Joint application of event-based calibration and dynamic identifiability analysis in rainfall–runoff modelling: implications for model parametrisation

Johannes Cullmann and Gunter Wriedt

ABSTRACT

The goal of this work is to comparatively evaluate the potential of both event-based automatic calibration and Dynamic Identifiability Analysis (DYNIA) as proposed by Wagener *et al.* This is combined with an investigation on a potential relation of *a priori* knowledge on event characteristics with optimal model parameters.

A joint application of DYNIA and automatic parameter estimation leads to implications considering the informational content of both methods. Optimal model parameters, identified on an event basis, are tested for statistical relations with physical characteristics of the rainstorm events (e.g. intensity). In this paper, we present results of a modelling study in the Rietholzbach catchment (Switzerland). We employed the hydrological model WaSiM-ETH, using a combined DYNIA (Dynamic Identifiability Analysis) and automatic parameter estimation (PEST) approach to investigate best parameter sets as well as parameter variability along the time series of the hydrograph. The results of the study indicate that the "drainage" parameter identifiability for the long-term simulation is linked to the event-based calibrated parameter for the flood events. However, the parameter sets obtained with single-event calibration could not be fully linked to the chosen characteristic features derived from the precipitation forecast.

Key words | *a priori* information, automatic calibration, identifiability of parameters, parameter optimisation

Johannes Cullmann (corresponding author)
German IHP/HWRP Secretariat,
Am Mainzer Tor 1,
56086 Koblenz,
Germany
E-mail: *cullmann@bafg.de*

Gunter Wriedt
Institute for Envoronment and Sustainability,
Joint Research Centre,
21020 Ispra,
Italy

INTRODUCTION

Generally, any model structure is merely an approximation of the portrayed natural process. Amongst others, this is one important motivation for calibrating model parameters when modelling a specific catchment. Applicability of simple conceptual models is limited to conditions represented by the data used for the calibration of their parameters. Process-oriented models are supposed to maintain system dynamics even beyond the range of calibration data. Parametrisation of hydrological models has been the subject of enormous scientific effort throughout the last decades. One main focus was scaling in space, i.e. finding more or less homogeneous regions to be portrayed by parameter ranges determined either by *a priori* knowledge or by calibration (Gupta *et al.* 1994; Gupta & Dawdy 1995; Post & Jakeman 1996). The results of these studies made a valuable contribution to our general understanding of parametrisation of both conceptual and process models.

However, experience in rainfall–runoff modelling resulted in the awareness that various best parameter sets apply to single events or periods of time (Beven & Freer 2001). For continuous simulations, the optimum parameter sets may not only be different for separate periods within the investigated series, they may also change in time concurrently to variations in boundary conditions and process characteristics. The event-based or subset-specific

doi: 10.2166/hydro.2008.055

variance of best model parameter sets may result from uncertainty of input data, observation data and equifinality of the system (Beven & Binley 1992). However, these parameter changes may also result from systematic changes of system behaviour, revealing model structural errors caused by inadequate process representation. We hypothesise that systematic relations of parameter optima and state variables exist in this case. It is therefore necessary to investigate the transient characteristics of best parameter sets and to link systematic changes to corresponding state indicators. *A priori* determination of the appropriate model parameters in relation to these transient dominant process controls has not been sufficiently addressed by the aforementioned studies and could be very relevant on the way towards a more successful hydrological modelling approach. This becomes clear from Merz & Blöschl (2003): on the basis of indicators such as timing, storm duration, rainfall depth, catchment state, etc., they divided floods into five classes, where different flood formation mechanisms (forces and runoff-generating processes) cause varying system response characteristics.

The DYNIA (Dynamic Identifiability Analysis) method presented by Wagener *et al.* (2003) is an approach to find suitable model structures and/or parameters in the context of transient dominant process controls and parameter uncertainty. Rearrangement of the time series data by state variables may reveal underlying relations explaining the temporal changes. DYNIA has been applied to simple conceptual models. Yet few studies have been published focusing on process-oriented, distributed models. Wriedt & Rode (2006) applied the DYNIA methodology to simulated runoff time series in a subcatchment of the Weisse Elster river, generated with the WaSiM-ETH model. They found a clear relation of the interflow parameter *dr* to observed discharge, suggesting state dependence of the parameter to catchment wetness or soil moisture state.

Cullmann *et al.* (2006) found evidence that characteristic criteria of the catchment preconditions–represented by a pre-event rainfall index–are correlated to runoff generation and runoff concentration parameters in the WaSiM-ETH model. They investigated a study area with a rather poor input data situation and were not able to clearly identify the source of the parameter dependence. The two studies indicate that relations between best parameter sets and specific characteristic criteria describe the transient character of optimal model parametrisation. However, more research is needed from a different and well-observed catchment in order to further understanding of the relations between best parameter sets and preconditions.

The objective of this work is to identify factors controlling event-based optimum parameter sets and temporal evolution of model parameters. This is done with the intention to link best parameters to *a priori* knowledge in order to make a first step towards flexible model parameters, which adapt to the governing process and the dominating preconditions reflected by the chosen characteristic criteria. We use a combined approach of event-based model calibration and dynamic identifiability analysis based on consistent data from the well-observed Rietholzbach catchment. As the application has exemplary character, the analysis was restricted to selected model parameters, which are well suited to show the potential of the combination of the two methods, albeit the parameters used in this study do not describe the full complexity of the model. The event-based calibration results in a number of optimal parameter sets, which are statistically analysed (see the third section) in the context of criteria describing the event characteristics. The dynamic identifiability analysis allows us to trace temporal evolution of optimum model parameters and to relate parameter changes to state variables (see the third section). The joint application of both methodologies offers an additional means of evaluating model parameter identifiability in the context of specific event characteristics reflected by the results of automatic calibration, i.e. the more general DYNIA method can be assessed on event scale by parameter estimation.

MATERIAL AND METHODS

Study area, data and model

The Rietholzbach drains a 3.18 km^2 hilly pre-alpine watershed with an average precipitation of 1,600 mm per year, generating a mean annual runoff of 1,046 mm. It is located in north-eastern Switzerland, in the centre of the Thur basin, with elevations ranging from 681 to 938 m a.s.l.

The land use mainly consists of pasture (73%), the rest being dominated by forest (23%) including a few settlement areas. The various soil types range from gley soils to more permeable brown soils and regosols with relatively large soil water storage capacities. The catchment is equipped with a meteorological station, continuous time domain reflectometry (TDR) soil moisture measurements, a lysimeter and a well-defined runoff profile at Rietholzbach gauging station. Datasets for the meteorological input parameters (temperature, humidity, wind, global radiation and precipitation) as well as for soil moisture at a single location (four different TDR probes) and runoff at the catchment outlet were available for the period 1981–1999. For hydrological modelling of the catchment we used the process-oriented model WaSiM-ETH (Schulla 1997; Schulla & Jasper 2001; Gurtz *et al.* 2003). This model serves as a basis for automatic calibration of up to three parameters affecting the runoff generation and propagation. The DYNIA analysis described later is carried out with the same model and parameter set.

Our investigations focused on three model parameters: *kd, ki* and *dr* (Table 1). *dr* [m^{-1}] is a so-called drainage density factor, which is a measure of lateral subsurface conductivity and controls the amount of interflow that can be generated by the model. Once the interflow is generated, the result of Equation (2) is used as input for the calculation of the recession according to Equation (3). *kd* [h] is controlling the recession constant of the direct runoff component, while *ki* [h] is the recession constant of the interflow component. Recession of the specific runoff components is described using an exponential equation, corresponding to a linear reservoir model. These parameters are catchment-specific and, according to previous analyses (Cullmann 2007), are the most sensitive parameters controlling runoff characteristics of the study area.

Table 1 | Parameters studied

Parameter	Description	Reference	Range
kd [h]	Recession constant for direct runoff	4	1–20
ki [h]	Recession constant for interflow	8	1–20
dr [m^{-1}]	"Drainage density", controls interflow volume	6	1–20

The implementation of these parameters in the model is given by the following equations:

$$\text{Direct runoff recession}(kd): \quad q_{\mathrm{d}}(t) = q_{\mathrm{d}}(t-1) \cdot e^{-(\Delta t/kd)} \tag{1}$$

Interflow generation (dr) :

$$q_{\mathrm{ifl}} = \mathrm{Min}\left\{ \begin{array}{c} (\Theta(\Psi) - \Theta_{\Psi=3.45}) \cdot \frac{\Delta z}{\Delta t} \\ K_s(\Theta) \cdot \Delta z \cdot dr \cdot \tan\beta \end{array} \right\} \tag{2}$$

$$\text{Interflow recession } (ki): \quad q_{\mathrm{ifl}}(t) = q_{\mathrm{ifl}}(t-1) \cdot e^{-(\Delta t/ki)} \tag{3}$$

where q_{d} = direct runoff, q_{ifl} = interflow, Θ = soil moisture content in actual layer, ψ = suction power, β = slope, K_s = saturated conductivity, Δz = thickness of soil layer and Δt = time step. For interflow generation, the two values are computed and the smaller is used for calculation.

Monte Carlo simulation and modified dynamic identifiability analysis

The DYNIA algorithm and modifications

Wagener *et al.* (2003) developed the DYNIA algorithm as an extension of the regionalised sensitivity analysis (Spear & Hornberger 1980; Hornberger & Spear 1981). The analysis is a tool to evaluate the results of Monte Carlo simulations. The basic idea is to calculate the probability distribution for individual model parameters for each simulation time step within a specified time frame (moving window). The results are visualised in a 2D plot of parameter values vs. time, where the parameter probability density is shaded in a grey scale. Only the best 10% of simulations, according to an appropriate support measure (here: sum of absolute errors, SAE), are included in the analysis.

The main results calculated with DYNIA are as follows:

- the probability density function of each model parameter for each time step;
- the 95% and 5% confidence limit of the probability density function;
- information content, which is defined as

$$I(t) = 1 - \frac{C_{0.95}(t) - C_{0.05}(t)}{\Delta P}$$

where $I(t)$ = information content at time t, ΔP = parameter range, and $C_{0.95}$ and $C_{0.05}$ are the parameter values of the 95% and 5% confidence limits of the parameter density function, respectively. The information content takes values between 0 and 1, where high values indicate a small confidence interval relative to the parameter range. Small confidence ranges (in relation to parameter range) express a high identifiability (or low uncertainty) of the individual parameters. All these results are given for each time step of the moving window procedure; thus it is possible to evaluate temporal changes in the criteria.

Runoff characteristics depend on catchment wetness state and antecedent rainfall intensity, including threshold or nonlinear behaviour. This leads us to the awareness that different model parameter values describe different system states. To assess such state-dependent changes of optimum model parameters, a modified DYNIA approach was developed.

In this modification the data are re-ordered against an observed state variable, instead of using the original time series prior to executing the DYNIA algorithm.

The original DYNIA approach interprets optimum model parameters as time series. In contrast to this, the modified approach allows for easily visualising optimum model parameters as dependents of state variables or other system characteristics. This makes it easy to quickly assess if discharge is related to, for example, soil moisture or rainfall intensity.

In the reordered series, neighbouring data is no longer in a temporal context. If enough data is included, the effects of individual events will be filtered out and only state-dependent relations remain, indicating that inter-event variability is not higher than possible state-dependent effects.

Simulation runs and analysis

Based on the calibrated WaSiM-ETH model for the Rietholzbach catchment, a sensitivity analysis and a Monte Carlo analysis are carried out. The simulations cover the calendar year 1994 on an hourly time step. DYNIA requires a continuous simulation. The number of Monte Carlo simulations and the high temporal resolution of a single simulation results in a high computation time, restricting the total simulation period. Therefore the simulations were carried out for the calendar year 1994 only, evaluating data from 01.04.1994–31.10.1994, excluding winter events affected by snow melt processes.

A simple sensitivity analysis (Table 2) was calculated for the parameters *dr*, *ki* and *kd* based on the 10% elasticity index defined according to

$$e_{10} = \frac{O_1 - O_0}{1.1P_0 - P_0} \cdot \frac{P_0}{O_0} = \frac{O_1 - O_0}{0.1 \cdot O_0}$$

where P_0 = parameter value, O_0 = model output for parameter P_0 and O_1 = model output for changed parameter $P_1 = 1.1P_0$.

The elasticity index is a measure of the change in model output, when changing a parameter by 10%. In the above notation, the index is normalised to correct for different units and value ranges of model output and parameter. It thus gives the relative change of the model output related to a change of the parameter by 10%.

The Monte Carlo simulation was carried out including all three parameters. 1,000 parameter sets were sampled from the parameter space using a Latin Hypercube approach. A list of parameters and the associated parameter range is given in Table 1. The DYNIA results of discharge are reordered against observed rainfall, observed and simulated soil moisture, a 10 h discharge gradient and a 3 h precipitation gradient. Where the 10 h discharge gradient is the change of discharge within a 10 h interval, the 3 h precipitation gradient is the change of precipitation within a 3 h interval.

A window size of 101 h was applied in the DYNIA analysis. A prior analysis was carried out to assess the effect of different window sizes on DYNIA results. Using the original DYNIA approach based on a time series of data, the size of the moving window may mask time-dependant

Table 2 | Local sensitivities of total runoff and the WaSiM calculated runoff components: base flow, interflow and direct runoff to the model parameters

Parameter	Local sensitivities given as normalised 10% elasticity index e_{10}			
	Total flow	Base flow	Interflow	Direct runoff
kd	0.0063	0	0	0.7154
ki	0.0714	0	0.0750	0
dr	0.2065	0.3313	0.2122	0.2349

parameter changes. If the window is too large, short term dynamics cannot be resolved; a small window, however, may emphasize short-term dynamics while masking slower dynamics. Using a state-dependent re-ordered time series, we no longer investigate temporal process dynamics and a larger window size is needed to smooth out irregular effects caused by the sorting, as adjacent data points originate from different, non-continuous time steps. On the other hand, reducing the window size would lead to an increased uncertainty in the DYNIA results.

For general characterisation of model fit we calculated different objective functions (sum of absolute errors, SAE, Nash–Sutcliffe index, NSC, and index of agreement, IoA). The objective functions are given by the following equations, where x represents the modelled flow, $\bar{y}$ is the average observed flow and y stands for observed flow:

$$\mathrm{SAE} = \sum_{1}^{t} |x - y|$$

$$\mathrm{NSC}^{*} = 1 - \frac{\sum_1^t (y - x)^2}{\sum_1^t (y - \bar{y})^2}$$

$$\mathrm{IoA}^{*} = 1 - \frac{\sum_1^t (x - y)^2}{\sum_1^t \left(\sqrt{(x - \bar{y})^2} + \sqrt{(y - \bar{y})^2}\right)^2}$$

Both IoA^* and NSC^* are subtracted from 1 in the presented work, leading to $\mathrm{NSC} = 1 - \mathrm{NSC}^*$ and $\mathrm{IoA} = 1 - \mathrm{IoA}^*$. The idea of this notation is to reverse the range of NSC from $[1, -\infty]$ to $[0, \infty]$ and IoA from [1, 0] to [0, 1], making 0 the optimum. This was comfortable in our case, because the DYNIA analysis we employed addresses the minimum value of the objective function to be the best model approximation.

Before applying the DYNIA algorithm, an overall sensitivity and uncertainty screening was made following the GLUE methodology (Beven & Binley 1992). Objective functions (SAE, NSC, IoA) were calculated for the complete time series of each simulation (as opposed to a specified time window in the DYNIA approach) and the dot plots were analysed. These dot plots are scatter plots of the objective function of each simulation against parameter value, one plot for each parameter. They can give a first indication if calibrated parameters refer to global optima or if model parameters are highly or poorly identifiable.

Event-based parameter estimation

For the calibration study the 36 largest rainstorm events have been chosen from the period 1981–1999. The events are characterised by seven criteria describing the precondition and the characteristic rainfall of each event: absolute length of the rainfall event ("Duration"); peak rainfall intensity ("Peak"); absolute volume of the rainstorm ("Volume"); a combined parameter where the peak rainfall intensity is divided by the absolute volume ("Form"); the time to rainfall peak divided by the peak intensity ("TI"), describing the initial phase of the rainfall event, mean precipitation intensity of the event ("MPI") and an exponential function describing the 14-day pre-event rainfall ("PF"). This criteria is calculated on the basis of the last 14 recorded values of flow y according to

$$\mathrm{PF} = \sum_{i=1}^{14} y_i e^{-(i/\tau)}$$

The variable τ is set so that the most recent time steps mainly account for PF in our case. The mean, minimum and maximum values of the criteria for all events are given in Table 3. The model was calibrated automatically using the PEST (parameter estimation) software package (Skahill & Doherty 2006).

Methodology

Two approaches for parameter estimation are implemented in PEST; both methods try to minimise an objective function that is represented by least squares. The standard method uses the Gauss–Marquardt–Levenberg (GML) algorithm. This nonlinear parameter estimation algorithm is fast and stable as it switches between the steepest gradient search and the Gauss–Newton approach, depending on a

Table 3 | Mean, minimum and maximum criteria of the events used in the calibration study

	Duration [h]	Peak [mm/h]	Volume [mm]	Form [h^{-1}]	TI [mm]	PF [mm]	MPI [mm h^{-1}]
Mean	14.2	10.6	41.7	0.3	0.9	43.1	3.7
Min	2.0	2.3	12.7	0.1	0.1	11.8	1.3
Max	30.0	32.8	93.5	0.9	6.1	97.4	8.4

scalar of the identity matrix (details in Mohamed & Walsh 1986). The drawback of this method is that the algorithm might converge to local minima, depending on the error response surface and the start values for the optimisation run. The second method supplied with the PEST suite is the global SCEUA (shuffled complex evolution–University of Arizona) search algorithm (Duan *et al.* 1992). This algorithm is capable of reducing the risk of "getting stuck" in a local minimum of the objective function. It is more costly in terms of machine time.

Simulation set-up

The PEST set-up for automatic calibration is event-based. WaSiM-ETH is calibrated for each of the 36 events independently, allowing at least three preceding months for model warm-up.

In order to check reliable convergence of the GML algorithm to the global minimum, we tested the GML method with various start values and compared them with results of the SCEUA method. This test was performed with data of a single event which we chose because its seven characteristic parameters are near to the mean of the overall considered event characteristics shown in Table 3.

The results shown in Table 4 confirm the stable optimisation capability of both SCEUA and GML methods for the parameters considered with the chosen start values used and the parameter ranges covered in this study. The difference between optimised parameters is marginal; therefore GML was chosen for further use in this study, as GML is significantly faster than SCEUA. Figure 1 supports this approach: the error surface for a wide parameter range shows a relatively smooth behaviour, indicating that GML is appropriate for parameter optimisation.

RESULTS

DYNIA

Sensitivity and uncertainty analyses

The local sensitivity analysis (Table 2) revealed that *dr* is the most sensitive of the selected parameters. It affects total runoff as well as the different flow components: base flow, interflow and direct runoff. The effect of *ki* on total runoff and interflow was only one-third of that of *dr*. The parameter *kd* controls direct runoff, but the sensitivity of total flow to this parameter is only marginal.

For the sum of absolute errors, the dot plots (Figure 2) suggest optimum parameter values of $kd \sim 15$, $ki \sim 1$ and $dr \sim 3$. Assuming a threshold of SAE = 200 for a "good" simulation, parameter uncertainty is very large, especially for *kd* and *ki*, including almost the entire parameter range. The parameter *dr* is most clearly defined. Except for *dr*, the reference parameter values (Table 1) differ from the optimum values of the dot plots. This also holds, less clearly, when we look at results evaluated with the NSC criterion. If IoA is analysed, the picture is quite undifferentiated (Figure 2) and does not allow for any conclusions; in further discussions we refer only to SAE. Differences between reference parameters (Table 1) and optimised parameters may result from the calibration approach, including manual procedures, visual interpretation and other objective functions. The parameter values of *kd* and *dr* are well within the *a priori* range. The parameter optimum for *ki* is found near the lower parameter

Table 4 | Test of convergence for the optimisation strategy (parameter estimation method used in the study is shaded)

Method	Start value			Optimisation parameter range			Optimised parameter		
	kd	*ki*	*dr*	*kd*	*ki*	*dr*	*kd*	*ki*	*dr*
SCEUA	*	*	*	1–40	1–40	1–40	3.84	9.73	6.11
SCEUA	*	*	*	$1-10^7$	$1-10^7$	$1-10^7$	3.84	9.73	6.11
GML	2	8	4	$1-10^7$	$1-10^7$	$1-10^7$	3.89	9.72	6.11
GML	2	6	4	1–60	1–40	1–80	3.89	9.72	6.11
GML	35.9	27.3	29.9	1–40	1–40	1–40	3.89	9.72	6.11

*70 random points in parameter space.

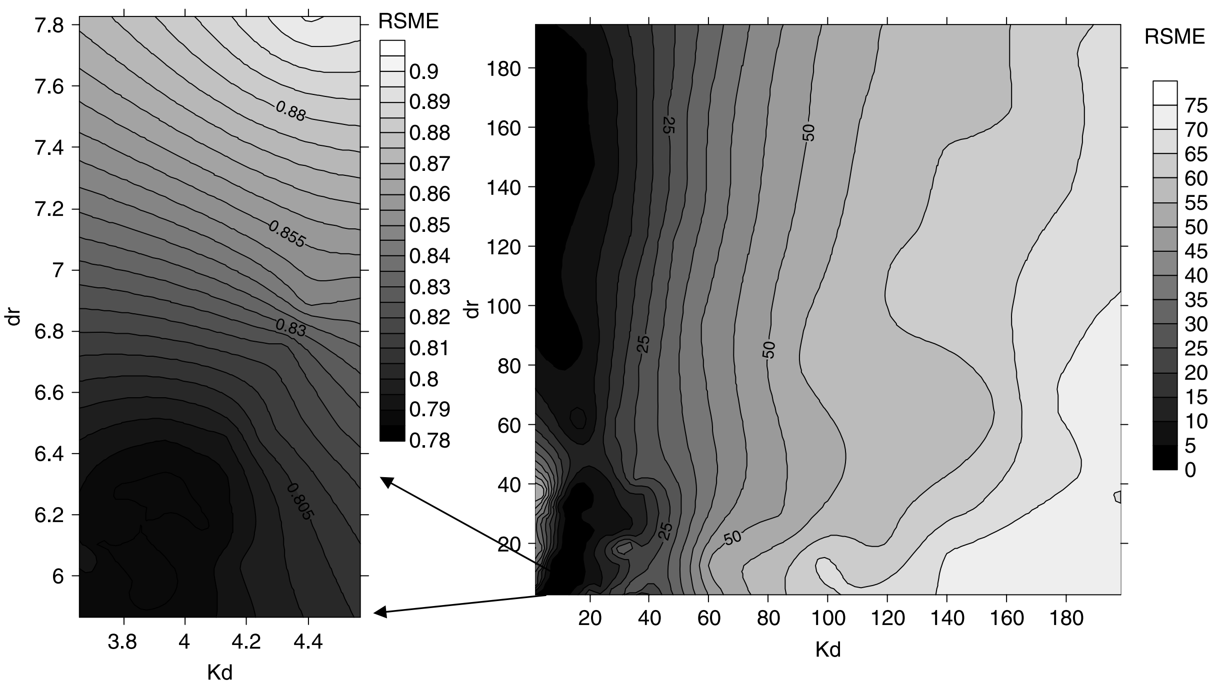

Figure 1 | Error surface of the parameters *ki* and *dr*.

boundary of the *a priori* estimation. It is also worth noticing that *ki* is higher than *kd* for the reference parameter set mentioned in Table 1; in our study this relation is inverse. This could be a result of the fact that optimal parameters are event-specific and the reference parameter set was determined in a modeling context which was focused on portraying the water balance, while we focus on peak flows in the present study.

Monte Carlo simulation

In the Monte Carlo simulation we evaluate the effect of three free model parameters, *dr*, *ki* and *kd*. Worst and best simulations range from 0.18–0.80 for the Nash–Sutcliffe index and 0.79–0.80 for the index of agreement (Table 5). If only the best 10% of simulations are considered, all parameter sets can be considered as good descriptors of the hydrograph, as the difference in results between the worst and best simulation is low (0.75–0.80 for *NSC*). Within the DYNIA, the best simulated values for each time step were used to formulate a hypothetical optimum simulation. This hypothetical optimum is not the best overall simulation, but the extraction of the best values for each time step, chosen from the set of all simulations. This is a theoretical value and serves as a means of comparison only. It is remarkable that even the least performing parameter set displayed the hydrograph quite well as shown in Figure 3.

A relation between optimum parameter ranges and state variables was found for the *dr* interflow parameter.

The DYNIA analysis indicates a linkage of *dr* parameter range to the process dynamics. During low flow periods, *dr* remains at low values (see the highly identifiable parameter ranges highlighted in Figure 4). This is the case mostly for periods where slow recession dominated the flow dynamics. From the sorted plot in Figure 5, generally, higher values of *dr* are identifiable during flood events. From this figure we also see that the identifiable parameter space fluctuates for values of intermediate and low flows. This could be an indicator for a model deficiency in portraying the highly dynamic processes just before a flood event and in the fast recession phase of the flow. This is also visible in Figure 4, where identifiability is low just after the main flood peak.

For low flow data in Figure 5, *dr* adopted values below 5, interrupted by various segments of fluctuating identifiability, where *dr* adopts values between 10–20. For increasing flows, the upper *dr* limit increases to 15 for runoff values of 1 mm. For runoff data >2 mm, *dr* confidence limits

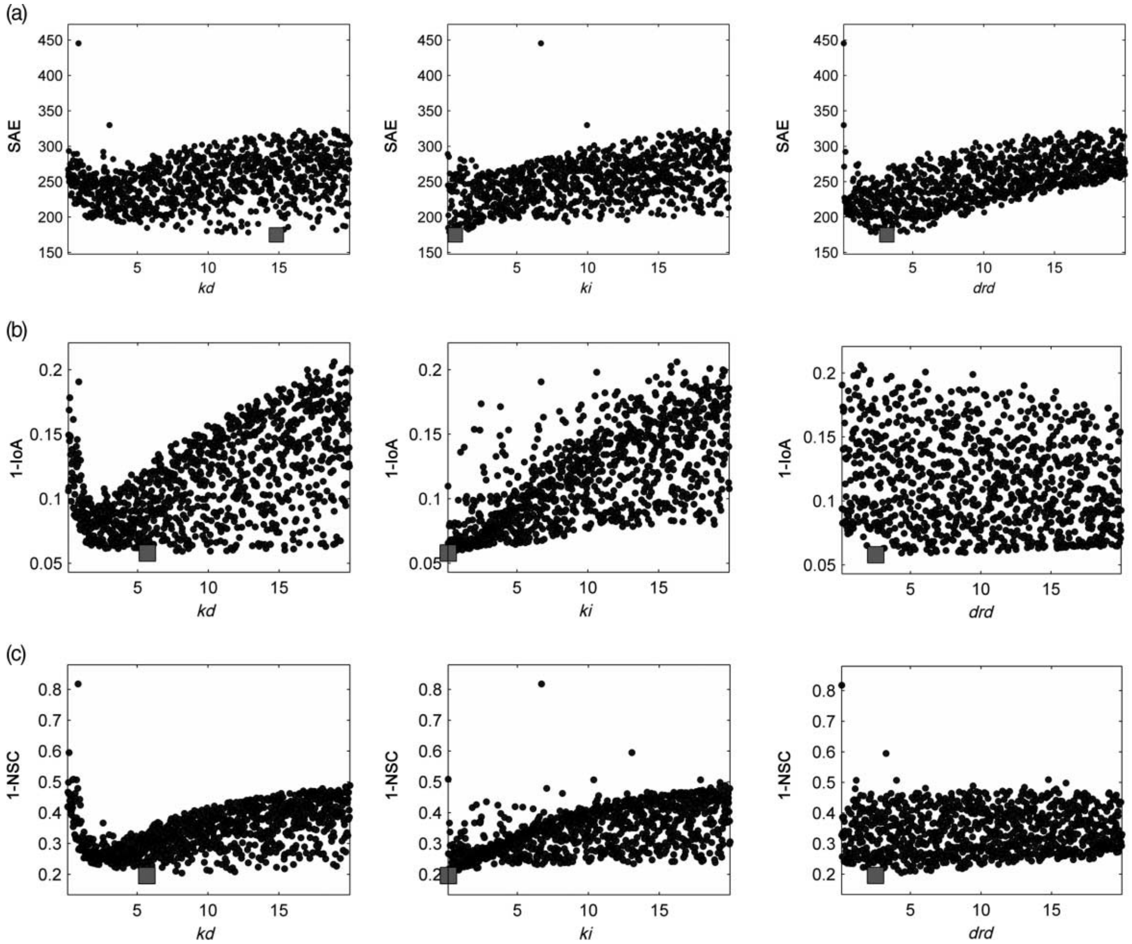

Figure 2 | Dot plots of the parameter space for different objective functions (SAE, IoA, NSC). The boxes indicate optimal parameter values.

cover the entire parameter range and large values for *dr* are most uncertain (also indicated by a decrease in information content, expressed by the diverging of the two lines in the graph that mark the upper and lower confidence limits). This generally indicates increasing uncertainty for extreme events. This might partly be a consequence of a larger data uncertainty for extreme events as well as a less accurate process description by means of the hydrological model.

These findings are similar to those of a previous study in the Weisse Elster river basin (Wriedt & Rode 2006), where a WaSiM-ETH-model was used to simulate daily runoff of a 170 km^2 subcatchment.

A second relation was found linking *dr* and observed soil moisture (Figure 6). For soil moisture exceeding 0.5, the optimum *dr* values range from 5 to 18 while the information content is low. For lower soil moisture values, high identifiability is given with *dr* values between 0 and 4, interrupted by a section of decreased identifiability and increased noise (steps 1,200–2,400). For the upper range of soil moisture values, the optimum *dr* range first increases and then decreases. It is worth noting that observed soil moisture was measured at a single point within the catchment, which might not be representative for the

Table 5 | Objective functions and corresponding parameter set numbers for the Monte Carlo simulation with three free parameters (*kd, ki, dr*), given for the best model fit, the worst model fit, the 90% quantile model fit and model fit for the DYNIA optimum

Objective function	Best parameter set	90% quantile parameter set	Worst parameter set	DYNIA optimum
SAE	175	206	445	88
IoA	0.94	0.93	0.79	0.96
NSC	0.80	0.75	0.18	0.88

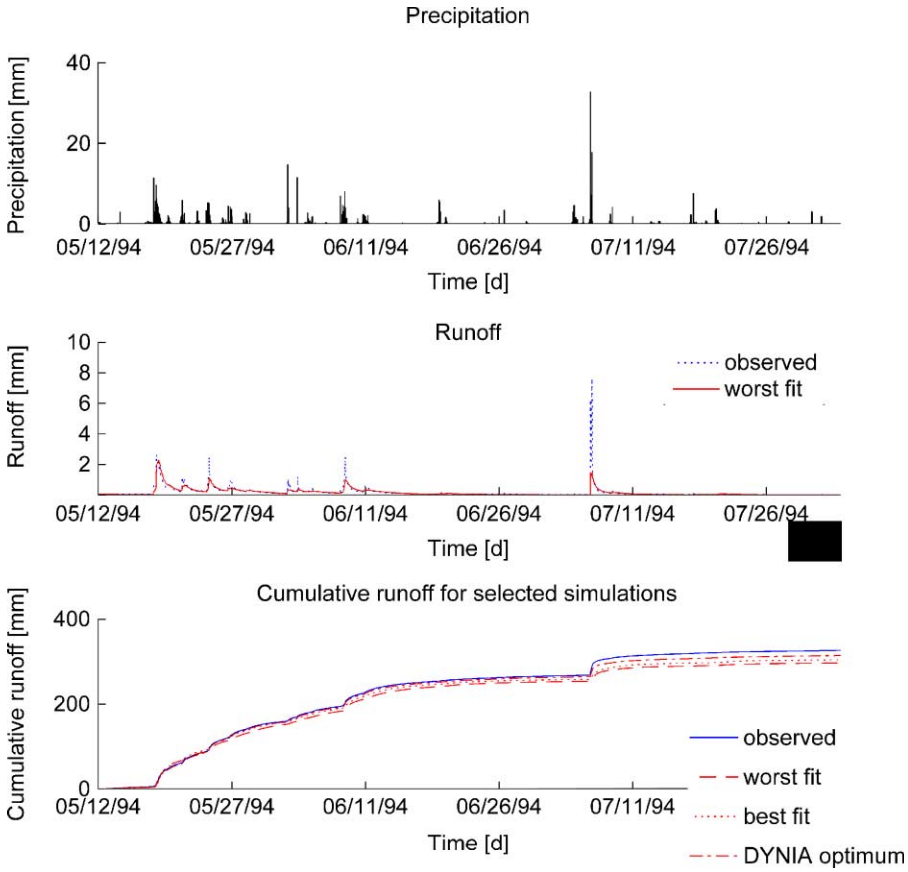

Figure 3 | Precipitation and runoff time series, cumulated runoff for observed data, best and worst simulation of Monte Carlo simulation 1 and DYNIA optimum.

distributed natural soil moisture dynamics. Data for time steps exceeding 4,700 refer to unobserved field data and can therefore not be interpreted in this analysis.

For comparison, we also included simulated soil moisture as sorting criteria for the dataset (Figure 7). A clear relation was found, indicating an increase of *dr* with increasing soil moisture. Interpreting these patterns, it must be considered that simulated and observed soil moisture are not directly comparable: simulated soil moisture is an averaged value of the entire catchment. Its behaviour is consistent with the process descriptions of the model and therefore a clear relation is a logical consequence.

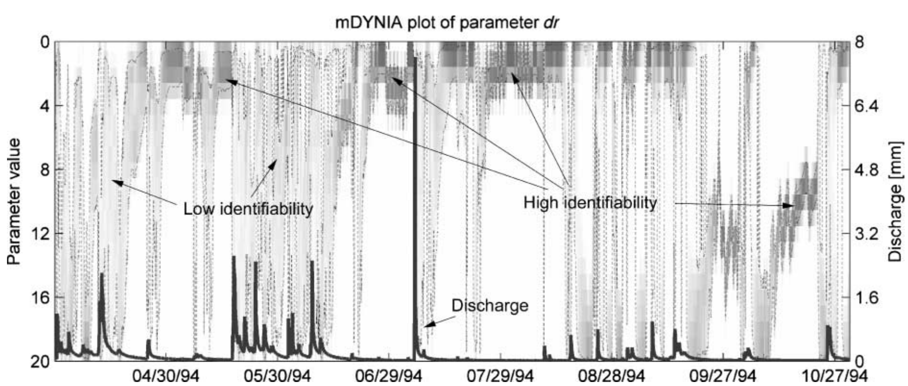

Figure 4 | DYNIA plot for parameter *dr*.

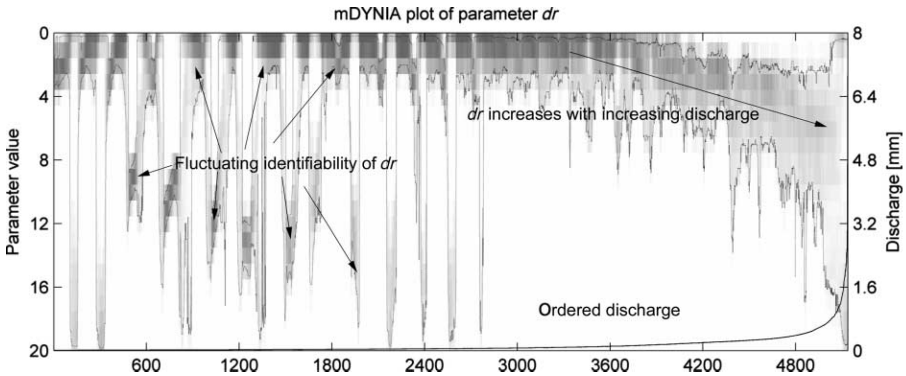

Figure 5 | DYNIA plot for parameter *dr*, time series reordered by observed discharge.

The Monte Carlo analysis did not reveal any clear relations for *ki* and *kd*. As *kd* controls fast surface runoff, increased identifiability is limited to single runoff peaks. During low flow situations the model generates no surface runoff, hence *kd* is not affecting the model result and high uncertainty exists, indicated by confidence ranges covering almost the complete parameter range. A similar behaviour was found for *ki*. Periods of high information content for the model parameter *ki* are also limited to runoff periods, but not as restricted to runoff peaks, as *ki* has a longer lasting effect on interflow recession. Neither *ki* nor *kd* show significant relations to system state variables, when applying re-ordered data series.

Parameter estimation

The parameter estimation presented in this study is characterised by large upper boundaries (the parameter estimation variant used in this study is shaded in Table 4). Nonetheless, the optimal parameters concentrate in the range of 1–100. Automatic parameter estimation of the parameters *ki, kd* and *dr* for all 36 events results in small RMSE, calculated as the square root of mean square error of peak flows (of the order of a few percent, Table 6), and it seems appropriate to suggest that PEST generally converges to reasonable parameter sets. However, some individual events are characterised by large RMSE values, exceeding the mean RMSE by at least one order of magnitude (the maximum values in Table 6). For all of these events one or several of the optimised parameters are forced to the lower boundary during auto-calibration. This is a strong indicator that either the model exhibits structural deficiencies in portraying the dominant processes of the considered events, or the meteorological data do not represent the measured flow data. Further, parameter sets leading to high RMSE values could result from local optima. However, this seems quite unlikely, because the error response surface shown in Figure 1 is quite smooth and gives reasonable results for

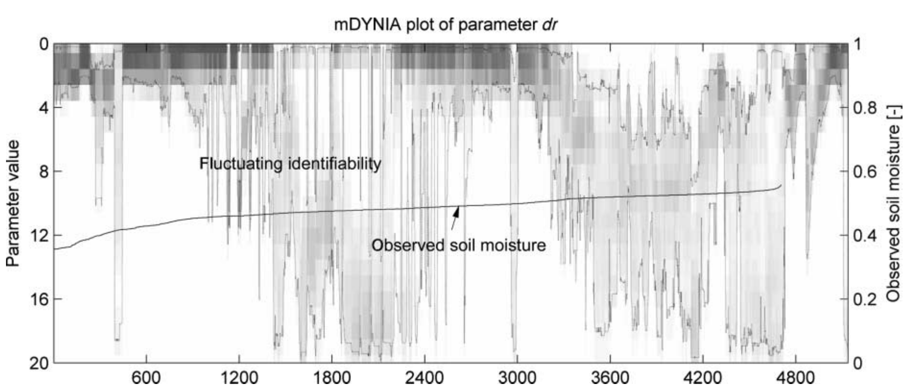

Figure 6 | DYNIA plot for paramater *dr*, time series reordered by observed soil moisture.

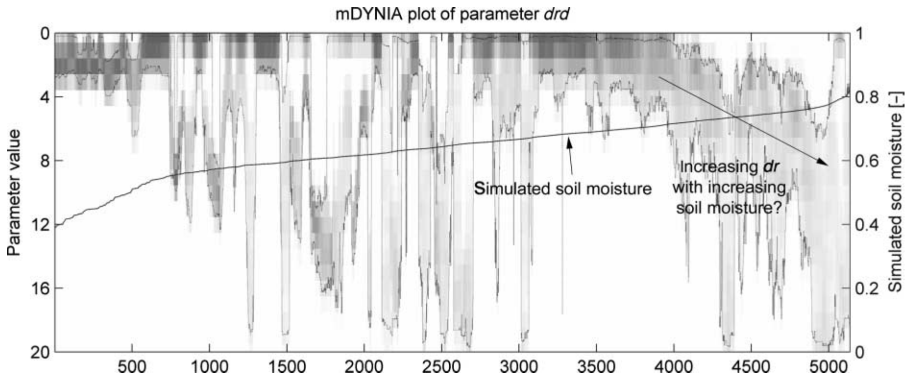

Figure 7 | DYNIA plot for parameter *dr*, time series reordered by simulated soil moisture (catchment average).

most of the considered events. In order to avoid the eventual effects of local optima, the parameter combinations of four events are excluded from subsequent analysis. For those events all parameter sets were optimised to the borders of the parameter ranges. We therefore assume that the data is wrong or the model is not able to handle the specific event dynamics properly.

The results of the automatic calibration are statistically analysed for eventual dependences of optimal parameter sets on the seven chosen characteristic criteria. A simple regression or functional relationship between the parameters and the criteria could not be established with satisfying results, therefore quantitative statements about the setting of model parameters according to *a priori* knowledge seems impossible. In order to be able to judge whether the optimisation yields qualitative information about the relation of event characteristics and parameters, the characteristic criteria are divided into three classes, namely low, medium and high, with an equal number of members for each class. The relation of classes and optimum parameter values is investigated using the Mann–Whitney *U* test, as the statistical requirements for an Analysis of Variance (ANOVA) are not fulfilled (i.e. the homogeneity of variances). The non parametric Mann–Whitney *U* test has less statistical power than ANOVA, but it accepts inhomogeneous variances. With this test, it is possible to compare two classes of criteria, concerning their mean values for a parameter. It measures how much the average rank of one class differs from the average rank of another class. The investigation is confined to the parameter *dr*, which is chosen because visual control of box plots for the classes indicated a possible relation between the three criteria classes and the parameter *dr*. Also, the studies of Cullmann *et al.* (2006) and Wriedt & Rode (2006) indicated a possible relation between characteristic criteria and *dr*. Because of possible parameter interaction of the aforementioned best sets, the optimisation was repeated for the parameter *dr* alone with fixed values of $ki = 8$ and $kd = 4$. These values were chosen in order to base the new exercise on the calibration that well represents general dynamics (the water balance) of the watershed. The general result of this second parameter estimation is shown in Table 7. The RMSE is calculated only for the peak flow value. The results of this optimisation are similar to the pattern found in Table 6. It is worth noting that optimal parameter values for *dr* and the espective error criteria (RMSE) are almost the same in Tables 6 and 7. This indicates that, in the presented study, *dr* is the dominant parameter in automatic parameter estimation.

Table 6 | Results of the joint optimisation of *dr*, *ki* and *kd* (RSME of peak discharge)

	dr	*ki*	*Kd*	RMSE [%]
Mean	27.4	8.1	14.88	1.6
Min	1	1	1	0
Max	235	42.8	177	16.26

Table 7 | Results of the estimation of *dr* with fixed *ki* and *kd* (RSME of peak discharge)

	dr	RMSE [%]
Mean	20.4	1.8
Min	1	0
Max	249.3	18.4

Table 8 | Optimized mean and median *dr* values for the grouped characteristic criteria

		Duration [h]	Peak [mm/h]	Volume [mm]	Form [h]	TI [mm]	PF [mm]	MPI [mm]
Mean	Low	12	16.4	13.4	13.4	13.6	6.7	14.5
	Medium	14.6	5.7	15.3	15.3	6.6	3	8.1
	High	11.9	14.7	8.4	8.5	17	6.6	14.4
Median	Low	5.7	5.7	5.1	5	8.5	5.9	5.9
	Medium	5.6	5	7.3	7.3	5.7	5.4	5.7
	High	6.1	9.3	7.4	7.4	6	5.9	8.5

The classification of the criteria and the corresponding optimised *dr* values for the three groups are shown in Table 8. The mean values of the *dr* parameter within the different criteria classes do not show any general behaviour. This could be a consequence of different types and magnitudes of errors occurring for the specific events considered in this study, leading to high intra-class standard deviation for *dr*. Differing behaviour for different criteria probably partly reflects the fact that different criteria are affected by process dynamics differently and thus must lead to different values. The median was additionally investigated. Here, generally the high criteria classes coincide with the highest *dr* values. Only for TI is this not the case. A statistical interpretation of the results is now possible by means of employing the above mentioned Mann–Whitney *U* test. Table 9 contains the results of this test which helps us to seek statistical proof for differences in *dr* for the combinations low/medium class and medium/high class. In the table, σ describes the level of significance, i.e. 0.15 describes a 15% significance level. The statistical dependence of *dr* values for high and low classes is of theoretical interest only, therefore it has not been investigated. From this statistical test we learn that the peak rainfall rate shows the most powerful statistical link to *dr* (Table 9). If we assume a 15% significance level, the difference between the medium and high classes are statistically significant (see shaded value in Table 9). Would we to assume a significance level of 10%, no statistical proof can be derived from the data at all. The statistics only indicate the strength of the relation between parameters and criteria. With a median of $dr = 9.3$ the high peak rainfall class exhibits the highest overall value for *dr*. The statistical test confirms the interpretation of Table 8. A qualitative relation links the optimal *dr* parameter value to some of the investigated *a priori* characteristic criteria. However, this link is not very strong and can only serve as a very rough description of dependences linking model parameters to event characteristics.

DISCUSSION

Event-specific parameter estimation

One of the main goals of this study was to investigate the relationship of best parameter sets and event preconditions in order to find alternatives to model updating with internal state variables. This would benefit all forecasting activities in headwater catchments and generally all forecasting operations dealing with weather predictions. The results of the event-specific automatic parameter estimation study are

Table 9 | Results of the Mann–Whitney *U* test for the median of *dr*

Criterion	Classes	σ
Duration [h]	Low/medium	0.85
	Medium/high	0.67
Peak [mm h^{-1}]	Low/medium	0.25
	Medium/high	0.13
Volume [mm]	Low/medium	0.2
	Medium/high	0.34
Form [h^{-1}]	Low/medium	0.19
	Medium/high	0.34
TI [mm]	Low/medium	0.38
	Medium/high	0.45
PF [mm]	Low/medium	0.84
	Medium/high	0.81
MPI [mm h^{-}]	Low/medium	1.0
	Medium/high	0.49

ambiguous. On the one hand, it was impossible to find quantitative functional relations linking the parameters *dr*, *ki* and *kd* to the chosen criteria. With the parameter sets found we were not able to perform ANOVA analysis because of the violation of the precondition of equal variances for the classes in question. It is not clear whether a larger database would allow parametric testing of dependences. For the investigated catchment it was therefore impossible to derive *a priori* best parameters according to generally available information, such as the precipitation forecast. On the other hand, a more detailed study of only one parameter (*dr*) revealed that there is a distinct qualitative relation between the event characteristics and the best parameter. For the case of peak rainfall intensity this relation is strong. The classified relation is also notable for the criteria Volume and Form (Table 9). The significance of these two criteria is higher for the test of low and medium classes. The results suggest that *dr* is positively related to the criteria Volume and Form. This might indicate that the WaSiM-ETH infiltration module causes systematic error, as it is the primary model component reacting to rainfall intensity. The parameter *dr* then serves to balance out this systematic error (with larger *dr* values for higher precipitation intensities) in the runoff generation of the WaSiM-ETH soil module.

Generally, applying statistics to calibrated parameters can lead to an improved understanding of the degree of dependence of model parameter sets to criteria derived from widely available information. This helps to improve model parametrisation strategies regarding the sensitivity to characteristic inputs–e.g. rainstorms of different intensities but equal volumes–not accounted for by rigid model structures. For further development of better *a priori* parameter setting, the possible parameter interaction remains to be examined to complete the results presented in this study.

DYNIA application

The more general DYNIA revealed a relation between the interflow parameter *dr* and observed discharge; as soon as runoff has a relevant interflow component, the optimum range of *dr* increases with discharge.

A similar relation was found for soil moisture. However, for ordered observed soil moisture (Figure 6), *dr* values follow a u-shaped relation for the upper spectrum, with values decreasing for highest soil moisture values. A relation between *dr* and simulated soil moisture indicates increasing parameter values for increasing soil moisture. The following aspects need to be considered for interpretation of these findings: First, observed soil moisture values result from tensiometer measurements at a single location in the catchment; they are indicative but not necessarily representative for the catchment. Second, simulated soil moisture values are spatially averaged model results; they are therefore representative for the catchment and can represent catchment wetness indices according to the model assumptions. As a consequence of this contradictory behaviour, the relation between observed soil moisture and parameter values should be further investigated with more representative data.

The observed relation of *dr* to discharge and soil moisture can be interpreted from a process-oriented view. As stated before, observed discharge can be taken as an integral measure of the catchment wetness state. Soil moisture values or groundwater levels (not considered here) can also be indicators of catchment wetness state. The wetness of a catchment or soil moisture levels considerably influence connectivity of subsurface flow pathways and conductivity of the soil matrix: the hydraulic conductivity of matrix flow is directly related to soil water content and increases considerably with soil moisture above field capacity; flow paths and high conductive areas will be linked if soil moisture increases. This would result in a situation, where more lateral subsurface flow can be produced at higher moisture values than at lower moisture values. Generally, the results obtained in this study lead to the suggestion that the model is not well enough describing the generation of direct runoff.

No clear relations were found for the parameters *kd* and *ki*, although the DYNIA suggested different values for low and high observed discharges. The effect of *dr* on runoff differs from those of the parameters *kd* and *ki*. While *dr* controls a runoff volume, *ki* and *kd* control runoff recession of surface and interflow, influencing temporal dynamics. *kd* is only effective during runoff peaks, where surface runoff is produced. Generally this source of uncertainty (different parameters lead to the same result as a consequence of intermittent processes) is often ignored and not discussed separately in the context of the actual uncertainty debate.

However, for fast flow components, the temporal resolution of the simulation may also play a key role in properly resolving surface runoff behaviour. More research is needed.

The *dr* parameter was forced to its upper boundary for peak flows. Further simulations with modified parameter range confirmed this finding. As only one parameter is considered in a single DYNIA, the optimisation of the parameter range compensates for all model deviations, neglecting the causal process relation of the parameter under consideration and the model deviations.

Combination of DYNIA and event-specific parameter estimation

Combining the results of the two different approaches for one year (1994) leads to the result shown in Figure 8. In some cases optimal DYNIA parameter identifiability differs significantly from parameters estimated for single events. One possible explanation for this is the DYNIA window size, which levels out peak errors in parameter identifiability. The DYNIA results alone should therefore be interpreted with care if single events are studied. Generally, however, from Figure 8 it also becomes clear that in most cases the DYNIA parameter identifiability coincides with the event-based optimal *dr* parameter.

DYNIA can be used to back up automatically estimated parameters. If the estimated parameters are within the identifiable parameter range found with DYNIA, those parameters are likely representative for the modelled process. It can be assumed that the model produces right results for the "right reasons", i.e. the model is not tuned with parameters which are far from a meaningful parameter space.

We found no considerable improvement of model results using the DYNIA methodology compared to the model with default parameters. The differences between the best and worst parameter sets of the Monte Carlo simulations were low, albeit wide parameter ranges were allowed. Although optimum model parameters revealed strong relations to state variables, the effect on the simulated hydrograph was low for the given model application.

The low quantitative effect reflects the careful model set-up, based on long-term observations in an intensively monitored research catchment. Also, runoff characteristics of the Rietholzbach catchment may contribute to the low quantitative effect; the catchment reacts fast on rainfall inputs due to the small size and the high relevance of surface runoff and lateral subsurface flows (interflow). Simulations differ most in the ability to reproduce peak discharge; the deviations between event-based estimated best parameters and DYNIA shown in Figure 8 confirm this statement.

The joint evaluation of the two methods applied to our test catchment reveals that state-dependent parameter changes need to be considered in continuous simulations. Such relations can be identified qualitatively. For the continuous simulation, *dr* depends on catchment moisture state, controlling interflow generation. At the event scale, runoff generation is controlled by the separation of direct runoff and interflow, thus leading to a stronger relation of *dr* and rainfall characteristics. Here, the event-specific calibration is able to reveal model-structure-dependent

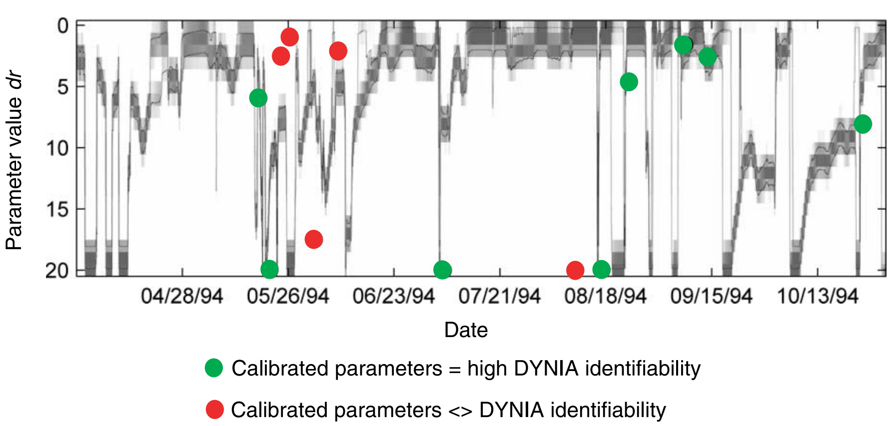

Figure 8 | DYNIA for *dr* compared to event-based estimation of best *dr* with PEST.

relations not detectable by the DYNIA of the continuous time series.

The event-based calibration and the DYNIA methodology complement one another. The joint interpretation of the two methods gives a reason to check the statements made so far and to deepen our understanding of the dependence of best model parameter behaviour in the runoff generation and runoff concentration modules of a process model.

CONCLUSIONS

Event-based automatic parameter estimation leads to the following conclusions:

- Qualitative relations between *dr* and Volume and Peak classes are evident. The criterion Form is also statistically relevant for the parameter value *dr*. This is a direct consequence of the combination of the two aforementioned criteria.
- The catchment pre-event rainfall index classes (PF) are not statistically meaningfully related to *dr*. This contradicts the findings of Cullmann *et al.* (2006).
- The results of Cullmann *et al.* (2006) are qualitatively confirmed by the results of the non-parametric testing, yet a functional relationship for *a priori* parameter variation needs further research in terms of the model's ability to portray processes in general.

The basic conclusions drawn from the modified DYNIA approach are as follows:

- The DYNIA analysis revealed a strong relation between the interflow parameter *dr* and observed discharge.
- The analysis fails for peak flows. Here, surface flow becomes the dominant runoff component, and runoff generation cannot be attributed to the interflow parameter *dr*, which would be forced to the upper boundary of the parameter range.
- No clear relations were found for the parameters *kd* and *ki*, although the DYNIA analysis suggested different values for low and high observed discharges.

The results of the study might indicate a way towards a new approach in hydrological modelling, incorporating generally available information about the rainfall characteristics and state variables to operate with transient, tailor-made model parameters. This parameter adaptation could compensate for model structural errors. At present, the qualitative information about parameters depending on *a priori* characteristics could serve as a base for a fuzzy-rule-based parameter setting of models.

Analogously to state-of-the-art model updating procedures in conceptual rainfall–runoff modeling, this approach could be used for application of process-oriented models for operative purposes. The advantage would be to maintain the general process relations of the model while being able to account for structural problems in a meaningful way. In contrast, automatic parameter or internal state updating would result in violation of the specific process relations. This is especially critical if the uncertainty of the forecast is a matter of concern. However, this approach cannot avoid the necessity for extensive and catchment-specific prior analysis of model behaviour inherent to any process-oriented model.

To benefit from the potential of the combined DYNIA and parameter estimation evaluation, further research is required. Here, especially, the effects of the DYNIA window size should be closely examined for comparison with event-based parameter estimation. The objective function of automatic parameter estimation might also play a substantial role for the optimal parameter set: this applies both for DYNIA and parameter estimation. To broaden our understanding of the identifiable parameter space it might be helpful to study more catchments with differing characteristics. This would broaden the range of investigated dominant process stages and thus sharpen the meaning of the joint analysis proposed in this paper.

ACKNOWLEDGEMENTS

We thank Karsten Jasper, who provided input data, and a calibrated WaSiM-ETH model for the Rietholzbach catchment.

REFERENCES

Beven, K. & Binley, A. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process.* **6**, 279–298.

Beven, K. & Freer, J. 2001 Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.* **249** (1–4), 11–29.

Cullmann, J. 2007 *Online Flood Forecasting on the Basis of a Synthesis of Process Modelling and Artificial Intelligence*. Dissertation, TU-Dresden.

Cullmann, J., Schmitz, G. H. & Mishra, V. 2006 Eine neue Strategie zur Hochwasservorhersage in schnellreagierenden Einzugsgebieten–Modellierungsaspekte (A new strategy for flood forecasting in fast reacting watersheds–aspects of modelling) *BFG-Veranstaltungen* 2006/3, Koblenz.

Duan, Q., Sorooshian, S. & Gupta, V. 1992 Effective and efficient global optimisation for conceptual rainfall-runoff models. *Water Resour. Res.* **28** (4), 1015–1031.

Gupta, V. K. & Dawdy, D. R. 1995 Physical interpretations of regional variations in the scaling exponents of. flood quantiles. *Hydrol. Process.* **9** (3/4), 347–361.

Gupta, V. K., Mesa, O. & Dawdy, D. R. 1994 Multiscaling theory of. flood peaks: regional quantile analysis. *Water Resour. Res.* **30** (12), 3405–3421.

Gurtz, J., Zappa, M., Jasper, K., Lang, H., Verbunt, M., Badoux, A. & Vitvar, T. 2003 A comparative study in modelling runoff and its components in two mountainous catchments. *Hydrol. Process.* **17**, 297–311.

Hornberger, G. M. & Spear, R. C. 1981 An approach to preliminary analysis of environmental systems. *J. Environ. Manage.* **12**, 7–18.

Merz, R. & Blöschl, G. 2003 A process typology of regional floods. *Water Resour. Res.* **39** (12), 1340.

Mohamed, J. L. & Walsh, J. 1986 *Numerical Algorithms*. Oxford Science Publications, Oxford.

Post, D. A. & Jakeman, A. J. 1996 Relationships between catchment attributes and hydrological response characteristics in small Australian mountain ash catchments. *Hydrol. Process.* **10** (6), 877–892.

Schulla, J. 1997 *Hydrologische Modellierung von Flussgebieten zur Abschätzung der Folgen von Klimaänderungen*. Dissertation., ETH Zürich, Switzerland.

Schulla, J. & Jasper, K. 2001 *Model Description WaSiM-ETH*. Internal report, IAC, ETH Zürich, Switzerland.

Skahill, B. E. & Doherty, J. 2006 Efficient accomodation of local minima in watershed model calibration. *J. Hydrol.* **329**, 122–139.

Spear, R. C. & Hornberger, G. M. 1980 Eutrophication in Peel Inlet, II: Identification of critical uncertainties via generalised sensitivity analysis. *Water Res.* **14**, 43–49.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S. & Gupta, H. V. 2003 Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrol. Process.* **17**, 455–476.

Wriedt, G. & Rode, M. 2006 Investigation of parameter uncertainty and identifiability of the hydrological model WaSiM-ETH. *Adv. Geosci.* **9**, 145–150.

First received 31 October 2007; accepted in revised form 29 June 2008