in most cases underestimated, they assumed a higher value of 417 mm/h.

The porosity n [m^3/m^3] was in all but three cases estimated to be in the range of 0.40 to 0.45. The models which used a smaller n were CMF (0.35), SIMULAT (0.34) and

WaSiM-ETH (0.38), all of them using the German soil definition (Adhoc AG Boden, 2005). The German soil definition, the estimators of Carsel and Parrish (1988) and of Saxton et al. (1986), and the analogy to Swedish sands do not require bulk density nor organic matter content, information which was not available in this case. The estimates of the water content at the wilting point varied from 0.045 to 0.090 [m^3/m^3] and the field capacity from 0.125 to 0.280 [m^3/m^3].

The hydraulic parameterization of the unsaturated zone was mostly done using the methods of Mualem (1976) and van Genuchten (1980) (Catflow, Hill-Vi, HYDRUS-2D) or that of Brooks and Corey (1964) (CoupModel, NetThales, SIMULAT). The empirical pore tortuosity/connectivity parameter L is usually assumed to be 0.5 (Mualem, 1976), but was varied in some HYDRUS-2D simulations because more recent studies revealed considerable deviations from this value (Schaap et al., 2001). The pore-size index λ as defined by Brooks and Corey is here expressed in terms of the α_{sg} , and n_{vG} parameters as defined by van Genuchten. If $\alpha \cdot h_b \gg 1$ then

$$\lambda = n_{\text{vG}} - 1 \quad (1)$$

WaSiM-ETH used the smallest n_{vG} (1.13) CoupModel a constant n_{vG} (1.42), HYDRUS-2D n_{vG} between 1.15 and 1.88, Catflow a soil specific n_{vG} (loamy sand: 2.28 and sandy clay loam: 1.48). The models CMF, Hill-Vi, and SIMULAT assumed a spatial variation of n_{vG} from 1.15 to 1.37, 1.37 to 3.57, and 1.56 to 2.33, respectively. NetThales, SWAT, and Topmodel did not account for unsaturated flow, nor did they use Richards equation for representing the unsaturated flow. In Topmodel, the flow between the unsaturated and saturated storage is controlled by one parameter representing the time delay per unit storage deficit (Gallart et al., 2007; Choi and Beven, 2007). The complete parameter sets are listed in the Supplement (<http://www.hydrol-earth-syst-sci.net/13/2069/2009/hess-13-2069-2009-supplement.pdf>).

3.7 Initial conditions

The initial conditions were not well defined, in particular the initial volumetric soil-water content $\theta(t_0)$ [m^3/m^3]. SIMULAT estimated the soil to be dry. Other models were run to initialize this variable and its spatial variation: Hill-Vi three times (0.20 ± 0.25) and CMF (0.22 ± 0.06), SWAT ($\theta(t_0) = 0.11 \pm 0.04$), and WaSiM-ETH ($\theta(t_0) = 0.27 \pm 0.05$) once. CMF used the 3-year rainfall record for the initialization run, with a wet year in 2008. Catflow was run twice to find stable initial conditions, in this case not for the soil-water content but for matric potential. Pre-runs were used to achieve quasi-steady-state conditions, which were then used as initial condition. WaSiM-ETH archived system-stable initial conditions of the whole model period using default values.

CoupModel initialized the soil moisture at field capacity. HYDRUS-2D was run with different $\theta(t_0)$. The wet scenar-

ios assumed a constant matric potential of -0.3 m, whereas the dry runs started with a matric potential of -1.0 m. When model runs were started, assuming dry soil, the discharge was too little to fill the lake at the outlet of the catchment within the first year. Since the presence of the lake was known to the modellers, such model runs were rejected. SIMULAT assumed a matric potential of -3 m at the bottom of the sand layer and decreasing values towards the soil surface assuming hydrostatic equilibrium. Topmodel used an initial vertical subsurface flow parameter of 0.017 mm/h per unit area which was estimated from the mean annual rainfall of 496 mm and the assumed runoff coefficient of 0.3 .

The groundwater levels were part of the initial data set but none of the models except SIMULAT made use of it, because the case of an “empty”, newly constructed catchment without initial groundwater was not considered, because it would lead to numerical problems. Therefore, Catflow, Hill-Vi, and WaSiM-ETH used a warm-up run for the formation of a groundwater table. HYDRUS-2D defined the groundwater table at 40 to 60 cm within a soil cover of constant thickness (1.90 m) (Fig. 3).

3.8 Water budget of the Chicken Creek

The measurements used to close the water budget of the Chicken Creek catchment were precipitation, discharge from the lake, lake storage change, and changes of the levels of the groundwater table. Soil moisture measurements were available from mid 2007 onwards. For reference, the potential evapotranspiration PET was calculated using grass-referenced Penman-Monteith using the standard parameterization (Allen et al., 1994) and the reference actual evapotranspiration AET was estimated using a modified Black approach (Black et al., 1969; DVWK, 1996). The continuous data by the Black approach were compared with some AET data by the Bowen Ratio method. The comparison showed a good agreement of the AET during summer months but an underestimation of AET during the windy seasons of spring and autumn.

The Chicken Creek catchment drains into a lake (Fig. 1). The gauge for measuring the catchment discharge is located at the outflow of the lake. The inflow into the lake is not monitored. Since several models did not consider the lake as a buffer compartment, we determined the catchment outflow into the lake by subtracting the observed lake storage changes and precipitation onto the lake from the measured lake outflow and added the evaporative losses from the lake. The back calculated inflow into the lake is the standard against which the modelled discharge is compared.

For the above calculation, we assume that the clay base prevents any vertical seepage. Vattenfall Europe Mining AG constructed the clay layer and tested the clay beforehand. The hydraulic conductivity of the clay is $2 \cdot 10^{-10}$ m/s. Using the maximum water level in the lake (2.50 m) and a clay layer thickness of 1.50 m, the losses through the clay would be in

Table 7. Time to set up the models and computation time.

model	model development (men-days)	computation time	computer performance
Catflow	5	9 h	2.0 GHz, Dual Core, 2 GB RAM
CMF	14 ^a	1 h	2.6 GHz, Quad Core
CoupModel	7	20 min	standard personal computer
Hill-Vi	15 ^a	15 min	3.16 GHz, Dual Core, 3 GB RAM
HYDRUS-2D	35	15–20 min ^b 12 h and more ^c	1.8 GHz, Dual Core, 1 GB RAM
NetThales	6	23 min	2.2 GHz, Dual Core, 2 GB RAM
SIMULAT	4	2 h	standard personal computer
SWAT	3	5 s	2.0 GHz, Dual Core, 2 GB RAM
Topmodel	2	>1 s	any personal computer
WaSiM-ETH	2	2.5 h	2.6 GHz

^a including code implementation. ^b standard run without numerical problems. ^c run with numerical problems.

the order of 17 mm/y. Precipitation into the lake were taken from the weather station data. The largest uncertainty results from the evaporation. This was calculated by the Dalton method including the Richter wind function (Richter, 1977) and a wind function for small water bodies (Penman, 1948; Nenov, 2009). The comparison with the measured declines of the lake levels during dry season showed a good agreement.

3.9 Computation time

Models, including the pre-calculations, were set up in one week, except for CMF, Hill-Vi, and HYDRUS-2D. The CMF and the Hill-Vi user needed to adjust the model to the specific needs of an artificial catchment. The HYDRUS-2D modeller applied the model in a catchment context. Since the model does not simulate surface runoff, direct runoff, e.g. due to frozen soil conditions, needed to be calculated before. Additional time was needed because the HYDRUS-2D modeller developed several scenarios. All computations were run on a standard personal computer. The fastest run was done by Topmodel which ran within one second. Similar was the runtime of SWAT (5 s). CoupModel, Hill-Vi, and NetThales used less than one hour and all other models needed more than one hour. Catflow used the maximum calculation time of 9 h. HYDRUS-2D simulations needed 15 to 20 min if no numerical problems were available. Numerical problems were due to saturation of surface-near cell which would produce overland flow which HYDRUS-2D is not able to simulate. This increased simulation times to 12 or more hours per run (Table 7).

4 Results

We first compare the predictions and observations in terms of the water budget, discharge, and groundwater levels. The predictions are presented for the three hydrological years from November through October (2005/2006, 2006/2007, and 2007/2008 only until 8 September 2008). These periods are referred to as the 1st, 2nd, and 3rd year.

4.1 Water budget

Below, the annual values of the 1st, 2nd, and 3rd year are reported as triplets (1st, 2nd, and 3rd year). Annual precipitation used as input was 373, 566, and 511 mm/y (Table 8a–c). All models used hourly data except HYDRUS-2D, where wind-corrected daily precipitation was used. CoupModel used wind-corrected hourly precipitation. In CMF, a 20% interception loss of the total precipitation (Table 4) was assumed.

The calculated reference PET was 779, 782, and 511 mm/y. PET, predicted by the ten model, ranges from 146 to 807 mm/y (1st year). The values for the 2nd and 3rd year vary in the same range. The reference AET, calculated by the modified Black method (Black et al., 1969; DVWK, 1996) was 163, 165, and 137 mm/y, which yields a ratio AET/PET of 0.21, 0.21, and 0.27. Only Hill-Vi predicted a similar behaviour. The other models systematically overestimated AET relative to PET.

CMF predicted the significantly lowest PET and AET, whereas Hill-Vi predicted a high PET but a low AET. Catflow produced AETs of 161, 170 and 163 mm/y assuming a vegetation cover of 5%, an LAI ranging between 1 and 2, a

Table 8a. Predicted and observed water budget of the Chicken Creek catchment for the 1st year.

	P (mm/y)	PET (mm/y)	AET (mm/y)	Discharge (mm/y)	Storage (mm/y)	Balance (mm/y)
Catflow	373	NA	161	249	−59	22
CMF ^b	298	146	88	208	−44	46
CoupModel	401	NA	437	12	−48	0
Hill-Vi	373	717	153	306	−63	−23
HYDRUS-2D	431	611	409–545	34–48	−158–−38	−5–22
NetThales	373	392	226	189	−38	−4
SIMULAT	373	680	239	189	25	−80
SWAT	373	807	350	76	−4	−49
Topmodel	373	570	271	94	0	8
WaSiM-ETH	373	700	283	107	0	−17
Chicken Creek	373	779	163	113 ^d	35	62

Table 8b. Predicted and observed water budget of the Chicken Creek catchment for the 2nd year.

	P (mm/y)	PET (mm/y)	AET (mm/y)	Discharge (mm/y)	Storage (mm/y)	Balance (mm/y)
Catflow	565	NA	170	262	80	53
CMF ^b	452	139	104	238	13	97
CoupModel	666	NA	563	27	76	0
Hill-Vi	565	718	156	346	58	5
HYDRUS-2D	635	602	520–579	19–67	27–33	1–17
NetThales	565	421	284	259	23	−1
SIMULAT	565	713	318	339	−9	−83
SWAT	565	815	409	145	18	−7
Topmodel	565	573	384	171	0	10
WaSiM-ETH	565	689	371	162	0	32
Chicken Creek	565	782	165	105	69	226

Table 8c. Predicted and observed water budget of the Chicken Creek catchment for the 3rd year.

	P (mm/y)	PET (mm/y)	AET (mm/y)	Discharge (mm/y)	Storage (mm/y)	Balance (mm/y)
Catflow	511	NA	163	258	55	35
CMF ^b	409	116	78	250	−39	120
CoupModel	563	NA	498	76	−11	0
Hill-Vi	511	588	128	329	44	10
HYDRUS-2D ^c	357	331	277–313	34–64	−9–7	2–26
NetThales	511	307	199	275	39	−2
SIMULAT	511	628	278	283	17	−67
SWAT	511	706	331	164	−4	20
Topmodel	511	486	294	198	NA	19
WaSiM-ETH	511	573	272	178	NA	61
Chicken Creek	511	674	137	113	162	99

^a until 8 September 2008. ^b 20% interception losses. ^c until 3 July 2008. ^d 69 mm were needed to fill up the lake.

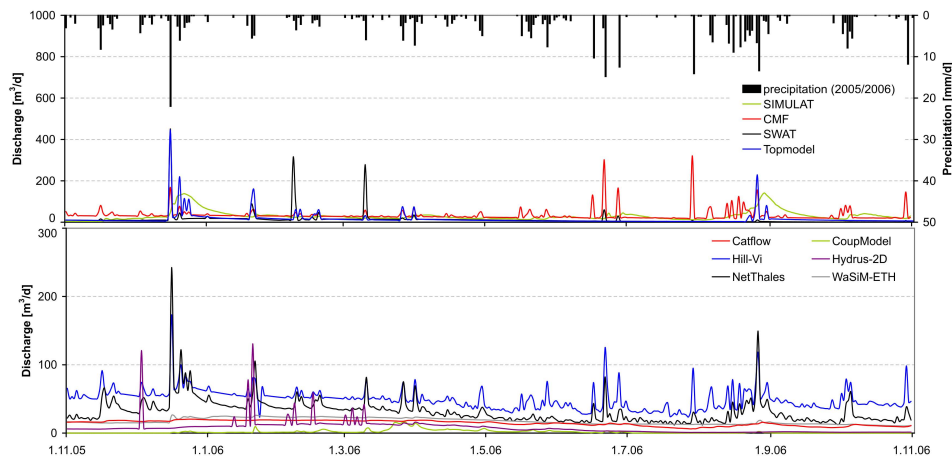


Fig. 4a. Predicted discharge for the hydrological year 2005/2006.

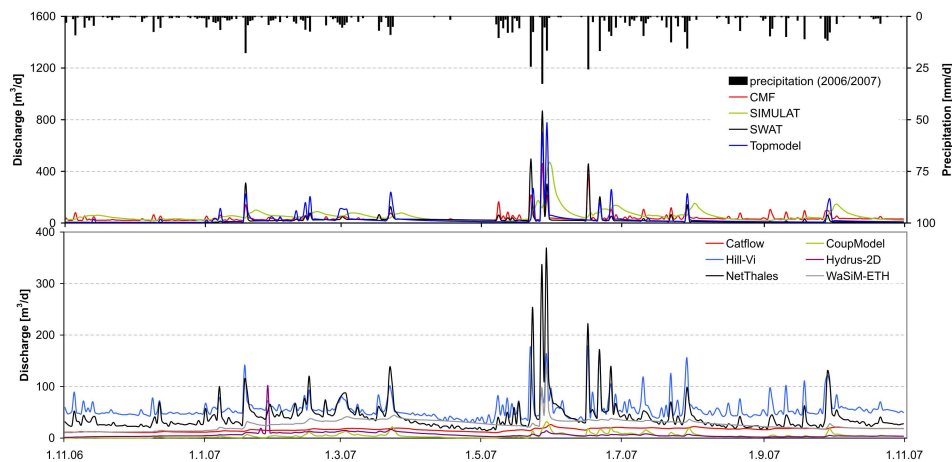


Fig. 4b. Predicted discharge for the hydrological year 2006/2007.

canopy height increasing in the course of the growing season from 13 to 40 cm, and a stomatal resistance of 200 s/m.

The measured discharge from the catchment was 113, 105, and 113 mm/y. The range of the ten discharge predictions was 12 to 306, 27 to 346, and 76 to 329 mm/y. Expressed as percentage of the measured discharge, the predicted discharge ranges from 10 to 221, 19 to 329, and 30 to 290% (Fig. 4a–c). The catchment was built by dumping relatively dry soil onto the clay base so that the groundwater gradually filled up after construction. At the end of the three years, the groundwater storage was 35, 69, and 162 mm, determined according to the water-table fluctuation method (Meinzer, 1923; Healy and Cook, 2002) using the means of porosity and groundwater table rise. Water storage in the unsaturated zone was not available as model input. The predicted storage changes (sum of ground and soil-water) varied between –63 and 25, –9 and 76, and –39 and 44 mm.

The modellers were unaware that the dumped soil material was relatively dry (see Sect. 3.7) and groundwater absent.

Most of them assumed an initial water content corresponding to field capacity or they estimated the soil-water contents from pre-runs. Therefore, the predictions cannot be directly compared with the observed data but can be placed there in relation to each other. All models, except SIMULAT, predicted a loss of soil- and groundwater for the first year. This is not surprising because the precipitation was less than the long-term mean.

The errors in the internal model mass balance ΔM_{error} [mm/y] are

$$\Delta M_{\text{error}} = P - \text{AET} - Q - \Delta S \quad (2)$$

with P being measured and AET, Q , and ΔS simulated entities (Table 8a–c). The CoupModel, Hill-Vi, HYDRUS-2D, NetThales, Topmodel and WaSiM-ETH produce a ΔM_{error} of less than 5% of P , Catflow 7%, and CMF, SIMULAT and SWAT more than 10%, and CMF up to 25%.

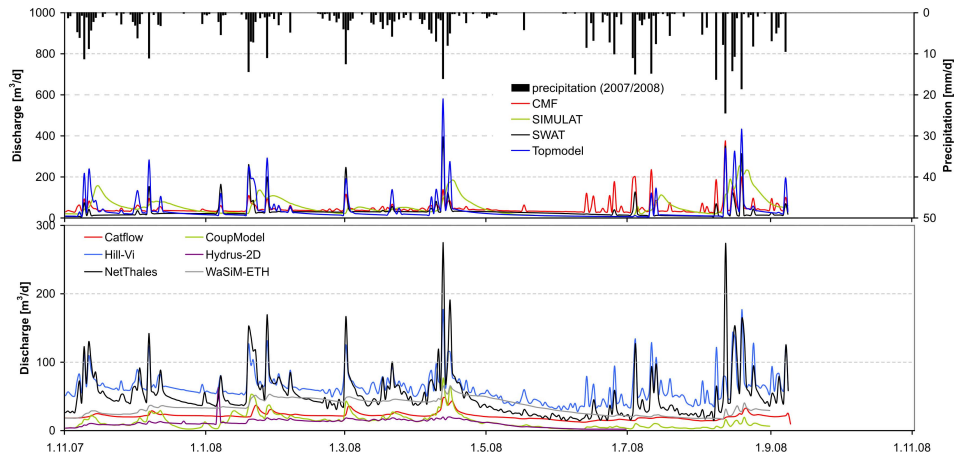


Fig. 4c. Predicted discharge for the hydrological year 2007/2008.

4.2 Discharge dynamics

The predicted discharge is illustrated in Fig. 4a–c for the three years. NetThales, SIMULAT and Hill-Vi produced a larger baseflow compared to the other models, that is 35, 25, and 50 m³/d, respectively. Hill-Vi used the Dupuit-Forchheimer assumption (Freeze and Cherry, 1979; Wigmosta and Lettenmaier, 1999) for saturated flow and a large K_{sat} of 90 mm/h. NetThales and SIMULAT used a K_{sat} of 50 and 75 mm/h, respectively. Catflow predicted a baseflow of 20 to 25 m³/d based on Richards equation using a large K_{sat} of 146 mm/h. SWAT and HYDRUS-2D showed a seasonally differing baseflow. SWAT predicted a winter baseflow of 5 m³/d, which increased up to 15 m³/d in spring. HYDRUS-2D consistently predicted a minimum baseflow of nearly zero in autumn and winter and a maximum in spring (10 to 20 m³/d). SWAT uses the Hooghoudt (1940) approach and a K_{sat} of 75 mm/h, whereas HYDRUS-2D the Richards equation and a K_{sat} of 54 mm/h. The other models (CoupModel, Topmodel, and WaSiM-ETH) predicted less than 10 m³/d baseflow. These three models use different flow equations (Hooghoudt (1940), time delay function, and linear storage approach, respectively) and a K_{sat} of 84, 58, and 118 mm/h, respectively. CMF predicted nearly no baseflow using Darcy’s law and the largest K_{sat} of 420 mm/h.

Figure 5 shows the discharge-frequency relationship. CMF, SWAT and Topmodel show the largest discharge rates, with CoupModel having the smallest and HYDRUS-2D somewhat higher rates. The sharpest reduction of Q_{max}/Q_{95} were predicted by CMF, SWAT and Topmodel, whereas in case of Catflow and Hill-Vi Q_{95} is only about half of Q_{max} . The range of the baseflow of all models is very narrow which show the very small difference between Q_{50} , Q_5 and Q_{min} . This also shows that all models estimate baseflow conditions during most of the year.

In all models except CoupModel, precipitation completely infiltrates into the soil. Catflow defined the discharge from

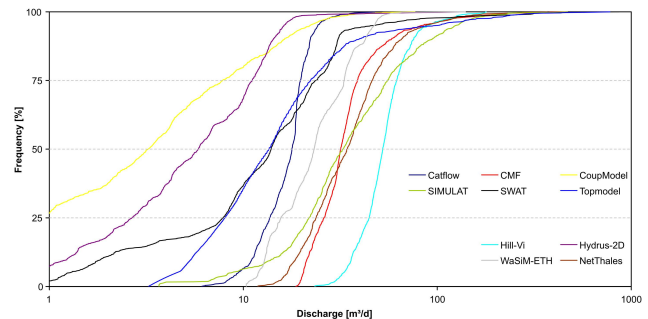


Fig. 5. Discharge-frequency relationship of the ten predictions.

the 0–100 cm as interflow assuming that the gullies are approximately 100 cm deep in the lower part of the slope into which the water enters. The lateral flow from 100–200 cm depth exiting the lower boundary of the catchment is defined as baseflow. The models with high subsurface flow routed more than 60% of the total discharge via baseflow (SIMULAT, Hill-Vi, and Catflow). SIMULAT does not calculate interflow because it is a single layer system. It only produces lateral flow in case an impermeable subsurface layer impedes vertical transport. NetThales does not make any distinction between baseflow and interflow. SWAT and Topmodel calculate about 40% surface and 60% subsurface flow.

NetThales and Topmodel predicted the most immediate and strongest response to precipitation. During intense spring or summer storms, their discharge often exceeded 400 m³/d, in a few cases even 800 m³/d (Fig. 4a–c), the latter being equivalent to about 12 mm/d on a catchment basis.

A strong response of up to 300 m³/d to precipitation events is predicted by SWAT and CMF but runoff is only simulated for very strong events. SIMULAT predicted also high discharges with a slow recession of up to one month after the strong events. Table 9a–c show that almost all of this discharge was simulated as baseflow. The discharge simulated

Table 9a. Discharge components predicted for the 1st year^a.

	runoff (mm/y)	interflow (mm/y)	baseflow (mm/y)	total discharge (mm/y)
Catflow		90	159	249
CMF				208
CoupModel	8		4	12
Hill-Vi	>1	305	306	
HYDRUS				34–48
NetThales				189
SIMULAT	>1	0	189	189
SWAT	27	51		76
Topmodel	31		63	94
WaSiM-ETH	0	83	24	107
Chicken Creek				113

Table 9b. Discharge components predicted for the 2nd year^a.

	runoff (mm/y)	interflow (mm/y)	baseflow (mm/y)	total discharge (mm/y)
Catflow		101	161	262
CMF				238
CoupModel	20		7	27
Hill-Vi	>1	346	346	
HYDRUS				19–67
NetThales				259
SIMULAT	>1	0	339	339
SWAT	61	84		145
Topmodel	75		96	171
WaSiM-ETH	2	138	22	162
Chicken Creek				105

by Hill-Vi during precipitation events was relatively slow compared to those of the other models and reached a maximum of 170 m³/d. HYDRUS-2D predicted some peak discharge rates in the 1st year but this model barely responded to the intensive events in the summer of the 2nd and 3rd year. Changing the *L*-factor (tortuosity) increased the response somewhat, but only negligibly compared to the much larger discharge of the other predictions. Catflow and CoupModel predicted the smallest response to the very strong summer events (Fig. 4a–c). CoupModel showed the lowest discharge of all models, whereas Catflow predicted mainly baseflow.

Predicted discharge of the other models is mainly interflow and baseflow. WaSiM-ETH and Hill-Vi are the only models which separate the discharge into all three components. Hill-Vi identified about 97% of the discharge as subsurface flow. WaSiM-ETH gave a similar result but with about 80% interflow, about 20% baseflow, and a very small amount of surface runoff. Although the hydraulic conductivity was larger than in Hill-Vi, most of the water did not reach groundwater table before it laterally discharged. Catflow predicted only interflow (40%) and baseflow (60%) using a higher hydraulic conductivity. Interflow was assumed to be released from the

Table 9c. Discharge components predicted for the 3rd year^a.

	runoff (mm/y)	interflow (mm/y)	baseflow (mm/y)	total discharge (mm/y)
Catflow		112	146	258
CMF				250
CoupModel	62		14	76
Hill-Vi	>1	329	329	
HYDRUS				34–64
NetThales				275
SIMULAT	>1	0	283	283
SWAT	57	112		164
Topmodel	94		104	198
WaSiM-ETH		148	30	178
Chicken Creek				113

^a no value is equal to no information.

upper 1 m of the soil so that it can enter the gullies. The clay dam developed a build-up of the groundwater table which resulted in groundwater discharge. SIMULAT quantifies interflow and baseflow, but interflow was not simulated at any time step. The clay dam had no influence on these predictions because the concentration time method does not consider any barrier. Figure 4a–c shows that the predicted subsurface flow of SIMULAT is baseflow given the long and slow recession of the discharge. CMF and NetThales did not provide information about the different discharge components.

The calculated direct runoff played a minor role for the total of the simulated discharge (Table 9a–c), seen in the 1st year, when no direct runoff was predicted at all. CoupModel produced the largest surface runoff in relative terms, about 80% of the total discharge because it simulated the second lowest total of discharge with a maximum direct runoff of 62 mm/y in the 3rd year. Topmodel simulated a larger direct runoff (95 mm/y) in this period, which was only about 40% of the predicted total discharge.

Although seven models included the clay dam into their model, the dam had a minor impact on the flow characteristics. CoupModel and CMF needed to allow a sand layer of at least 0.5 m for numerical reasons. HYDRUS-2D simulated its discharge caused by the clay wall but had numerical problems during some simulations due to saturation of grid cells near to the surface which would produce surface runoff. The main problem was that HYDRUS-2D is not able to handle surface runoff.

4.3 Groundwater levels

The observed groundwater dynamic is typical for Central Europe with a groundwater table rise after the winter period and a drawdown during the vegetation period, despite the sparse vegetation cover of the catchment (Fig. 6). Both observation wells were influenced by the clay dam. The water-table

fluctuations of the two neighbouring observation wells are closely linked.

Figure 6 illustrates the groundwater fluctuations at the observation wells F4 and L4 and the corresponding predictions of Catflow, CMF, Hill-Vi, HYDRUS-2D, and WaSiM-ETH. Observation wells F4 and L4 were chosen because they are located in the central part of the catchment (Fig. 6) and are also represented by the 2-D models (Catflow and HYDRUS-2D). The measured average groundwater level exhibited an increasing trend over the three years. This is also evident from the positive storage term in the water budget (Table 8a–c). Since there was no information on the initial soil-water contents, the soil-water storage was handled differently by the various modellers (see Sect. 4.7). The same applies to the groundwater storage. Surprisingly, none of the modelling groups used the information that initially there was no groundwater present.

The fluctuations of the groundwater level predicted at the two observation wells were similar. This indicates that K_{sat} at the two locations is similar (see the Supplement <http://www.hydrol-earth-syst-sci.net/13/2069/2009/hess-13-2069-2009-supplement.pdf>). The predicted groundwater tables did not show any influence of the clay dam. The groundwater fluctuations F4 and L4 predicted by CMF, Hill-Vi and WaSiM-ETH were fairly similar and showed small variations and no seasonal trend. CMF predicted a groundwater table drawdown of about 50 cm in the 1st year, a rise of 50 cm in the 2nd year and a nearly constant water table height in the 3rd year. Hill-Vi states a non-seasonal fluctuation of about 30 cm. WaSiM-ETH gave only a single average groundwater table height for the whole catchment. During the first year, the simulated average groundwater table height dropped by 50 cm and remained constant afterwards. A constant groundwater table height within a catchment throughout the year is the result of a balance between recharge and discharge at all times. All three models used K_{sat} . Hill-Vi predicted the highest discharge but used the lowest K_{sat} of the three models. It reported that the discharge was almost entirely subsurface flow but it did not provide direct information on the groundwater flow. The estimated initial groundwater table level was near the surface. WaSiM-ETH predicted the lowest baseflow of 22 to 30 mm/y and used the lowest K_{sat} of the three models. The total porosity of all three models was 0.38.

Catflow and HYDRUS-2D were the only models which showed a seasonal fluctuation of the groundwater table. Catflow showed a maximum amplitude of 80 cm with rapid changes. This is a consequence of the model structure because in these models a grid cell is either completely saturated (=groundwater) or not. The use of a cell thickness of 20 cm produced groundwater table jumps of 20 cm. The groundwater tables by HYDRUS-2D are calculated for six scenarios. The fluctuations of HYDRUS-2D are the largest of all models and exceeded the measured fluctuations. The amplitude was about 1 m and was constant throughout

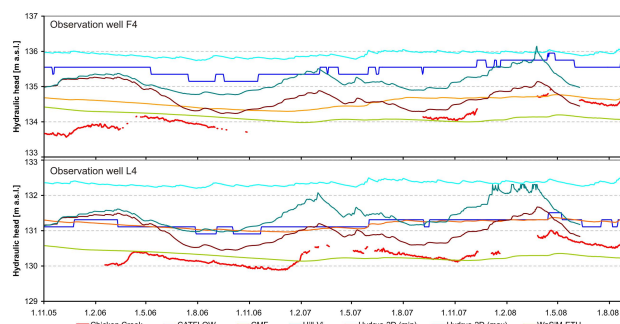


Fig. 6. Predicted and measured hydraulic heads at the observation wells F4 and L4.

the simulated period. The two scenarios by HYDRUS-2D (Fig. 6) were calculated with two different L -factors, the lower groundwater table being predicted using an L -factor of 0.5 and the higher for $L=-0.78$. Both scenarios have been started with the same initial groundwater level and developed differently during the 1st year. The fluctuated groundwater tables of HYDRUS-2D is nearly parallel to the 2nd and the 3rd year.

Catflow and HYDRUS-2D simulate the same fluctuation pattern. The difference in the amplitude is due to the different K_{sat} . Catflow assumed a K_{sat} , which is three times as large (146 mm/h) as in HYDRUS-2D (54 mm/h). Neither Catflow nor HYDRUS-2D predicted the sharp groundwater table rise toward the end of each winter period or the long and very slow drawdown during spring, summer, and fall months.

5 Discussion

5.1 Water budget

The errors in the measured mass balance, ΔM_{error} , were large. In the second year, the error was 40% of P . The large errors are due to the fact that the actual evapotranspiration AET was not measured but estimated according to Black et al. (1969; DVWK, 1996). This approach was developed for bare soils and neglects the effect of vegetation. Additionally, the influence of soil-water storage on AET is neglected. The error in the first year was mainly due to the neglected soil-water storage changes, whereas, the error in the last year was due to AET of a denser and taller vegetation.

The $\Delta M_{\text{error}} \approx 10\%$ for SWAT is due to the fact that SWAT was not designed for small catchments. Therefore, the representation of detailed processes within an artificial, newly constructed small catchment caused relatively large errors. CMF had the highest ΔM_{error} (up to 25%), probably because it is a recently developed code under construction.

Using the grass-referenced potential evapotranspiration, PET most likely overestimates the role of the sparse vegetation in the Chicken Creek, so PET is likely to be smaller

than presented in Table 8a–c. Although most models predicted a PET in the order of 600 to 800 mm/y, few values were surprisingly smaller, e.g. the 139 mm/y (CMF) and the 421 mm/y (NetThales, 2nd year, Table 8b), despite the fact that most groups used the Penman-Monteith method and that the information on vegetation coverage was available. Therefore, the differences originate from the parameter choice, in the case of CMF due to the constant PET-independent interception and the time-invariant sparse vegetation for all three years (LAI=0.1, plant height 10 cm). The models, which did not account for any vegetation, predicted both about 600 mm/y (HYDRUS-2D and Topmodel). The PET of SIMULAT (680 mm/y) using sparse vegetation and stomatal resistances of 50 s/m was nearly the same as that of WaSiM-ETH (700 mm/y), which only used evaporation and stomatal resistances of 150 s/m. Only the two models, which did not use the Penman-Monteith method, predicted a larger PET. SWAT calculated the highest PET using the Hargreaves equation and Hill-Vi the second highest (Turc equation). In SWAT a relatively well established plant cover (maximum LAI=2.68, plant height 50 cm) and the lowest stomatal resistance were assumed. The Turc equation, calculated by Hill-Vi yields a grass-referenced PET, excluding additional information about the vegetation. It overestimates the role of the vegetation and, therefore, PET.

Evidently, the plant parameterization was in this case, of minor importance because the catchment was left for a natural slow invasion of plants. The choice of the pore size index (van Genuchten parameter n_{vG}) had the main impact on estimating the actual evapotranspiration.

5.2 Soil parameterization

The range of predicted AET was, in all cases, a function of PET and of the soil-water status. Since the range of PET is similar for most models, the differences must originate from the differing soil-water content. The main inputs and parameters, beside precipitation (P), were the available field capacity, the hydraulic conductivity K_{sat} , and the α -parameter of the soil-water characteristics $\theta(h)$ and the unsaturated conductivity $K(\theta)$ function. The groups used different pedo-transfer functions (Table 5) to estimate the hydraulic soil parameters from soil texture.

The mean saturated hydraulic conductivity was predicted within a small range from 54 to 146 mm/h. Only CMF used a much larger K_{sat} (417 mm/h). The van Genuchten parameter n_{vG} varied from 1.13 to 2.28. The lowest value is used by WaSiM-ETH and introduces a small reduction of $K(S_e)$ on small changes in S_e . This leads to a larger water holding capacity in the top layer of the soil. Therefore, AET in WaSiM-ETH is considerably larger than in Catflow, which uses the largest n_{vG} -parameter. The low AET of HillVi is a consequence also of the parameterization and of the model structure. Due to the assumed saturated hydraulic conductivity and low L , only small amounts of water are stored in

the unsaturated zone, which reduces the water content dependent AET. The influence of the assumed K_{sat} can also be seen from the AET predicted by CMF. Infiltration sensitivity changes the level of the water table in Hill-Vi leaving only a shallow unsaturated zone, which results in a reduced AET. CoupModel used the second lowest n_{vG} and calculated the second highest AET. HYDRUS-2D predicted the largest AET using a low to intermediate n_{vG} . Different to other models, the groundwater flow is limited by the clay dam which results in a rising groundwater table near the dam. Therefore, the upper soil also has on average a higher soil moisture than in the other models. The changes of L from the standard value 0.5 to -0.78 resulted in a lower AET. NetThales calculated an AET just below the mean of all models using the smallest K_{sat} . NetThales neglects the vertical redistribution of water within the vertical soil column so that there is no unsaturated flow. The water holding capacity in NetThales corresponds to the available field capacity of the sand (0.11). SIMULAT and SWAT predicted an AET just above the mean. Their K_{sat} are very similar. SIMULAT uses an n_{vG} in the upper range.

The hydraulic soil properties affect discharge as well. The infiltration calculations (Table 3) are based on K_{sat} as limiting parameter. For example, Catflow predicted a large baseflow using a large K_{sat} and an intermediate n_{vG} . The chosen hydraulic conductivities of all models except CoupModel resulted in either no or little surface runoff and high infiltration, which leads mainly to interflow and baseflow. The saturated zone contributes mostly to baseflow. Models with a larger water holding capacity – due to a small pore size index n_{vG} – predicted large AETs and, therefore, lower baseflow because the water was primarily lost by evapotranspiration (e.g. CoupModel and HYDRUS-2D) and vice-versa (e.g. Hill-Vi and SIMULAT). In CoupModel, the water was stored for a long time in the upper soil compartment and resulted in a high AET and often also in surface runoff due to the saturation. Therefore, CoupModel predicted the highest amount of surface runoff and the second lowest total discharge.

In Topmodel, the partitioning between baseflow and surface runoff is mainly controlled by m , the shape factor of the recession curve, and the maximum root zone storage deficit. Contrary to the latter, m is often a very sensitive parameter. Errors in the estimation of m , therefore, severely affect the partitioning. It describes a catchment behaviour, which is difficult to relate to measurable catchment characteristics (other than discharge) and, therefore, it is not trivial to “a priori” estimate the value of m . It can be stated that the recession curve depends on porosity and storage capacity, so that overestimating the porosity and storage capacity would obviously result in underestimating surface runoff and vice-versa. In this study, the lack of a physical interpretation of m may be seen as a problem for applying Topmodel because there is no analogue for this kind of a catchment in the literature.

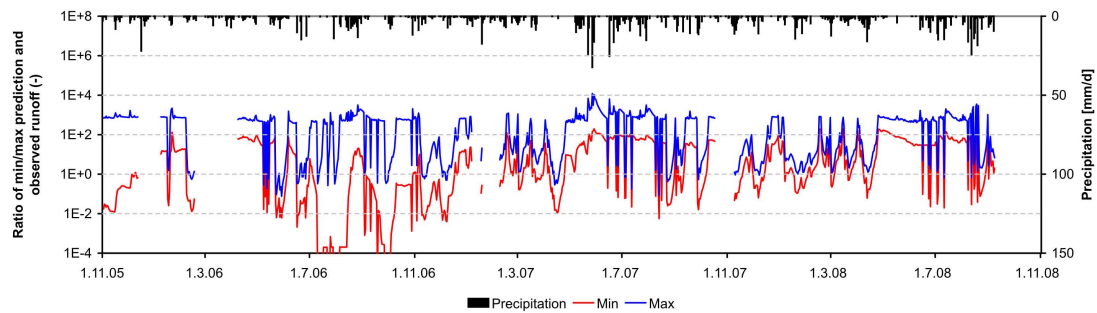


Fig. 7. Comparison of measured discharge to the maximum and minimum predicted discharge.

5.3 Discharge

NetThales did not simulate the vertical redistribution of water. All water infiltrated into the soil. Overland flow is only as a result of soil saturation. This results mainly in the sub-surface flow, which is equal to baseflow because water can only laterally be redistributed using a nonlinear kinematic flow routine.

Figure 7 shows the ratio of the maximum discharge predicted by the ten models and measured discharge as well as the ratio of minimum discharge and measured discharge. It is obvious that the predicted maxima as well as the minima of predicted baseflow (e.g. December 2005, April 2006, October 2007; about 300 and 80 times, respectively) are much larger than the observed baseflow. However, the maximum predicted discharge during the strong spring and summer events (e.g. 27 May 2007, 16 June 2007, 12 August 2008) was larger than measured (about 400 times). The models, which predicted low discharge at those times, underestimate the observed discharge. Only the models predicting the maximum discharge showed similar results to the observed during precipitation events with small intensities, but they predicted no more discharge than their baseflow. The minimum discharges underestimate the observed discharges by a factor of about 100. This also indicates that the predicted surface runoff is underestimated and baseflow overestimated.

Another important reason for the high discharges is the initial water content and groundwater level. The initial water content and the groundwater level were too high prior to the warm-up runs, resulting in a constant outflow of the catchment.

A process which was neglected by most modellers was snow melt and frozen soil conditions but it was evident in the first winter (9 January–7 February 2006). The four models (CoupModel, SIMULAT, SWAT, and WaSiM-ETH) which included the frozen conditions were using three different process-based methods (energy balance, temperature index, or degree day method). No model predicted the substantial discharge during the melt periods. The frozen soil period lasted 30 days and 18.7 mm precipitation occurred during that time. The maximum observed discharge was $55 \text{ m}^3/\text{d}$.

The following winter periods were warmer with shorter and more moderate soil frost.

5.4 Groundwater

The differences between observations and predictions in the subsurface storage were large in both the groundwater and unsaturated zone. Neglecting the given initial groundwater data and overestimated initial soil-water content resulted in the situation that none of the models predicted the observed soil- and groundwater storage, which in turn caused several problems and uncertainties in the predictions. Too high groundwater tables resulted in an overestimation of the evaporation in the first year. HYDRUS-2D, for instance, simulates the highest groundwater table and calculates the highest AET/PET ratios (67% and 95%) of all models. The capillary rise from the shallow and relatively wet unsaturated zone enhanced evapotranspiration (Table 8a–c). The low AET/PET ratio of about 20% (Hill-Vi) and the high groundwater table do not correlate. This is based on the fact that Hill-Vi always evaporates from the unsaturated zone independent from the depth to groundwater table. The AET predicted by Catflow was lower than in most other predictions. Catflow did not output PET but it must be at least in the range of 600–720 mm/y because the modeller used the Penman-Monteith equation, a dense vegetation, and the highest stomatal resistance. Therefore, neither the vegetation nor the capillary rise allows the water to evaporate, so that the AET calculated by Catflow must be lower compared to other model predictions. This is a surprising result especially since the roots are nearly as long as the distance down to the groundwater.

The groundwater tables of CMF and Catflow run almost parallel. At position F4 there is a difference of one meter, whereas the groundwater table at position L4 is almost the same. Since the calculation methods are the same and the hydraulic conductivity in both models is large, the groundwater discharge is fast with very low fluctuations.

WaSiM-ETH was the only model, which did not use the Richards and Darcy equations. The predicted groundwater fluctuations are similar to those predicted by Catflow and CMF but they used a lower hydraulic conductivity. This

indicates that the linear storage function is flexible enough to simulate the strong changes during recharge periods. All models have one feature in common: the amplitudes of the groundwater table fluctuations and, therefore, also the estimated porosity is not correct. Catflow, CMF, Hill-Vi, and WaSiM-ETH estimate the drainable porosity to be too large and HYDRUS-2D used a too low a drainable porosity.

5.5 Modeller experience

An important issue is the scientific background and the personal experience of the modellers. For instance, the CoupModel user used soil parameters from Swedish sands. The NetThales modellers concluded that the initially estimated runoff coefficient of 70% was too large, so they increased the effective root zone to make the runoff coefficient more realistic. Similarly, some results of the modellers indicate that apparently they had no prior experience with barren, sandy soils as they significantly overestimated AET. Also the underestimation of surface runoff was related to specific prior experience. A visual inspection of soil texture and erosion gullies from photos did provide evidence that surface runoff could be important, but only the CoupModel user accounted for it. It was obviously difficult to combine conflicting evidence like, for instance, large hydraulic conductivities and gully formation at the same time.

6 Lessons learned for future modelling studies in ungauged basins

While the comparative study mainly focused on the technical aspects of estimating discharge in the artificial catchment, the modelling group has indeed learned important lessons during the comparative assessment that have wider implications. When estimating runoff in an ungauged catchment, one has to decide on three main things: model structure, model parameters and model inputs (our boundary conditions) (Blöschl, 2005). Most previous model comparisons have attempted to identify the relative merits of alternative approaches to each of the three. Intercomparison studies in gauged catchments have generally focused on whether one model structure is superior to the other (e.g. Naef, 1981; Goodrich, 1990; Reed, 2004). In contrast to these studies, we found that the more complex models do not perform better than their simpler counterparts which were usually traced back to parameter identifiability problems.

Intercomparison studies for the ungauged catchment case have attempted to find optimum parameter estimation methods (Parajka et al., 2005; Oudin et al., 2008) including pedo-transfer functions (Wösten et al., 2001). These studies have never highlighted the role of the modeller; in fact have tried to treat the intercomparison as subjectively as possible as if the modellers were interchangeable. The results of this paper, in contrast, indicate that the modeller is an intrinsic part of the modelling study and has a major bearing on the model

results – even more so in ungauged catchments where there are more degrees of freedom in making modelling decisions. This is indeed a new paradigm in hydrological modelling. The question of *what method* to choose seems to be less important than the hydrological community has previously thought, and the question of *who* makes the choices seems to be more important. This study has shown that the modeller is just part of the entire modelling process, his/her prior experience in the field and his/her experience with the chosen model. Not only should this be recognised in comparative studies but, in fact, in any simulation exercise. This study has shown that it is not sufficient to choose the right modelling code, it is equally important to choose the modeller with the best intuition for the system to be modelled.

The second general lesson learned during this study was the role of dominant processes. The artificial catchment was chosen for the comparison in view of the overwhelming hydrological process complexity one usually encounters in catchments. The hope was that, in an artificial catchment, there will be minimum heterogeneity and processes will be simple, so modelling will be much easier than in catchments that have evolved over a longer time period. The results of the models were expected to converge, and this would be a good starting point to proceed for a model comparison in natural, more challenging catchments. This model comparison showed that we have been overly optimistic. Even a small artificial catchment exhibits heterogeneities which lead to similar modelling problems as in natural catchments. We did find, similar to natural catchments, that knowledge about the dominant processes is essential and can vastly improve the modelling results (Seibert and McDonnell, 2002). This knowledge was found to be very useful even if it is purely of a qualitative nature, as can be obtained by visual inspection of catchment photos to judge erosion marks and canopy features that might be indicative for surface runoff and root water uptake. Clearly, this type of information will help the modeller to make more informed choices on model structure and better guesses of model parameters. This may also help in the modelling strategies along the lines of the “dominant process concept” (DCP) to avoid overparameterisation by focusing on the most important processes occurring at the scale of the system to be modelled (Grayson and Blöschl, 2000b). Equally important, this finding highlights the need for field visits. While in the scientific literature on the problem of ungauged basins (PUB) field trips are usually downplayed, this study has made it very clear that, while a catchment may have no historic discharge data, there may be a wealth of qualitative information available that gives clues on the catchment dynamics (Blöschl, 2005). There is a lot that a hydrologist can gather by walking through a catchment and searching for complementary, qualitative data that can substantially contribute to process understanding and, hence, improve the model predictions (Fenicia et al., 2008). It is suggested that qualitative field evidence should play a more prominent role in hydrological modelling of ungauged basins.

The two lessons learned from this model/modeller inter-comparison are closely linked. The experience of a modeller is crucial in the (subjective) process of deciding upon the dominant processes that seem to be sufficiently important to be incorporated into the model. The cumulated experience will also play an important role in how different pieces of evidence from, for example, field inspections, will modify the initial conceptual understanding. One could say, in a Bayesian framework, that the different modellers differ in their prior estimates (on the conceptual understanding) causing also posterior estimates/perceptions to differ unless the new evidence (from field visits, for example) overrides the prior understanding.

7 Conclusions

Ten modelling groups used ten different catchment models to predict the major hydrological variables of the small artificial catchment Chicken Creek based on the same small data set. The observed discharge was not known to the modellers. This constellation of a minimal set of easily accessible data mimics the modeller's situation when confronted with predicting the response of ungauged catchments. The various modellers had a different background and were used to applying their models to catchments of quite different scales and features. This affected the choice of the optional model structures and parameters. For an "a priori prediction" the modeller's experience turned out to be very relevant. The discharge predictions differed in a wide parameter mainly due to a different process understanding, catchment conceptualization, and different parameter estimation.

The initial soil-water content was not available. Most of the models estimated it on the basis of pre-runs or by assuming a certain field capacity. Also, the initial groundwater situation was determined by pre-runs, despite the fact that it was part of the provided data. This influenced the predictions because the soil and groundwater compartments of the models were already full at the beginning of the simulation, contrary to what was observed in the catchment. Therefore, none of the models reproduced the steadily increasing groundwater table.

Catflow, HYDRUS-2D, WaSiM-ETH and, up to a certain degree, also SIMULAT and Hill-Vi are based on calculations using Richards equation. However, the predictions vary in a wide range. The largest differences were predicted in the case of the discharge with an extremely wide range in peak flow from 15 to 840 m³/d. This was mainly an effect of the estimated hydraulic soil properties. The models which predicted a low actual evapotranspiration predicted a higher runoff. These models mainly used a small pore size index (van Genuchten parameter n_{vG}). Surface runoff was also under-predicted by most models. The observations – not known to the modellers – show that surface runoff is, however, the main flow component. "A posteriori", it is ob-

vious that this process must have been dominant in absence of vegetation and on unconsolidated soils, which dry out in summer and freeze during winter. Hence, understanding the dominant processes appears to be essential for any "a priori" prediction of catchment response.

None of the models included the influence of the subsurface clay dam on how the lower catchment area was being drained. Predicted groundwater discharge was, therefore, larger than measured. Neglecting the aerial photo information of the gully network, resulted in underestimating direct runoff. Only CoupModel integrated this information and predicted the highest direct runoff. After visiting the catchment, the modellers immediately revised their view on the catchment processes occurring in this particular case. Hence, on-site information, even when it is purely of a qualitative nature, is very conclusive.

Choosing parameters involves a degree of guess-work and guessing is an art. The hydraulic conductivities and in particular the infiltration rates were too large, which overemphasized subsurface flow. The guessing-algorithms – pedotransfer functions or estimation by analogy to other (Swedish) soils – were less important relative to the assumptions of flow routing. The large infiltration capacities kept the soil-water storage on a high level, which resulted in a large actual evapotranspiration.

One important reason why the predictions suffered was from misjudging the initial conditions. Assuming an initially wet catchment instead of a dry one, resulted in wrong storage functions and too large discharge rates, or in other cases, in an overestimated actual evapotranspiration.

Some process assumptions were ineffective for reproducing the hydrological behaviour of the Chicken Creek catchment. The most obvious is the failure when using the energy balance, the temperature index, and the degree-day method to predict snow accumulation and soil frost, or the over-prediction of the potential evapotranspiration using the Turc method in case of a sparse vegetation.

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). The authors thank Vattenfall Europe Mining AG for providing the research site. We acknowledge the contribution by Ivo Strahm (Eawag) who carried out the HYDRUS-2D simulations. We explicitly acknowledge the positive role of the reviewers who challenged us to make the essence of this prediction exercise as visible as possible.

Edited by: J. Seibert

References

- Adhoc AG Boden: *Bodenkundliche Kartieranleitung*, 5th edn., Hannover, Germany, 438 p., 2005.
- AG Boden: *Bodenkundliche Kartieranleitung*, 4th edn., Hannover, Germany, 392 p., 1994.
- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration. Guidelines for computing crop water requirements, *Irrigation and Drainage Paper*, FAO, Rome, 300 pp., 1998.
- Alvenäs, G. and Jansson, P.-E.: Model for evaporation, moisture and temperature of bare soil: calibration and sensitivity analysis, *Agric. For. Met.*, 88, 47–56, 1997.
- 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.
- Barbour, S. L., Boese, C., and Stolte, B.: Water balance for reclamation covers on oil sands mining overburden piles, *Canadian Geotechnical Conference*, 313–319, 2001.
- Beven, K.: Changing ideas in hydrology – The case of physically-based models, *J. Hydrol.*, 105, 157–172, 1989.
- Beven, K., Lamb, R., Quinn, P., Romanowicz, R., and Freer, J.: Topmodel, in: *Computer Models of Watershed Hydrology*, Colorado, USA, 627–668, 1995.
- Beven, K. J. and Kirkby, M. J.: A physically based variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, 24, 43–69, 1979.
- Beven, K. J.: *Rainfall-runoff modelling: The Primer*, John Wiley & Sons, Chichester, 372 pp., 2001.
- Black, T. A., Gardner, W. R., and Thurtell, G. W.: The prediction of evaporation, drainage and soil-water storage for a bare soil., *Soil Sci. Soc. Amer. Proc.*, 33, 655–660, 1969.
- Blöschl, G.: Rainfall-runoff modelling of ungauged catchments, article 133, in: *Encyclopedia of Hydrological Sciences*, edited by: Anderson, M. G., John Wiley & Sons, Chichester, 2061–2080, 2005.
- Bormann, H.: *Hochskalieren von prozessorientierten Wassertransportmodellen – Methoden und Grenzen*, Reihe Geowissenschaften, Herbert-Utz-Verlag – Wissenschaft München, 164 pp., 2001.
- Bormann, H.: Sensitivity of a regionally applied soil vegetation atmosphere scheme to input data resolution and data classification, *J. Hydrol.*, 351, 154–169, 2008.
- Bronstert, A., Bárdossy, A., Bismuth, C., Buiteveld, H., Disse, M., Engel, H., Fritsch, U., Hundecha, Y., Lammersen, R., Niehoff, D., and Ritter, N.: Multi-scale modelling of land-use change and river training effects on floods in the Rhine basin, 2007.
- Brooks, R. H. and Corey, A. T.: Hydraulic properties of porous media, *Colorado State University, Fort Collins, Colorado*, 27 pp., 1964.
- Carsel, R. F. and Parrish, R. S.: Developing Joint Probability Distributions of soil-water Retention Characteristics, *Water Resour. Res.*, 24, 755–769, 1988.
- 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.
- Choi, H. T. and Beven, K.: Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of topmodel within the glue framework, *J. Hydrol.*, 332, 316–336, 2007.
- Diekkrüger, B. and Arning, M.: Simulation of water fluxes using different methods for estimating soil parameters, *Ecol. Model.*, 81, 83–95, 1995.
- DVWK: *Ermittlung der Verdunstung von Land- und Wasserflächen*, Merkblätter, Kommissionsbetrieb Wirtschafts- und Verlagswesen Gas und Wasser mbH, Bonn, 135 pp., 1996.
- Feddes, R. A., Kowalik, P. J., and Zaradny, H.: Simulation of field water use and crop yield, in: *Simulations Monograph*, Pudoc, Wageningen, 188 pp., 1978.
- Fenicia, F., McDonnell, J. J., and Savenije, H. H. G.: Learning from model improvement: On the contribution of complementary data to process understanding, *Water Resour. Res.*, 44, W06419, doi:10.1029/2007WR006386, 2008.
- Freeze, R. A. and Cherry, J. A.: *Groundwater*, Prentice-Hall, Englewood Cliffs, NJ, 604 pp., 1979.
- Gallart, F., Latron, J., Llorens, P., and Beven, K.: Using internal catchment information to reduce the uncertainty of discharge and baseflow predictions, *Adv. Water Resour.*, 30, 808–823, 2007.
- Gassmann, P. W., Reyes, M. R., Green, C. H., and Arnold, J. G.: The soil and water assessment tool: historical development, applications and future research directions, *T. ASAE*, 50, 1211–1250, 2007.
- 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, 2009, <http://www.hydrol-earth-syst-sci-discuss.net/6/1769/2009/>.
- Giertz, S., Diekkrüger, B., and Steup, G.: Physically-based modelling of hydrological processes in a tropical headwater catchment (West Africa) – process representation and multi-criteria validation, *Hydrol. Earth Syst. Sci.*, 10, 829–847, 2006, <http://www.hydrol-earth-syst-sci.net/10/829/2006/>.
- Goodrich, D. C.: Geometric simplification of a distributed rainfall-runoff model over a range of basin scales, Ph.D. Thesis, The University of Arizona, 361 pp., 1990.
- Grayson, R. B. and Blöschl, G.: *Spatial Patterns in Catchment Hydrology: Observations and Modelling*, Cambridge University Press, Cambridge, UK, 404 pp., 2000a.
- Grayson, R. B. and Blöschl, G.: Summary of pattern recognition and concluding remarks, in: *Spatial Patterns in Catchment Hydrology*, edited by: Grayson, R. B. and Blöschl, G., Cambridge University Press, Cambridge, UK, 355–367, 2000b.
- Gu, W.-Z. and Freer, J.: Patterns of surface and subsurface runoff generation, *IAHS Publications*, 229, 265–273, 1995.
- Gustafsson, D., Stähli, M., and Jansson, P.-E.: The surface energy balance of a snow cover: comparing measurements to two different simulation models, *Theor. Appl. Climatol.*, 70, 81–96, 2001.
- Hansen, D. P., Jakeman, A. J., Kendall, C., and Gu, W.-Z.: Identification of internal flow dynamics in two experimental catchments, *Math. Comput. Simulat.*, 43, 367–375, 1997.
- Hargreaves, G. L., Hargreaves, G. H., and Riley, J. P.: Agricultural benefits for Senegal River Basin, *J. Irrig. and Drain. Engr.*, 111, 113–124, 1985.
- Healy, R. W. and Cook, P. G.: Using groundwater levels to estimate recharge, *Hydrogeol. J.*, 10, 91–109, 2002.
- Hölzel, H. and Diekkrüger, B.: Hydrological analyses as a prerequisite for soil erosion modeling – Landscape related studies in a mesoscale hydrological catchment, in: *Landform – structure, evolution, process control*, International Symposium on Land-

- form, in press, 2008.
- Hooghoudt, S. B.: Bijdragen tot de kennis van enige natuurkundige grootheden van de grond, *Versl. Landb. Onderz.*, 42, 449–541, 1940.
- Jansson, P.-E. and Halldin, S.: Model for the annual water and energy flow in a layered soil, *Comparison of Forest and Energy Exchange Models*, Copenhagen, 145–163, 1979.
- 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, 2001.
- Jasper, K.: Hydrological Modelling of Alpine River Catchments using Output Variables from Atmospheric Models, ETH Zurich, 138 pp., 2005.
- Kendall, C., Mc Donnell, J. J., and Gu, W.-Z.: A look inside “black box” hydrograph separation models: a study at the Hydrohill catchment, *Hydrol. Process.*, 15(10), 1877–1902, 2001.
- Kraft, P., Vaché, K. B., Breuer, L., and Frede, H.-G.: A solute and water flux library for catchment models, *Proceedings of the iEMSs Fourth Biennial Meeting: International Congress on Environmental Modelling and Software Barcelona*, 2008.
- Kroes, J. G., van Dam, J. C., Groenendijk, P., Hendriks, R. F. A., and Jacobs, C. M. J.: SWAP version 3.2., Theory description and user manual, Alterra, Wageningen, 262 pp., 2008.
- Lindenmaier, F., Zehe, E., Dittfurth, A., and Ihringer, J.: Process identification at a slowmoving landslide in the Voralberg Alps, *Hydrol. Process.*, 19, 1635–1651, 2005.
- Lohammar, T., Larsson, S., Linder, S., and Falk, S. O.: FAST – simulation models of gaseous exchange in Scots pine, in: *Structure and Function of Northern Coniferous Forests – An Ecosystem Study*, edited by: Persson, T., *Ecological Bullentins Stockholm*, 505–523, 1980.
- Lundmark, A. and Jansson, P.-E.: Generic soil descriptions for modelling water and chloride dynamics in the unsaturated zone based on Swedish soils, *Geoderma*, 150, 85–95, 2009.
- Maurer, T.: Physikalisch begründete, zeitkontinuierliche Modellierung des Wassertransports in kleinen ländlichen Einzugsgebieten, Universität Karlsruhe, 1997.
- Meinzer, O. E.: The occurrence of groundwater in the United States with a discussion of principles, *US Geol. Surv. Water-Supply Paper*, 489, 321 pp., 1923.
- Monteith, J. L.: Evaporation and environment, in: *The Company of Biologists, The State and Movement of Water in Living Organisms*, 19th Symp. Soc. Exp. Biol., Cambridge, UK, 205–234, 1965.
- Monteith, J. L. and Unsworth, M. H.: *Principles of environmental physics*, edited by: Arnold, S. E., London, UK, 291 pp., 1990.
- Mualem, Y.: A new model for predicting the hydraulic conductivity of unsaturated porous media, *Water Resour. Res.*, 12, 513–522, 1976.
- Naef, F.: Can we model the rainfall-runoff today? *Hydrological Sciences Bulletin*, 26, 281–289, 1981.
- Nenov, R.: Determination of the evapotranspiration from the artificial catchment Hühnerwasser, Chair of Hydrology and Water Management, Brandenburg University of Technology, Cottbus, 100 pp., 2009.
- Nicolau, J.: Runoff generation and routing on artificial slopes in a Mediterranean continental environment, *Hydrol. Process.*, 16, 631–647, 2002.
- Niehoff, D., Fritsch, U., and Bronstert, A.: Land-use impacts on storm-runoff generation: scenarios of land-use change and simulation of hydrological response in a meso-scale catchment in SW-Germany, *J. Hydrol.*, 267, 80–93, 2002.
- Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments, *Water Resour. Res.*, 44, W03413, doi:10.1029/2007WR006240, 2008.
- Parajka J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, *Hydrol. Earth Syst. Sc.*, 9, 157–171, 2005.
- Penman, H. L.: Natural evaporation from open water, bare soil and grass. *Proc. Roy. Meteorol. Soc. A*, 193, 120–145, 1948.
- Peschke, G.: Moisture and Runoff Components from a Physically Founded Approach, *Acta Hydrophys.*, 31, 191–205, 1987.
- Plate, E. and Zehe, E.: *Hydrologie und Stoffdynamik kleiner Einzugsgebiete: Prozesse und Modelle*. Schweizerbart, 366 pp., 2008.
- Rawls, W. J. and Brakensiek, D. L.: Prediction of soil-water properties for hydrologic modeling, in: *Proceedings of Symposium on Watershed Management*, ASCE, 293–299, 1985.
- Richter, D.: Zur einheitlichen Berechnung der Wassertemperatur und der Verdunstung von freien Wasserflächen auf statistischer Grundlage, *Abh. Meteor. Dienst der DDR*, 16, 35 pp., 1977.
- Ritchie, J. T.: A model for predicting evaporation from a row crop with incomplete cover, *Water Resour. Res.*, 8, 1204–1213, 1972.
- Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., and Seo, D. J.: Overall distributed model intercomparison project results, *J. Hydrol.*, 298, 27–60, 2004.
- Romano, N. and Santini, A.: Water retention and storage: Field, in: *Methods of Soil Analysis*, edited by: Topp, J. H. D. a. G. C., *SSSA Book Series No. 5*, Madison, WI, USA, 721–738, 2002.
- Saxton, K. E., Rawls, W., Romberger, J., and Papendick, R.: Estimating generalized soil-water characteristics from texture, 1031–1035, 1986.
- Schaap, M. G., Leij, F. J., and van Genuchten, M. T.: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions, *J. Hydrol.*, 251, 163–176, 2001.
- Schulla, J. and Jasper, K.: *Model Description WaSiM-ETH*, ETH Zrich, Zrich, 181 pp., 2007.
- Seibert, J. and McDonnell, J. J.: On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, 38(11), 1241, doi:10.1029/2001WR000978, 2002.
- Simunek, J., Sejna, M., and van Genuchten, M. T.: *The HYDRUS2 Code for Simulating the Two-Dimensional Movement of Water, Heat, and Multiple Solute in Variably-Saturated Porous Media*, edited by: Service, U. S. S. L. A. R., US Department of Agriculture, Riverside, California, USA, 251 pp., 1999.
- Sivapalan, M., Takeuchi, K., Franks, S., Gupta, V., Karambiri, H., Lakshmi, V., Liang, X., McDonnell, J., Mendiondo, E., O’Connell, P., Oki, T., Pomeroy, J., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences, *Hydrolog. Sci. J.*, 48, 857–880, 2003.
- Smith, R. E. and Parlange, J. Y.: A parameter-efficient hydrologic infiltration model, *Water Resour. Res.*, 14, 533–538, 1978.
- Stähli, M., Jansson, P.-E., and Lundin, L.-C.: Preferential water flow in a frozen soil – a two-domain model approach, *Hydrol.*

- Process., 10, 1305–1316, 1996.
- Stähli, M. and Gustafsson, D.: Long-term investigations of the snow cover in a subalpine semi-forested catchment, *Hydrol. Process.*, 20, 411–428, 2006.
- Turc, L.: Évaluation des besoins en eau irrigation, l'évapotranspiration potentielle, *Ann. Agron.*, 12, 13–49, 1961.
- Vaché, K. and McDonnell, J. J.: A process-based rejectionist framework for evaluating catchment runoff model structure, *Water Resour. Res.*, 42, W02409, doi:10.1029/2005WR004247, 2006.
- van Dam, J. C., Huygen, J., Wesseling, J. G., Feddes, R. A., Kabat, P., van Walsum, P. E. V., Groenendijk, P., and van Diepen, C. A.: Theory of SWAP version 2.0. Simulation of water flow, solute transport and plant growth in the Soil-Water-Atmosphere-Plant environment, Wageningen University, Wageningen, 165 pp., 1997.
- van Genuchten, M. T.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sciences Society of America*, 44, 892–898, 1980.
- Weiler, M. and McDonnell, J. J.: Virtual experiments: A new approach for improving process conceptualisation in hillslope hydrology, *J. Hydrol.*, 285, 3–18, 2004.
- Weiler, M. and McDonnell, J. J.: Testing nutrient flushing hypotheses at the hillslope scale: A virtual experiment approach, *J. Hydrol.*, 319, 339–356, 2006.
- Weiler, M. and McDonnell, J. J.: Conceptualizing lateral preferential flow and flow networks and simulating the effects on gauged and ungauged hillslopes, *Water Resour. Res.*, 43, W03403, doi:10.1029/2006WR004867, 2007.
- Wigmosta, M., Vail, L., and Lettenmaier, D. P.: Distributed hydrology-vegetation model for complex terrain, *Water Resour. Res.*, 30, 1665–1679, 1994.
- Wigmosta, M. and Lettenmaier, D. P.: A comparison of simplified methods for routing topographically driven subsurface flow, *Water Resour. Res.*, 35, 255–264, 1999.
- 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, 2001.
- Zehe, E. and Flüßler, H.: Preferential transport of isoproturon at a plot scale and a field scale tile-drained site, *J. Hydrol.*, 247, 100–115, 2001a.
- Zehe, E. and Flüßler, H.: Slope scale variation of flow patterns in soil profiles, *J. Hydrol.*, 247, 116–132, 2001b.
- Zehe, E. and Bloeschl, G.: Predictability of hydrologic response at the plot and catchment scales: Role of initial conditions, *Water Resour. Res.*, 40, W10202, doi:10.1029/2003WR002869, 2004.
- Zehe, E., Becker, R., Bárdossy, A., and Plate, E.: Uncertainty of simulated catchment runoff response in the presence of threshold processes: Role of initial soil moisture and precipitation, *J. Hydrol.*, 315, 183–202, 2005.