Reusser, D. E., Buytaert, W., Zehe, E. (2011): Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test. - Water Resources Research, 47, W07551

DOI: 10.1029/2010WR009947

Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test

D. E. Reusser,[1,2] W. Buytaert,[3] and E. Zehe[4]

Received 27 August 2010; revised 10 June 2011; accepted 15 June 2011; published 29 July 2011.

[1] The quest for improved hydrological models is one of the big challenges in hydrology. When discrepancies are observed between simulated and measured discharge, it is essential to identify which algorithms may be responsible for poor model behavior. Particularly in complex hydrological models, different process representations may dominate at different moments and interact with each other, thus highly complicating this task. This paper investigates the analysis of the temporal dynamics of parameter sensitivity as a way to disentangle the simulation of a hydrological model and identify dominant parameterizations. Three existing methods (the Fourier amplitude sensitivity test, the extended Fourier amplitude sensitivity test, and Sobol's method) are compared by applying them to a TOPMODEL implementation in a small mountainous catchment in the tropics. For the major part of the simulation period, the three methods give comparable results, while the Fourier amplitude sensitivity test is much more computationally efficient. This method is also applied to the complex hydrological model WaSiM-ETH implemented in the Weisseritz catchment, Germany. A qualitative model validation was performed on the basis of the identification of relevant model components. The validation revealed that the saturation deficit parameterization of WaSiM-ETH is highly susceptible to parameter interaction and lack of identifiability. We conclude that temporal dynamics of model parameter sensitivity can be a powerful tool for hydrological model analysis, especially to identify parameter interaction as well as the dominant hydrological response modes. Finally, an open source implementation of the Fourier amplitude sensitivity test is provided.

Citation: Reusser, D. E., W. Buytaert, and E. Zehe (2011), Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test, *Water Resour. Res.*, *47*, W07551, doi:10.1029/2010WR009947.

1. Introduction

[2] Rainfall-runoff models have become important tools to represent and test our knowledge about the processes in a hydrological catchment. One of the most important aims in model building is to keep the model structure as parsimonious as possible, to aid calibration and uncertainty analysis, and to avoid parameter interaction and lack of identifiability.

[3] For this purpose, it is necessary to identify dominant hydrological processes and to parameterize them adequately in the model as functional components. This is often not straightforward. Depending on the hydrological context (e.g., rainfall driven, energy driven, or occurrence of snowmelt) different processes will be active in the hydrological system at different moments in time. Ideally, in a parsimonious model with low parameter interaction, this should be reflected in the model structure, with different model components dominating simulated dynamics over time. Hence, we expect simulation results to be most sensitive to variations of exactly those parameters that belong to the corresponding model component. For instance, we expect a good model to be sensitive to variation of snowmelt parameters during snowmelt periods, but rather insensitive in snow free periods.

1.1. Sensitivity Analysis for Temporal Dynamics

[4] Sensitivity analysis (SA) assesses the impact of model parameters on the model outcome, and is therefore a convenient tool to assess model behavior and particularly the importance of certain parameterizations within the model.

[5] Classically, SA is of most interest in the context of model calibration. The goal is then to determine the most important parameters for the calibration process as well as the unimportant parameters that may be fixed at a predefined value. Therefore, in hydrology, sensitivity is most often calculated for some objective function, for example the root mean square error RMSE or the Nash-Sutcliffe

[1] Institute of Earth and Environmental Sciences, University of Potsdam, Potsdam, Germany.
[2] Potsdam Institute for Climate Impact Research, Potsdam, Germany.
[3] Department of Civil and Environmental Engineering, Imperial College London, London, UK.
[4] Institute for Water and Environment, TU München, Munich, Germany.

0043-1397/11/2010WR009947

coefficient of efficiency. In contrast, we do not approach the question of sensitivity from a calibration point of view. By analyzing temporal dynamics of parameter sensitivity (TEDPAS) of model output variables, such as discharge, groundwater level or snow water equivalents, we can quantify which model components dominate the simulation response. This information can then be used as an indicator for dominant processes in the catchment, as well as the functioning of the model. In other words the output quantity can be a summary statistic of model performance, but it can also be the model output at a given point in time. The same SA methods can be applied for both approaches, with the main difference that for TEDPAS, SA is performed for each time step individually. The model runs required for the sensitivity analysis only need to be performed once, and it is possible to use these model runs to compute the sensitivity of any output quantity. TEDPAS as an analytic tool for identification of dominant model components has been reported before by *Sieber and Uhlenbrook* [2005] and *Cloke et al.* [2008].

[6] Using TEDPAS as an analytic tool is related to the dynamic identifiability analysis introduced by *Wagener et al.* [2003]. However, the two methods serve a different purpose. Identifiability analysis aims at identifying parameters that can be confined by given observations. It is a necessary, but not sufficient condition that parameters must be sensitive to be identifiable. Nonidentifiable but sensitive parameters occur for instance, when parameters are strongly correlated as for the Nash cascade where a decrease of one parameter can be compensated by increasing the other [*Bárdossy*, 2007].

1.2. Sensitivity Analysis Methods

[7] A wide range of SA methods exist. Many methods characterize local gradients at a given point in parameter space by assessing the response of the model output to a small variation of single parameters (the so-called one at a time method). This is a sensible approach if SA is used in the context of model calibration. The main disadvantage of this method (and other local SA methods) is that information is available for this very specific location in the parameter space only, which is usually not representative of the physically possible parameter space. To overcome this problem global SA methods have been proposed, where multiple locations in the physically possible parameter space are evaluated. Global methods may be used without prior calibration of the model, which may reduce the total computing time required considerably as calibration often requires a large number of model runs. Global SA with regression-based methods rests on the estimation of linear models between parameters and model output. The method provides good estimates of parameter sensitivity for nearly linear models, but fails if the model output shows nonlinear (especially nonmonotonic) dependence on model parameters, which is very common for hydrological models. Regional sensitivity analysis (RSA) [*Hornberger and Spear*, 1981] and derived methods approach the question by comparing an initial distribution of model parameters to the distribution after conditioning of the model output to observations. This approach is more suited to find identifiable parameters than to find sensitive parameters (see above). Finally, a number of methods are based on ANOVA-like analysis of the dependence of the model output variance to simultaneously modified parameters (partial variance–based methods):

$$V = \sum_i V_i + \sum_{i<j} V_{ij} + \cdots + V_{1,2,3,\ldots,n} \quad (1)$$

V is the total variance, V_i is the variance caused by parameter θ_i (first-order variance), V_{ij} is the covariance caused by θ_i (second-order variance) and θ_j and higher-order terms show the variance contribution from multiple parameters. Sensitivities in terms of partial variance are then calculated by dividing by the total variance V. Therefore, all such defined sensitivities add up to 1. Variance-based methods result in reliable estimates of sensitivities also for strongly nonlinear models, as was often demonstrated using examples where the analytical solution can be computed [e.g., *Saltelli and Bolado*, 1998]. The main drawback of these methods is that they are not easy to implement and the required number of model runs is very high (usually >1000) for most approaches. The most important variants of this method are Sobol's method [*Sobol*, 2001] and the (extended) Fourier amplitude sensitivity test ((E)FAST) [*Schaibly and Shuler*, 1973; *Cukier et al.*, 1973, 1975; *Fang et al.*, 2003; *Saltelli and Bolado*, 1998].

[8] Some of the recent studies applying SA to rainfall runoff, flood inundation, and water quality models are listed in Table 1. Eight out of the 18 studies use variance-based methods. In seven studies, on the order of 10,000 model runs were computed to calculate sensitivities, which is impossible for computationally expensive models. To our surprise, we were unable to find an application of FAST or EFAST to rainfall-runoff modeling.

[9] The selection of the appropriate method for analyzing parameter sensitivity depends strongly on the goal of the sensitivity analysis [*Saltelli et al.*, 2006] (Figure 1). Three types of application and their potential results are illustrated in Figures 1a, 1b, and 1c using an arbitrary precipitation pulse and a three-parameter model. First, if the correct values of a parameter can be fixed from additional, independent data before calibration, then which parameter causes the greatest reduction in variance (called factor prioritization setting by *Saltelli et al.* [2006])? This use is illustrated in Figure 1a). In Figure 1 (left), the distribution of parameter values (normalized to the range between 0 and 1) is shown using parallel coordinates [*Wegman*, 1990]. The grey sets illustrate the physically possible parameter space while the highlighted sets indicate the parameter range, to which P2 is fixed from independent data without calibration. In reality, there is no true parameter set because the parameters are not observable [*Beven*, 2002; *Zehe et al.*, 2007] or cannot be identified [*Klaus and Zehe*, 2010]; however, we assume perfectly determinable parameters for the illustration of the factor prioritization setting.

[10] Figure 1 (right) shows the distribution of the model output for the sampled parameter space in light grey. The red shaded area shows the greatly reduced variance after fixing P2. The black line shows the observations. Factor prioritization is used to identify the relevant model components for a certain time step or to identify periods with high information content for the calibration of these parameters [*Saltelli et al.*, 2006].

[11] A second use of SA is the identification of parameters that can be fixed at any value in their possible range without significantly reducing the output variance.

Table 1. Recent Sensitivity Analysis Studies in Surface Hydrology and Water Quality Modeling[a]

Study	Model	Parameters	Method	Runs	FOPV	Evaluated Model Output
Deflandre et al. [2006]	QUESTOR	5–12 parameters: chemical reaction constants and oxygen exchange constants	EFAST	500	x	Nash-Sutcliffe coefficient of efficiency
Cloke et al. [2008]	ESTEL-2D	9 parameters: moisture, hydraulic conductivity, Brooks-Corey and van Genuchten parameters, storage parameters, upslope pressure, river stage, rainfall	MMGSA[b]	<1280	x	fuzzy membership (overall, $f(t)$); sum squared errors
van Werkhoven et al. [2009]	SAC-SMA	14 parameters for upper (3) and lower (5) zones, partition (3), and percolation (3)	Sobol's method	7.5×10^6	x	4 objective functions
van Werkhoven et al. [2008b]	SAC-SMA	78 cells × 14 parameters for upper (3) and lower (5) zones, partition (3), and percolation (3)	Sobol's method	4×10^6	x	RMSE
van Werkhoven et al. [2008a]	SAC-SMA	14 parameters for upper (3) and lower (5) zones, partition (3), and percolation (3)	Sobol's method	130,000	x	RMSE, TRMSE, SFDCE, ROCE
Wagener et al. [2009]	SAC-SMA	78 cells × 14 parameters for upper (3) and lower (5) zones, partition (3), and percolation (3)	Sobol's method	not reported	x	RMSE, TRMSE, ROCE
Tang et al. [2007b]	SNOW-17–SAC-SMA	5 parameters for SNOW (precipitation correction, melting factors (2), wind, and SCA parameter) and 13 parameters for SAC-SMA (see above)	Sobol, RSA, ANOVA, PEST	≤10,000	x	RMSE of discharge, RMSE of Box-Cox transformed discharge
Pappenberger et al. [2008]	HEC-RAS	6 parameters: input quality (1), 3 parameters for roughness, parameter for numerical solution, downstream initial slope	multiple[c]	not reported	x	mean absolute error; Nash-Sutcliffe measure
Pappenberger et al. [2006]	HEC-RAS[d]	3 surface roughnesses, 3 model input parameters, and 1 parameter for numerical solution	SARS-RT, correlation, RSA	3000		inundation performance measure
Demaria et al. [2007]	VIC (variable infiltation capacity)	10 parameters: base flow, layer thickness, hydraulic conductivity, infiltration, Brooks-Corey parameter	RSA+ (Freer96), scatterplots	60,000		RMSE, RMSE Box-Cox, ARE of discharge
McIntyre et al. [2003]	WaterRAT	26 parameters related to chemical processes	RSA	10,000		water quality concentrations
Sieber and Uhlenbrook [2005]	TAC^D	20 parameters: precipitation (2), snow (7), soil (13), runoff (22), and routing (4)	RSA, regression with Latin hypercube sampling	400		q(t), Nash-Sutcliffe of q and log(q)
Christiaens [2002]	MIKE SHE	soil hydraulic parameters	regression with Latin hypercube sampling	25		multiple[e]
Muleta and Nicklow [2005]	SWAT	35 for snow (6), sediment (5), evaporation (5), routing (10), and base flow (9)	regression with Latin hypercube sampling	300		RMSE (discharge, sediment yield)
Van Griensven et al. [2006]	SWAT	41 parameters related to snow, soil, groundwater, geomorphology, evaporation, channel flow, runoff, erosion, and crop	Morris modified: multiple OAT with Latin hypercube sampling	not reported		total sums and sum squared errors of flow, sediment, and nutrients
Foglia et al. [2009]	TOPKAPI	35 (thickness of soil, hydraulic conductivity, water content, Manning roughness)	OAT local SA	71		discharge (composite sensitivity)
Benke et al. [2008]	2C-model	5 parameters: store shape (2), evaporation (1), and maximum discharge (2)	stepwise fixing of parameters, local SA	30,000		annual discharge
Cullmann et al. [2006]	WaSiM-ETH	6 soil module–related parameters	OAT local SA	13		peak discharge

[a]ARE, absolute relatvie error; FOPV, first-order partial variance; MMGSA, multimethods global sensitivity analysis; QUESTOR, Quality Evaluation and Simulation Tool for River systems; ROCE, runoff coefficient error; SARS-RT, sensitivity analysis based on regional splits and regression trees; SFDCE, slope of the flow duration curve error; TAC^D, tracer aided catchment model, distributed; TRMSE, RMSE after Box-Cox transformation.
[b]MMGSA is a combination of Sobol, K-L entropy, and Morris method.
[c]Sobol, Kullback-Leibler entropy, Morris, RSA, and regression method.
[d]Plus two simple model structures for testing.
[e]Cumulative and peak discharge, average base flow, average groundwater elevation, average soil moisture.

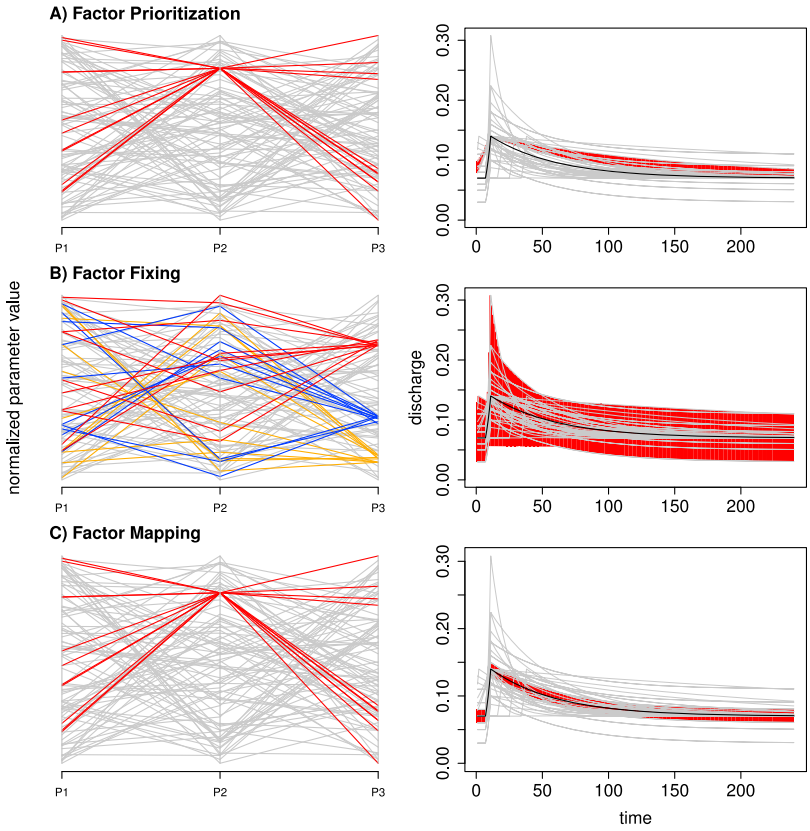

Figure 1. Illustration of the different purposes of sensitivity analysis (SA) [*Saltelli et al.*, 2006]: (a) factor prioritization investigates the most influential parameter, (b) factor fixing investigates the least influential parameter, and (c) factor mapping is related to calibration and GLUE-like procedures. (left) Possible and selected parameter sets are shown using parallel coordinates. (right) Measured discharge (black), possible simulation runs (light grey), and the remaining variability (red) are shown to visualize SA settings. The three different selected parameter sets for factor fixing (Figure 1b) result in the same remaining variability, and only the red set is shown.

The remaining parameters explain the variance (factors fixing setting). This is useful to exclude irrelevant parameters from the model calibration and set their values to an arbitrary one. Figure 1b shows that P3 may be set to any value in the full range, without reducing the variance in the model output significantly. The variance in the model output covers the full range without depending on whether P3 is selected from one of the ranges around 0.1 or 0.3 or 0.7.

[12] Finally, SA can also be used in a third context [*Saltelli et al.*, 2006, p. 1117]: "Sometimes practitioners want to analyze input factors with respect to their capacity to produce realization of the model output Y within a given region, e.g., between bounds, or above a threshold. This leads to a factors' mapping (FM) setting, whose question is which factor is mostly responsible for producing realizations of Y in the region of interest"?

[13] One of the best known applications of this is the generalized likelihood uncertainty estimation method (GLUE). Parameters are sampled from a prior distribution, evaluated against a likelihood measure such as the Nash-Sutcliffe efficiency and subsequently rejected or accepted. The accepted (also know as behavioral) parameters (Figure 1c) are then used to generate prediction boundaries. Sensitivity analysis can here be seen as a by-product, because the width of the posterior parameter distributions is an indication of their sensitivity.

[14] As "best practice" to determine the dominant model components *Saltelli et al.* [2006] suggested to use measures based on first-order partial variance (FOPV). As stated above, partial variance–based methods belong to the global sensitivity analysis (SA) methods, which determine the parameter sensitivity for an entire region in parameter space with distributed sampling techniques in parameter space. As indicated in equation (1), FOPV measures the sensitivity caused by each single parameter, thus parameter interactions do not need to be analyzed for FOPV.

1.3. Advantages of FAST

[15] There are several methods to compute FOPV sensitivities. FAST was originally developed for the analysis of chemical reaction systems, providing a computationally efficient way to compute FOPV [*Schaibly and Shuler*, 1973; *Cukier et al.*, 1973, 1975] As for all partial variance–based methods, FAST is able to reliably estimate sensitivities of parameters also for nonlinear models and is therefore well suited for hydrological models. Multiple sensitivity measures are reported to give contradicting results for the same application [*Tang et al.*, 2007a; *Cloke et al.*, 2008]. (This is not very surprising, if for example local and global sensitivities, regional SA and variance-based methods are compared). In contrast, results for FOPV appear to show better

comparability. For example, *Saltelli and Bolado* [1998] demonstrated the equivalence of sensitivities computed using Sobol's method with FAST. *Saltelli and Bolado* [1998] concluded that FAST is computationally much more efficient, requiring for example only 150 runs to determine reliably the sensitivity of 6 parameters. This may help to overcome problems with high computational expenses when calculating TEDPAS such as those of *Sieber and Uhlenbrook* [2005]. However, FAST has some limitations which make the method unsuitable for certain types of problems. Results from FAST are not accurate for discrete parameter values [*Saltelli et al.*, 2000; *Frey and Patil*, 2002]. Also parameter interactions cannot be detected by the FAST method. The focus of this study is on factor prioritization and therefore, FOPV may be appropriate and parameter interactions are not required according to *Saltelli et al.* [2006].

[16] As argued, FOPV SA is an appropriate method to analyze TEDPAS in order to quantify which model components dominate the simulation as an indicator for dominant processes in the catchment. However, for complex hydrological models we need a highly efficient method to estimate sensitivities. Therefore this paper aims at (1) implementing the computationally efficient original FAST method, (2) investigating whether FAST is applicable to rainfall-runoff models, (3) comparing FAST to existing implementations of FAST, EFAST, and Sobol's method using a lumped (computationally inexpensive) hydrological model, (4) applying FAST to a computationally expensive hydrological model (WaSiM-ETH), and (5) showing how the resulting sensitivities are a useful diagnostic tool by performing a qualitative model validation and we will identify interactions among parameters in WaSiM-ETH which result in problems during model calibration.

2. Methods and Study Area

2.1. Fourier Amplitude Sensitivity Test

[17] As with other global SA methods, for FAST, the ranges in which parameters are varied according to the physically possible parameter space, are usually determined with pretests including expert knowledge, model documentation or model runs. All SA methods have in common that the parameters are modified between the model runs j = 1, 2, 3, …, N (which we denote as the model run domain). FAST is based on the fact that the model output in this model run domain can be expanded into a Fourier series. The coefficients of the Fourier series can then be used to estimate the mean expected model outcome as well as the variance. If individual parameters are varied with specific frequencies, the corresponding Fourier coefficients allow estimation of the partial variance or model parameter sensitivity. This constitutes the first step of the method (section 2.1.1): generating the Fourier parameter set for the sensitivity test. This is exemplified in Figure 2, with the labeled arrows indicating the analysis steps. Figure 2 shows the three relevant domains (time domain, model run domain, and model run frequency domain) in three boxes. The top box shows how simulation results and parameter sensitivity change with time. Also presented is the simple toy model consisting of a time-dependent, weighted average of parameters a and b: $o = an + b\,(1 - n)$ with $n = t/50$. The bottom left box shows the model run domain, where different model runs constitute the x axis. The left plot in the model run domain shows how parameters a and b are varied with frequencies ω = 3 and 7, respectively, in the model run dimension (j = 1, 2, 3, …, N) and varied according to a uniform distribution in the range −0.5–0.5. Note that the order of the parameter sets along the model run dimension j needs to be maintained for the evaluation method to work.

[18] The model is then evaluated for each of the parameter sets (the second step, arrow 2). Figure 2 shows the exemplary evaluation of the very simple model for the three time steps (t = 0, 25, 50) in the three plots in the bottom right box of the model run domain.

[19] To retrieve the information from the different frequencies, i.e., to analyze the sensitivity of the model output for the different parameters, the model output is Fourier transformed in the model run dimension j = 1, 2, 3, …, N (the third step, arrow 3; section 2.1.2). For our example, the results for the Fourier transformation for the three time steps are shown in the model run frequency domain in Figure 2. The fraction of the variance in the model run dimension that can be explained by a certain parameter is proportional to the Fourier coefficient for the corresponding frequency and its multiples (see *Cukier et al.* [1978] for further details). In the example we observe Fourier coefficients above 0.02 only for frequencies 3 and 9 for t = 0, frequencies 3, 7, 9, and 21 for t = 25, and frequencies 7 and 21 for t = 50. Thus, the variance at t = 0 can be fully explained by parameter a (frequency 3 and its multiples). Corresponding statements are possible for t = 25 and t = 50. Finally, the results can be presented in the time domain (Figure 2, top box).

[20] The three steps are described in more detail in sections 2.1.1 and 2.1.2. The method is available both as part of SimLab and as software package [*Reusser*, 2008] for the open source data analysis language R [*R Development Core Team*, 2008].

2.1.1. Generation of the Parameter Set and Model Evaluation

[21] As stated, when generating the Fourier parameter set, we want to modify each parameter with a different frequency among the model runs. Generation of the parameter set can be subdivided into selection of appropriate frequencies ω_i, the generation of a value set S with uniform distribution between −0.5 and 0.5 and a transformation of S into the actual Fourier parameters θ.

[22] *Cukier et al.* [1975] present a table with suggested frequencies ω_i that are mutually independent (see *Cukier et al.* [1975] and *McRae et al.* [1982] for further details) and have mutually independent multiples up to order 4. The higher the order, the smaller the error of the numerical approximation of the FAST method (for further details, see *Cukier et al.* [1975]. With higher number of parameters, the number of required model runs (also presented by *Cukier et al.* [1975]) increases in order to assure independence of frequencies. Therefore, the selected frequencies $\omega(i)$ and the required sampling size N (j = 1,…,N) depend on the number of parameters n (i = 1,…,n). The selected frequencies together with the model run index are then converted into a supporting variable $S(j, i)$ (equation (2)), which varies at the appropriate frequency in the range −0.5–0.5. Calculation of S (j, i) was initially proposed as an exponential function [*Cukier et al.*, 1973], which has the disadvantage of resulting in a distribution that over emphasizes low and high values

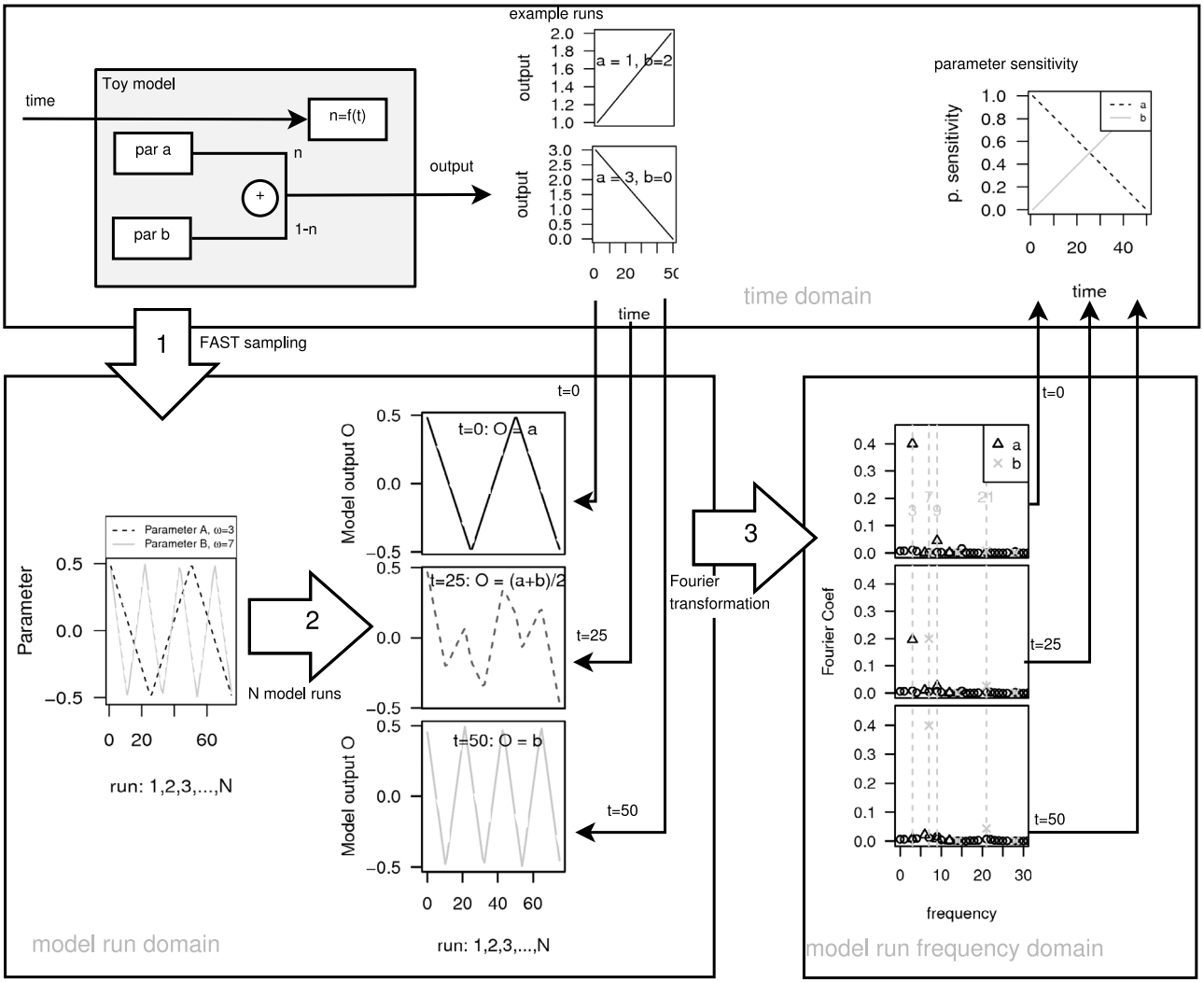

Figure 2. Illustration of the Fourier amplitude sensitivity test (FAST) with a simple toy model: $o = na + (1 - n)b$ with $n = 1 - t/50$ for $t = 0, \ldots, 50$. In the first step (arrow 1), parameters are sampled according to a predefined sampling scheme for multiple model runs (model run domain). The second step includes running the model for each time step (arrow 2). Because of the special sampling design, parameter sensitivities can be calculated with Fourier transformation in the third step (arrow 3, model run frequency domain). Applying FAST for each time step allows calculation of TEDPAS.

[*Saltelli et al.*, 1999]. *Saltelli et al.* [1999] proposed to use the function shown in equation (2) which results in a uniform distribution of $S(j, i)$. The final transformation of $S(j, i)$ to the actual Fourier parameters $\theta(j, i)$ has undergone some development since the publication of the original method. The transformation based on the cumulative density function $F(\theta)$ of the parameter as shown in equation (3) was presented by *Fang et al.* [2003]. Compared to the original method [*Cukier et al.*, 1978] this method has an advantage if nonuniformly distributed parameters are used. We used uniform distributions for all parameters with ranges as shown in Tables 2 and 3. Note that in this case, the transformation procedure of *Fang et al.* [2003] does not provide an improvement of the method compared to the original method of *Cukier et al.* [1978].

$$S(j, i) = \arcsin(\sin\{\omega_i(\pi/N)[(2j - N - 1)/2]\})/\pi \qquad (2)$$

where $j = 1, \ldots, N$, $i = 1, \ldots, n$,

$$\theta(j, i) = F_i^{-1}(S(j, i) + 0.5) \qquad (3)$$

with F_i^{-1} being the inverse of the cumulative density function for parameter i.

2.1.2. Analysis of Parameter Sensitivity

[23] For each model time step t, a model output series $M = y(j, t)$ was transformed with fast Fourier transformation resulting in a power spectrum. The variance σ_i^2 that could be explained by a certain parameter i was calculated from the sum of the power in the spectrum for the frequencies $\omega(i)$, $2\omega(i)$, $3\omega(i)$, $4\omega(i)$ (for further details on the reason for using higher-order frequencies to the order of 4, see *Cukier et al.* [1975]). Whereas the total variance σ^2 was calculated as the sum of the power spectrum over all frequencies. The sensitivity of model output y on parameter i is then calculated as the partial variance, which is the ratio σ_i^2/σ^2.

Table 2. Parameter Ranges for TOPMODEL

Name	Symbol	Units	Minimum	Maximum	FAST Frequency
Initial subsurface flow	qs_0	M	1e-5	6e-5	19
Soil transmissivity (log transformed)	ln *Te*	log(m^2 /h)	−7e-1	−4e-1	59
Shape of the transmissivity curve	*m*		1e-2	4e-2	91
Initial root zone storage deficit	Sr_0	m	1e-3	4e-2	113
Maximum root zone storage deficit	Sr_{max}	m	1e-1	1e+0	133
Unsaturated zone time delay[a]	*td*		−3e+0	1e+0	143
Channel flow velocity	*vr*	m/h	8e+2	2e+3	149
Surface hydraulic conductivity	k_0	m/h	1e-3	1e-2	157
Capillary drive	CD		1e-1	1e+0	161

[a]The unsaturated zone time delay is dimensionless if it is negative and h/m when it is positive.

2.2. EFAST

[24] We used the implementation of EFAST from the software package SimLab [*Saltelli et al.*, 1999; *Saltelli and Bolado*, 1998]. In EFAST, total order sensitivity can also be determined. This sacrifices the efficiency of FAST to obtain simultaneously the FOPV with a limited number of runs, thus only a fraction of the runs is available to estimate the sensitivity for a certain parameter. While the basic idea to vary parameters with a certain frequency in the model run dimension remains, some modifications to the algorithm need to be introduces in order to asses total order sensitivity. The reader is referred to *Saltelli et al.* [1999] and *Saltelli and Bolado* [1998] for further details.

2.3. Sobol's Method

[25] For Sobol's method, a special sampling scheme is applied as well. For a given sampling size N and n parameters, a sub sample size N_s is calculated as $N_s = N/2n + 2$. Parameters θ_i are then sampled randomly for two sub sample sets M_1 and M_2, each consisting of N_s independent parameter sets. Variances are then estimated by evaluating the model for parameter sets, where one parameter in M_1 is replaced by the corresponding parameter of M_2, thereby assessing the effect of changing this single parameter. For further details, see *Sobol* [2001] and *Saltelli* [2002].

2.4. Study Regions

2.4.1. Huagrahuma Catchment, Ecuador

[26] The Huagrahuma catchment is located in the southern Ecuadoran Andes, as part of the Paute river basin (Figure 3). The geology consists of Cretaceous and early Tertiary lavas and andesitic volcanoclastic deposits, shaped and compacted by glacier activity during the last ice age [*Hungerbühler et al.*, 2002]. The hydraulic conductivity of the bedrock is low, particularly compared to the hydraulic conductivity of the thin layer of volcanic ashes that constitute the soil layer [*Buytaert et al.*, 2005]. On average, the soil layer is about 80 cm thick, with some bedrock outcroppings at convex locations and hilltops [*Buytaert et al.*, 2006a]. No deep aquifers are present, and water flow is restricted to overland flow and subsurface flow in the soil layer above the bedrock. The vegetation of the Huagrahuma site consists of noetropical alpine grasses and shrubs and some low statured cloud forest. The climate regime is bimodal, with a average annual precipitation of around 1300 mm y^{-1} but a very low seasonality. Precipitation is characterized by frequent low-intensity events (drizzle), resulting in around 75% of wet days throughout the year.

2.4.2. Weisseritz Catchment

[27] The catchment of the Wilde Weisseritz upstream of the gauging station Ammelsdorf (49.3 km^2) served as a second case study. The catchment is situated in the eastern Ore Mountains at the Czech-German border (Figure 3) and has an elevation of 530 m to about 900 m asl. Slopes are gentle with an average of 7°, 99% are <20°; calculated from a 90 m digital elevation model [*SRTM*, 2002]. Soils are mostly shallow Cambisols of 1 to 2 m thickness. Land use is dominated by forests (≈30%) and agriculture (≈50%). The climate is moderate with mean temperatures of 11°C and 1°C for the periods April–September and October–March, respectively. Annual precipitation for this catchment is 1120 mm/yr for the two years of the simulation period from 1 June 2000 until 1 June 2002. During winter, the catchment usually has a snow cover of up to about 1 m for 1 to 4 months with high flows during the snowmelt period (Figure 5 shows the pronounced peaks

Table 3. Parameters of the Model WaSiM-ETH Used for the Sensitivity Analysis

Parameter Name	Process	Symbol	Units	Range	FAST Frequency
Temperature limit for snowmelt	snowmelt	T_{m0}	°C	−2–2	41
Difference between snow/rain temperature limit and temperature limit for snowmelt (the first is always higher)	snow accumulation	$T_{R/S}$	°C	0–2	67
Temperature melt index	snowmelt	C_0	mm/°C/day	0.7–2	105
Fraction on snowmelt which is surface runoff	snowmelt	c_{melt}		0.2–0.5	145
TOPMODEL regionalization parameter	base flow	*m*		0.005–0.04	177
Scaling factor for transmissivities	base flow	T_{korr}		0.005–0.4	199
Scaling factor for vertical flow	base flow	K_{korr}		800–8000	219
Recession constant for surface runoff single linear storage	surface runoff	k_D	h	1–120	229
Maximum content of the interflow storage	interflow	SH_{max}	mm	1–150	235
Recession constant for interflow runoff single linear storage	interflow	k_H	h	50–300	243
Precipitation intensity limit	fast infiltration	P_{limit}	mm/h	0.2–20	247

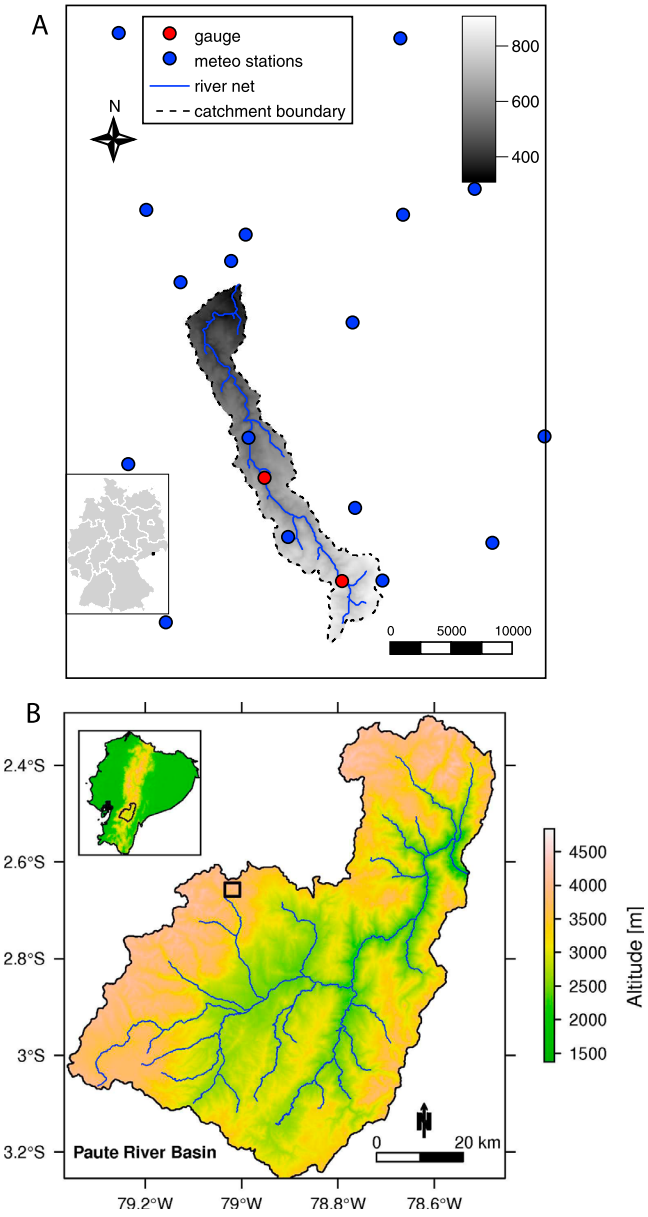

Figure 3. Maps of (a) the Wilde Weisseritz catchment (scales in m; between 50° 40′ and 50° 49′ northern latitude and 13° 35′ and 13° 45′ eastern longitude) and (b) the Huagrahuma catchment within the Paute basin, Ecuador.

during spring). High flows can also be induced by convective events during summer. *WASY Gesellschaft für wasserwirtschaftliche Planung und Systemforschung mbH and Internationales Hochschulinstitut Zittau* [2006] conclude from their analysis based on topography, soil types, and land use that subsurface stormflow is likely to be the dominant process. Meteorological data including precipitation, temperature, wind speed, humidity, and global radiation for 11 surrounding climate stations was obtained from the German Weather Service (Deutscher Wetter Dienst, Climatological data for 11 climate stations around the Weisseritz catchment, 2007). Discharge data, as well as data about land use and soil was obtained from the state office for environment and geology (Landesamt für Umwelt und Geologie Sachsen, Data about land use, soils, discharge, and the digital elevation model, 2007).

2.5. Hydrological Models

2.5.1. TOPMODEL

[28] The hydrological model TOPMODEL is used in this study [*Beven and Kirkby*, 1979]. TOPMODEL is a frequently used model, is based on simple physical approximations, and is well documented in the literature (for an overview, see *Beven et al.* [1995] and *Beven* [1997, 2001]). It has been applied to a wide range of catchments, including regionalization studies [e.g., *Ibbitt et al.*, 2000; *Bastola et al.*, 2008]. The choice of TOPMODEL as a good model structure for the hydrology of the páramo ecosystem is based on extensive field experience [*Buytaert et al.*, 2006b; *Buytaert and Beven*, 2009, 2011]. The steep topography induces large spatial differences in soil moisture and tendency for the generation of overland flow, which are captured by the topographic index. Additionally, the absence of a dry season, and the marked drop of soil hydraulic conductivity in nonsaturated conditions result in continuously wet soils (>60 vol% [*Buytaert et al.*, 2005]). Field research has shown that also in dry periods a saturated soil layer exists above the bedrock, even on steep slopes [*Buytaert et al.*, 2005]. This suggests that the variation in the contributing area is minimal, and that the entire catchment contributes to base flow most of the time.

[29] Finally, the high porosity and low bulk density (typically below 0.6 g/cm^3) [*Buytaert et al.*, 2006a] give rise to easily compressible soils. Bulk density tends to rise and hydraulic conductivity tends to fall with depth [*Buytaert et al.*, 2006a], giving support to the use of a nonlinear transmissivity profile. The TOPMODEL assumption of an exponential function of the storage deficit appears to give a good representation of the recession curves in these catchments.

[30] The model has seven parameters and two initialization values (Table 2). Here qs_0 and Sr_0 initialize the initial subsurface flow per unit area and the initial root zone storage deficit, respectively. The surface hydraulic conductivity (k_0) and the capillary drive (CD) are only used in the infiltration excess routine. The maximum root zone storage deficit (Sr_{max}) is part of the root zone equations, the unsaturated zone time delay (td) controls the flow from unsaturated to saturated zone, while the areal average of the transmissivity (ln Te, log transformed), and the rate of decline of transmissivity with increasing storage deficit (m) are related to the saturated subsurface flow. vr is the river flow velocity. An explicit euler scheme is used to update model states. The model is run with 15 minute time steps.

2.5.2. WaSiM-ETH

[31] WaSiM-ETH is a modular, distributed model (J. Schulla and K. Jasper, Model description WaSiM-ETH, available at http://www.wasim.ch/downloads/doku/wasim/wasim_2007_en.pdf) and was used for the Weisseritz catchment with a regularly spaced grid of 100 m resolution, with 115 × 147 cells to cover the catchment. The model provides methods for the interpolation of meteorological input data. For each cell, a surface runoff storage and an interflow storage are parametrized with the corresponding linear recession constants and a maximum storage size for the interflow storage (see Table 3). The precipitation intensity limit defines a threshold, above which macro pore flow is active and

rainfall enters the lower soil storage directly. Interception (leaf area index–dependent simple bucket), evapotranspiration (Penman-Monteith) and snow (temperature-index approach) are also included as modules. Four parameters of the snow module were investigated more closely. Snow accumulation is determined by the snow-rain temperature limit. The temperature melt index defines the amount of snow melted for each degree and hour the temperature is above the snowmelt limiting temperature (third parameter). Finally, the fraction of snow melt which builds surface runoff is the fourth parameter. The unsaturated zone is described for each subbasin on the basis of the TOPMODEL approach [*Beven and Kirkby*, 1979]. The TOPMODEL regionalization parameter m determines how strong the gradient in the saturation deficit is due to differences in the topographic index. m also enters the equations for the vertical flow q_v (equation (5)) and the base flow Q_B (equation (4)). Vertical flow and base flow are both calibrated with the scaling factors T_{korr} and K_{korr}. Channel flow is routed with a simple storage to account for diffusion.

$$Q_B = T_{\mathrm{korr}} e^{-\gamma} e^{-S_m/m} \tag{4}$$

$$q_v = K_{\mathrm{korr}} k_f e^{-S_i/m} \tag{5}$$

where γ is the mean value of the topographic index, a constant for a given basin, k_f the saturated hydraulic conductivity, S_m and S_i the mean and local saturation deficit for a subbasin, model state variables.

[32] WaSiM-ETH was set up and run 487 times with hourly time steps, the number of required runs (see section 2.1.1) for sensitivity analysis with 11 varying parameters. WaSiM uses an explicit euler time stepping scheme [*Clark and Kavetski*, 2010] to update model states. The set of resulting discharge time series $y(j, t)$, one for each of the N parameter sets was then further analyzed to calculate sensitivities.

3. Results

3.1. Comparison of Sensitivity Analysis Methods With TOPMODEL

[33] Sensitivities were calculated with the following SA methods: (1) Sobol in SimLab 3.2.6 (n = 5632), (2) EFAST in SimLab 2.2 (n = 5000), (3) FAST (SimLab 3.2.6, n = 1289), and (4) FAST (R package [*Reusser*, 2008], n = 487). The number of model runs for EFAST and Sobol was selected as a balance between the reduction of numerical artifacts (e.g., first-order sensitivities outside the possible range from 0 to 1, random fluctuations of TEDPAS) and computation time, while the minimum requirements as suggested by the implementation were used for FAST. The 487 runs for the method in the R package are reported by *Cukier et al.* [1975, 1978] (with the corresponding frequencies listed in Table 2), while the minimum requirement of 1289 simulation runs for SimLab 3.2.6 is undocumented.

[34] Mean absolute difference (MAD) for parameter sensitivities calculated with different sensitivity algorithms are in general smaller than 0.07 (partial variance, unit free) with two exceptions. First for the unsaturated zone time delay td, comparing results from Sobol's method to the other results, MAD is between 0.18 and 0.25. Second, MAD is up to 0.27 when comparing results generated with the EFAST method to the other methods.

[35] The reason for the two exceptions becomes apparent in Figure 4. For each method (Figures 4a–4d), two graphs show the sensitivity of the modeled discharge for different parameters. The top graph shows the initial conditions (qs_0, Sr_0), root zone storage ($Sr_{\max}$), and transmissivity-related parameters (m, ln Te), while the bottom graph shows the remaining parameters. With Sobol's method, the modeled discharge depends on td after the initial period, while sensitivity determined for td with the other methods is low. With EFAST, the parameter with the highest sensitivity shows higher sensitivity compared to the other methods (most clearly visible for m toward the end of the simulation). In addition, with EFAST ln Te has some influence on modeled discharge for June and July.

[36] For all three methods, the TEDPAS is similar for these four variables: initial conditions (qs_0 and Sr_0) are dominant until middle of the simulation, thereafter m and vr are of highest importance. Scrutinizing the 4 parameters qs_0, Sr_0, m, and vr, 90% of the time, the same parameter dominates in all of the methods (rank 1 in sensitivities).

[37] As explained in section 2, the sensitivity is reported as partial variance that can be explained by this parameter at this time step. For example, a value of around 0.7 for parameter m during June indicates that 70% of the observed variation between the model runs $j = 1, \ldots, N$ can be explained by this parameter. The sum over all parameter sensitivities never exceeds 1.0 but may be lower because of the numeric approximation [*Cukier et al.*, 1975] or when parameter interactions are of importance (nonadditive models [*Saltelli et al.*, 2006]). Figure 4e shows the 25 best (selected according to RMSE) modeled discharge time series in black and the measured time series in gray.

3.2. FAST WaSiM-ETH

[38] Sensitivities for the computationally expensive WaSiM-ETH model were calculated with the FAST method only (FAST frequencies are listed in Table 3). Figure 5 shows TEDPAS of the modeled discharge on the annual time scale. Figures 5a–5c show the sensitivity of the modeled discharge for different parameters, grouped according to the different model components. Figure 5a shows the snowmelt-related parameters. The three saturation deficit–related parameters, m, T_{korr} and K_{korr}, are shown in Figure 5b. Figure 5c shows the remaining parameters, k_D, k_H, $SH_{\max}$, and P_{limit}. Last, Figure 5d shows the 25 best (selected according to RMSE) modeled discharge time series (in black) and the measured time series (in grey).

[39] We observe that snow-related parameters are important during winter and spring, which is as expected. Also, saturation deficit–related parameters are unimportant for discharge during snowmelt periods. Note that the plots only show first-order effects. Influence of interacting parameters are not visible from these plots. Therefore, in order to exclude the influence of a parameter, higher-order terms (or total order sensitivity) need to be calculated. TEDPAS of the three saturation deficit–related parameters is highly correlated with correlation coefficients of >0.62. This suggests a strong interaction of the these parameters, as will be further discussed in section 4. The correlation between $T_{R/S}$ and T_{m0} is also high with a coefficient of 0.73, while correlation coefficients between all the other time series are

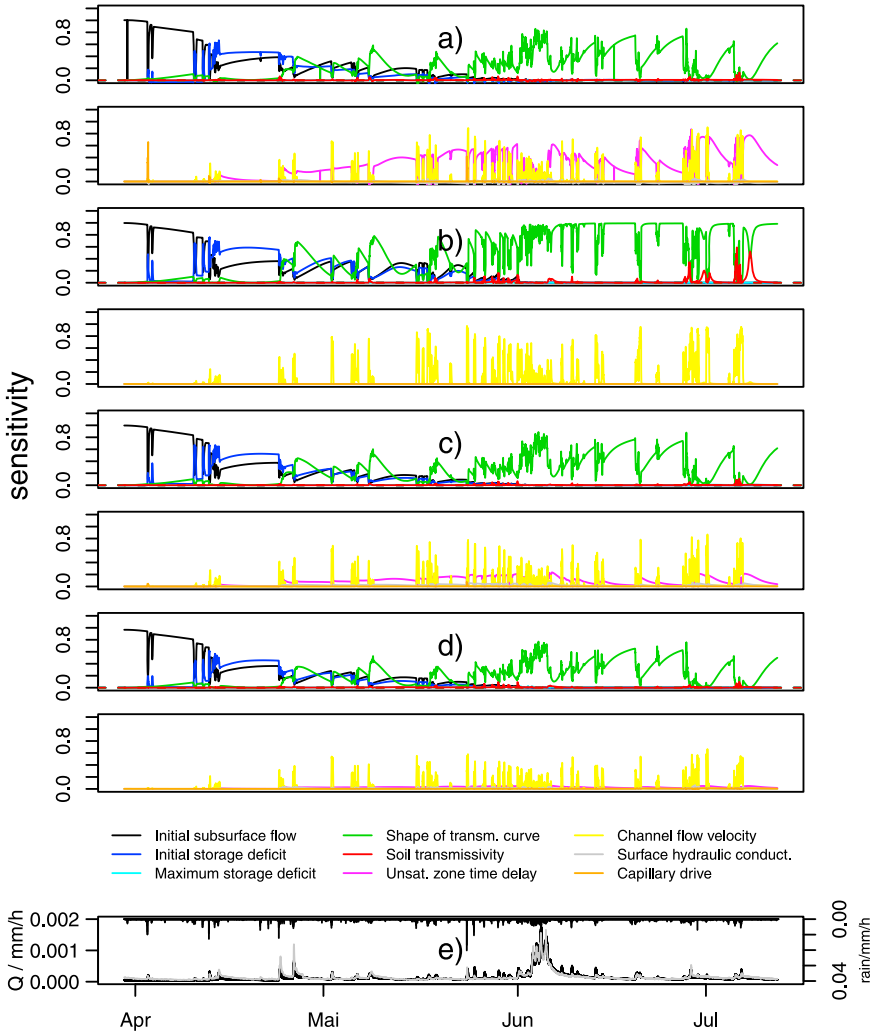

Figure 4. Parameter sensitivities of TOPMODEL, with (a) Sobol's method (SimLab 3.2.6, n = 5632), (b) EFAST (SimLab 2.2, n = 5000), (c) FAST (SimLab 3.2.6, n = 1289), and (d) FAST (R package, n = 487). Initial conditions, transmissivity, and root zone storage–related parameters are shown always in the top panel, and the remaining parameters are shown in the bottom panel. (e) Discharge for the 25 best simulation runs and the observation.

<0.5. For TEDPAS at the event time scale, see *Reusser and Zehe* [2011].

4. Discussion

4.1. Comparing Sensitivity Methods for TOPMODEL

[40] All 3 SA methods result in very similar TEDPAS for four important variables. Differences exist for two parameters: the model appears to be more sensitive for ln *TE* in July with EFAST. For *td* sensitivity is high with Sobol's method. With FAST the model shows only minor sensitivity for the two parameters.

[41] We suggest that four potential sources for such differences exist: (1) a rough response surface of the discharge dependent on the parameters from the explicit Euler scheme, as suggested by *Kavetski and Clark* [2010], causing small parameter variations to have a potentially high effect on sensitivity, (2) the possibility of interferences causing errors for FAST, as suggested by *Saltelli and Bolado* [1998], (3) the different sampling schemes, and (4) the different algorithms to compute the partial variance might cause differences in the estimated sensitivities.

[42] We identified a rough response surface (source 1) of the discharge when varying the initial storage deficit Sr_0 and the unsaturated zone time delay *td* by visual inspection (not shown; details are available from the corresponding author). The modeled discharge shows a pronounced increase if the time delay *td* approaches 0, while the dependence of the discharge on the initial storage deficit is erratic. We also identified some effect on sensitivities when interchanging the FAST sampling frequencies ω_i (results not shown, details available from the corresponding author) which may be caused by interferences (source 2) [*Saltelli and Bolado*, 1998] or the rough surface [*Kavetski and Clark*, 2010]. However, the identification of the dominant process from the sensitivities was not affected by these effects, as indicated by stable rank 1 assignment. Thus, while this could be a possible explanation for differences between the different SA methods, it does not affect the value of the FAST method

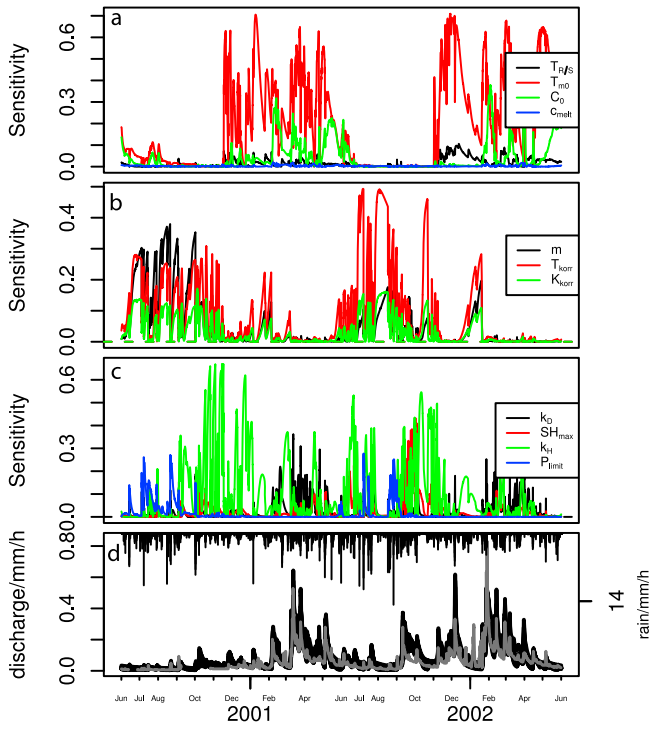

Figure 5. TEDPAS for WaSiM-ETH. The parameter sensitivity of the modeled discharge for (a) snow model–related parameters, (b) saturation deficit–related parameters, and (c) the remaining parameters, k_D, k_H, SH_{max}, and P_{limit}. The sensitivity is reported as partial variance that can be explained by the corresponding parameter. (d) The 25 best modeled discharge time series (in black) and the measured time series (in grey).

for the identification of dominant model components from TEDPAS.

[43] The sampling schemes (source 3) are tightly linked to the evaluation algorithms (source 4) for FAST and Sobol's method, making it challenging to disentangle these two possible sources of the differences. However, we may still note a few facts. First, the sampling density is about a factor of 5 or more higher for Sobol's method (n = 5632) compared to the other methods. This also holds for EFAST (n = 5000), as in EFAST only a fraction of the model runs can be used to determine the FOPV for a certain parameter. With the higher sampling density for Sobol's method, we have a higher chance that the pronounced increase in discharge when td approaches 0 is sampled appropriately compared to the other methods. Second, the fundamental principle for the evaluation algorithm is the same for the three FAST-based methods. Thus, the differences observed by the three methods are caused by the sampling scheme. The main differences are the higher sensitivity for the soil transmissivity and the shape of the transmissivity curve for EFAST compared to the other two and the higher unsaturated zone time delay of FAST (SimLab) compared to the other two.

[44] While the pronounced increase of the modeled discharge with the time delay td approaching 0 in connection with the higher sampling density for Sobol's method and the sampling scheme as the source of the differences between the three variants of the FAST method indicate the sampling (and curse of dimensionality) to be the main reason, the ultimate identification of the difference remains an issue for further research.

4.2. TEDPAS of TOPMODEL

[45] The results of the temporal sensitivity analysis for TOPMODEL on the Huagrahuma catchment are very much as expected. Although the effect of the initialization values decreases over time, several thousands of time steps are required for the effect to die out. This highlights the importance of a warm-up period when using the model in prediction mode. Model parameter *m*, which defines the shape of the transmissivity curve, is known to be a sensitive parameter [e.g., *Buytaert and Beven*, 2009], with an effect over the entire recession curve. The channel velocity parameter *vr* has a major effect on the time to peak flow, and to a lesser extent on the shape of the steep part of the recession curve. The observation that the sensitivity of *vr* shows a peaky behavior, with high sensitivity related to precipitation events is therefore physically plausible.

[46] Other parameters, particularly Sr_{max} and *td* are known to be relatively insensitive in the ecosystem. Evapotranspiration is nearly independent of soil moisture in the continuously wet grasslands. The organic soils also tend to have a very high soil moisture, accelerating the flow from the unsaturated to the saturated zone [*Buytaert et al.*, 2006b; *Buytaert and Beven*, 2009]. Therefore, the relatively high sensitivity of *td* in the Sobol method is not very clear, but may be related to interaction between *td* and *m*. It is indeed possible that an artificially high time delay between the unsaturated and saturated zone compensates for too quick saturated flow, which is controlled by both ln *Te* and *m*. The peaks in the sensitivity of CD are related to the very rare occurrence of infiltration excess overland flow in the study catchment. Hence, this model routine is nearly always inactive, apart from a few occurrences of intense precipitation events.

[47] Finally, it is interesting to note that the sensitivity of the parameters does not change during the major high-flow event in early June. This suggests that the similar model structures are operational during this period as during the dry periods and that the parsimonious model structure performs well for a relatively wide range of hydrological conditions. This is in agreement with results from *van Werkhoven et al.* [2008a], who also find a comparatively small variability in the patterns of dominating hydrological processes for a given catchment in their interannual comparison of parameter sensitivity for a number of catchments along a hydroclimatic gradient in the United States. However, these patterns of dominating hydrological processes are also dependent on the conditions of a given year. Thus, *van Werkhoven et al.* [2008a] summarize their results in a table presenting parameter sensitivity for three types of catchments (dry, medium wet, and wet) and three climatic conditions (dry, mid and wet years).

4.3. TEDPAS of WaSiM-ETH

[48] Checking the yearly patterns in TEDPAS is a first assessment to verify the model structure. A first pattern on the annual scale is, as expected, the model showing high sensitivities for snow-related parameters during winter and spring. However, simulated discharge shows some sensitivity for snow-related parameters approximately until end

of June, although in reality snow cover was completely depleted by the end of May. We, thus, suggest to revise the parameter range of $C0$ used for sensitivity analysis for this catchment for subsequent analyses.

[49] A second pattern on the annual scale is the first-order insensitivity of discharge for saturation deficit–related parameters during snowmelt periods. This is plausible since discharge is (by definition) dominated by meltwater during these periods.

[50] This is in agreement with results from *van Werkhoven et al.* [2008a], who also find plausible, typical patterns of sensitivity for the Sacramento Soil Moisture Accounting Model for certain combinations of hydroclimatic and model storage conditions: "In each case the respective conditions for sensitivity reflect conditions when the storage is actively impacting model predictions and reasonably represents the expected dominant processes."

[51] Plausibility checks are also possible on the event time scale by checking the sequence of parameter sensitivity and comparing it to expectations on the basis of model design [*Reusser and Zehe*, 2011; *Sieber and Uhlenbrook*, 2005]. Performing TEDPAS has two additional benefits. First, TEDPAS is also a valuable tool for calibration of model parameters. The fraction of meltwater contributing to overland flow c_{melt} will hardly ever be well identifiable, because the sensitivity of simulated discharge for this parameter is always smaller than 2% despite the large range for c_{melt} of 20%–50%. Note that this result needs to be confirmed with calculation of higher-order sensitivities in order to definitely exclude any influence of c_{melt}.

[52] As stated before, sensitivity is a necessary but not sufficient condition for identifiability [*Wagener et al.*, 2003]: parameters may show high sensitivity but be poorly identifiable if compensatory effects between two parameters make them interdependent. Such compensatory effects of parameters may be detected (second benefit), indicated by a highly correlated sensitivity of the model output for multiple parameters. Correlated model parameters can be a major source for poor identifiability in hydrological modeling [*Bárdossy*, 2007]. In the case study, we observe correlated parameters for the saturation deficit–related parameters, which will complicate proper identification of these parameters during calibration. *Brun et al.* [2001] demonstrate how to derive a set of identifiable parameters from such a set of correlated parameters.

[53] We see TEDPAS and identifiability analysis [*Wagener et al.*, 2003] as complementary analytical methods, with TEDPAS indicating dominant model components, while identifiability analysis shows how the parameters have to vary in time to reproduce observed quantities, which is another important piece of information to identify model structural errors.

5. Conclusions

[54] We demonstrated that SA can provide valuable information for improved model understanding, which goes beyond the most often selected approach to date, where the most influential parameters for calibration are determined.

[55] For our case study we found that TEDPAS is consistent with expectations on both the annual and event time scale. In addition, SA may enhance calibration because time periods of high parameter sensitivity are the relevant periods for calibration. On the basis of the understanding of the importance of parameters, a priori assumptions may be revised and field experiments may be guided. Finally, the method allows us to detect compensatory effects of parameters, which we found for the saturation deficit–related parameters of WaSiM-ETH.

[56] An optimal use of SA methods needs highly efficient methods. We applied such a highly efficient SA method called FAST to two rainfall-runoff models. FAST allows us to determine global sensitivity for parameters with only a limited number of model runs. The current case study required around 150 and 490 runs for 6 and 11 parameters, respectively. Our analysis of parameter sensitivities for WaSiM-ETH would not have been possible without this very efficient sampling scheme. In addition, on the basis of the efficient calculation of sensitivities from model output variables, entire time series of parameter sensitivity can be calculated (e.g., for discharge).

[57] That extended SA presented here is only a first step in obtaining better model understanding. The ultimate goal is to gain insight into model behavior by answering the three research questions: (1) During which periods is the model reproducing observed quantities and dynamics? (2) What is the nature of the error in times of poor model performance? (3) Which components of the model are causing this error? The first two research questions may be answered using the time series of grouped errors (TIGER) method [*Reusser et al.*, 2009], while the third question is answered using SA methods as presented here. This approach closely relates to the framework for diagnostic model evaluation proposed by *Gupta et al.* [2008].

[58] **Acknowledgments.** The study has been funded as part of OPAQUE (operational discharge and flooding predictions in head catchments), a project within the BMBF-Förderaktivität "Risikomanagement extremer Hochwasserereignisse" (RIMAX). W.B. was funded by a Marie Curie Intra-European Fellowship at the University of Lancaster during the implementation of TOPMODEL for the Huagrahuma catchment. We would like to thank Jenny Eckart for her support with the data preprocessing for WaSiM-ETH. A major part of the analysis was carried out with the open source statistical software R and contributed packages, and we would like to thank its community.

References

Bárdossy, A. (2007), Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci.*, *11*(2), 703–710.

Bastola, S., H. Ishidaira, and K. Takeuchi (2008), Regionalisation of hydrological model parameters under parameter uncertainty: A case study involving TOPMODEL and basins across the globe, *J. Hydrol.*, *357*(3–4), 188–206, doi:10.1016/j.jhydrol.2008.05.007.

Benke, K. K., K. E. Lowell, and A. J. Hamilton (2008), Parameter uncertainty, sensitivity analysis and prediction error in a water-balance hydrological model, *Math. Comput. Modell.*, *47*(11–12), 1134–1149.

Beven, K. J. (1997), TOPMODEL: A critique, *Hydrol. Processes*, *11*(9), 1069–1085.

Beven, K. J. (2001), *Rainfall-Runoff Modelling: The Primer*, John Wiley, Chichester, U. K.

Beven, K. J. (2002), Towards a coherent philosophy for modelling the environment, *Proc. R. Soc. London, Ser. A*, *458*(2026), 2465–2484, doi:10.1098/rspa.2002.0986.

Beven, K. J., and M. Kirkby (1979), A physically based, variable contributing area model of basin hydrology, *Hydrol. Sci. J.*, *24*(1), 43–69.

Beven, K. J., R. Lamb, P. Quinn, R. Romanowicz, and J. E. Freer (1995), TOPMODEL, in *Computer Models of Watershed Hydrology*, pp. 627–668, Water Resour. Publ., Highlands Ranch, Colo.

Brun, R., P. Reichert, and H. Künsch (2001), Practical identifiability analysis of large environmental simulation models, *Water Resour. Res.*, *37*(4), 1015–1030.

Buytaert, W., and K. Beven (2009), Regionalization as a learning process, *Water Resour. Res.*, *45*, W11419, doi:10.1029/2008WR007359.

Buytaert, W., and K. Beven (2011), Models as multiple working hypotheses: Hydrological simulation of tropical alpine wetlands, *Hydrol. Processes*, *25*, 1784–1799, doi:10.1002/hyp.7936.

Buytaert, W., G. Wyseure, B. De Bièvre, and J. Deckers (2005), The effect of land-use changes on the hydrological behaviour of Histic Andosols in south Ecuador, *Hydrol. Processes*, *19*, 3985–3997, doi:10.1002/hyp.5867.

Buytaert, W., J. Deckers, and G. Wyseure (2006a), Description and classification of nonallophanic Andosols in south Ecuadorian alpine grasslands (páramo), *Geomorphology*, *73*(3–4), 207–221, doi:10.1016/j.geomorph.2005.06.012.

Buytaert, W., R. Célleri, B. De Bièvre, F. Cisneros, G. Wyseure, J. Deckers, and R. Hofstede (2006b), Human impact on the hydrology of the Andean páramos, *Earth Sci. Rev.*, *79*(1–2), 53–72, doi:10.1016/j.earscirev.2006.06.002.

Christiaens, K. (2002), Use of sensitivity and uncertainty measures in distributed hydrological modeling with an application to the MIKE SHE model, *Water Resour. Res.*, *38*(9), 1169, doi:10.1029/2001WR000478.

Clark, M. P., and D. Kavetski (2010), Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes, *Water Resour. Res.*, *46*, W10510, doi:10.1029/2009WR008894.

Cloke, H., F. Pappenberger, and J.-P. Renaud (2008), Multi-method global sensitivity analysis (MMGSA) for modelling floodplain hydrological processes, *Hydrol. Processes*, *22*, 1660–1674.

Cukier, R. I., C. M. Fortuin, K. E. Shuler, A. G. Petschek, and J. H. Schaibly (1973), Study of sensitivity of coupled reaction systems to uncertainties in rate coefficients. 1. Theory, *J. Chem. Phys.*, *59*(8), 3873–3878.

Cukier, R. I., J. H. Schaibly, and K. E. Shuler (1975), Study of sensitivity of coupled reaction systems to uncertainties in rate coefficients. 3. Analysis of approximations, *J. Chem. Phys.*, *63*(3), 1140–1149.

Cukier, R. I., H. B. Levine, and K. E. Shuler (1978), Non-linear sensitivity analysis of multi-parameter model systems, *J. Comput. Phys.*, *26*(1), 1–42.

Cullmann, J., V. Mishra, and R. Peters (2006), Flow analysis with WaSiM-ETH model parameter sensitivity at different scales, *Adv. Geosci.*, *9*, 73–77.

Deflandre, A., R. J. Williams, F. J. Elorza, J. Mira, and D. B. Boorman (2006), Analysis of the QUESTOR water quality model using a Fourier amplitude sensitivity test (FAST) for two UK rivers, *Sci. Total Environ.*, *360*(1–3), 290–304.

Demaria, E. M., B. Nijssen, and T. Wagener (2007), Monte Carlo sensitivity analysis of land surface parameters using the variable infiltration capacity model, *J. Geophys. Res.*, *112*, D11113, doi:10.1029/2006JD007534.

Fang, S. F., G. Z. Gertner, S. Shinkareva, G. X. Wang, and A. Anderson (2003), Improved generalized Fourier amplitude sensitivity test (FAST) for model assessment, *Stat. Comput.*, *13*(3), 221–226.

Foglia, L., M. C. Hill, S. W. Mehl, and P. Burlando (2009), Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function, *Water Resour. Res.*, *45*, W06427, doi:10.1029/2008WR007255.

Frey, H. C., and S. R. Patil (2002), Identification and review of sensitivity analysis methods, *Risk Anal.*, *22*(3), 553–578.

Gupta, H. V., T. Wagener, and Y. Q. Liu (2008), Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, *Hydrol. Processes*, *22*, 3802–3813.

Hornberger, G. M., and R. C. Spear (1981), An approach to the preliminary-analysis of environmental systems, *J. Environ. Manage.*, *12*(1), 7–18.

Hungerbühler, D., M. Steinmann, W. Winkler, D. Seward, A. Egüez, D. Peterson, U. Helg, and C. Hammer (2002), Neogene stratigraphy and Andean geodynamics of southern Ecuador, *Earth Sci. Rev.*, *57*(1–2), 75–124, doi:10.1016/S0012-8252(01)00071-X.

Ibbitt, R., R. Henderson, J. Copeland, and D. Wratt (2000), Simulating mountain runoff with meso-scale weather model rainfall estimates: A New Zealand experience, *J. Hydrol.*, *239*(1–4), 19–32, doi:10.1016/S0022-1694(00)00351-6.

Kavetski, D., and M. P. Clark (2010), Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, *Water Resour. Res.*, *46*, W10511, doi:10.1029/2009WR008896.

Klaus, J., and E. Zehe (2010), Modelling rapid flow response of a tile-drained field site using a 2D physically based model: Assessment of equifinal model setups, *Hydrol. Processes*, *24*, 1595–1609, doi:10.1002/hyp.7687.

McIntyre, N. R., T. Wagener, H. S. Wheater, and S. C. Chapra (2003), Risk-based modelling of surface water quality: A case study of the Charles River, Massachusetts, *J. Hydrol.*, *274*(1–4), 225–247.

McRae, G. J., J. W. Tilden, and J. H. Seinfeld (1982), Global sensitivity analysis—A computational implementation of the Fourier amplitude sensitivity test (FAST), *Comput. Chem. Eng.*, *6*, 15–25.

Muleta, M., and J. Nicklow (2005), Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model, *J. Hydrol.*, *306*(1–4), 127–145, doi:10.1016/j.jhydrol.2004.09.005.

Pappenberger, F., I. Iorgulescu, and K. J. Beven (2006), Sensitivity analysis based on regional splits and regression trees (SARS-RT), *Environ. Modell. Software*, *21*(7), 976–990.

Pappenberger, F., K. J. Beven, M. Ratto, and P. Matgen (2008), Multi-method global sensitivity analysis of flood inundation models, *Adv. Water Resour.*, *31*, 1–14.

R Development Core Team (2008), R: A Language and Environment for Statistical Computing, R Found. for Stat. Comput., Vienna.

Reusser, D. E. (2008), Implementation of the Fourier amplitude sensitivity test (FAST).

Reusser, D. E., and E. Zehe (2011), Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity, *Water Resour. Res.*, doi:10.1029/2010WR009946, in press.

Reusser, D. E., T. Blume, B. Schaefli, and E. Zehe (2009), Analysing the temporal dynamics of model performance for hydrological models, *Hydrol. Earth Syst. Sci.*, *13*(7), 999–1018, doi:10.5194/hess-13-999-2009.

Saltelli, A. (2002), Making best use of model evaluations to compute sensitivity indices, *Comput. Phys. Commun.*, *145*(2), 280–297.

Saltelli, A., and R. Bolado (1998), An alternative way to compute Fourier amplitude sensitivity test (FAST), *Comput. Stat. Data Anal.*, *26*(4), 445–460.

Saltelli, A., S. Tarantola, and K. P.-S. Chan (1999), A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics*, *41*(1), 39–56, doi:10.2307/1270993.

Saltelli, A., S. Tarantola, and F. Campolongo (2000), Sensitivity analysis as an ingredient of modeling, *Stat. Sci.*, *15*(4), 377–395.

Saltelli, A., M. Ratto, S. Tarantola, and F. Campolongo (2006), Sensitivity analysis practices: Strategies for model-based inference, *Reliab. Eng. Syst. Safety*, *91*(10–11), 1109–1125.

Schaibly, J. H., and K. E. Shuler (1973), Study of sensitivity of coupled reaction systems to uncertainties in rate coefficients. 2. Applications, *J. Chem. Phys.*, *59*(8), 3879–3888.

Sieber, A., and S. Uhlenbrook (2005), Sensitivity analyses of a distributed catchment model to verify the model structure, *J. Hydrol.*, *310*(1–4), 216–235.

Sobol, I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simul.*, *55*(1–3), 271–280, doi:10.1016/S0378-4754(00)00270-6.

SRTM (2002), Shuttle Radar Topography Mission (SRTM) Elevation Data Set, http://seamless.usgs.gov, U. S. Geol Surv., Boulder, Colo.

Tang, Y., P. Reed, T. Wagener, and K. van Werkhoven (2007a), Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci.*, *11*(2), 793–817.

Tang, Y., P. Reed, K. van Werkhoven, and T. Wagener (2007b), Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis, *Water Resour. Res.*, *43*, W06415, doi:10.1029/2006WR005813.

Van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan (2006), A global sensitivity analysis tool for the parameters of multi-variable catchment models, *J. Hydrol.*, *324*(1–4), 10–23.

van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2008a), Characterization of watershed model behavior across a hydroclimatic gradient, *Water Resour. Res.*, *44*, W01429, doi:10.1029/2007WR006271.

van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2008b), Rainfall characteristics define the value of streamflow observations for distributed watershed model identification, *Geophys. Res. Lett.*, *35*, L11403, doi:10.1029/2008GL034162.

van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2009), Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models, *Adv. Water Resour.*, *32*(8), 1154–1169, doi:10.1016/j.advwatres.2009.03.002.

Wagener, T., N. R. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Processes*, *17*, 455–476.

Wagener, T., K. van Werkhoven, P. Reed, and Y. Tang (2009), Multiobjective sensitivity analysis to understand the information content in streamflow observations for distributed watershed modeling, *Water Resour. Res.*, *45*, W02501, doi:10.1029/2008WR007347.

WASY Gesellschaft für wasserwirtschaftliche Planung und Systemforschung mbH and Internationales Hochschulinstitut Zittau (2006), Schätzung dominanter Abflussprozesse mit WBS FLAB, technical report, Berlin.

Wegman, E. (1990), Hyperdimensional data analysis using parallel coordinates, *J. Am. Stat. Assoc.*, *85*(411), 664–675.

Zehe, E., H. Elsenbeer, F. Lindenmaier, K. Schulz, and G. Bloschl (2007), Patterns of predictability in hydrological threshold systems, *Water Resour. Res.*, *43*, W07434, doi:10.1029/2006WR005589.

W. Buytaert, Department of Civil and Environmental Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.

D. E. Reusser, Institute of Earth and Environmental Sciences, University of Potsdam, D-14476 Potsdam, Germany. (reusser@pik-potsdam.de)

E. Zehe, Institute for Water and Environment, TU München, Arcisstr. 21, D-80333 Munich, Germany.