

Bayesian inference of uncertainties in precipitation-streamflow modeling in a snow affected catchment

J. J. Koskela,¹ B. W. F. Croke,² H. Koivusalo,¹ A. J. Jakeman,² and T. Kokkonen¹

Received 27 December 2011; revised 14 August 2012; accepted 2 October 2012; published 10 November 2012.

[1] Bayesian inference is used to study the effect of precipitation and model structural uncertainty on estimates of model parameters and confidence limits of predictive variables in a conceptual rainfall-runoff model in the snow-fed Rudbäck catchment (142 ha) in southern Finland. The IHACRES model is coupled with a simple degree day model to account for snow accumulation and melt. The posterior probability distribution of the model parameters is sampled by using the Differential Evolution Adaptive Metropolis (DREAM_(ZS)) algorithm and the generalized likelihood function. Precipitation uncertainty is taken into account by introducing additional latent variables that were used as multipliers for individual storm events. Results suggest that occasional snow water equivalent (SWE) observations together with daily streamflow observations do not contain enough information to simultaneously identify model parameters, precipitation uncertainty and model structural uncertainty in the Rudbäck catchment. The addition of an autoregressive component to account for model structure error and latent variables having uniform priors to account for input uncertainty lead to dubious posterior distributions of model parameters. Thus our hypothesis that informative priors for latent variables could be replaced by additional SWE data could not be confirmed. The model was found to work adequately in 1-day-ahead simulation mode, but the results were poor in the simulation batch mode. This was caused by the interaction of parameters that were used to describe different sources of uncertainty. The findings may have lessons for other cases where parameterizations are similarly high in relation to available prior information.

Citation: Koskela J. J., B. W. F. Croke, H. Koivusalo, A. J. Jakeman, T. Kokkonen (2012), Bayesian inference of uncertainties in precipitation-streamflow modeling in a snow affected catchment, *Water Resour. Res.*, 48, W11513, doi:10.1029/2011WR011773.

1. Introduction

[2] It is the nature of conceptual rainfall-runoff models that at least some of the model parameters must be calibrated against observations. Values of these parameters cannot be measured in the field because they aggregate properties of the whole catchment, which is described as a lumped system by the model. For several decades automatic calibration schemes for the estimation of these parameters against historical data and the related problems and challenges have received much attention [e.g., *Sorooshian and Dracup*, 1980; *Duan et al.*, 1992]. More recently focus has turned into utilization of multiple sources and measures of information in model calibration [e.g., *Gupta et al.*, 1998; *Seibert and McDonnell*, 2002;

Gupta et al., 2008; *Efstratiadis and Koutsoyiannis*, 2010] and into better understanding of the uncertainties affecting the calibration and model simulations in a predictive mode [e.g., *Beven and Binley*, 1992; *Vrugt et al.*, 2009a]. Some recent contributions attempt to take into account in the calibration procedure all the factors that may contribute to uncertainty in the parameter estimates [*Kavetski et al.*, 2002; *Ajami et al.*, 2007; *Renard et al.*, 2010; *Renard et al.*, 2011].

[3] In rainfall-runoff modeling there are four major sources of uncertainty that can be linked to the total uncertainty of model parameters and ensuing predictions/simulations. These are well recognized as being with respect to inputs, model structure, parameters and output calibration data. Several approaches have been devised to approximate the total uncertainty of rainfall-runoff simulations. Very popular among these approaches is the Generalized Likelihood Uncertainty Estimation (GLUE) procedure [*Beven and Binley*, 1992; *Beven and Freer*, 2001; *Beven*, 2006]. To improve model performance, however, it would be essential to know how much different sources contribute to total uncertainty. Such information can guide research efforts and measurement activity can then be allocated to the most influential factors. The contribution of the different sources of uncertainty cannot be distinguished by using GLUE because total uncertainty is aggregated into parameter uncertainty within the

¹Department of Civil and Environmental Engineering, School of Engineering, Aalto University, Aalto, Finland.

²Integrated Catchment Assessment and Management Centre, Fenner School of Environment and Society, National Centre for Groundwater Research and Training, Australian National University, Canberra, ACT, Australia.

Corresponding author: J. J. Koskela, Department of Civil and Environmental Engineering, School of Engineering, Aalto University, PO Box 15300, FI-00076 Aalto, Finland. (jarkko.koskela@aalto.fi)

equifinality concept. Furthermore, the subjectively chosen thresholds between behavioral and nonbehavioral models and common use of informal likelihood functions in GLUE have led to an active debate about whether it gives trustworthy estimates of uncertainties in rainfall-runoff modeling [e.g., *Mantovan and Todini*, 2006; *Beven et al.*, 2007, 2008; *Stedinger et al.*, 2008].

[4] In boreal regions snow processes are an important part of the hydrological cycle. To our knowledge specific snow modeling research focusing on uncertainty and in particular separation of different sources of uncertainties related to snow modeling are scarce. *Rutter et al.* [2009] compared 33 snowpack models in different hydrometeorological and forest canopy conditions. The comparison indicates that model structure uncertainty plays an important part in total uncertainty of snow modeling. Furthermore, their research reveals that it is more difficult and uncertain to model snow processes at forested sites than open sites. *Franz et al.* [2010] used Bayesian model averaging (BMA) to combine results of 12 snow models to assess uncertainty in hydrologic prediction. They found out that different snow models perform best at different locations and time periods. They suggest that consideration of multiple models would provide useful information for probabilistic hydrologic prediction. *He et al.* [2011] studied the parameter uncertainty of the SNOW17 model using Generalized Sensitivity Analysis and Differential Evolution Adaptive Metropolis (DREAM) in 12 contracting study sites in the U.S. They showed that parameter uncertainty of the model depends on the study site and that uncertainty ranges of some model parameters show correlation with forcing data (precipitation and temperature). Uncertainty in input variables, model structure and snow observations were not explicitly addressed in the research. Although it seems that other specific studies on snow modeling using Bayesian inference are missing, there are individual published uncertainty studies where a snow process scheme is included in the hydrological model. *Kavetski et al.* [2006b] applied the Bayesian Total Error Analysis (BATEA) framework in the two case studies in U.S. (French Broad River and South Branch Potomac River) using rainfall and runoff measurements (without snow observations) as the calibration data. Snow process parameters identified with the standard least squares calibration were affected when the input uncertainty model was included and the hydrological model was recalibrated. Also *Clark and Vrugt* [2006] pointed out that input uncertainty may seriously bias the parameter estimates in snow modeling. These studies highlight the need for addressing input uncertainty in the calibration of a snow model.

[5] It is well established that the estimates of the model parameters and the confidence limits of streamflow simulation are affected when data uncertainty is taken into account in the calibration of conceptual rainfall-runoff models [e.g., *Huard and Mailhot*, 2008; *Croke*, 2009; *Thyer et al.*, 2009]. *Kavetski et al.* [2002, 2006a] pointed out three possible outcomes when errors in the input variables are neglected in the calibration of hydrological models:

[6] 1. Parameter bias caused by possible input errors is likely to be different from catchment to catchment and thus regionalization of parameters to ungauged catchments may be significantly confounded.

[7] 2. Biased parameters may yield biased predictions and, because input uncertainty affects the structure of parameter uncertainty, the confidence limits of the parameters are likely to be erroneous if input uncertainty is neglected in calibration.

[8] 3. Ignorance about different types of errors in hydrological models prevents proper analysis of model error and model adequacy that may help to improve the model.

[9] *Kavetski et al.* [2002, 2006a, 2006b] and *Kuczera et al.* [2006] proposed the BATEA framework to separate and take into account all the uncertainty sources in model calibration. In BATEA, input (precipitation) uncertainty is included by using additional latent variables that are multipliers for individual storm events and calibrated simultaneously with model parameters. *Vrugt et al.* [2008] extended the same input uncertainty model by introducing more vague priors (uniform) for the multipliers and use of the DREAM in parameter sampling. One of the problems in the application of latent variables is the number of additional parameters, which may lead to computational problems in calibration when the number of precipitation events (and thus latent variables) increases considerably. *Ajami et al.* [2007] tried to decrease the number of parameters in the input uncertainty model of BATEA by using multipliers sampled from the same distribution for each rain observation in their Integrated Bayesian Uncertainty Estimator (IBUNE)-framework. (see discussion by *Renard et al.* [2009] and *Ajami et al.* [2009]). *Reichert and Mieleitner* [2009] presented a method to account for input uncertainty that can be seen as a generalization of the storm-dependent latent variables. Their approach is based on stochastic time-dependent variables.

[10] None of the multiplicative precipitation error models is able to “correct” rainfall observations if rainfall occurred but was not recorded by the gauges or more generally if events were very poorly sampled. By using a case study in New Zealand *McMillan et al.* [2011] conclude that multiplicative error model is consistent with observations and can be used in hydrological modeling where an appropriate minimum rainfall intensity threshold is respected. They also show the dependency of the rainfall error structure on the data time step and point out that common independence assumption of consecutive rainfall multipliers may be flawed for models operating with short (subdaily) timesteps. In addition, separation of storm events in the precipitation series is a practical problem when the latent variables are used to account for rainfall uncertainty. *Renard et al.* [2010] presented a recent example of how to overcome this issue by assigning an individual multiplier (latent variable) for each rainy day. Use of daily multipliers, however, still does not solve problems connected to nondetection of rainfall and of how to separate the influence of model structure and missing information (e.g., precipitation intensity during a time step used in the model). Furthermore, latent variables will most likely be affected by model structure error unless it is specified separately (parameter values may also be biased with respect to their physical interpretation by model structure error), and this leads to the question of how to represent model structure error.

[11] There are several methods for approximating structural uncertainties in rainfall-runoff modeling:

[12] 1. It is possible to use multiple models and approximate model uncertainty based on the range of simulation

results [e.g., *Refsgaard et al.*, 2006]. As an example of a technique BMA has been used in different disciplines to combine results of different models. *Franz et al.* [2010] for example used BMA in studying snow model uncertainty.

[13] 2. It is possible to account for the bias in the model with a method based on stochastic, time-dependent parameters [*Kuczera et al.*, 2006; *Reichert and Mieleitner*, 2009]

[14] 3. Model structure uncertainty can also be taken into account by fitting an ARMA-model to the model residuals [e.g., *Schaefli et al.*, 2007; *Vrugt et al.*, 2009a; *Laloy et al.*, 2010].

[15] The generalized likelihood function of *Schoups and Vrugt* [2010] adopted in this study applies autoregressive model to account for model structural uncertainty. Autocorrelation in the residuals of model output variables is a typical indication of model structural problems. Autocorrelation is caused, e.g., by the storages in conceptual models that may unrealistically smooth estimated runoff series. However, systematic errors in precipitation and streamflow observations can lead to similar problems and autocorrelated errors in the model simulation results. We assume in this study that the ARMA-model accounts solely for model structural errors, because the other sources of errors are separately handled. This study is motivated by the need to assess how valid the above assumption is in a snow-affected catchment and whether different sources of errors can be separated in the inference.

[16] This study adopts the input uncertainty model presented by *Kavetski et al.* [2002] in a snow-fed basin in conjunction with Bayesian inference. It extends the studies of *Kavetski et al.* [2006ab], *Vrugt et al.* [2008] and *Schoups and Vrugt* [2010] by utilizing snow observations in addition to streamflow records in calibration. To our knowledge it is the first time that the generalized likelihood function of *Schoups and Vrugt* [2010] that uses an autoregressive part to account for model structural uncertainties has been applied with an input uncertainty model. *Renard et al.* [2010] concluded that given rainfall-runoff time series alone, input and structural errors cannot be separated unless one has good prior knowledge about the magnitude of the errors. Based on a case study *Renard et al.* [2011] concluded that inclusion of independent estimates of rainfall and runoff uncertainties could overcome the ill-posedness of the inference. A dense rain gauge network enabled geostatistical conditional simulation for independent estimation of rainfall uncertainties prior to Bayesian inference. Compared to *Renard et al.* [2010, 2011] this study adopts a different strategy for describing model structure uncertainty and extends their study by introducing snow water equivalent (SWE) observations as another state variable in the inferencing. The hypothesis is that the use of informative priors for precipitation multipliers, e.g., through geostatistical analysis of rain gauges, is not necessary, when rainfall-runoff data is complemented with additional hydrological information such as SWE data used in this case study.

[17] The conceptual rainfall-runoff model IHACRES combined with a degree-day snow model is used to simulate SWE and streamflow in a small boreal catchment in southern Finland. The aim is to determine unbiased parameter estimates for the combined degree-day snow model and IHACRES by taking into account both input uncertainty and model structure uncertainty in the model calibration. At

the same time our goal is to reveal precipitation uncertainty from the streamflow and SWE data. Thus we are doing “hydrology backward” [*Kirchner*, 2009; *Vrugt et al.*, 2008]. Bayesian multiresponse calibration is adopted to utilize both streamflow and SWE observations. Precipitation uncertainty is separately handled in the calibration of the model, whereas input air temperature, streamflow and SWE measurements are considered to be free of error.

2. Rudbäck Catchment and Data Available

[18] The Rudbäck catchment in southern Finland has an area of 142 ha (Figure 1). The elevation in the area is undulating and ranges from 5 m to 65 m. The bedrock is exposed on the hilltops and the downslope areas have shallow depths of mostly silty and sandy till underlined by impermeable bedrock. Some layers of clay and peat are found in the lowland areas. The catchment is covered by a managed forest stand dominated by Norway Spruce (*Picea abies* L. Karst.) with a mixture of mainly Scots Pine (*Pinus sylvestris* L.) and Birch (*Betula* sp.). The catchment includes a number of stand management units where stand density, age and tree species distributions vary. More details on the site information are published by *Lepistö* [1996] and *Lepistö and Kivinen* [1997].

[19] The climate of the southern coast of Finland is temperate with cold, wet winters and several cycles of snow accumulation and melt. Mean annual precipitation (uncorrected) during 1991–1996 was 700 mm with 15–25% of snowfall. Mean annual air temperature is about 5°C. Daily discharge values at the outlet of the catchment were available for the period 1997–2000.

[20] Micrometeorological variables were measured at a height of 2 m above ground in a clearcut area of 3 ha (Figure 1). The variables were processed to estimate daily air temperature and precipitation for the current study. Precipitation was corrected for gauging errors according to the operational recommendations of *Førland et al.* [1996]. Missing meteorological measurements were supplied from the nearby weather station (Vihti) operated by the Finnish Meteorological Institute. The processing of the weather data is described by *Koivusalo* [2002]. Snow depth was manually measured at 22 points in the forest, and at 12 points in the clearcut area (Figure 1). Bulk snow density was measured at three points in both forested and open sites by weighing cylindrical snow samples. Snow measurements were taken once or twice per week depending on the occurrence of snowfalls and snowmelt. An average of the field and forest measurements was used in the current study. Details of the snow and meteorological measurement are given by *Koivusalo and Kokkonen* [2002, 2003].

3. Methods

3.1. Model

[21] The rainfall-runoff model used in this research is the parsimonious conceptual model IHACRES [*Jakeman et al.*, 1990; *Jakeman and Hornberger*, 1993] available, e.g., in the hydromad software package [*Andrews et al.*, 2011]. The model consists of two modules: a nonlinear module that converts rainfall to effective rainfall and a linear routing module. There are different versions of the nonlinear

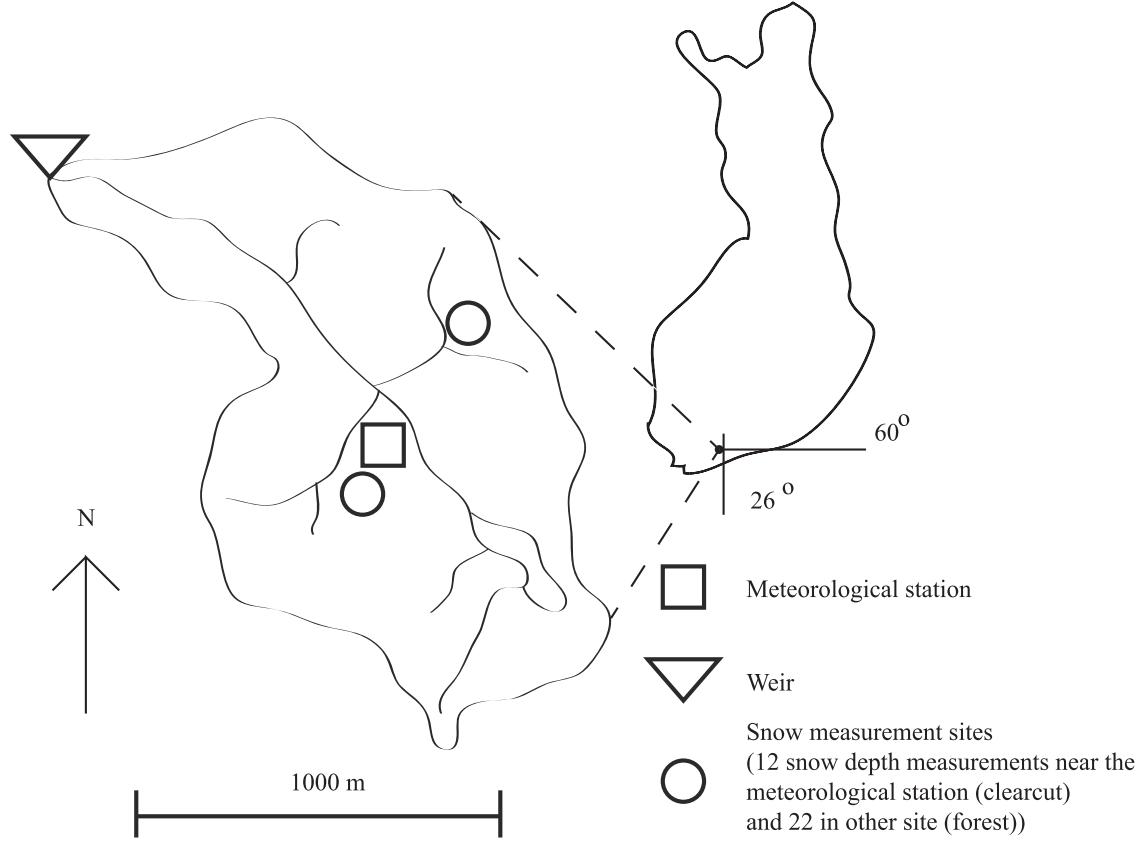


Figure 1. The 142 ha Rudbäck catchment in southern Finland and locations of the hydrometeorological measurement sites used in the study.

module but here the Catchment Moisture Deficit (CMD) version [Croke and Jakeman, 2004, 2005] is used. The linear model consists of single or multiple linear storages that are connected either in series or parallel. To account for snow processes in the catchment, a snow module using a degree-day factor approach is attached to the model. The model structure is illustrated in Figure 2.

[22] The mass balance equations for the snow module are

$$\frac{dI}{dt} = P_s + F - m \quad (1)$$

for ice (I [mm]) in the snowpack and

$$\frac{dL}{dt} = P_r + m - F \quad L \leq L_{\max} \quad (2)$$

for liquid water (L [mm]) in the snowpack. P_r [mm d⁻¹] and P_s [mm d⁻¹] are the corrected rainfall and snowfall, respectively, m [mm d⁻¹] is the rate of melting and F [mm d⁻¹] the rate of freezing in the snowpack. Snowmelt discharge is the excess of liquid water in the snowpack

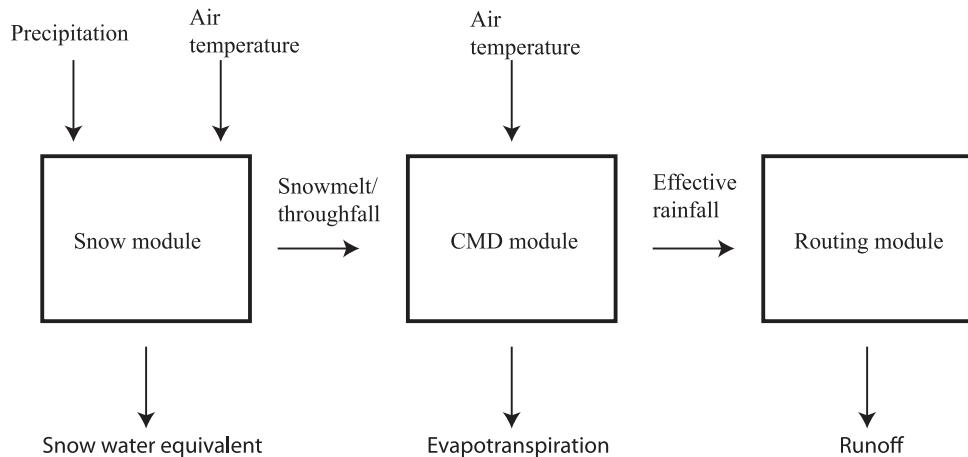


Figure 2. Structure of the hydrological model.

compared to the maximum liquid water retention capacity of the snowpack (L_{max} [mm]), which is represented simply as

$$L_{max} = r_{cap}I \quad (3)$$

where r_{cap} [-] is the retention parameter.

[23] The precipitation is ruled to be either in the form of snow or in the form of rain in the snow module based on the parameter T_{thres} [$^{\circ}\text{C}$].

$$f_r = \begin{cases} 0 & T < T_{thres}, \\ 1 & T \geq T_{thres} \end{cases}, \quad (4)$$

$$f_s = 1 - f_r, \quad (5)$$

where f_r [-] is the fraction of rainfall, f_s [-] the fraction of snowfall, T_{thres} is the threshold temperature, and T [$^{\circ}\text{C}$] is the measured air temperature. To account for the effect of the canopy on interception, rainfall and snowfall are adjusted by using correction factors c_r [-] and c_s [-] i.e.,

$$P_r = c_r f_r P_{obs}, \quad (6)$$

$$P_s = c_s f_s P_{obs}, \quad (7)$$

where P_{obs} [mm d^{-1}] is the daily precipitation measurement. In the degree-day model, snowmelt is linearly related to air temperature above a given threshold temperature. In this application, the parameter T_{thres} (equation (4)) is used as this threshold. The rate of melting (m) is then

$$m = \min(I, \max(0, k_d(T - T_{thres}))), \quad (8)$$

where k_d [mm $\text{d}^{-1} ^{\circ}\text{C}^{-1}$] is the degree-day factor. Similarly the rate of freezing (F) in the snowpack is

$$F = \min(L, \max(0, k_f(T_{thres} - T))), \quad (9)$$

where k_f [mm $\text{d}^{-1} ^{\circ}\text{C}^{-1}$] is the degree-day factor for freezing. Thus the snow model contains 6 parameters (T_{thres} , c_r , c_s , k_d , k_f , r_{cap}).

[24] The winter snowmelt discharge or summer throughfall is the input to the CMD module of the IHACRES model that converts this input into effective rainfall U [mm d^{-1}]. The mass balance step of the CMD module is

$$\frac{dM}{dt} = -P_O + E_A + U, \quad (10)$$

where M [mm], P_O [mm d^{-1}], E_A [mm d^{-1}] and U are the catchment moisture deficit, the snowmelt discharge/throughfall, evapotranspiration, and effective rainfall, respectively. In the CMD-module the response of the catchment dM to a small input dP is assumed to be dependent on M alone

$$\frac{dU}{dP} = 1 - \min\left[1, \left(\frac{M}{d}\right)^b\right], \quad (11)$$

where the threshold parameter d [mm] is used to define the catchment moisture deficit above which part of the input becomes effective rainfall. The value of d is set to 200 mm [Croke and Jakeman, 2004] and the shape parameter b [-] must be calibrated. Note that the introduction of the b parameter is a modification to the original CMD module. To satisfy the conditions listed in Croke and Jakeman [2004], b should be greater than, or equal to, one. It is also assumed that E_A is only a function of M and potential evapotranspiration (E_P [mm d^{-1}]) with a threshold at $M = fd$ where f [-] is another model parameter.

$$E_A = E_P \min\left\{1, \exp\left[2\left(1 - \frac{M}{fd}\right)\right]\right\}. \quad (12)$$

[25] Potential evapotranspiration is assumed to be a simple function of daily air temperature

$$E_P = \begin{cases} eT & T > 0^{\circ}\text{C} \\ 0 & T \leq 0^{\circ}\text{C} \end{cases}, \quad (13)$$

where e [mm $\text{d}^{-1} ^{\circ}\text{C}^{-1}$] is a potential evaporation coefficient. Thus the number of calibrated parameters in the CMD module is three (b , e , f).

[26] Finally, effective rainfall is the input to the routing module. The number of parameters in the linear module depends on the number of reservoirs used. Here the linear module was identified to consist only of a single reservoir i.e., base flow and quick flow components could not be separated by using the available data. Thus only a single parameter τ [d] is needed as the time-constant of the reservoir. The mass balance of the storage (S [mm]) of the linear reservoir is

$$\frac{dS}{dt} = -kS + U, \quad (14)$$

$$k = 1 - \exp\left(-\frac{1}{\tau}\right). \quad (15)$$

3.2. Time-Stepping Scheme

[27] Recent studies point out that the numerical approximations of differential equations in hydrological models have an effect on the inference and that it can dwarf the uncertainty caused by model structure [Clark and Kavetski, 2010; Schoups et al., 2010]. In the present research Monte Carlo Markov Chain (MCMC) based sampler is used to sample the posterior probability density functions (pdf) of the parameters and latent variables. Thus the model must be executed thousands of times to reach convergence and calculation speed is of great importance. Daily data are available for both meteorological forcing and streamflow, but only about weekly data for SWE. To achieve sufficient speed in the calculations an explicit Euler method is used to solve the differential equations in the snow model (equations (1) and (2)) and the linear storage routing module of the IHACRES (equation (14)). The nonlinear module (equation (11)) is solved analytically, and the solution coded into the model. To minimize the impact of the numerical solution scheme on daily model output series, computationally, hourly substepping is used within each daily modeling time step, with

the hourly results aggregated to the daily scale. Hourly precipitation is the daily value divided by 24.

3.3. Calibration and Uncertainties in Modeling

[28] Bayesian inference is used to calibrate the model parameters. By using the Bayesian inference with MCMC sampling it is possible to estimate the uncertainties of parameter estimates. In the Bayesian approach the posterior pdf of the parameters is sampled by the MCMC algorithm after convergence being reached. In this study, parameters are conditioned simultaneously on both streamflow and SWE observations. It is expected that errors in the simulated SWE are independent and follow a normal distribution. For streamflow, however, assumptions of Gaussian and independent errors are often violated [e.g., Kuczera, 1983]. To account for the case with correlated, heteroscedastic and non-Gaussian errors Schoups and Vrugt [2010] presented a generalized likelihood function. They proposed a model

$$\Phi_p(B)e_{e2,i} = \sigma_{e2,i}a_i, \quad (16)$$

for the residuals e_{e2} , where a_i follow a skew exponential power distribution $\text{SEP}(0,1,\xi,\beta)$ with zero mean, unit variance, and with the parameters ξ and β accounting for the skewness and the kurtosis. The standard deviation of the residuals $\sigma_{e2,i}$ can be explicitly defined to account for heteroscedasticity, $\Phi_p(B)$ is an autoregressive polynomial with the parameters ϕ_p , and B is the backshift operator. Details of this approach can be found in Schoups and Vrugt [2010] where they show that the log likelihood function with these assumptions is

$$l(\theta|\hat{Y}) = n_2 \log \frac{2\sigma_\xi \omega_\beta}{\xi + \xi^{-1}} - \sum_{i=1}^{n_2} \log (\sigma_{e2,i}) - c_\beta \sum_{i=1}^{n_2} |a_{\xi,i}|^{2/(1+\beta)}, \quad (17)$$

where n_2 is the number of streamflow observations. The variable $a_{\xi,i}$ in equation (17) is

$$a_{\xi,i} = \xi^{-\text{sign}(\mu_\xi + \sigma_\xi a_i)} (\mu_\xi + \sigma_\xi a_i), \quad (18)$$

and the equations for the variables μ_ξ , σ_ξ , c_β and ω_β that are functions of β and ξ are given in Appendix A of Schoups and Vrugt [2010] (equation (A2) for ω_β , equation (A3) for c_β , equation (A5) for μ_ξ and equation (A6) for σ_ξ). To reduce the heteroscedasticity of the residual series the Box-Cox transformation is adopted here for the observed and modeled streamflow [Box and Cox, 1964].

$$\hat{Y}_2(Q, \lambda_1, \lambda_2) = \begin{cases} \frac{[(Q + \lambda_2)^{\lambda_1} - 1]}{\lambda_1} & \lambda_1 \neq 0 \\ \ln(Q + \lambda_2) & \lambda_1 = 0 \end{cases}, \quad (19)$$

where Q [mm d^{-1}] is streamflow and λ_1 and λ_2 are the transformation parameters. In Schoups and Vrugt [2010] heteroscedasticity was handled with the assumption that

$$\sigma_{e2,i} = \sigma_0 + \sigma_1 Q_i, \quad (20)$$

where both σ_0 and σ_1 were calibrated along with model parameters. The Box-Cox transformation, however, was found to be more suitable to account for heteroscedasticity in this case study. Thus σ_1 was set a value of 0 and σ_{e2} for Box-Cox transformed streamflow became a constant σ_0 .

[29] The parameters β , ξ , σ_0 and the autoregressive model parameters ϕ_p are calibrated along with the model parameters. We assume that errors of streamflow and snow models are uncorrelated and therefore the log likelihood function is

$$\begin{aligned} l(\theta|\hat{Y}_1, \hat{Y}_2) = & -\frac{n_1}{2} \ln(2\pi) - \frac{n_1}{2} \ln(\sigma_{e1}^2) - \frac{1}{2\sigma_{e1}^2} \sum_{j=1}^{n_1} [y_{1,j}(\theta) - \hat{y}_{1,j}]^2 \\ & + n_2 \log \frac{2\sigma_\xi \omega_\beta}{\xi + \xi^{-1}} - \sum_{i=1}^{n_2} \log(\sigma_{e2,i}) - c_\beta \sum_{i=1}^{n_2} |a_{\xi,i}|^{2/(1+\beta)}, \end{aligned} \quad (21)$$

where \hat{Y}_1 and \hat{Y}_2 are measured SWE and streamflow observations and θ contains all the parameters. Douglas et al. [1976] pointed out that it is possible that the other response (e.g., y_1) will dominate the parameter estimation procedure if n_1 is much greater than n_2 in equation (21). Whether the larger number of streamflow observations dominates the parameter estimation in our case study is discussed later.

[30] The Differential Evolution Adaptive Metropolis algorithm (DREAM_(ZS)), is used here to sample the posterior pdf of the parameters. DREAM is an efficient MCMC sampler that simultaneously runs multiple Markov chains. During the process the whole parameter space is explored and the algorithm automatically adjusts the scale and orientation of the proposal distribution. The original DREAM is well documented in the studies of Vrugt et al. [2008, 2009a]. The DREAM_(ZS) used is a modification of that algorithm. It uses an archive of past states to generate candidate points in each individual chain. Such sampling is more efficient than optimal random walk Metropolis [ter Braak and Vrugt, 2008]. DREAM_(ZS) maintains efficient convergence and needs less parallel Markov Chains than the original DREAM. Schoups and Vrugt [2010] presented a detailed analysis of the performance of DREAM_(ZS) with artificial data in the context of hydrological modeling. In the present application chains are initialized by using the Latin hypercube method to achieve different starting points for the chains. DREAM_(ZS) is recently discussed by Laloy and Vrugt [2012]. The R -statistic of Gelman and Rubin [1992] is used to check whether the chains have converged. Similarly to Vrugt et al. [2009b], the threshold value of R is 1.2 for each chain.

[31] At the beginning of the model calibration process, the basic model is calibrated with two precipitation multipliers (c_r and c_s in equations (6) and (7)) using the Gaussian assumptions for both streamflow and the SWE residual series (Approaches A1 and A2 later in the text; see Table 1). Subsequently, streamflow errors are assumed to be non-Gaussian and autocorrelated (Approach B1). In these cases the multipliers c_r and c_s describe both precipitation (input) uncertainty and the effect of the canopy on snow and streamflow. Finally (Approach B2), an individual multiplier is assigned to each identified precipitation event and the streamflow residuals are assumed to be non-Gaussian and autocorrelated. For this purpose the daily precipitation

Table 1. Different Model Inference Setups Used in the Study^a

Model Setup	Separate Input Uncertainty	T_{thres}, k_d	r_{cap}	k_f	e	$f, b, \tau, \sigma_{e1}, \sigma_{e2}$	C_r, C_s	p_i	β	ξ	ϕ_1	All the Calibrated Parameters Well-Identified?
A1	No	C	C	C	C	C	C	1	0	1	0	No
A2	No	C	0	0	0.17	C	C	1	0	1	0	Yes
B1	No	C	0	0	0.17	C	C	1	C	C	C	Yes
B2	Yes	C	0	0	0.17	C	1	C	C	C	C	No

^aThe letter C indicates that the parameter is calibrated and a number indicates the fixed value used for the parameter.

series was partitioned into discrete events separated by at least 3 consecutive days without precipitation. In the calibration period (1 August 1997 to 31 July 1999) there were 43 such precipitation events and in the so-called validation period (1 August 1999 to 31 December 2000) there were 33 events. Thus a latent variable p was introduced for each event and equations. (6) and (7) in the model were replaced by

$$P_{r,i} = p_i f_r P_{obs,i} \quad i = 1, \dots, 43, \quad (22)$$

$$P_{s,i} = p_i f_s P_{obs,i} \quad i = 1, \dots, 43. \quad (23)$$

[32] The separation of the storm events using the threshold defined above reveal several types of precipitation events in the time series. Some events have short duration with a low amount of accumulated precipitation (<5 mm). On the other hand, some events last for a long time and accumulate large precipitation depths (>60 mm). Large precipitation events occur especially during the winter snowfall season. In the validation period (33 events) uncertainty of the precipitation is presented by sampling the precipitation multiplier from a posterior of a valid class. These classes are based on the 43 precipitation events of the calibration period that were classified based on the event precipitation sums. For example, if in the validation period there is a rain event that is a member of class 4 (Table 2), uncertainty is presented by sampling from the posterior of that class. The classification into six classes (Table 2) is rough as only the accumulated precipitation is used to classify different types of storm events. It would be possible, for example, to use the length of storm events or the peak intensity of precipitation as the classification criterion. However, it was not reasonable to use a larger number of classes than six as the calibration sample size comprised only 43 storm events.

[33] Uniform prior distributions are assumed for the model parameters and their ranges are presented in Table 3.

Table 2. Description of the Different Kinds of Storm Events

Class Number	Description of the Storm Event	Number of Events in Calibration
1	Rain, precipitation sum < 5 mm	11
2	Rain, 5 mm < precipitation sum < 20 mm	9
3	Rain, 20 mm < precipitation sum < 60 mm	11
4	Rain, precipitation sum > 60 mm	3
5	Snow, precipitation sum < 50 mm	6
6	Snow, precipitation sum > 50 mm	3

The parameters λ_1 and λ_2 of the Box-Cox transformation (equation (19)) were estimated a priori and set to $\lambda_1 = 0$ and $\lambda_2 = 1$. Error variances σ_{e1}^2 and σ_{e2}^2 (equation (21)) are estimated along with the model parameters. A uniform initial distribution $U[0.1, 2.0]$ was used for the latent variables following *Vrugt et al.* [2008]. *Kavetski et al.* [2002] questioned the use of uniform priors for precipitation multipliers. In their BATEA-framework the prior of the multipliers is usually assumed to be Gaussian. *Renard et al.* [2010, 2011] showed that informative priors about input and output uncertainty are needed in rainfall-runoff modeling to separate different sources of uncertainties in prediction and to avoid ill-posedness of the inference. Such information for precipitation could be provided by geostatistical analysis of rain gauges [*Renard et al.*, 2011]. The precipitation estimate used in this study is based on measurements of a single rain gauge located inside the 142 ha catchment. The uncertainty in the precipitation estimates is thus a combination of representativeness of this gauge measurement for the whole catchment and accuracy of rain gauge measurement itself. There is no a priori information available about the distribution of the error characteristics. *McMillan et al.* [2011] found out in their case study that lognormal multiplier distribution (used, e.g., by *Kavetski et al.* [2006b]) presents a relatively close approximation about the true distribution. However, *McMillan et al.* [2011] did not address snowfall.

[34] The use of uniform priors does lose information about the precipitation events as only timing is expected to be correct a priori and this may lead to ill posedness of the inference [*Kavetski et al.*, 2006a; *Renard et al.*, 2010]. *Vrugt et al.* [2008], however, were able to infer multipliers from rainfall-runoff data by using flat priors. In this study we decided to apply uniform priors for precipitation multipliers and focus on the usability of the SWE observations, in addition to rainfall-runoff data, for providing sufficient amount of information for the inference. We hypothesized that the use of informative priors for precipitation multipliers, e.g., through geostatistical analysis of rain gauges, is not necessary, when rainfall-runoff data is complemented with sufficient amount of additional hydrological information. In this study such additional data is provided through SWE observations.

3.4. Predictive Uncertainty

[35] The results of the Bayesian inference contain all the necessary information to estimate predictive uncertainty of the model for both streamflow and SWE [*Vrugt et al.*, 2009a]. When the posterior pdfs of the parameter estimates are available, pdfs are sampled 10,000 times to produce a

Table 3. Prior Ranges of the Parameters

Parameter	Description	Minimum	Maximum	Unit
T_{Thres}	Threshold temperature of the degree day model	-1	1	°C
k_d	Degree day factor for melting	1	5	mm d ⁻¹ °C ⁻¹
k_f	Degree day factor for freezing	0.001	2	mm d ⁻¹ °C ⁻¹
r_{cap}	Parameter for the water retention capacity in snow	0.001	0.2	—
c_r	Correction factor for measured rainfall	0.1	2.0	—
c_s	Correction factor for measured snowfall	0.1	2.0	—
b	Power parameter in CMD module	1.1	100	—
f	The ratio between the stress and flow threshold (CMD)	0.01	3.0	—
e	Potential evaporation coefficient (CMD)	0.01	0.3	mm d ⁻¹ °C ⁻¹
τ	Time-constant of the linear reservoir	0.1	20	d
σ_{e1}^2	Error variance, snow water equivalent	0.1 ²	20 ²	(mm) ²
σ_{e2}^2	Error variance, streamflow	0.1	1 ²	(mm d ⁻¹) ²
β	Kurtosis parameter	-1	1	—
ξ	Skewness parameter	0.1	10	—
ϕ_1	Lag-1 autocorrelation coefficient	0	1	—

large set of model outputs. Results for the outputs are then analyzed and 90% confidence limits are estimated. This uncertainty is a representation of the parameter uncertainty only and thus does not contain unexplained error (e.g., output uncertainty). To estimate the total predictive uncertainty the remaining error is expected to be additive. For each outcome of SWE, the residual error

$$\varepsilon_i \sim N(0, \sigma_{e1}^2) \quad (24)$$

is added to the outcome. After that SWE confidence limits are produced similarly to the estimation of model output uncertainty described above.

[36] To approximate total uncertainty limits for modeled streamflow, an additive error approach is used as described in *Schoups and Vrugt* [2010]. First an independent residual series from an assumed skew exponential power distribution is generated and then this series is manipulated to obtain the required autocorrelation properties (see equation (16)). The manipulated series is then used as an additive error series for Box-Cox transformed modeled streamflow. Finally the modeled streamflow series is obtained through an inverse Box-Cox transformation (see equation (19)). Results for these outputs are then analyzed and 90% confidence limits are estimated.

[37] The effect of uncertainties in the latent variables on model output is quantified in the calibration period by sampling their posterior pdf, similarly to the model parameters as described above. In the validation period, uncertainty caused by precipitation is revealed by sampling the precipitation multiplier from a posterior of a valid class. For example, when there is a rain event in class 4 (Table 2) in the simulation period, uncertainty is presented by sampling from the posterior of that class.

4. Results and Discussion

4.1. Assuming Gaussian Errors

[38] With the assumptions of $\beta = 0$, $\xi = 1$ and an independent error series equation (21) simplifies to the “classical” Gaussian likelihood function where both error series (SWE and streamflow) are independent and identically distributed according to a normal distribution with zero mean. This assumption was the first step of the analysis. Calibration of the model under this assumption (Approach A1 in Table 1)

led to the conclusion that the number of model parameters can be decreased. The marginal posterior of the snow model parameter r_{cap} (equation (3)) indicated a value of $r_{cap} = 0$, thus rendering k_f (equation (9)) redundant in the model. DREAM_(ZS) converged in 9.8×10^4 model runs. To speed up the convergence of the MCMC chains the values of r_{cap} and k_f were fixed to zero. When calibration yields a r_{cap} value of zero the description for liquid water storage in the snowpack is disregarded as there is not enough information in the data to identify its characteristics. *Kokkonen et al.* [2006] have shown that measured SWE values can be equally well reproduced without the liquid water storage. Furthermore, the value of e was not identified properly. The value was therefore estimated a priori by using the Penman-Monteith equation and the available net radiation estimates. As a result the parameter value (e in equation (13)) was fixed to $e = 0.17$.

[39] Recalibration of the simplified model (Approach A2 in Table 1) led to the marginal posterior distribution estimates of the parameters shown in Figure 3. DREAM_(ZS) using 9 parallel chains converged in 4.1×10^4 model runs. All parameters are well-identified and show narrow posterior ranges compared to their uniform priors (Table 3). The precipitation multipliers c_r and c_s (equations (6) and (7)) are equally well identified and indicate that precipitation must be scaled down due to canopy interception losses. In a catchment nested inside the current study area, *Koivusalo and Kokkonen* [2002] measured interception losses in the canopy and found that the average summer and winter season losses were 29 and 26%, respectively. The estimated multipliers in the current study are in line with the reported interception losses.

[40] It is well known that the snow model parameters T_{thres} and k_d (equation (8)) are correlated [*Clark and Vrugt*, 2006; *Kokkonen et al.*, 2006]. This was confirmed by studying the parameter values of MCMC chains after the convergence. The larger the value of T_{thres} was the larger was also the value of k_d . In Finland *Kuusisto* [1984] reported that degree-day factors in forests range from 1.8 to 3.4 mm d⁻¹ °C⁻¹. The posterior of the calibrated k_d (mode 2.3 mm d⁻¹ °C⁻¹) in this study is well within this range.

[41] In Figure 4 the estimates of streamflow and in Figure 5 the estimates of SWE are shown with uncertainty bounds. Uncertainty bounds in Figure 5 are not shown for the summer periods with no snow cover. Lower values of the 90% total uncertainty bounds were fixed to zero, although the assumptions used resulted in negative lower prediction

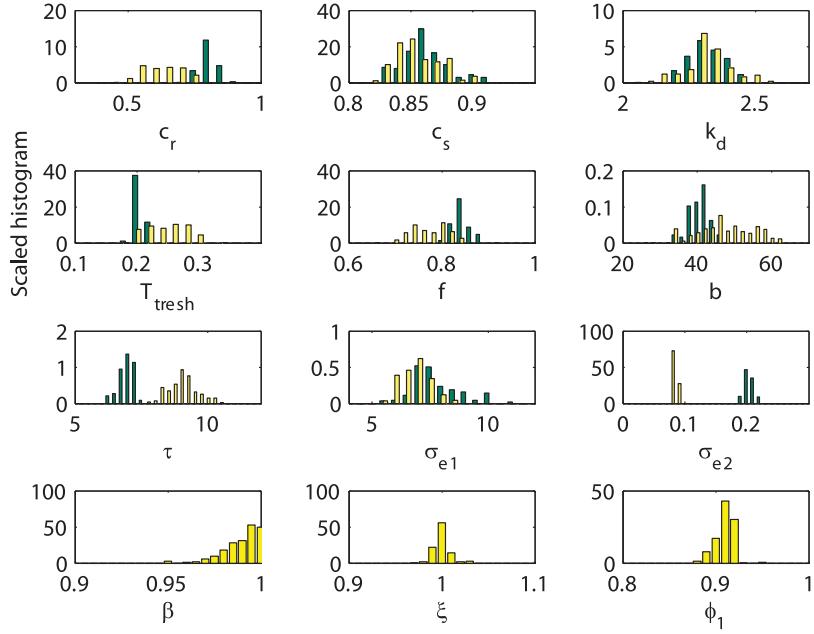


Figure 3. Scaled histograms of estimated posterior model parameters in Approach A2 (dark) and B1 (light).

bounds for both near-zero flows and SWE. The assumption of a constant additive error leads to an estimated large range of uncertainty bounds and high uncertainties at low values of flow and SWE.

[42] The parameter uncertainty was found to be only a small fraction of the total uncertainty (90% uncertainty bounds in Figures 4 and 5). Only 12% of the streamflow observations (Figure 4) are inside the parameter uncertainty

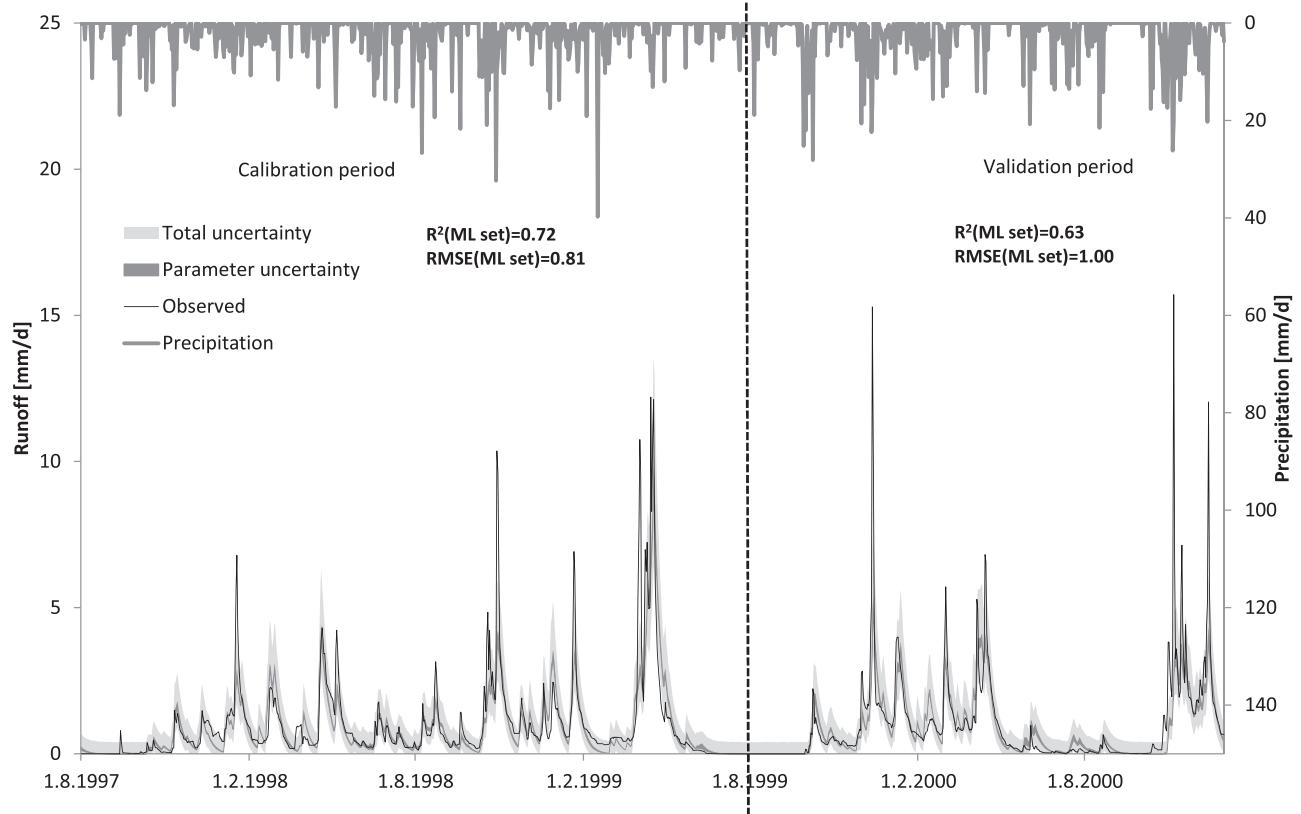


Figure 4. Uncertainty bounds (90%) for simulated streamflow in Approach A2 and measured precipitation.

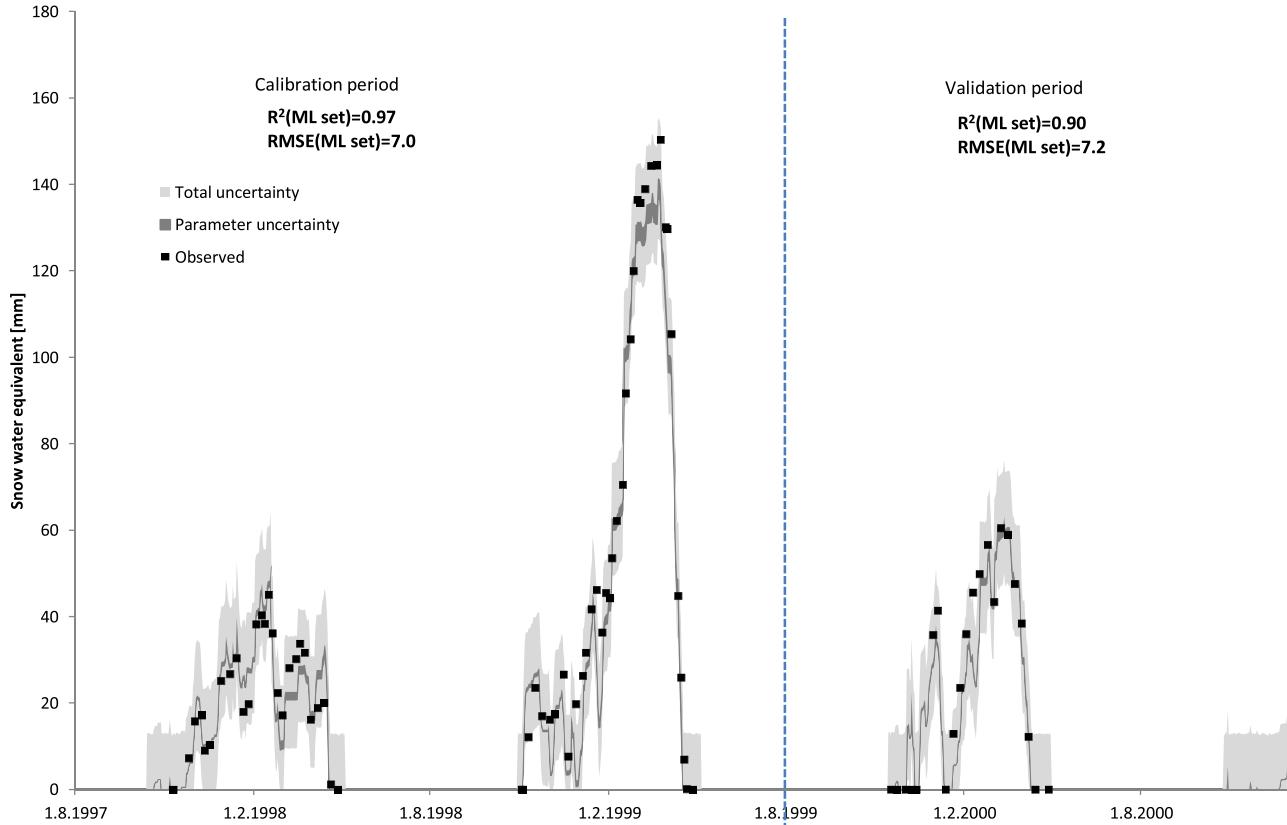


Figure 5. Uncertainty bounds (90%) for simulated snow water equivalent in Approach A2.

bounds in the calibration period and 13% in validation. The 90% bounds for total uncertainty (consisting of parameter and remaining additional residual error), however, cover 90% of the streamflow observations both in calibration and validation. It is obvious that uncertainty is overestimated for low flows and underestimated for peak flows (Figure 4). Most streamflow observations outside the uncertainty bounds are peak events. For the case of SWE (Figure 5) 13% of the observations in the calibration period are inside the parameter uncertainty and 94% inside the 90% total uncertainty bounds. In validation these figures are 27% for the parameter uncertainty and 91% for the total uncertainty.

[43] The following assumptions were made about the model errors. It was assumed that the residuals are normally distributed, the model errors are homoscedastic, and

the errors are mutually independent showing no autocorrelation. Furthermore, it was assumed that the snow model errors are not correlated with the errors in modeled streamflow. This last assumption was satisfied as no significant correlation was found between snow and streamflow model errors (significance level $\alpha = 0.05$).

[44] The assumptions related to the snow model can be visually assessed using graphs shown in Figure 6. Kolmogorov-Smirnov test [Massey, 1951] revealed that the normality assumption cannot be rejected (Figure 6a) and Figure 6c does not show evident heteroscedasticity. There is correlation between consecutive residuals, however (Figure 6b). The correlation coefficient is 0.57 for consecutive residuals, where the time-lag between observations is not constant. The obvious reason for the presence of autocorrelation is the

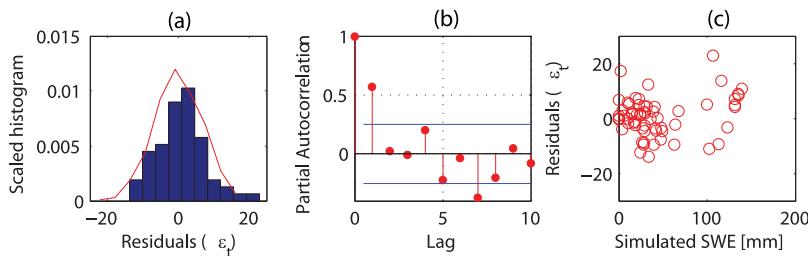


Figure 6. Residual analysis of the SWE results using the maximum likelihood parameter set in Approach A2. (a) The histogram is based on the observed errors and the red line is the theoretical density based on the assumed distribution. (b) Partial autocorrelation of residual series. (c) Residuals as a function of simulated SWE.

fact that errors during the accumulation period influence the later predictions until snowpack is melted. If the snow model misses one snowfall event at the start of the snow season, the following error will be present in the modeled SWE throughout the winter and thus causes the autocorrelation. In addition, the presence of autocorrelation can be a symptom of inadequacy of the snow model to account for key snow processes or inaccuracies in the observations of meteorological input variables and SWE. It should be noted that the snowfall measurement is more prone to observation errors than the rainfall measurement. The correction factor for snowfall gauging error typically exceeds the factor for rain catch error, and the snowfall correction factor can vary in different weather conditions [Førland *et al.*, 1996].

[45] In addition to autocorrelation in SWE, there is severe autocorrelation in the streamflow residuals (Figure 7b) in the calibration period, suggesting that the lumped model structure is not able to account for all the hydrological variability in the catchment. Autocorrelation is slightly more severe for periods with snow cover and snowmelt compared to the summer months. As there is no immediate streamflow response to precipitation during the snow cover periods, such autocorrelation is expected. However, even during the summer time autocorrelation remains significant. This may be caused by the inability of the model to generate streamflow peaks as fast as the measured peaks occur. The catchment contains soil covered lowland areas and exposed bedrock at hilltops and it is likely that the runoff generation mechanism changes when runoff producing areas extend from near stream areas toward areas covered with thin soil layers or even exposed bedrock [Koivusalo and Kokkonen, 2003]. The rapid generation of high discharge peaks reflects the distribution of soil types as well as variable antecedent soil moisture conditions within the catchment. The routing model contains only a single linear storage with a time-constant of approximately 7 days, which may not be sufficient for variable hydrological conditions. If there were more observations (longer period) available another routing storage could become identifiable and model structure problems could decrease.

[46] In addition to the presence of autocorrelation in the streamflow error, the normality assumption in the residuals is violated (Kolmogorov-Smirnov test, significance level $\alpha = 0.05$) and Figure 7c reveal heteroscedasticity in the residuals. The residual histogram (Figure 7a) indicates that an assumption of a skewed distribution with a high kurtosis might be more appropriate for the remaining residual error.

Since error sources are not handled separately in the calibration, it is uncertain whether the violations of the assumptions about the model residuals are caused by the model structure, incorrect input variables, or errors in output observations. The highly significant correlations between the model residuals, however, are an indication of the model structure being an important source of uncertainty. The well-known consequences of violations of residual assumptions include biased parameters and inaccurate confidence limits, as recently demonstrated by Thyre *et al.* [2009] and Schoups and Vrugt [2010]. To fix these problems the generalized likelihood function (equation (21)) is applied next with the assumptions of autocorrelated and non-Gaussian errors for streamflow.

4.2. Generalized Likelihood Function for Autocorrelated and Non-Gaussian Errors (Approach B1)

[47] In addition to the model parameters 5 additional parameters are now included. These represent statistical properties of the assumed error distributions (σ_{e1} , σ_{e2} , β , ξ , ϕ_1). These parameters were calibrated at the same time with the model parameters. Parameter ξ frees the symmetry assumption of the streamflow residual errors and β allows estimation of different values for the kurtosis of the distribution. It was assumed that an AR1-model with lag-1 parameter ϕ_1 can be used to represent the model structural error in the remaining residual series and to remove the autocorrelation from the residual errors that was present in Approach A2.

[48] DREAM_(ZS) with 12 parallel chains converged in 7.0×10^4 model runs. The resulting posterior parameter distributions are shown in Figure 3. All the parameters are well identified. The posterior distributions are narrow compared to their uniform priors (Table 3). Compared to Approach A2 there are clear differences in the modes of the posterior distributions (e.g., c_r , τ and σ_{e2} in Figure 3). In addition, the posterior distributions are now wider for T_{thresh} and b . The posterior values of β and ξ are close to unity indicating a symmetric distribution with a high kurtosis for the remaining residuals (also seen in Figure 8a). The lag-1 autocorrelation coefficient ϕ_1 attains values close to 0.9. The autocorrelation of the posterior residual series is small (Figure 8b). Also the scaled histogram of the remaining residuals visually closely matches the assumed distribution (Figure 8a) but the two-sample Kolmogorov-Smirnov test rejects the assumption that observed residuals would come from the assumed skewed exponential power distribution. The test was performed by using observed residuals

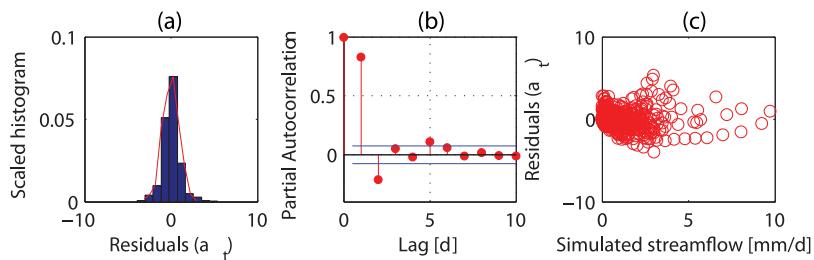


Figure 7. Residual analysis of the streamflow results using the maximum likelihood parameter set in Approach A2. (a) The histogram is based on the observed errors and the red line is the theoretical density based on the assumed distribution. (b) Partial autocorrelation of residual series. (c) Residuals as a function of simulated streamflow.

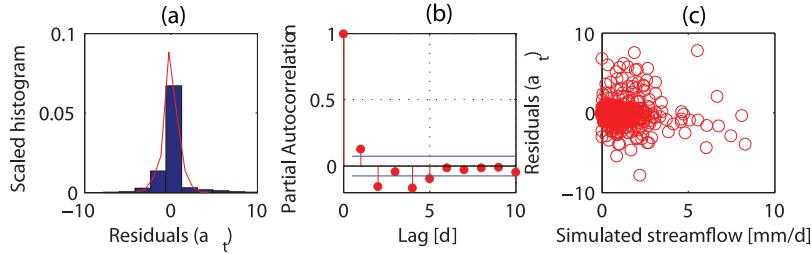


Figure 8. Residual analysis of the streamflow results using the maximum likelihood parameter set in Approach B1 and 1-day-ahead simulations. (a) The histogram is based on the observed errors and the red line is the theoretical density based on the assumed distribution. (b) Partial autocorrelation of residual series. (c) Residuals as a function of simulated streamflow.

and an equally sized random sample from $\text{SEP}(0, 1, \xi, \beta)$ and a significance level $\alpha = 0.05$. By using the White test [White, 1980] with a significance level $\alpha = 0.05$ the null-hypothesis of homoscedasticity (Figure 8c) must be rejected. There was no significant ($\alpha = 0.05$) correlation between snow and streamflow errors.

[49] In this inference setup, the model contains an AR1-component. The model can be used, however, also without the AR1-component. In the following the “Simulation mode” refers to batch-computation of the model for the entire length of the simulation period starting from initial values of the state variables and using the air temperature and precipitation series as input. In this mode the AR1 component is inactivated in the model and no information about the residuals of

runoff series is utilized in the simulation. The other operational mode is the “1-day-ahead simulation mode,” which includes the AR1-component and utilizes the residual series of streamflow from the previous computation time steps. Results from both of these simulations are presented and discussed.

[50] By using the maximum likelihood parameter set and 1-day-ahead simulation, the Nash-Sutcliffe efficiency R^2 for streamflow is 0.86 for the calibration period and 0.68 for the validation period. These values are clearly improved compared to Approach A2 (0.72 and 0.63). At the same time R^2 for SWE is 0.98 for calibration and 0.88 for validation. In Figure 9 the confidence limits (90%) for the 1-day-ahead simulated streamflow can be observed to be narrow. The

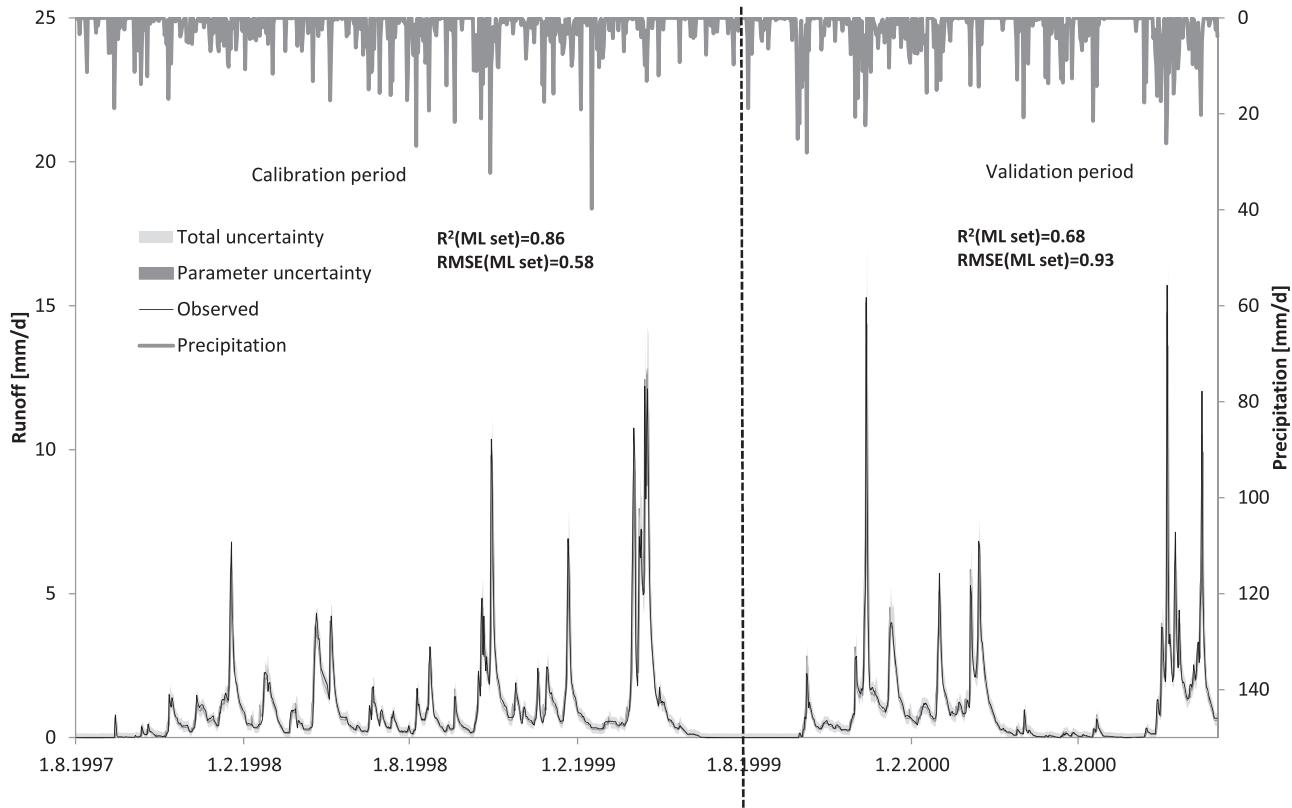


Figure 9. Uncertainty bounds (90%) for 1-day-ahead simulated streamflow in Approach B1 and measured precipitation.

parameter uncertainty limits contain 23% of the streamflow observations during the calibration period and 29% in validation. The 90% bounds for total uncertainty contain 88% of the observations in calibration and 89% in validation. These numbers are similar to Approach A2 but now the confidence limits are much narrower (compare Figures 4 and 9). The AR1-model was used to account for the structural uncertainty in the model. As shown the AR1-model has significantly narrowed the uncertainty bands without negatively affecting the percentage that describes the streamflow observations within the uncertainty limits. For SWE (Figure 10) the percentage of the streamflow observations within the 90% limits for total uncertainty is 92% for calibration and 86% for validation. Both of these percentage values are smaller compared to the values for Approach A2.

[51] In the simulation mode (i.e., when knowledge of previous streamflow errors are not utilized) the snow model is still working well. The rest of the model (runoff generation), however, is not working properly. The R^2 value of the maximum likelihood parameter set for streamflow drops to 0.63 in calibration (compared to 0.86 in 1-day-ahead simulation mode) and 0.53 in validation (compared to 0.68, respectively). The 90% bounds for the total uncertainty also become much wider. The main reason for the inferior performance is that with the addition of the AR1-model the parameter c_r seems to attain too small a value in calibration. This raises the question whether the addition of the AR-component has actually led to unbiased parameter estimates. Instead, it seems that the parameter ϕ_1 , accounting for the model structural errors, has interacted with c_r ,

that describes the rainfall uncertainty. In any case, the results suggest that the model should not be used for purposes other than those corresponding to the calibration objectives (in this case 1-day-ahead simulation).

[52] Next we study whether the problems above related to parameter c_r and violation of the residual assumptions can be ameliorated by using latent variables for the precipitation events. In addition, we examine whether or not there is enough information in the data to infer storm-dependent precipitation uncertainty during model calibration.

4.3. Using Separate Precipitation Uncertainty Model

[53] The inclusion of latent variables in Approach B2 (Table 1) led to a solution where most of the latent variables and some model parameters were poorly identified. The DREAM_(zs) algorithm converged smoothly (6.7×10^5 model runs) although in Approach B2 there were 5 model parameters, 5 statistical parameters and 43 latent variables in the model. The posterior distributions of the unidentified latent parameters are spread throughout their whole prior range. There is not enough information contained in the streamflow and SWE observations to simultaneously identify all the latent variables, the model structure uncertainty (ϕ_1) and the model parameters.

[54] The marginal posterior parameter estimates are shown in Figures 11 and 12. Changes in the calibrated model parameter values compared to Approach A2 and B1 (Figure 3) are obvious. For example, the posterior distribution of parameter f (equation (12)) is now almost uniform as long as f is greater than $f = 0.9$. In addition, b and τ have wide posterior

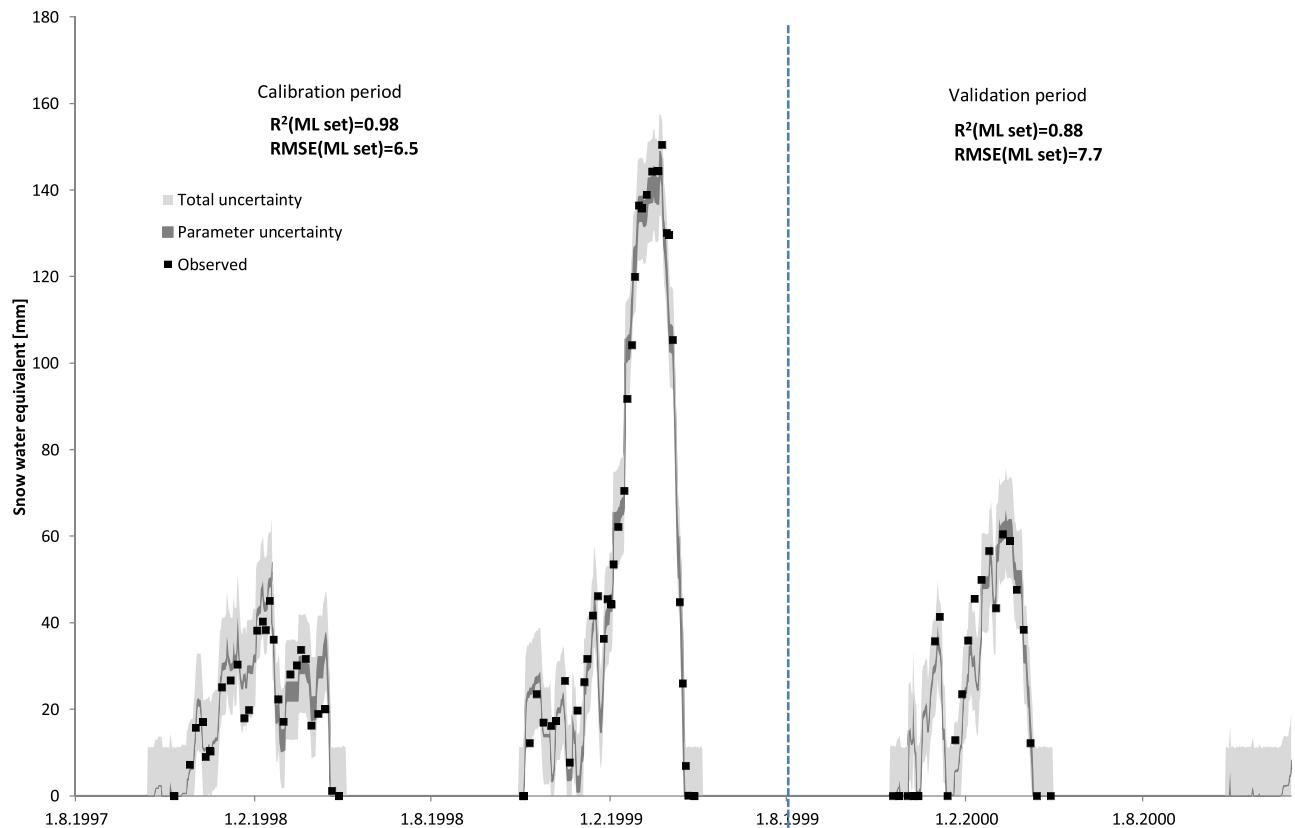


Figure 10. Uncertainty bounds (90%) for snow water equivalent in Approach B1.

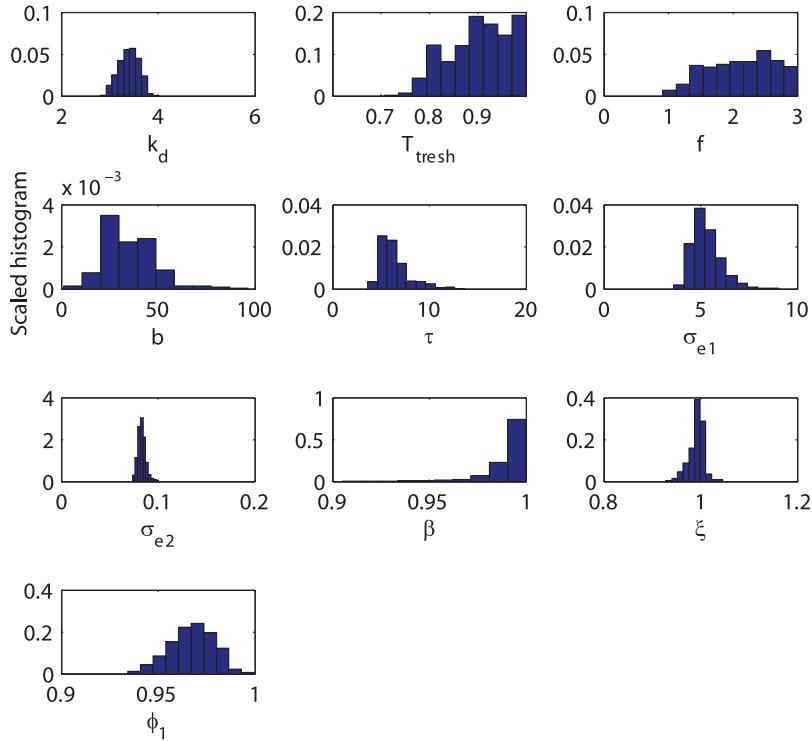


Figure 11. Scaled histograms of estimated posterior model parameters in Approach B2.

distributions. This is caused by correlations between some rainfall multipliers and the parameters b and τ (Figure 13), all of which are now influencing the water balance of the model. Furthermore, the posterior distribution of the degree-day factor k_d has clearly widened and the posterior of the threshold temperature T_{thresh} is at the upper limit of its assumed prior. Correlations between the model parameters and the latent variables (Figure 13) indicate that the latent variables, in addition to representing input uncertainty, are trying to compensate for model structural problems. Table 4

shows that the posteriors of the snow model parameters k_d and T_{thresh} are also still correlated.

[55] Most of the precipitation multipliers are not identified from the data and their posteriors are spread over the entire prior range (Figure 12). The well-identified latent variables are predominantly the ones describing the uncertainty of large precipitation events (class 4, class 6). Indeed, all the variables that describe rain events with accumulated rainfall less than 20 mm are not identified from the data. This is caused by the fact that either there is no

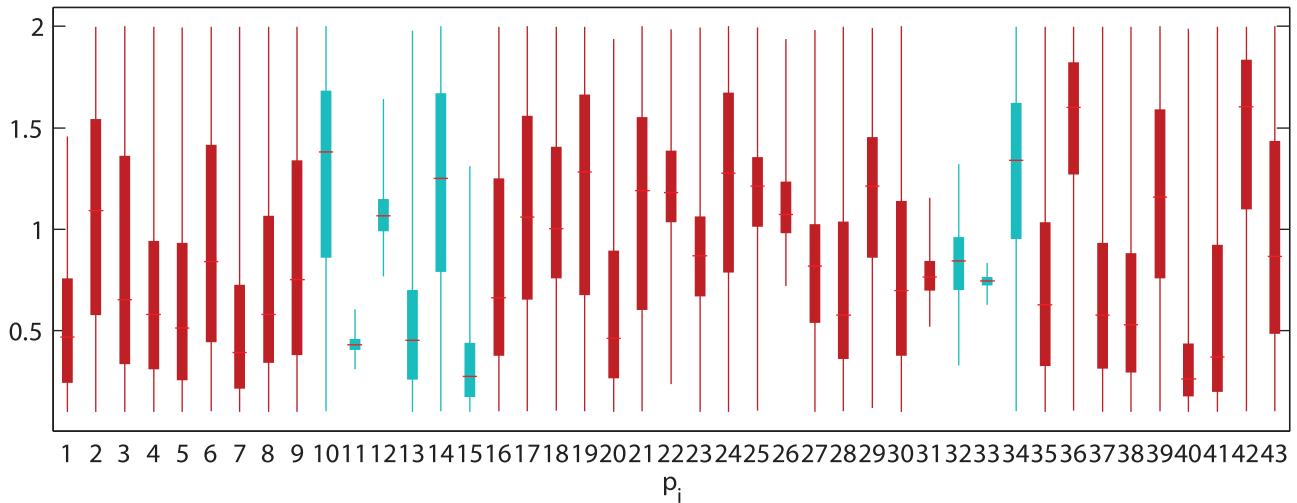


Figure 12. Box plot of the posterior values of the precipitation multipliers in the calibration period in Approach B2. Plot displays lower, median and upper quartile and whiskers extend to the most extreme data point. Snowfall events are represented with lighter colors.

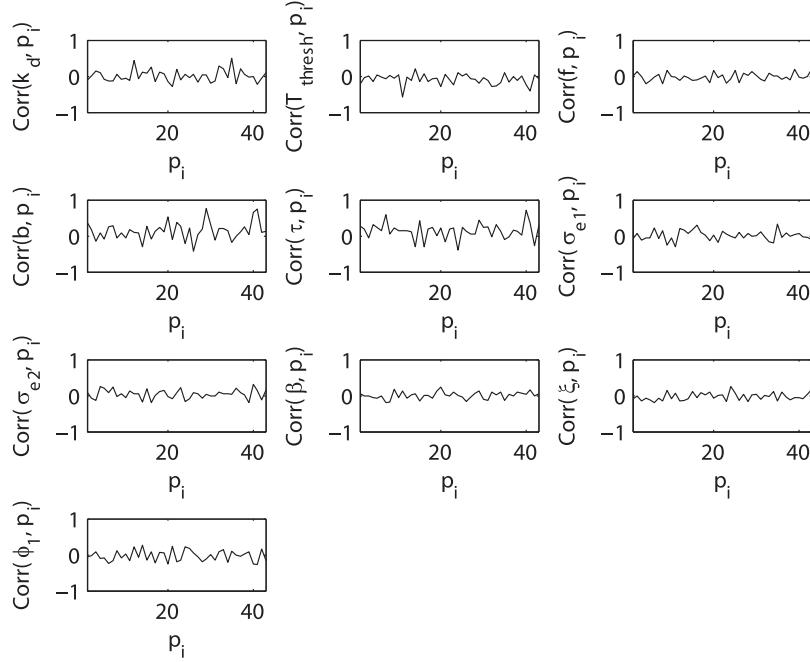


Figure 13. Correlations between model and latent parameters in Approach B2.

response in streamflow and SWE records to these minor rain events, the weight of small and short events in the objective function is negligible, or the autoregressive component is accounting for both input and model structure uncertainty. Thyer *et al.* [2009] proposed a method to remove unidentifiable latent variables from the model to speed up the convergence of MCMC chains.

[56] Figure 14 shows the histograms of the latent variables inferred from the data for each class presented in Table 2. It is obvious that precipitation uncertainty cannot be inferred from the data except for the large events. The use of distributions of class 1, class 2 and class 5 events, for example (Figure 14), to sample latent variables in validation leads to wide parameter and precipitation uncertainty limits for SWE. Similar effect is not seen in the modeled streamflow because it is simulated only 1-day-ahead with the inclusion of the AR1-model. Poorly identified storm event multipliers indicate that the modeler has only information about the timing of the storm event and no information about the accumulated precipitation of the event itself. If there were more informative priors available for the precipitation uncertainty, this

type of problem could be avoided [Renard *et al.*, 2010, 2011]. Such information for the most of the storm events cannot be inferred from the data in this case study.

[57] For Approach B2 the model is still not able to satisfactorily simulate all the flow peaks and streamflow recessions during the calibration period. The Nash-Sutcliffe efficiency R^2 is 0.86 in the calibration period but only 0.59 in the validation period. The unexplained variance of streamflow remains similar compared to that of Approach B1. For SWE the posterior of the unexplained variance is slightly narrower in B2 than B1 and the mode has decreased a little.

[58] In Approach B2 (Table 1) 26% of the streamflow observations in the calibration period are inside the 90% uncertainty bounds caused by the parameter and precipitation uncertainty, and 42% in validation. The 90% bounds for total uncertainty, however, cover 90% of the streamflow observations both in calibration and in validation. For SWE 55% of the observations in calibration are inside the uncertainty bounds caused by the parameter and precipitation uncertainty, and 82% in validation. These figures for the 90% bounds for total uncertainty are 95% for both

Table 4. Correlation Matrix of Model Parameters and Statistical Parameters in Approach B2

	k_d	T_{thresh}	f	b	τ	σ_{e1}	σ_{e2}	β	ξ	ϕ_1
k_d	1.00	0.46	0.09	-0.07	-0.03	0.14	-0.06	-0.07	0.01	0.04
T_{thresh}	0.46	1.00	0.15	-0.35	-0.38	0.16	-0.23	-0.12	0.08	0.01
f	0.09	0.15	1.00	0.04	-0.05	-0.15	-0.09	0.13	0.40	-0.09
b	-0.07	-0.35	0.04	1.00	0.58	0.01	0.18	0.13	-0.01	-0.19
τ	-0.03	-0.38	-0.05	0.58	1.00	-0.02	0.29	-0.03	-0.11	-0.15
σ_{e1}	0.14	0.16	-0.15	0.01	-0.02	1.00	-0.12	-0.10	-0.13	0.08
σ_{e2}	-0.06	-0.23	-0.09	0.18	0.29	-0.12	1.00	-0.04	-0.03	-0.16
β	-0.07	-0.12	0.13	0.13	-0.03	-0.10	-0.04	1.00	-0.01	0.06
ξ	0.01	0.08	0.40	-0.01	-0.11	-0.13	-0.03	-0.01	1.00	0.08
ϕ_1	0.04	0.01	-0.09	-0.19	-0.15	0.08	-0.16	0.06	0.08	1.00

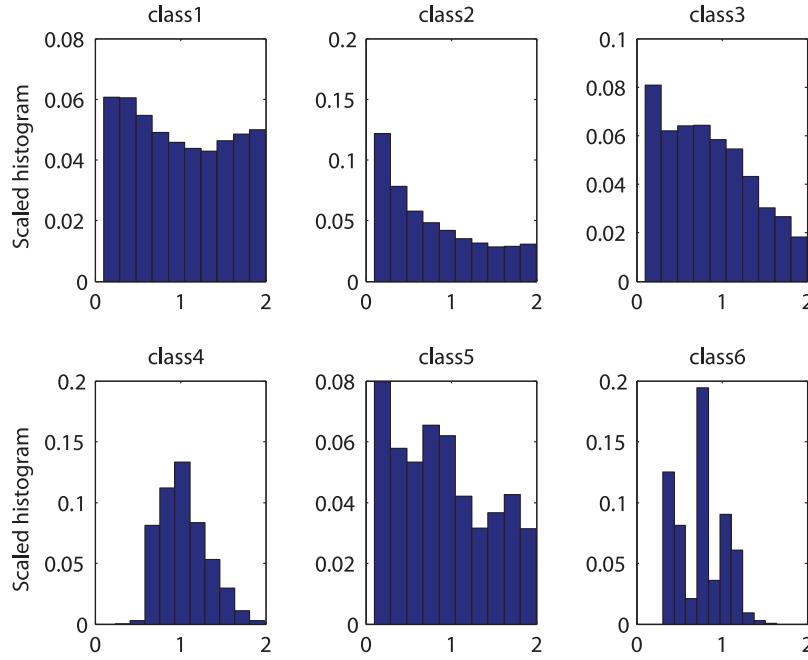


Figure 14. Histograms of latent variables inferred from the data for each storm event class.

calibration and validation. The uncertainty limits for SWE are overestimated.

[59] No significant ($\alpha = 0.05$) correlation was found between the residuals of snow and streamflow models. Statistical properties of the streamflow residual series (ML parameter estimates) are shown in Figure 15. The remaining autocorrelation is small (Figure 15b). The null hypothesis of the homoscedasticity is however, rejected by using the White test with significance level $\alpha = 0.05$. The scaled histogram of the observed residuals is visually a good presentation of the assumed distribution (Figure 15a). However, the two-sample Kolmogorov-Smirnov test rejects the null hypothesis that these two samples (Figure 15a) would arise from the same distribution. Further analysis reveals that the observed residual series contains a large number of zero residuals corresponding to time periods of zero flow. Thus the observed distribution is more highly peaked (high kurtosis) than the assumed $SEP(0,1,\xi,\beta)$. By repeating the two-sample Kolmogorov-Smirnov test for a time period from 1 December 1997 to 31 May 1999 that does not contain

zero flows, the null hypothesis could not be rejected. This observation points out the need for further consideration in the formulation of the likelihood function in case studies having a large number of zero or close to zero flows.

[60] Properties of the SWE residual series can be seen from the diagnostic plot (Figure 16). The residual series follows the assumed normal distribution (Kolmogorov-Smirnov test with significance level $\alpha = 0.05$). Compared to Approach A2 (Figure 6) the remaining autocorrelation between consecutive (lag-1) errors is clearly less significant and residuals are visually homoscedastic.

[61] Similarly to Approach B1 the model does not work properly in the simulation mode (i.e., when the AR1-model is not utilized). The R^2 value of the ML parameter set for streamflow is only 0.53 in the calibration period. This is caused by the effect of the latent variables and the AR1-model on model parameters. In validation the result depends on the selection criteria of the latent variables. If the mean value of the respective class is used, the R^2 value for streamflow is only 0.28 in the validation period.

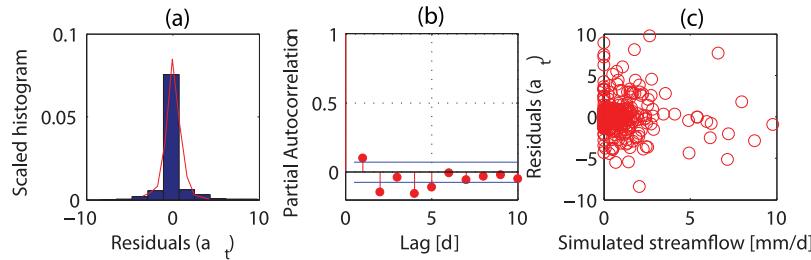


Figure 15. Residual analysis of the streamflow results using the maximum likelihood parameter set in Approach B2 and 1-day-ahead simulations. (a) The histogram is based on the observed errors and the red line is the theoretical density based on the assumed distribution. (b) Partial autocorrelation of residual series. (c) Residuals as a function of simulated streamflow.

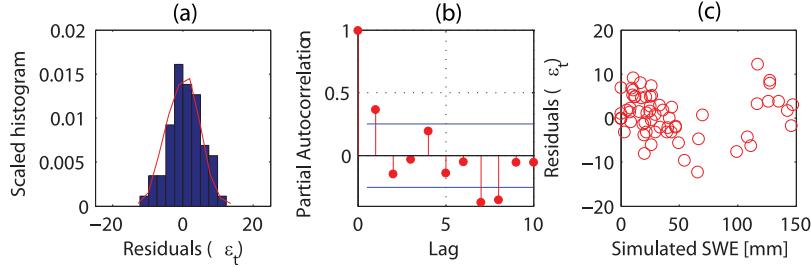


Figure 16. Residual analysis of the SWE results using the maximum likelihood parameter set in Approach B2. (a) The histogram is based on the observed errors and the red line is the theoretical density based on the assumed distribution. (b) Partial autocorrelation of residual series. (c) Residuals as a function of simulated SWE.

[62] Both *Vrugt et al.* [2009a] and *Kavetski et al.* [2006b] were able to infer the precipitation uncertainty along with model parameter uncertainty from the streamflow data. *Vrugt et al.* [2009a] used uniform priors for the latent variables in a similar manner to the present case study, but they were not trying to simultaneously estimate model structural uncertainties. Here, some model parameters could not be estimated at the same time with the latent variables. There was not enough information in the streamflow and SWE observations to simultaneously identify model parameters, model structural uncertainty, and uncertainties in precipitation observations. In the recent studies *Renard et al.* [2010, 2011] concluded that input and structural errors cannot be separated in the results (by using rainfall-runoff data) unless one has good prior knowledge about the magnitude of the errors. Model structural uncertainty was defined here differently, but the results are in line with *Renard et al.* [2010, 2011]. Furthermore, this case study shows that addition of observations from another state variable (SWE) does not unmask the sources of uncertainty in a snow-fed basin and more informative priors for precipitation must be used for proper inference. As suggested by earlier studies, e.g., *Seibert* [2000], it is possible to decrease the parameter uncertainty and to improve the model structure by using groundwater observations in calibration. There are groundwater observations available in the Rudbäck catchment, and thus this research line is interesting. However, this would require major changes in the model structure to describe groundwater levels and it was left for future research.

[63] During the snowmelt season the measured air temperature affects the snowmelt rate and thus streamflow peaks. In this study, input temperatures were assumed to be free-of-error but the effect of the temperature uncertainty will be the focus of subsequent studies. There are also indications that the degree-day factor in the model is dependent on the season [Kuusisto, 1984]. Thus it would be interesting to see whether model performance would improve if k_d was treated as a stochastic time-dependent variable.

[64] The log likelihood function (equation (21)) is affected by performance of both snow and streamflow simulations. The components in equation (21) that account for snow and streamflow errors were of the order of same magnitude. Furthermore, the unexplained variances were intuitively reasonable. Thus it is fair to say that neither one of the data sets is dominating the inference although there is clearly a higher number of streamflow measurements compared to snow

measurements in equation (21). This observation obviously cannot be generalized and it might be necessary to use nondimensionalizing to avoid scale issues in other case studies.

5. Conclusions

[65] The approach presented in the BATEA-framework of *Kavetski et al.* [2002] was used in this study with noninformative priors on the latent variables to account for precipitation uncertainty in the calibration of a conceptual model in a snow-fed catchment in southern Finland. The IHACRES rainfall-runoff model was coupled with a degree-day snow model to represent snow accumulation and melt. Model parameters were simultaneously calibrated against snow water equivalent (SWE) and streamflow observations by using a formal Bayesian approach with the posterior distribution sampled using the DREAM_(ZS) algorithm.

[66] The results suggest that the original degree-day model needs to be simplified because some of its parameters were not identifiable in the calibration. Furthermore, it is concluded that there was not enough information in the streamflow and SWE observations to simultaneously identify all the model parameters, the model structural uncertainty, and the latent variables describing precipitation uncertainty. The use of an autoregressive model to account for the model structural uncertainty affected the input uncertainty parameters leading to poor performance of the model in simulation mode. The introduction of the latent variables was found to have a clear effect on the calibrated model parameters. Both the mode of the parameters and wideness of the posterior distributions changed. This was caused by the mutual correlation between the latent variables and the model parameters. When the residual assumptions are satisfied using formal Bayesian approach, the inference should lead to unbiased parameter estimates (assuming the accounting of model structure error is sufficiently correct). The high value of the lag-1 autocorrelation coefficient ϕ_1 (for streamflow) is a strong evidence of structural problems in the model. The hypothesis that structural problems are the main cause of uncertainties in this case study is supported by the correlation between the latent variables and the model parameters, and by the relatively small improvement in model fit after the introduction of the latent variables. The hypothesis that the model structure is the main source of uncertainty cannot be confirmed, however. The evident interaction between the different sources of

uncertainties in the calibration exercise means that separation of individual uncertainties was not successful.

[67] The generalized likelihood function was used for the first time in combination with latent variables to account for precipitation uncertainty. DREAM_(ZS) still converged smoothly although the number of calibration parameters was high. The present study did not reveal any general problems in combining these two methods, and hence encourages future research in this area. It was observed, however, that availability of a priori information on the uncertainty estimates is desirable. The case study also revealed that the generalized likelihood function may not be flexible enough for hydrological conditions having long periods of zero flow.

[68] To conclude, more precise a priori knowledge of the input uncertainty as well as a better understanding and a description of the hydrological processes within the catchment are needed, rather than mining such information from the snow and runoff data. Based on this case study, the hypothesis that snow data would provide enough new information, compared to rainfall-runoff data alone, to let us use flat priors for precipitation multipliers must be rejected. Uncertainty estimates can be very inaccurate if error models are misspecified or if the uncertainties in observed streamflow, model structure, or model inputs are not adequately known a priori. This is caused by the fact that different components representing different sources of uncertainties interact in the inferencing procedure. In any case, application of Bayesian techniques can provide useful insights into the limitations of both model structure and data, and thus assist in the development of models having a suitable structure and complexity for the system of interest. It must be kept in mind, however, that posterior checks on the statistical assumptions adopted in defining the likelihood function should be conducted as a part of meaningful Bayesian inference.

[69] **Acknowledgments.** Funding to this research from the Maa-ja vesitekniikan tuki ry and Sven Hallin foundation is greatly acknowledged. Daily runoff data from the operational small catchment of Rudbäck was kindly provided by the Finnish Environment Institute and meteorological reference data by the Finnish Meteorological Institute. Jasper Vrugt is acknowledged for providing the code for the DREAM_(ZS)-algorithm and for the generalized likelihood function. We also wish to thank the reviewers whose in-depth comments made this article considerable better.

References

- Ajami, N. K., Q. Duan, and S. Sorooshian (2007), An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction, *Water Resour. Res.*, 43, W01403, doi:10.1029/2005WR004745.
- Ajami, N. K., Q. Duan, and S. Sorooshian (2009), Reply to comment by B. Renard et al. on “An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction,” *Water Resour. Res.*, 45, W03604, doi:10.1029/2008WR007215.
- Andrews, F. T., B. F. W. Croke, and A. J. Jakeman (2011), An open software environment for hydrological model assessment and development, *Environ. Modell. Softw.*, 26, 1171–1185, doi:10.1016/j.envsoft.2011.04.006.
- Beven, K. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, 320, 18–36.
- Beven, K. J., and A. M. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, 6, 279–298.
- Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems, *J. Hydrol.*, 249, 11–29.
- Beven, K., P. Smith, and J. Freer (2007), Comment on “Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology,” by P. Mantovan and E. Todini, *J. Hydrol.*, 338, 315–318, doi:10.1016/j.jhydrol.2007.02.023.
- Beven, K. J., P. J. Smith, and J. E. Freer (2008), So just why would a modeler choose to be incoherent?, *J. Hydrol.*, 354, 15–32.
- Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *J. R. Stat. Soc. Ser. B*, 26, 211–243.
- 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.
- Clark, M. P., and J. A. Vrugt (2006), Unraveling uncertainties in hydrologic model calibration: Addressing the problem of compensatory parameters, *Geophys. Res. Lett.*, 33, L06406, doi:10.1029/2005GL025604.
- Croke, B. F. W. (2009), Representing uncertainty in objective functions: Extension to include the influence of serial correlation, paper presented at 18th World IMACS/MODSIM Congress, Cairns, Australia.
- Croke, B. F. W., and A. J. Jakeman (2004), A catchment moisture deficit module for the IHACRES rainfall-runoff model, *Environ. Modell. Softw.*, 19, 1–5.
- Croke, B. F. W. and A. J. Jakeman (2005), Corrigendum to “A catchment moisture deficit module for the IHACRES rainfall-runoff model” [Environ. Model. Softw. 19(1) (2004) 1–5], *Environ. Modell. Softw.*, 20(7), 977. doi:10.1016/j.jhydrol.2006.07.005.
- Douglas, J. R., R. T. Clarke, and S. G. Newton (1976), The use of likelihood functions to fit conceptual models with more than one dependent variable, *J. Hydrol.*, 29(1–2), 181–198.
- Duan, Q., V. K. Gupta, and S. Sorooshian (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28, 1015–1031.
- Efstratiadis, A., and D. Koutsoyiannis (2010), One decade of multi-objective calibration approaches in hydrological modelling: A review, *Hydrol. Sci. J.*, 55(1), 58–78.
- Førland, E. J., P. Allerup, B. Dahlström, E. Elomaa, T. Jónsson, H. Madsen, J. Perälä, P. Rissanen, H. Vedin, and F. Vejen (1996), Manual for operational correction of Nordic precipitation data, *Rep. 24/96*, 66 pp., Norwegian Meteorol. Inst., Oslo.
- Franz, K. J., P. Butcher, and N. K. Ajami (2010), Addressing snow model uncertainty for hydrologic prediction, *Adv. Water Resour.*, 33, 820–832, doi:10.1016/j.advwatres.2010.05.004.
- Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci.*, 7(4), 457–472.
- Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, 34(4), 751–763, doi:10.1029/97WR03495.
- Gupta, H. V., T. Wagener, and Y. Liu (2008), Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, *Hydrol. Processes*, 22(18), 3802–3813, doi:10.1002/hyp.6989.
- He, M., T. S. Hogue, K. J. Franz b, S. A. Margulis, and J. A. Vrugt (2011), Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes, *Adv. Water Res.*, 34, 114–127.
- Huard, D., and A. Mailhot (2008), Calibration of hydrological model GR2M using Bayesian uncertainty analysis, *Water Resour. Res.*, 44, W02424, doi:10.1029/2007WR005949.
- Jakeman, A. J., and G. M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, 29, 2637–2649.
- Jakeman, A. J., I. G. Littlewood, and P. G. Whitehead (1990), Computation of the instantaneous unit hydrograph and identifiable components flow with application to two small upland catchments, *J. Hydrol.*, 117, 275–300.
- Kavetski, D., S. W. Franks, and G. Kuczera (2002), Confronting input uncertainty in environmental modelling, in *Calibration of Watershed Models*, *Water Sci. Appl.*, vol. 6, edited by Q. Duan et al., pp. 49–68, AGU, Washington, D. C.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006a), Bayesian analysis of input uncertainty in hydrological modelling: 1. Theory, *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006b), Bayesian analysis of input uncertainty in hydrological modeling: 2. Application, *Water Resour. Res.*, 42, W03408, doi:10.1029/2005WR004376.
- Kirchner, J. W. (2009), Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, *Water Resour. Res.*, 45, W02429, doi:10.1029/2008WR006912.
- Koivusalo, H. (2002), Process-oriented investigation of snow accumulation, snowmelt and runoff generation in forested sites in Finland, PhD Thesis,

- 88 pp., *Water Resour. Publ. TKK-VTR-6*, Helsinki Univ. of Technol., Espoo, Finland.
- Koivusalo, H., and T. Kokkonen (2002), Snow processes in a forest clearing and in a coniferous forest, *J. Hydrol.*, 262, 145–164.
- Koivusalo, H., and T. Kokkonen (2003), Modelling runoff generation in a forested catchment in southern Finland, *Hydrol. Processes*, 17(2), 313–328.
- Kokkonen, T., H. Koivusalo, T. Jakeman, and J. Norton (2006), Construction of a degree-day snow model in the light of the ten iterative steps in model development, in *Proceedings of the iEMSs Third Biennial Meeting: Summit on Environmental Modelling Software*, [CD ROM], edited by A. Voinov et al., Int. Environ. Modell. Software Soc., Burlington, VT.
- Kuczera, G. (1983), Improved parameter inference in catchment models. I. Evaluating parameter uncertainty, *Water Resour. Res.*, 19(5), 1151–1162, doi:10.1029/WR019i005p01151.
- Kuczera, G., D. Kavetski, S. Franks, and M. Thyer (2006), Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, *J. Hydrol.*, 331(1–2), 161–177, doi:10.1016/j.jhydrol.2006.05.010.
- Kuusisto, E. (1984), Snow accumulation and snowmelt in Finland, in *Publ. Water Res. Inst.* 55, 149 pp., Natl. Board of Waters, Helsinki, Finland.
- Laloy, E., and J. A. Vrugt (2012), High-dimensional posterior exploration of hydrologic models using multiple-try DREAM_(ZS) and high-performance computing, *Water Resour. Res.*, 48, W01526, doi:10.1029/2011WR010608.
- Laloy, E., D. Fasbender, C. Bielders (2010), Parameter optimization and uncertainty analysis for plot-scale continuous modeling of runoff using a formal Bayesian approach, *J. Hydrol.*, 380, 82–93, doi:10.1016/j.jhydrol.2009.10.025.
- Lepistö, A. (1996), Hydrological processes contributing to nitrogen leaching from forested catchments in Nordic conditions, in *Monogr. Boreal Environ. Res.* 1, 72 pp., Finnish Environ. Inst., Helsinki, Finland.
- Lepistö, A., and Y. Kivinen (1997), Effects of climatic change on hydrological patterns of a forested catchment: A physically based modeling approach, *Boreal Environ. Res.*, 2, 19–31.
- Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, 330, 368–381, doi:10.1016/j.jhydrol.2006.04.046.
- Massey, F. J. (1951), The Kolmogorov-Smirnov test for goodness of fit, *J. Am. Stat. Assoc.*, 46(253), 68–78.
- McMillan, H., B. Jackson, M. Clark, D. Kavetski, and R. Woods (2011), Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models, *J. Hydrol.*, 400(1–2), 83–94, doi:10.1016/j.jhydrol.2011.01.026.
- Refsgaard, J. C., J. P. van der Sluijs, J. Brown, and P. van der Keur (2006), A framework for dealing with uncertainty due to model structure error, *Adv. Water Resour.*, 29, 1586–1597.
- Reichert, P., and J. Mieleitner (2009), Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, *Water Resour. Res.*, 45, W10402, doi:10.1029/2009WR007814.
- Renard, B., D. Kavetski, and G. Kuczera (2009), Comment on “An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction” by Newsha K. Ajami et al., *Water Resour. Res.*, 45, W03603, doi:10.1029/2007WR006538.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, 46, W05521, doi:10.1029/2009WR008328.
- Renard, B., D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S. W. Franks (2011), Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation, *Water Resour. Res.*, 47, W11516, doi:10.1029/2011WR010643.
- Rutter, N., et al. (2009), Evaluation of forest snow processes models (Snow-MIP2), *J. Geophys. Res.*, 114, D06111, doi:10.1029/2008JD011063.
- Schaeafi, B., B. D. Talamba, and A. Musy (2007), Quantifying hydrological modeling errors through a mixture of normal distributions, *J. Hydrol.*, 332, 303–315.
- Schoups, G., and J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resour. Res.*, 46, W10531, doi:10.1029/2009WR008933.
- Schoups, G., J. A. Vrugt, F. Fenicia, and N. C. van de Giesen (2010), Corruption of accuracy and efficiency of Markov chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models, *Water Resour. Res.*, 46, W10530, doi:10.1029/2009WR008648.
- Seibert, J. (2000), Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, *Hydrol. Earth Syst. Sci.*, 4, 215–224.
- Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeller in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, 38(11), 1241, doi:10.1029/2001WR000978.
- Sorooshian, S., and J. A. Dracup (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models—correlated and heteroscedastic error cases, *Water Resour. Res.*, 16(2), 430–442, doi:10.1029/WR016i002p00430.
- Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resour. Res.*, 44, W00B06, doi:10.1029/2008WR006822.
- ter Braak, C. J. F., and J. A. Vrugt (2008), Differential evolution Markov chain with snooker updater and fewer chains, *Stat. Comput.*, 18(4), 435–446, doi:10.1007/s11222-008-9104-9.
- Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, 45, W00B14, doi:10.1029/2008WR006825.
- Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, 44, W00B09, doi:10.1029/2007WR006720.
- Vrugt, J. A., C. J. F. ter Braak, H. V. Gupta, and B. A. Robinson (2009a), Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, *Stoch. Environ. Res. Risk Assess.*, 23, 1011–1026, doi:10.1007/s00477-008-0274-y.
- Vrugt, J. A., C. J. F. ter Braak, C. G. H. Diks, D. Higdon, B. A. Robinson, and J. M. Hyman (2009b), Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling, *Int. J. Nonlinear Sci. Numer. Simul.*, 10(3), 273–290.
- White, H. (1980), A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, *Econometrica*, 48(4), 817–838.