

Quantile hydrologic model selection and model structure deficiency assessment: 1. Theory

Saket Pande¹

Received 10 November 2011; revised 18 June 2013; accepted 7 July 2013.

[1] A theory for quantile based hydrologic model selection and model structure deficiency assessment is presented. The paper demonstrates that the degree to which a model selection problem is constrained by the model structure (measured by the Lagrange multipliers of the constraints) quantifies structural deficiency. This leads to a formal definition of model structure deficiency (or rigidity). Model structure deficiency introduces a bias in the prediction of an observed quantile, which is often not equal across quantiles. Structure deficiency is therefore diagnosed when any two quantile predictions for a given model structure cross since unequal bias across quantiles result in quantile predictions crossing. The analysis further suggests that the optimal value of quantile specific loss functions order different model structures by its structure deficiencies over a range of quantiles. In addition to such novelties, quantile hydrologic model selection is a frequentist approach that seeks to complement existing Bayesian approaches to hydrological model uncertainty.

Citation: Pande, S. (2013), Quantile hydrologic model selection and model structure deficiency assessment: 1. Theory, *Water Resour. Res.*, 49, doi:10.1002/wrcr.20411.

1. Introduction

[2] Current practice of uncertainty assessment of hydrologic models hypothesizes potential sources of errors, assumes that it obey certain distribution types and nests these distributions within a Bayesian inference framework [Kavetski *et al.*, 2006; Thyer *et al.*, 2009; Schoups and Vrugt, 2010; Smith *et al.*, 2010]. Bayesian inference therefore allows simultaneous modeling of uncertainties due to model and measurement errors. These methods are powerful and yield useful insights for improving model structures. A validation of assumptions is generally made by Q-Q plots by mapping observed quantiles to prediction quantiles for a variable of interest [Thyer *et al.*, 2009; Schoups and Vrugt, 2010]. A Q-Q plot verifies whether the prediction quantiles follow the observed quantiles, thereby assessing the applicability of the model assumptions.

[3] An extension of quantile regression to hydrologic model selection is proposed, which aims to identify a model, for a given model structure, by minimizing a loss function that asymmetrically penalizes the positive and negative residuals. Here a residual is defined as the difference between a prediction and the observed [Koenker and Basset, 1978]. The penalty determines the quantile of the observed data at which the model is being estimated. This contrasts likelihood methods that model an entire distribution by assuming a likelihood function. The likelihood methods do

not model 1 quantile at a time. It is equivalent to estimating a model such that the prediction of a variable of interest is as close as possible to a desired quantile of its observations. Observed quantiles may exactly be predicted when the model structure contains the truth. Quantile hydrologic model estimation presented here can be seen as equivalent to an inverse approach to Q-Q plot verification. A model is selected to match an observed quantile as closely as possible, instead of using the quantile to judge how well a model (selected independently using another inference method) replicates that quantile. The underlying motivation is to compare two model structures in terms of its deficiencies in representing the underlying processes (“truth”). In contrast to Bayesian approaches to model selection [such as Kavetski *et al.*, 2006; Thyer *et al.*, 2009; Schoups and Vrugt, 2010] where various sources of errors can be explicitly modeled, no assumption on the cumulative distribution of the residuals is made, where a distribution of residuals is due to unknown measurement errors and model structure deficiency.

[4] A Bayesian approach [Marshall *et al.*, 2006; Kavetski *et al.*, 2006; Schoups and Vrugt, 2010] to model selection is limited in certain aspects. The model parameters sampled based on a formal likelihood function is from a posterior parameter distribution only if the underlying processes belong to the model space or that the model space is fully specified [Davidson and MacKinnon, 2004, p. 399]. It is only then that the posterior distribution can be assumed to be proportional to the likelihood function based on Bayes rule and hence it is only then that model estimation (selection) can be based on a likelihood function. This is critical in studying model structure deficiency (in the sense of how limited a model structure is in representing the underlying processes). Innovating a complex error model such as in Schoups and Vrugt [2010] may ameliorate such

Companion to Pande [2013], doi:10.1002/wrcr.20422.

¹Department of Water Management, Delft University of Technology, Delft, Netherlands.

Corresponding author: S. Pande, Department of Water Management, Delft University of Technology, Delft NL-2628CN, Netherlands. (s.pande@tudelft.nl)

concerns in practice. However, a simple model error that cannot be represented by the family of additive skew exponential power distributed [DiCiccio and Monti, 2004] error is sufficient to show the limitations of even such complex error models. For example, when the error due to model structure deficiency is correlated with model predictions [Pande *et al.*, 2012a], it leads to an effect that is different from the heteroscedasticity effect. Any estimation technique that ignores its presence (dependence of error on model predictions) leads to a biased model estimate [Heckman, 2005]. Yet another limitation is that a posterior density is conditional on data, which for small samples can itself be uncertain due to sampling uncertainty [Pande *et al.*, 2009]. This though equally holds for the method proposed here.

[5] A Bayesian approach is superior to the proposed method when its assumptions on the error structure hold. This is because the assumptions on the error model structure define the likelihood function, which when valid yield the “true” parameter values of a hydrological model at the likelihood maximum. For example, Schoups and Vrugt [2010] assume that the error distribution belongs to a family of additive skew exponential power distribution [DiCiccio and Monti, 2004]. The method proposed in the paper makes no assumption on the structure of uncertainty due to underlying processes or measurement errors. This makes it difficult to isolate the uncertainty due to model structure from measurement uncertainty. However, more often than not, the assumptions on error structure (not just distributional but also how the model error enters the assumed error structure) do not hold. It is in this respect, i.e., of not distinguishing between different sources of error, that the presented method is similar in essence to the generalized likelihood uncertainty estimation (GLUE) methodology [Beven *et al.*, 2008]. The measurement uncertainty may, however, be isolated from model structure uncertainty by using noise (due to measurement error) adapted data based on measurement error benchmarking studies [McMillan *et al.*, 2012].

[6] Thus a motivation behind this paper is to propose a model selection and deficiency assessment approach that is at least not constrained by the requirement to possess “strong” a priori information about reality [Vapnik, 2002, p. 118]. Quantile hydrological model selection and assessment of model structure deficiencies based on it is therefore proposed. Its central idea is that a bias in model estimation by a method that does not assume any error model contains useful information on model structure deficiencies. Further, such an assessment is holistic when it is over the entire range of predictions (such as quantiles of flows with quantiles ranging from 0 to 1) of a model structure. It employs a loss function of Koenker and Bassett [1978], based on absolute deviations, as an objective function for estimating models that removes the need to identify quantiles of an observed time series.

[7] A deficient model structure constrains how well a quantile of observed variable of interest can be modeled. Different model structures may constrain its prediction of the same quantile in different manner, introducing different bias in predicting observed quantiles over a range of quantile values. The paper demonstrates that quantile model selection incorporates quantile specific bias due to model

structure deficiencies in the asymmetric loss function. The loss function thereby allows an ordering of model structures based on their flexibility to model a quantile. Further, model structure deficiencies may induce two quantile predictions of a model structure to cross, yielding a useful diagnosis of structure deficiency. The methodology in the paper thus provides both quantile model predictions for a given model structure and insights into model structure deficiencies for a collection of model structures.

[8] Quantile hydrological model selection is not the same as a standard quantile regression where the underlying model space is of linear functions. Though the standard quantile regression is also a quantile model selection problem, its model space is restricted (since it is linear). Thus, the extension of quantile model selection to a hydrological model space is nontrivial. This is where the need to formally analyze the properties of quantile “hydrological” model selection arises. One property that is crucial is the noncrossing of quantile predictions [Koenker and Bassett, 1978; Keyzer and Pande, 2009]. The conditions under which quantile predictions do not cross therefore need to be made explicit. Its formal treatment is beneficial as it formalizes the notion of model bias due to model structure deficiencies and the conditions reveal that if quantile hydrological predictions cross, it is due to model structure deficiency. It also reveals that bias in predicting observed quantiles due to structure deficiencies is independent of model parameter dimensionality and is time invariant. These are two strong properties that further allow us to compare different structures in terms of its structural deficiencies.

[9] This paper develops the theory of quantile hydrological model selection and deficiency assessment. Its companion paper [Pande, 2013] implements the theory in detail and studies cases of a parsimonious dryland model developed for western India [Pande *et al.*, 2012a, 2011, 2010], model structures for Guadalupe river basin [Schoups and Vrugt, 2010] and validates the performance of quantile model selection and deficiency assessment on French Broad River basin data using a flexible model structure.

[10] The paper is organized as follows. Section 2 first introduces the methodology, with implementations on a linear regression model, on a simple three-parameter hydrological model with a threshold and two case studies with complex hydrological models as examples. The latter three studies also compare and contrast the approach with Bayesian and point statistics approaches to model selection to elucidate the utility of the approach. A formal analysis of quantile hydrological model selection is then presented in section 3 that expands upon and generalizes the observations made in section 2. Section 4 then discusses the formal results, finally concluding with section 5.

2. Methodology

2.1. An Example of Quantile Regression

[11] Consider a data generating processes DGP1,

$$\begin{aligned} y_i &= 0.5x_i + 0.2 + \varepsilon_i, \\ \varepsilon_i &= \eta x_i \end{aligned}$$

where, x_i is independently and uniformly distributed, η is normally distributed with mean 0 and variance 0.25, and i

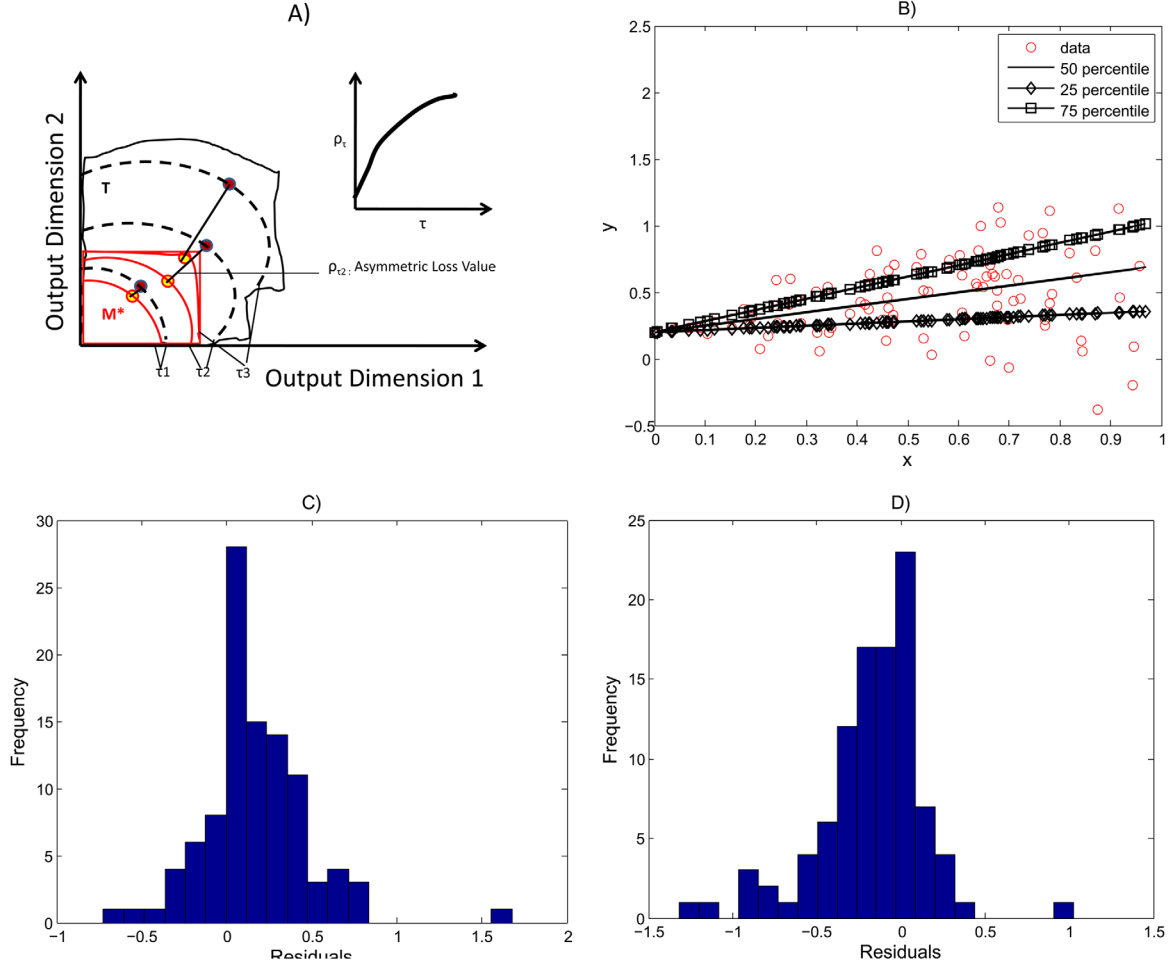


Figure 1. Quantile model selection describes the (a) concept behind quantile model selection, T is the true model output space while M^* indicates optimized model output space as a result of chosen structure (b) three different quantile (25, 50, and 75 percentiles) linear models on data generated by DGP1 are obtained such that the ratio of number of positive to negative residuals are $1/4$, 1 , and $3/4$ respectively. (c and d) The frequency distributions of the residuals for 25 and 75 percentile models, shown in Figure 1b, showing that ratio of positive to negative residuals is approximately $1/4$ and $3/4$ respectively.

indexes data points with $i=1, \dots, N$ where N is the sample size.

[12] Let us assume that one can regress a τ quantile specific parametric (linear) function $\hat{y}_\tau = a_\tau x + b_\tau$ by minimizing a certain objective function (to be discussed in section 2.2) S_τ , such that the frequency ratio of resulting positive residuals, i.e., $y_i - \hat{y}_{i,\tau} > 0$, to negative residuals, i.e., $y_i - \hat{y}_{i,\tau} \leq 0$, is $\tau/(1-\tau)$. This is described by Figures 1b–1d. Figure 1b shows the data set and displays three linear functions corresponding to $\tau=0.25, 0.5$, and 0.75 . Figures 1c and 1d show the frequency distribution of residuals corresponding to $\tau=0.25$ and 0.75 . Note that the estimation of parameters $\{a_\tau, b_\tau\}$ such that $\text{frequency}(\hat{y}_\tau \leq y)/\text{frequency}(\hat{y}_\tau > y) = \tau/1-\tau$ is equivalent to finding a prediction model \hat{y}_τ that matches the τ th quantile of observed y (since $\text{frequency}(\hat{y}_\tau \leq y) = \tau$). Note that this example is a case when the model structure (a set of linear functions) contains the “truth” since DGP1 is a linear function with a random intercept that has variance pro-

portional to x^2 . Thus the model structure used in not binding (is not constraining the predictions or is not deficient). At the same time note that the three quantile predictions do not cross. This may indicate that the quantile predictions do not cross when a model structure is not binding. Quantile specific parameter estimation, under no constraints posed by the model structure (here model structure is class of linear functions), is therefore equivalent to an inverse method of quantile matching that Q-Q plots otherwise aim to verify.

[13] Figure 1a further describes a situation when a given model structure is deficient. It shows a two-dimensional output space, wherein each axis represents a dimension corresponding to a data point. Thus, the output space is N dimensional when N is the sample size. T is the “true” model output space, which is a collection of all possible output points mapped by nature as a result of all possible input forcings x . Let the dashed lines in T represent its three quantiles (say 0.25, 0.5, 0.75 quantiles). A quantile observation y conditional on a given input forcing x (shown by red circles in T) is located on these iso-quantile lines of

T. Consider a model structure $\mathfrak{S} = \{m(\alpha) : \alpha \in \Lambda\}$ as a collection of several models $m(\alpha)$ which result from a certain choice of parameter values. This equivalently represents a model structure output space. Figure 1a represents optimized model output space M^* such that some measure of distance between observed quantile and modeled quantile prediction is minimum for each quantile and for a given input forcing \mathbf{x} . This norm measures the distance between two points in the output space and is shown in Figure 1a by the magnitude of lines connecting points on iso-quantile lines of T and M^* . A nonlinear monotonic function of this measure is called asymmetric loss function ρ_τ in this paper (also referred to as the loss function of *Koenker and Basset* [1978]).

[14] We show in the following section that two model structures that differ in its deficiencies to encapsulate T have different ρ_τ curves. The closeness of ρ_τ curve to 0 can judge the least deficient model structure in a pair. This forms a basis to compare different model structures.

2.2. Implementation of Quantile Model Selection on Arbitrary Model Structure

[15] Consider an observed data set $\{y_i, \mathbf{x}_i\}_{i=1, \dots, N}$ where $y_i \in \mathbb{R}$, $\mathbf{x}_i \in \mathbb{R}^M$. Here N is the sample size and M is the dimensionality of \mathbf{x}_i . Let \mathbf{X} be a $N \times M$ matrix with \mathbf{x}_i being the i th row. Let $F = \{f(\boldsymbol{\beta}; \mathbf{x}), \boldsymbol{\beta} \in \Lambda\}$ be a class of functions whose parameter set $\boldsymbol{\beta}$ needs to be estimated from the observed data set. A τ -quantile specific function and corresponding parameters are estimated by minimizing an asymmetrically weighted loss function [*Koenker and Basset*, 1978] ρ_τ ,

$$\sum_{i=1}^N [(1-\tau)|y_i - f(\boldsymbol{\beta}; \mathbf{x}_i)|_- + \tau|y_i - f(\boldsymbol{\beta}; \mathbf{x}_i)|_+].$$

Here,

$$|y_i - f(\boldsymbol{\beta}; \mathbf{x}_i)|_- = \begin{cases} |y_i - f(\boldsymbol{\beta}; \mathbf{x}_i)| & \text{if } y_i - f(\boldsymbol{\beta}; \mathbf{x}_i) < 0, \\ 0 & \text{otherwise} \end{cases},$$

and

$$|y_i - f(\boldsymbol{\beta}; \mathbf{x}_i)|_+ = \begin{cases} |y_i - f(\boldsymbol{\beta}; \mathbf{x}_i)| & \text{if } y_i - f(\boldsymbol{\beta}; \mathbf{x}_i) \geq 0, \\ 0 & \text{otherwise} \end{cases}.$$

[16] The above estimation can alternatively be formulated as (QE1) [see *Keyzer and Pande*, 2009],

$$\begin{aligned} W(\tau; \mathbf{y}, \mathbf{X}) = \min_{\boldsymbol{\beta}, \boldsymbol{\varepsilon}_i^-, \boldsymbol{\varepsilon}_i^+} \sum_{i=1}^N [(1-\tau)\boldsymbol{\varepsilon}_i^- + \tau\boldsymbol{\varepsilon}_i^+] \\ \text{st} \\ y_i - f(\boldsymbol{\beta}; \mathbf{x}_i) - \boldsymbol{\varepsilon}_i^+ \leq 0 \\ -y_i + f(\boldsymbol{\beta}; \mathbf{x}_i) - \boldsymbol{\varepsilon}_i^- \leq 0 \\ \boldsymbol{\varepsilon}_i^+, \boldsymbol{\varepsilon}_i^- \geq 0 \end{aligned}$$

[17] This formulation can be extended to conceptual water balance models, which is the scope of this paper, as in the following simple generalization. Let S_t denote the storage of a reservoir and let its outflux be a function of the storage denoted by $f(S_t, \mathbf{k})$. Here $\mathbf{k} \in \mathbf{K}$ represents a set of parameters (for example, slow and fast runoff coefficients),

\mathbf{K} represents the range of parameters and corresponds to a particular model structure. Let $\{y_t, x_t\}_{t=1, \dots, N}$ represent observed data set where $y_t \in \mathbb{R}_+$, $x_t \in \mathbb{R}_+$ represents observed outflux and input forcing, respectively, at time t . Let $\mathbf{x} = \{x_t\}_{t=1, \dots, N}$ represent the input forcing vector and let S_o be the initial soil moisture condition. A τ -quantile specific function and the corresponding parameters based on outflux observations can be estimated by the program (QE2):

$$\begin{aligned} W(\tau; \mathbf{y}, \mathbf{x}) = \min_{\mathbf{k} \in \mathbf{K}, \boldsymbol{\varepsilon}_t^-, \boldsymbol{\varepsilon}_t^+, S_t} \sum_{t=1}^N [(1-\tau)\boldsymbol{\varepsilon}_t^- + \tau\boldsymbol{\varepsilon}_t^+] \\ \text{st} \\ y_t - f(S_t, \mathbf{k}) - \boldsymbol{\varepsilon}_t^+ \leq 0 \\ -y_t + f(S_t, \mathbf{k}) - \boldsymbol{\varepsilon}_t^- \leq 0 \\ S_{t+1} - S_t = x_t - f(S_t, \mathbf{k}) \\ \boldsymbol{\varepsilon}_t^+, \boldsymbol{\varepsilon}_t^- \geq 0 \end{aligned}$$

2.3. A Comparison of Quantile Model Selection With Bayesian and Point Statistics Based Inference

[18] Three case studies now examine quantile model selection and contrast it with Bayesian and point statistics based inference. Two of these studies are synthetic in nature. The section concludes with a synthesis of observations from the three case studies and conceptually illustrates how quantile model selection incorporates structural deficiency assessment. These arguments are then formalized in the theory presented in section 3.

[19] A common theme across the three case studies is the inference (or the estimation) of models from deficient model structures. This is achieved in the first two case studies by synthetically generating time series of streamflow based on models that are more complex than the model structure(s) used for inference. In the third case study, a real data set is used to infer models using two model structures under an assumption that a model structure is always deficient.

[20] The first case study synthetically generates a streamflow data set using a linear reservoir model with a threshold on a synthetic rainfall time series. It then uses a linear reservoir model (without a threshold) structure to (i) infer quantile models (one for each quantile for a range of quantiles between 0 and 1) using quantile model selection, (ii) infer posterior distributions of the recession parameter (the only parameter of the linear reservoir model) using three different likelihood measures, and (iii) compare and contrast the two approaches in terms of the diagnosis of structural deficiencies. In particular, the case study discusses the validity of the three likelihood measures as probability measures and the crossing of quantile predictions as a diagnostic of structure deficiency.

[21] The second synthetic case study builds upon the first in complexity. It uses the climatic forcing of the French Broad river basin over 5 years and uses a more complex conceptual rainfall runoff model (with multiple parameters and states) to generate a synthetic streamflow time series. Two subcase studies are examined: without additive noise and with 10% heteroscedastic Gaussian additive noise on the synthesized streamflow. Two model structures (a linear reservoir model and a linear reservoir model with a threshold) are then used to infer models. The inference is again

based on quantile model selection and Bayesian inference using a general likelihood function. Three approximations of the marginal likelihood that are often used for Bayesian model selection are used. The two inference methods are again compared and contrasted with particular attention to (i) how the crossing of quantile predictions is associated with the degree to which a model structure is deficient (measured by approximation of bias in predicting quantiles that result from the deficiencies), (ii) how the crossing of quantile predictions identifies the quantile locations of model structure deficiencies, and (iii) how the ordering of model structures based on the loss function of quantile model selection ranks model structures and contrast it with the orderings based on Bayesian measures and point statistics.

[22] The third case study adds more complexity to the second case study by using the French Broad river basin data set and inferring models using two complex conceptual rainfall runoff model structures (with multiple parameters and states), a linear reservoir with a threshold model structure and a linear reservoir (without a threshold) model structure. All four model structures are such that they are nested. A comparative analysis between the approaches as in the second synthetic case study is again performed.

[23] In all the three case studies, the asymmetric loss functions are minimized using the SCE-UA algorithm [Duan *et al.*, 1992]. SCE-UA searches for a global optimum by independently (but periodically shuffled) evolving m complexes each containing p parameter sets based on operations such as expansion, contractions and reflection. Readers are referred to Duan *et al.* [1992] for additional details. For the study m is fixed at 20, $p = 41$ with a convergence criteria of 0.1% (change in objective function) and the search is terminated after 100,000 objective function evaluations if no convergence is achieved.

2.3.1. Inference of a Linear Reservoir Model on Data Synthetically Generated by a Thresholded Linear Reservoir Model

[24] This section discusses quantile model selection and contrasts it with Bayesian inference based on three likelihood measures. Both the approaches infer a linear reservoir model (a one parameter model) on a synthetic data set generated by a linear reservoir model with a threshold (a three parameter model).

[25] Figure 2a displays the linear reservoir with a threshold model, with two recession parameters, for slow (k_1 [1/T]) and fast runoff (k_2 [1/T]) respectively and a threshold S_{\max} [L]. A data set, D , of total flow of length $N = 50$ is generated (observations representing the underlying processes) by assuming $\{k, k_1, S_{\max}\} = \{0.1, 0.5, 3\}$ and forcing the model with a triangular type rainfall. Figure 2b shows the precipitation forcing, total and overland flows where the latter represents the fast component of the flow when the storage exceeds S_{\max} . The max (and similarly min) operators are smoothed as $\max(a, b) \rightarrow \log(\exp(a) + \exp(b))$. The threshold is therefore smoothed.

[26] A linear reservoir without a threshold model structure is then used to identify models that best represent the underlying processes. A linear reservoir with a threshold model structure contains the linear reservoir without a threshold model structure (since the latter is the former with infinite threshold). Data that are generated by a linear

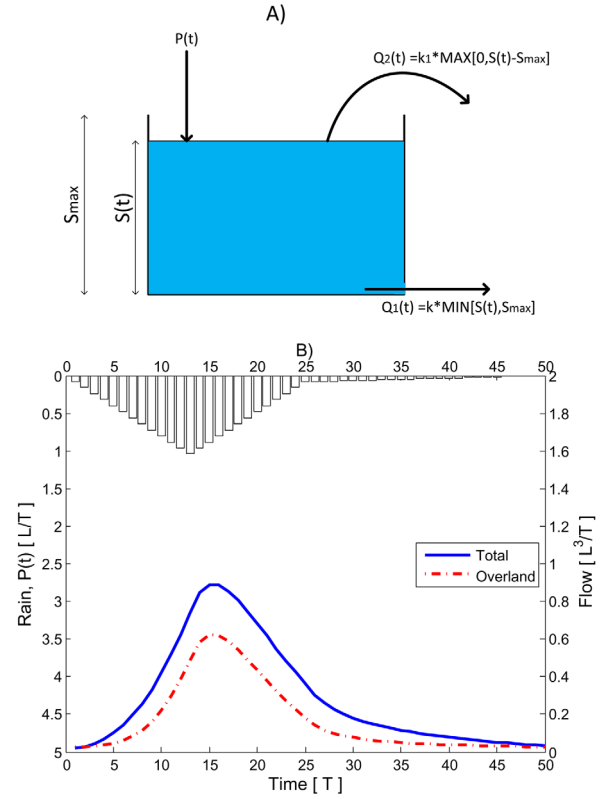


Figure 2. The synthetic case study of a thresholded linear reservoir model: (a) the thresholded linear reservoir model and (b) simulated synthetic data, D , that includes rainfall and flow over $T = 50$ time steps.

reservoir with a finite threshold model is used to infer a linear reservoir model (that is without a threshold). The linear reservoir model structure without a threshold is therefore deficient in representing the data. Such structure deficiency is nontrivial and nonideal especially because the error (residual) structure is complex. As is shown later in this section, even a complex likelihood function is unable to replicate it. No noise is added to the data for two reasons: (1) the added noise can only represent measurement errors since the structure error has already been represented and the quantile model selection does not distinguish between structure and measurement errors (though it can be done based on a priori specification of measurement errors based on benchmarking studies (such as McMillan *et al.* [2012]) and (2) the induced error structure is complex enough, additional complexity by adding noise to it is a relative distraction.

[27] We first consider a Bayesian approach by using three likelihood functions: Gaussian, Laplace, and the Generalized Likelihood (GL) function of Schoups and Vrugt [2010]. The Kernel density Independence Sampling based Monte Carlo Scheme (KISMCS) is used for sampling parameters from the likelihood functions. Further details of KISMCS are provided in Appendix A. We note that parameter, θ , sampling based on any likelihood function $L(\theta, D)$ is a sampling from its corresponding posterior $P(\theta|D)$ only when the likelihood function is fully specified [Davidson and MacKinnon, 2004, p. 399]. Here full specification

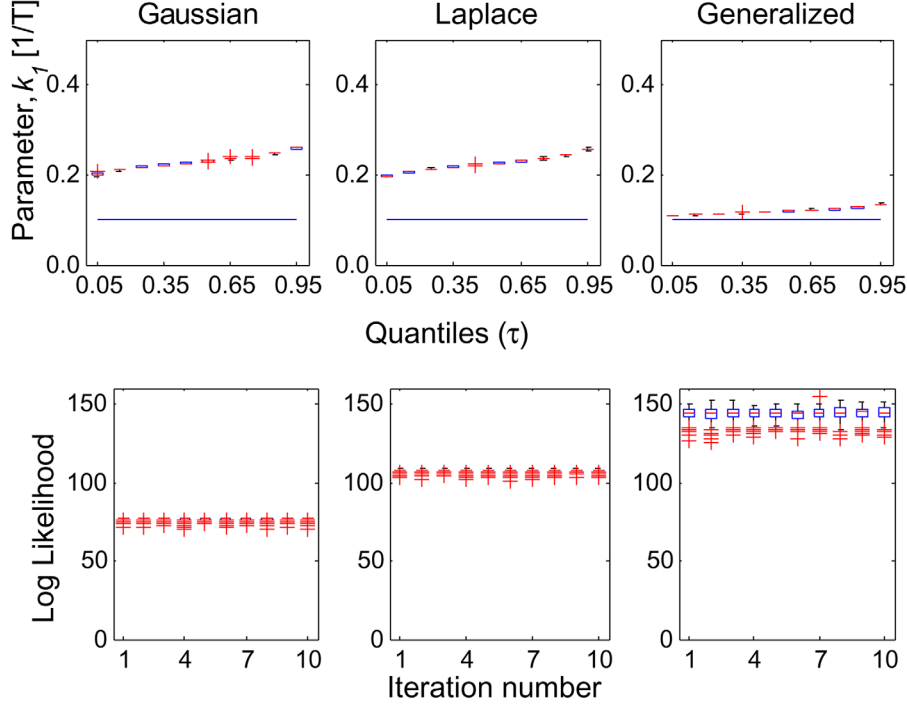


Figure 3. Sampled distribution of a single reservoir model’s parameter, k_1 [1/T], using three likelihood function (Gaussian, Laplace, and Generalized) on data D. 10 quantiles (0.05:0.10:0.95) of the parameter distribution over 10 MCMC simulations are shown in the first row. The blue line in each of the three subplots represents the true parameter value ($= 0.1$). The second row show the log-likelihood values of the sampled parameter points once KISMCS has converged for each of its 10 simulations.

means that even if the model (as represented by a parameter set θ) belongs to a deficient model structure, the likelihood function specifies the missing information (the deficiency) correctly. It is widely accepted that estimation of a model based on an incorrectly specified (or misspecified) likelihood function leads to results that are often meaningless or misleading [see, e.g., White, 1982; Berk, 1966; Fisher, 1922].

[28] The description of error, made by a single parameter linear reservoir without a threshold model in representing a thresholded three parameter linear reservoir model (the underlying process), by Gaussian or Laplace distribution is not its full specification. The general likelihood function of Schoups and Vrugt [2010] offers a better alternative but still it is not a full specification. It ignores the correlation between model prediction and residuals when the model structure is deficient. In practice though and in the synthetic study here, it apparently leads to biased parameter estimates (though smaller bias than when using the other two likelihood functions) of the single reservoir model. Thus one may conjecture that a model specification with GL will grow weaker when the bias effect due to model prediction-residuals correlation grows stronger relative to other structural deficiency effects.

[29] Figure 3 corroborates these statements. It shows the distribution over 10 runs of KISMCS of 10 quantiles of the sampled parameter of the linear reservoir model inferred from the synthetic data D. Note that D is generated by a linear reservoir with a threshold model with parameters $\{k, k_1, S_{\max}\} = \{0.1, 0.5, 3\}$. Since the linear reservoir

model with a threshold subsumes a linear reservoir model, the “true” recession parameter of the linear reservoir subsumed in D is $k = 0.1$.

[30] The second row of Figure 3 displays the log-likelihood values, solely to demonstrate consistent convergence of KISMCS. The Gaussian and Laplace likelihood function lead to parameter sampling that appear to be more biased than GL based, indicating better but not complete specification of structure deficiency by the GL function. The “true” value of the linear reservoir model is indicated by the blue line. Again we note that the “true” value of the linear reservoir without a threshold model when the underlying “truth” is a thresholded linear reservoir corresponds to the slow flow component since it is only the slow flow component of the total observed flow that can possibly be identified by a linear model conceptualization.

[31] All three likelihood functions yield point (and true) estimates of the parameters when the thresholded linear reservoir model structure (the “true” structure) is used instead of a linear reservoir model structure (without a threshold) (not shown here).

[32] Finally Figure 4 shows the performance of quantile model selection using the single linear reservoir without a threshold model structure on the data set D. We note that its implementation does not require full specification since one is interested in relating the loss function to structural deficiencies. It is sufficient that the loss function measures some distance to reality. Figure 4a shows that 0.05–0.95 quantile prediction coverage does not cover all the observations as a direct consequence of structural deficiencies.

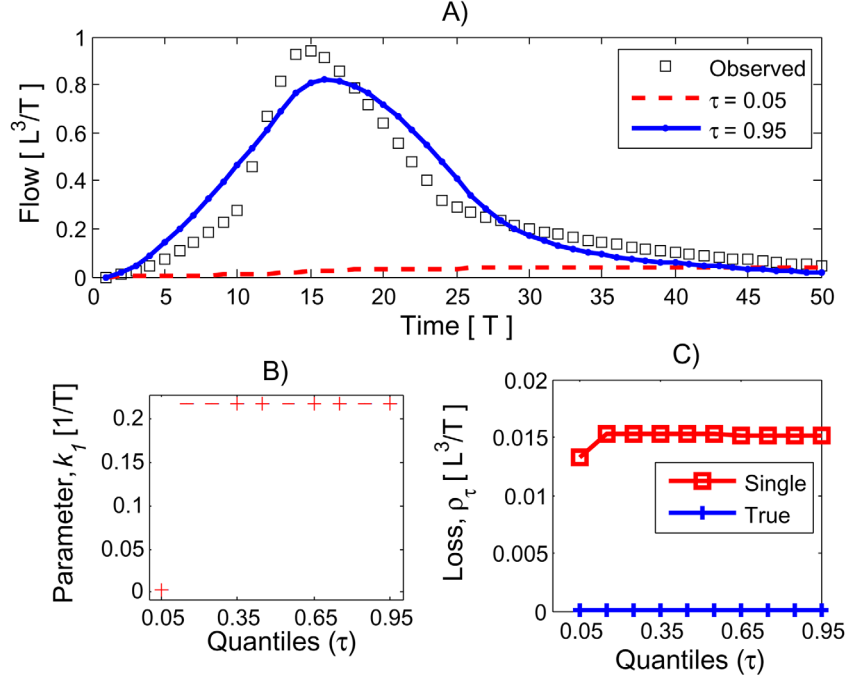


Figure 4. Performance of quantile model selection using the linear reservoir structure on data D over 10 simulations of SCE-UA: (a) the 5th and the 95th quantile prediction that cross around the time index of 44 for one of the simulations, (b) the estimated parameter values for 10 quantiles (0.05:0.10:0.95), and (c) the asymmetric loss function values for the two model structures (“truth” = the thresholded linear reservoir model, “linear” = linear reservoir model) over the quantiles.

[33] Further, the fifth-percentile prediction crosses the 95th percentile prediction. The observations contain information on the thresholding behavior of the flow that the model structure (of a linear reservoir without a threshold) is unable to replicate. This restricts the predictions of its best performing models at certain quantiles. The degree to which the structure binds (restricts) the prediction of its best performing model at a quantile captures the essence of model structure deficiency. These restrictions appear to be different at low flows and high flows, implying that structural deficiency is different at different flow quantiles. The low flows and high flows are inaccurately predicted, due to uneven deficiencies over quantiles, to the extent that low flow predictions cross the high flow predictions. The crossing of quantile predictions thus diagnoses structural deficiency.

[34] It needs to be emphasized here that structural deficiency need not always lead to the crossing of quantiles. The connection between the crossing of quantile predictions and structural deficiency is further analyzed in the following subsections and formalized in section 3.

[35] Figures 4b and 4c show the distributions of quantiles over 10 simulations and the asymmetric loss function values for the “truth” (i.e., when a linear reservoir model with a threshold model structure is used for inference) and the linear reservoir without a threshold model structure. Figure 4c points to the possibility that as one reduces structural deficiency at each quantile, through structural improvement, the loss function moves towards zero at each quantile. The parameter distribution in Figure 4b is also interpretable, being the parameter values corresponding to the respective quantile predictors.

[36] Quantile model selection provides true estimates of the parameters when the true model structure is used (not shown here) and the estimates are constant across quantiles.

2.3.2. Inference Using Deficient Model Structures on Data Synthetically Generated by a Complex Rainfall-Runoff Model

[37] The French Broad river basin data climatic forcing data of 5 years (1970–1975) is used to generate a synthetic streamflow time series. A complex conceptual rainfall runoff model with multiple states and parameters is used to generate the synthetic streamflow (see Appendix C for the description of model structure 3 used for streamflow generation, its structure and the parameter values used). In order to focus solely on the impact of nonlinear mapping of effective precipitation to streamflow, the evaporation scheme of the complex model is suppressed. It is done by forcing the evaporation flux of the model to be equal to the minimum of observed precipitation and potential evaporation. Further, two sets of synthetic streamflow data sets are generated: one without any additive noise and one in which 10% heteroscedastic Gaussian noise that is added to the generated streamflow.

[38] The linear reservoir without a threshold model structure (structure 1 in Appendix C) and the linear reservoir with a threshold model structure (structure 2 in Appendix C) are then used to infer quantile models (note that a quantile model is a model inferred using a model structure at that quantile; thus multiple quantile models are obtained for each model structure). Nine quantiles values ranging from 0.1 to 0.9 are considered.

[39] Figure 5 compares the quantile model predictions of the two model structures. Both the models structures have

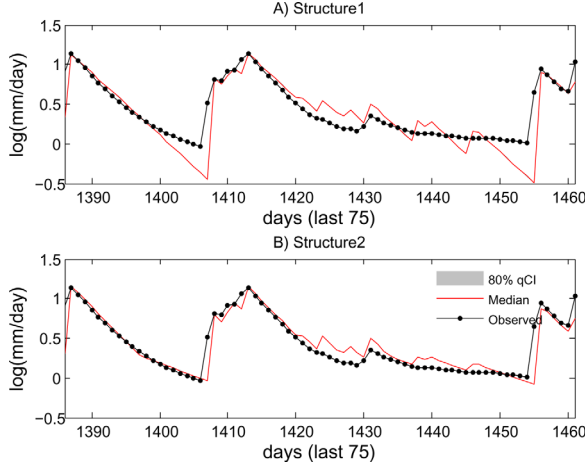


Figure 5. Synthetic case study. The time series performances of quantile models selected from (a) model structure 1 (linear reservoir without a threshold model structure) and (b) model structure 2 (linear reservoir with a threshold model structure) are shown. Both the structures have evaporation defined as $\min(\text{precipitation}, \text{potential evaporation})$. The y axis is in log scale. 80% quantile confidence interval, median model predictions, and observations for the last 75 days of 1970–1975 calibration period are shown. The 80% qCI for the two model structures are tight and overlap with the median prediction.

tight quantile confidence interval (the confidence quantile interval is 80% since the quantiles range from 0.1 to 0.9). The incapacity of the model structure 1 to replicate the observed low flows is evident. Meanwhile, both the model structures slightly overpredict medium flows (falling and rising parts of the time series). An underprediction (overprediction) of an observed quantile is equivalent to predicting the quantile with a positive (negative) bias. It is necessary to emphasize that this bias is not the bias that is estimated by averaging the errors in predicting a time series over its length. It is a quantile specific bias that represents the error that a model makes in replicating the observed quantile of a variable of interest. See, for example, equation (3.3a) for its formal specification. Figure 6 shows an approximate estimation of the bias in predicting various quantiles (formal specification of the bias is provided in section 3 and the details of how the approximation is estimated are provided in Appendix D). It shows that the main difference between the model structures is the positive bias of model structure 1 at low quantiles. The result suggests that the absence of a threshold affects the performance of a model at low flows. This is also observed in Figure 5a when contrasted with the performance of model structure 2 (a linear reservoir with a threshold model structure) in Figure 5b. Both the model structures appear to have biases of same (negative) sign in predicting medium quantiles. This is possibly due to the absence of multiple reservoirs. We note here that the estimated bias is an approximation (a first-order one, see Appendix D). The consideration of now ignored higher order terms may further differentiate the two model structures in terms of their structural deficiencies. Nonetheless, Figure 6 clearly demonstrates that the

structure 1 is more deficient than the structure 2 due to the absence of a threshold though both the model structures are deficient.

[40] The asymmetric loss function contains complete information about model structure deficiencies at various quantiles (this is formally shown in section 3). Figure 7 shows the asymmetric loss functions for the two structures. It shows that structure 1 is more deficient than structure 2 at all considered quantiles. This additional evidence supports the argument that structure 1 is more deficient than structure 2 at least in predicting low flows. Point statistics such as coverage probability (percentage of observations that lie within the 80% qCI), Nash-Sutcliffe, standard bias (the mean error in predicting a time series) and mean absolute error are also considered (shown in the figure). They are calculated on prediction models that are obtained from the respective model structures by minimizing mean absolute error (a standard performance metric). These statistics also support the argument that model structure 1 is more deficient than model structure 2.

[41] The question whether structural deficiency leads to quantile prediction crossing is now considered. It is not counterintuitive that quantile predictions may cross over time when different predictions have different biases. A deficient model structure though need not necessarily lead to its quantile predictions to cross over time. However, if quantile predictions do cross then the model structure is surely deficient (this is formally shown to hold in section 3). Do quantile predictions cross in either of the structural deficiency cases considered in this section?

[42] In order to demonstrate whether the quantile predictions cross, the number of quantile prediction crossing is calculated based on the ordering of quantiles predictions at each (daily) time step (over 6 years of data). If quantile predictions do not cross then the predictions have the same

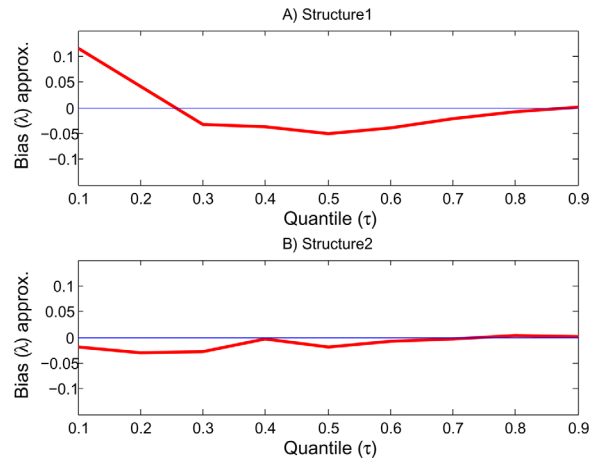


Figure 6. Synthetic case study. The first-order bias (λ) approximations of the model structures defined in Figure 5 across a range of quantile values between 0 and 1 are shown. Structure 1 has positive bias at low quantiles (low flows). Both the model structures have negative bias at medium quantiles (medium flows). Model structure 2 has negative bias in replicating low quantiles (low flows) as well. Structure 1 appears to be more deficient than structure 2 at nearly all the quantiles.

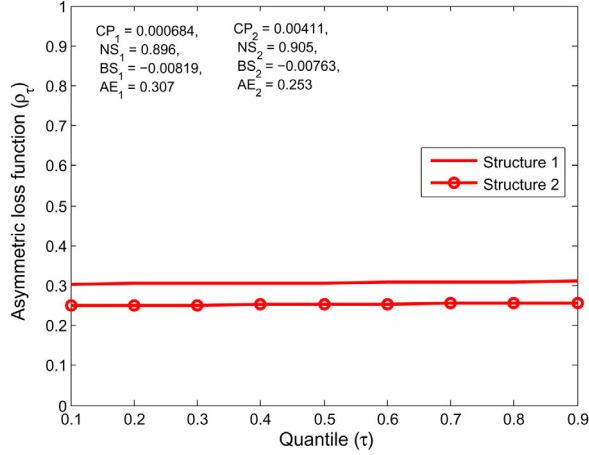


Figure 7. Synthetic case study. Asymmetric loss function (ρ_τ) for model structures defined in Figure 5 at various quantile values τ . The loss function values are quantile-wise closer to 0 for structure 2 than structure 1. Also shown are the traditional statistics of median predictors from the two model structures. The subscripts correspond to a model structure; CP = fraction of observations covered by the 80% qCI, NS = Nash-Sutcliffe, BS = standard bias, AE = $0.5 \times$ mean absolute error.

order as the set of quantiles values $\{10\%, 20\%, 30\%, \dots, 90\%\}$. That is, at a considered time step the 10% quantile prediction is below the 20% quantile prediction, the 20% quantile prediction is below the 30% quantile prediction and so on. If quantile predictions cross then the ordering of the predictions is one possible permutation of the set of quantile values. Thus if quantile predictions do not cross at a given time step then the number of quantile prediction crossings is 0. The maximum number of quantile prediction crossings at a particular time step is the number of permutations of the set of quantile values. A kernel density estimate of the number of quantile crossings with flow magnitudes at corresponding time steps is then created to demonstrate the “density” of quantile crossing at different flow levels. Since model structure deficiency is a necessary but not a sufficient condition for quantile prediction crossing, the latter can only in certain circumstances be used to assess the differences in structural deficiencies of two model structures. The “density” of quantile prediction crossing of two structures is however expected to be different if the difference in respective deficiencies is large.

[43] Figure 8 plots these densities for the two model structures. The quantile prediction crossings are dense at low flows for both the structures.

[44] Finally, three Bayesian model selection criteria are estimated for the two model structures. These criteria are BIC, harmonic mean of the log-likelihood values of parameter sets sampled from the posterior distribution [Kass and Raftery, 1995] and the marginal likelihood approximation of Chib and Jeliazkov [2001] used in Marshall et al. [2006]. Details of Bayesian model structure selection criteria are described in Appendix E. KISMCS is used to sample points from the posterior parameter distributions of the two model structures. The GL function is used (as discussed in section 2.3.1).

[45] Table 1 compares the quantiles of the obtained posterior distribution using KISMCS as well as the Bayesian model selection criteria for the two model structures. It also provides the parameter values and the asymmetric loss function of quantile prediction models (that use quantile model selection to infer the prediction models). We here re-emphasize that the quantiles of the posterior distribution are not pointwise comparable with the parameters of the quantile prediction models. The ranges of the obtained parameters can, however, be compared. The “true” parameter values are $K_s = 25$ days and $K_f = 4$ days. One expects to retrieve these values or atleast one of them if the problem of model selection is well specified. For example, consider the case of Bayesian inference of a model using a linear reservoir model structure without a threshold. If the likelihood function specifies all the processes that the linear reservoir model structure does not consider (such as the overland flow, routing of the total flow and percolation from the unsaturated zone), then the parameter value corresponding to the slow flow ($K_s = 25$ days) can be retrieved. But if the likelihood function does not completely specify all the missing processes, biased valued may be obtained.

[46] Table 1 demonstrates that biased parameter estimates are achieved through Bayesian inference. In the linear reservoir without a threshold model structure, the estimated parameter distribution lies between 25 and 4 days (true values of the parameters corresponding to the fast and slow responses). Such biased performance, though to a lesser extent, persists in the case of linear reservoir with a threshold model structure. The parameter ranges in this case appear to be closer to the true values. The three Bayesian criteria are close to each other for each of the considered model structures and support the preference of the linear reservoir with a threshold model structure over the linear reservoir model structure (without a threshold).

[47] The inferred parameter distributions shift to lower magnitudes under 10% heteroscedastic errors. The

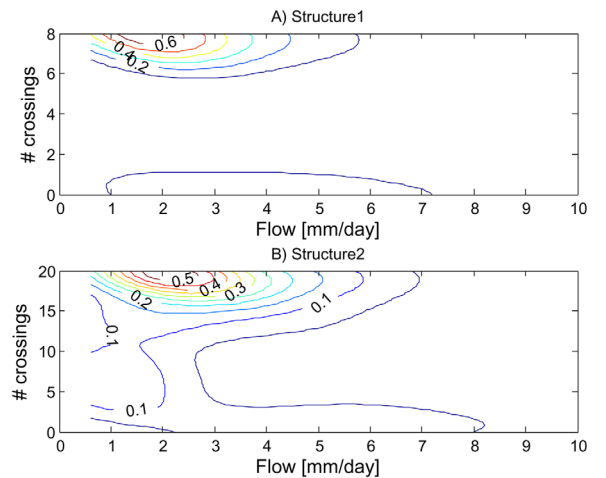


Figure 8. Synthetic case study. Kernel density estimates of the number of crossing of quantile predictions for model structures as defined in Figure 5. It displays the density of quantile crossing with the magnitude of flows. The quantile crossings are dense at low flows for both the models, indicating that both the model structures are deficient.

Table 1. A Comparison of Bayesian Inference Using a GL Function With Quantile Model Selection for Two Model Structures (1 = Linear Reservoir Without a Threshold Model Structure; 2 = Linear Reservoir With a Threshold Model Structure) on 6 Years of Synthetically Generated Daily Streamflow^a

	No Noise			10% Heteroskedastic Gaussian Noise		
	10th Percentile	50th Percentile	90th Percentile	10th Percentile	50th Percentile	90th Percentile
<i>Bayesian Inference With a General Likelihood (GL) Measure</i>						
Model structure 1						
K_s	15.53	16.40	17.30	10.27	10.79	11.27
BIC = -684.9 HM1 = -667.1 HM2 = -725.8				BIC = -1431.1 HM1 = -1412.0 HM2 = -1465.9		
Model structure 2						
K_s	25.36	28.97	33.49	16.66	20.17	24.13
K_f	9.40	9.70	9.90	6.05	6.45	6.84
BIC = -500.1 HM1 = -479.4 HM2 = -541.1				BIC = -1278.1 HM1 = -1298.5 HM2 = -1330.2		
	No Noise			10% Heteroskedastic Noise		
	10th Percentile	50th Percentile	90th Percentile	10th Percentile	50th Percentile	90th Percentile
<i>Quantile Model Selection</i>						
Model structure 1						
ρ_r	0.20	0.20	0.20	0.23	0.23	0.24
K_s	6.04	6.03	6.01	6.03	6.02	6.02
Model structure 2						
K_s	14.14	14.23	14.28	14.83	14.86	14.88
K_f	5.33	5.33	5.33	5.47	5.46	5.46
ρ_r	0.17	0.17	0.17	0.20	0.20	0.20

^aThe data are generated using a complex rainfall-runoff model. Two cases of no noise and 10% heteroscedastic Gaussian noise are considered. See section 2.3.2 for further description.

parameter distribution for the linear reservoir without a threshold model structure remains between 4 and 25 days. Meanwhile, the inferred parameter ranges for the linear reservoir with a threshold model structure appear to have moved closer to the true parameter values. The Bayesian model selection criteria again support the preference for the linear reservoir with a threshold model structure over the linear reservoir without a threshold model structure. The criteria values under heteroscedastic noise case are lower for the two model structures than the corresponding criteria values under no noise case. This is attributable to higher noise to signal ratio in the former case than in the latter case.

[48] Quantile model selection infers recession parameters in a relatively robust manner (across the noise levels) for the two model structures. This is because the 80% quantile ranges, i.e., the difference between the 10th and the 90th percentile, of the parameters are small. The estimated parameters of the linear reservoir without a threshold model structure corresponding to the 10%, 50%, and the 90% quantile models are closer to 4 days (true parameter value corresponding to fast flow). Meanwhile, in the case of linear reservoir with a threshold model structure, the quantile parameter estimates corresponding to fast (overland) flow are closer to 4 days and the quantile parameter estimates corresponding to slow flow are closer to 25 days. Nonetheless, the quantile parameter estimates for both the model structures and noise levels remain biased as a result of inherent structural deficiencies. The asymmetric loss functions for both zero noise and 10% heteroscedastic noise cases support the preference of the linear reservoir with a threshold model structure over the linear reservoir without a threshold model structure. For a given model structure, the asymmetric loss function increases in magnitude with noise level for each quantile value. This is possibly due to increasing noise to signal ratio.

[49] Quantile model selection did not assume any noise model when inferring parameter distributions under the 10% heteroscedastic noise case. Yet quantile model selection is robust in inferring the parameters of the two model structures. However, a general description of noise in quantile model selection may isolate hydrological model structural deficiencies better than quantile model selection without a noise model.

2.3.3. Inference Using Model Structures With Increasing Complexity on French Broad Basin Data

[50] The two model structures that are used for inference and the model structure that is used to generate the synthetic data in the previous section are 3 of the 4 model structures used in this section. The fourth model structure is the flexible model structure described in Appendix C that models evaporation as a nonlinear function of unsaturated zone storage. Thus, the structure complexity varies gradually in terms of its nonlinearity. The simplest structure is the linear reservoir without a threshold model structure (model structure 1) with evaporation defined as the minimum of daily precipitation and potential evaporation. A fundamental nonlinearity is introduced by considering a linear reservoir with a threshold model structure (model structure 2). It also has evaporation defined as minimum of precipitation and potential evaporation. The third model structure (model structure 3) has multiple reservoirs, smooth (thresholded) transformation of precipitation to overland flow (with a thresholded response as a particular case) and defines evaporation as the minimum of precipitation and potential evaporation. It is more complex than the first and second model structures. The fourth model structure (model structure 4) is most complex of all the structures. Note that the 4 model structures are nested in the sense that model structure 1 is a special case of model structure 2, model structure 2 is a special case of model

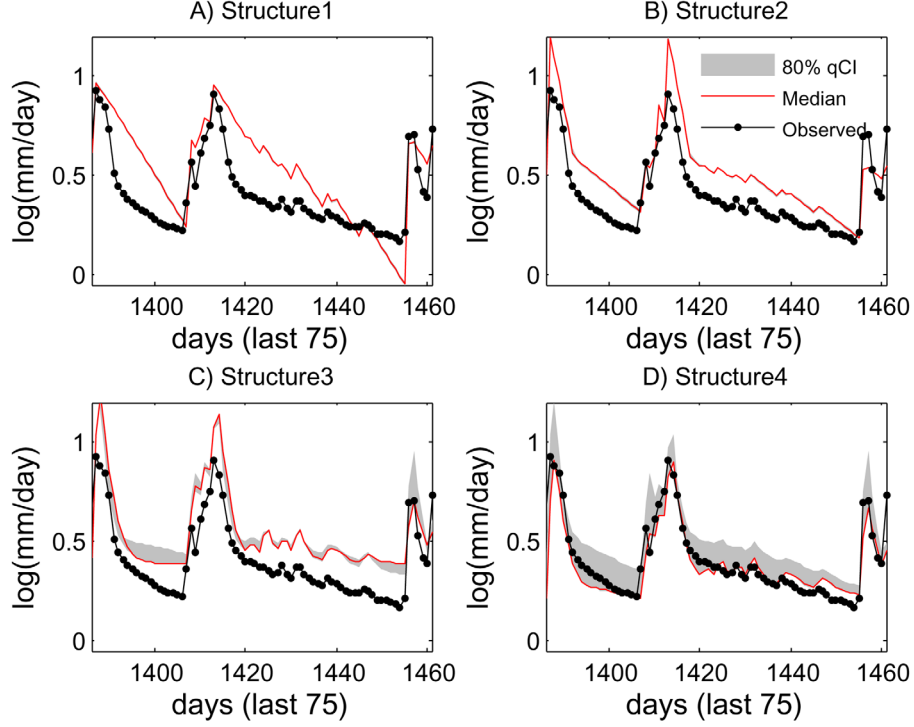


Figure 9. French Broad river basin case study. The time series performances of quantile models selected from (a) model structure 1 (linear reservoir without a threshold model structure), (b) model structure 2 (linear reservoir with a threshold model structure), (c) model structure 3 (a complex model structure with multiple states but with $\text{evaporation} = \min(\text{precipitation}, \text{potential evaporation})$), and (d) model structure 4 (same as model structure 3 except that the evaporation is a nonlinear function of a storage that represents the unsaturated zone) are shown. The 80%qCI gradually increases in width with increasing complexity of model structures from 1 to 4. The y axis is in log scale. 80% quantile confidence interval, median model predictions and observations for the last 75 days of 1970–1975 calibration period are shown.

structure 3 and model structure 3 is a special case of model structure 4. The daily streamflow, precipitation and potential evaporation data of French Broad River basin from 1970 to 1975 are used. The KISMCS sampler with General Likelihood measure (used in section 2.3.1) and 3 Bayesian model selection criteria used in the previous sections are used for Bayesian inference. Point statistics such as coverage probability (percentage of observations that lie within the 80% qCI), Nash-Sutcliffe, standard bias (the mean error in predicting a time series) and mean absolute error are also considered. They are calculated on prediction models that are obtained from the respective model structures by minimizing mean absolute error (a standard performance metric).

[51] Figure 9 plots the performance of quantile model selection for the four model structures over the last 75 days of 1970–1975 period. The deficiencies of model structures to explain the observations reduce with increasing complexity. Note that quantile predictions of model structures 1 and 2 are tight as indicated by its 80% quantile confidence interval (80% qCI). However, structure 1 is most deficient in explaining medium to low flows. Model structure 3 has slightly looser 80%qCI but appears to have similar structural deficiency as model structure 2. Finally, model structure 4 is the least deficient structure with widest 80% qCI.

[52] Figure 10 plots the first order approximation of bias in quantile prediction (formal specification of the bias is provided in section 3 and the details of how the approximation is estimated are provided in Appendix D). The first three model structures show bias in predicting low to medium quantiles while structure 4 has relatively minor bias in predicting any of the quantiles. Overall, model structure deficiency reduces as one move from structure 1 to structure 4 for nearly all quantiles. Further, model structures 1 and 2 appear to have lower bias than structure 3 in predicting lowest 2 quantiles. It may be due to the approximate nature of bias estimation (higher-order terms may reveal additional differences).

[53] Figure 11 plots the asymmetric loss functions for the four model structures. The asymmetric loss function at a given quantile contains full information about structural deficiency in predicting the quantile. This is formally shown in section 3 (see Proposition 4). The figure demonstrates that model structure 1 is the most deficient structure while model structure 4 is the least deficient model structure. Further, model structure 2 is less deficient than model structure 1 for all the quantiles. The model structure 3 is less deficient than model structure 2 at lower quantiles while they are indistinguishable in their deficiencies at higher quantiles. The lower deficiency of model structure 3

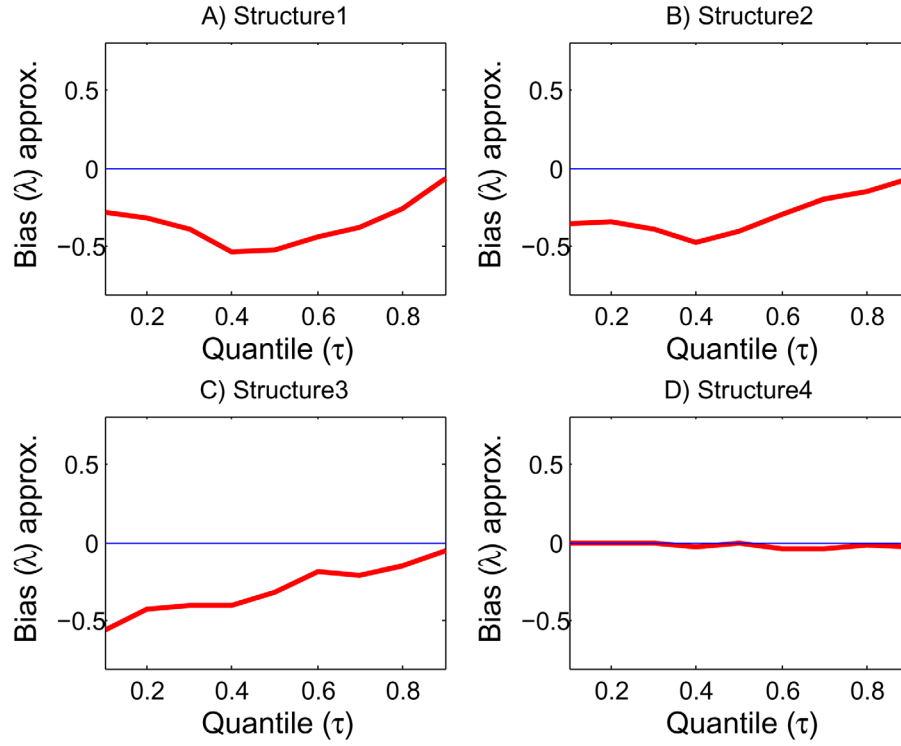


Figure 10. French Broad river basin case study. The first-order bias (λ) approximations of model structures as defined in Figure 9 across a range of quantile values between 0 and 1 are shown. The bias gradually decreases with increasing complexity especially at low to medium quantiles.

can be attributed to a distributed representation of rainfall-overland flow thresholding behavior as well as richer conceptualizations of percolation, slow flow and flow routing.

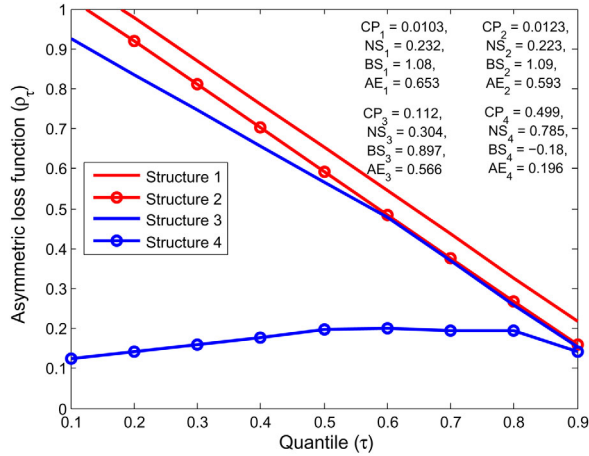


Figure 11. French Broad river basin case study. Asymmetric loss function (ρ_τ) for model structures defined in Figure 9 at various quantile values τ . The loss function values are quantile-wise closest to 0 for structure 4 and farthest for structure 1. The asymmetric loss function thus orders structures 1, 2, 3, and 4 as decreasing in structural deficiency. Also shown are the traditional statistics of median predictors from the 4 model structures. The subscripts correspond to a model structure; CP = fraction of observations covered by the 80% qCI, NS = Nash-Sutcliffe, BS = standard bias, AE = 0.5 * mean absolute error.

The model structure deficiencies of structure 4 are significantly lower than the other three structures for nearly all the quantiles. This can be attributed to the evaporation scheme that the first three model structures suppress.

[54] The three Bayesian criteria (Table 2) further support the argument that deficiency decreases from structure 1 to structure 4. The point statistics, i.e., coverage probability (percentage of observations that lie within the 80% qCI), Nash-Sutcliffe, standard bias (the mean error in predicting a time series) and mean absolute error (Figure 11), also suggest the same. It is worth mentioning that Nash-Sutcliffe and standard bias statistics suggest that structure 1 is marginally less deficient than structure 2 while the other two statistics suggest the opposite. The piecewise linear recession limb of the observed streamflow (in log-scale) around the day 1400 supports the latter (Figure 9), that a linear reservoir response is indeed not an adequate description.

[55] Figure 12 plots the density of the number of quantile crossings with the magnitudes of the observed flows. The density plots of quantile predictions reinforce the argument that the model structures are deficient at low flow quantiles. The quantile predictions cross at low flows for all the 4

Table 2. Bayesian Statistics for the Four Model Structures on French Broad River Basin Case Study

	Structure 1	Structure 2	Structure 3	Structure 4
BIC	-497.80	-369.73	23.34	196.55
HM1	-479.39	-349.37	36.33	214.47
HM2	-537.00	-426.00	-32.71	167.55

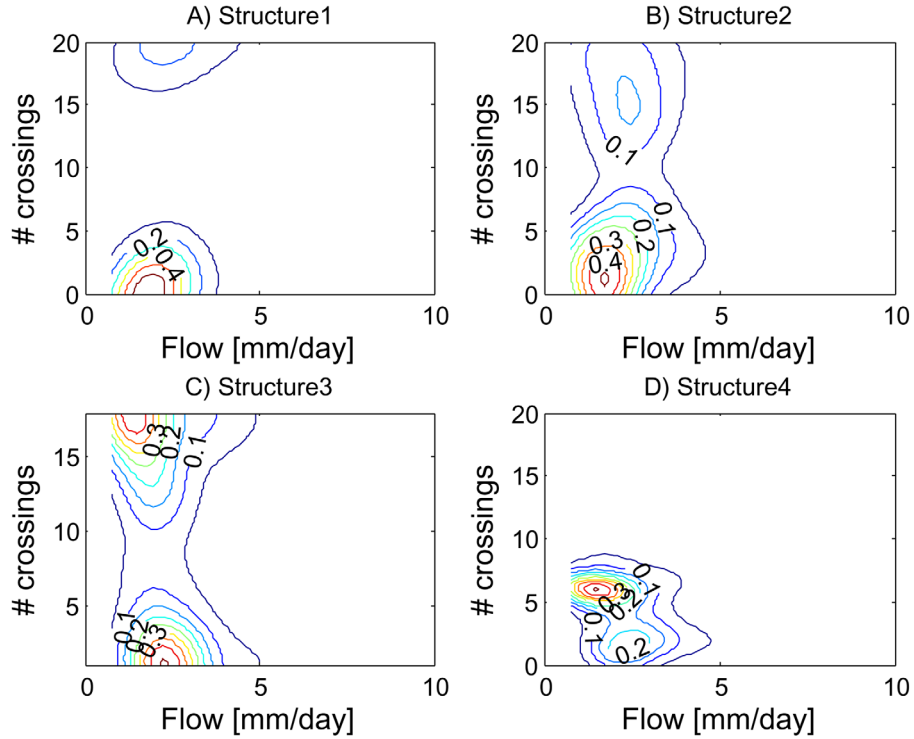


Figure 12. French Broad river basin case study. Kernel density estimates of the number of crossing of quantile predictions for model structures as defined in Figure 9. It displays the density of quantile crossing with the magnitude of flows. The quantile crossings are dense at low flows for all the models, though structure 4 appears to have lower quantile crossings at low flow.

model structures though the number of quantile predictions are low for structure 4.

2.3.4. Synthesis of Observations and a Conceptual Description of Deficiency Assessment

[56] The comparative assessment of model structures based on first order bias estimation, asymmetric loss function and density of quantile crossing correctly reveals the gradient of structure deficiencies over both synthetic and real data sets. The quantile-specific deficiency assessment provides local (in distribution) information on deficiencies than a traditional statistics based method such as Nash-Sutcliffe. The quantiles may be associated with the probability with which different processes combine. It is in this sense that quantile model selection based deficiency assessment may locally assess structure deficiencies. It also distinguishes between precision and accuracy of a model structure in replicating observations of a variable of interest. A simple model structure, such as a linear reservoir model structure, has tight 80% quantile confidence interval. This leads to poor coverage probability of predictions. However, its structural deficiencies (bias in predicting quantiles) depend on the underlying processes as well.

[57] The assessment of model structure deficiencies based on quantile model selection is corroborated both by traditional point statistics such as Nash-Sutcliffe, mean absolute error and standard bias (mean of time series error of a median prediction from the observed) as well as Bayesian model selection criteria. The traditional statistics are calculated on a model selected by minimizing mean absolute error: a traditional performance metric. The

Bayesian selection criteria are calculated on the posterior parameter distribution using KISMCS (a Markov Chain Monte Carlo sampler) and a General Likelihood measure (used in section 2.3.1). It also provides quantile specific information on structure deficiency and does not assume full specification, unlike traditional point statistics or Bayesian based inference methods. Quantile specific model parameters are also interpretable as those that correspond to a model that attempts to predict the observed quantile. Therefore, the parameter ranges obtained from Bayesian inference and quantile model selection are not always identical.

[58] The model structures considered in previous sections that increase in complexity are nested. Hence a more complex structure can never be more deficient than a less complex (but a nested) model structure since the latter is a particular case of the former. A more complex model structure is as deficient as the nested but less complex model structure in the worst case. In better cases, it is less deficient due to the nonoverlapping structure that is missing from the less complex structure that enables it to approximate the underlying processes better.

[59] The case studies infer a least deficient model structure at a particular quantile from a set of candidate structures based on the ordering of the asymmetric loss functions of candidate structures at the quantile. One may argue that large complexity of model structures may lead to incorrect inferences. However, complexity can only misguide structure selection on small sample size; a large sample size (as considered in the presented case studies) is sufficient for accurately identifying structure deficiencies.

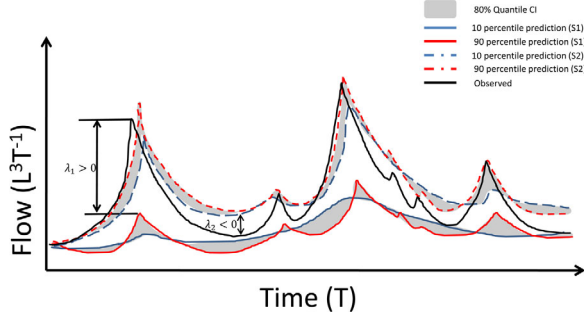


Figure 13. An illustration of bias (λ) and quantile crossing for two model structures S1 and S2. Structure S1 is deficient and underpredicts high flows (high quantiles), thus $\lambda_1 > 0$ at high quantiles. Structure S2 is deficient and overpredicts low flows, thus $\lambda_2 < 0$ at low quantiles. A deficient model structure may be such that bias has opposite signs at high and low flows. The quantile predictions for a given model structure may also end up crossing due to its deficiencies.

This is because the effect of structure complexity on the variation of its finite sample performance vanishes when the sample size is sufficiently large (based on a law of large numbers type argument).

[60] Figure 13 illustrates two extremes of model structure deficiency in streamflow prediction. A model structure is unable to appropriately model low flows on one end while the other model structure is unable to model high flows on the other end. This deficiency of a model structure to represent either low or high flows is also partly reflected in its narrow qCI (quantile confidence interval).

[61] It also shows that two different quantile predictions cross for a given model structure as a result. Figure 14 conceptualizes the reason behind it for a simple case when the length of a time series of a variable of prediction interest is more than 2. Only two dimensions of the prediction (output) space are shown. It shows the joint cumulative distribution $F(\mathbf{y})$ of observing a prediction variable of interest \mathbf{y} , where \mathbf{y} is a 2-dimensional vector, $\mathbf{y} = \{y(1), y(2)\}$. Thus for two given quantiles $\tau = 0.2$ and $\tau = 0.3$, (probable) observations of \mathbf{y} as a function of input forcing vector \mathbf{x} can be traced. These traces are represented by the dashed lines on the cumulative distribution surface. Two red circles on these two traces represent probable observation of \mathbf{y} for a given value of input forcing (say \mathbf{x}_0). Different observations for the same value of input forcing (\mathbf{x}_0) are allowed to represent the stochastic influence of unobserved biotic or abiotic variables on the variable of prediction interest. Let the model structure that is used to infer quantile models of the underlying process be more deficient in explaining $\tau = 0.2$ than $\tau = 0.3$. This results in biases $\lambda^{\tau=0.2}$ and $\lambda^{\tau=0.3}$ such that $\lambda^{\tau=0.2} > \lambda^{\tau=0.3}$. The inferred quantile models then “at best” model quantiles $0.2 + \lambda^{\tau=0.2}$ and $0.3 + \lambda^{\tau=0.3}$ instead of 0.2 and 0.3. The order of (probable) predictions at quantiles $\tau = 0.2$ and $\tau = 0.3$ thus ends up being the reverse of the order of quantile values $\{0.2, 0.3\}$. This is shown by the projections of $0.2 + \lambda^{\tau=0.2}$ and $0.3 + \lambda^{\tau=0.3}$ quantiles on the $y(1)$ – $y(2)$ plane. This reverse ordering due to the bias that is introduced by model structure deficiencies leads quantile predictions of a given model

structure to cross. A formal treatment of the quantile predictions crossing is provided in section 3 that proves that it is a consequence of model structure deficiency.

[62] The asymmetric loss function value at a particular quantile contains full information of bias due to structural deficiency at the quantile. The loss function is minimized when inferring a quantile model from a given model structure. Thus, the loss function value at a given quantile is higher for a model structure that is constraining the minimization of the loss function more than another model structure. Since the constraints posed by a model structure is due to its structural deficiencies, the loss function values for a set of candidate structures at a given quantile orders the structures in terms of its deficiencies. This has been demonstrated in previous sections and it is further formalized in section 3.

3. A Formal Analysis of Quantile Model Selection

[63] We now present a formal analysis of QE2 that has been defined in section 2.2. The formal analysis defines the problem of quantile model selection problem as a constrained minimization problem, where in the constraints are posed by the model structure. Model structure constraints are therefore also formalized. The analysis then proceeds

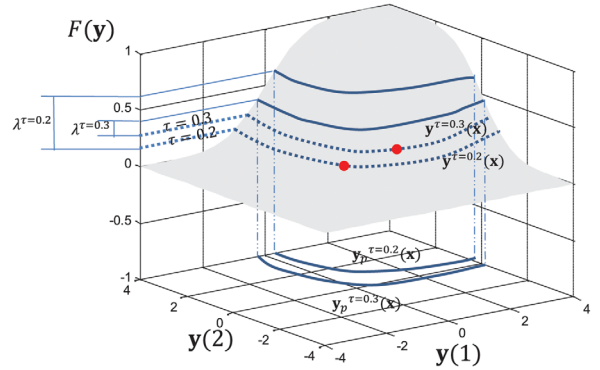


Figure 14. An illustration of quantile prediction crossing for a deficient model structure. The \mathbf{y} plane represents a space of predictions at two consecutive time steps. A joint cumulative distribution $F(\mathbf{y})$ of observing a prediction variable of interest \mathbf{y} , where \mathbf{y} is a 2 dimensional vector, $\mathbf{y} = \{y(1), y(2)\}$. Thus for two given quantiles $\tau = 0.2$ and $\tau = 0.3$, (probable) observations of \mathbf{y} as a function of input forcing vector \mathbf{x} can be traced (dashed lines). Two red circles on these two traces represent probable observation of \mathbf{y} for a given value of input forcing (say \mathbf{x}_0). Different observations for the same value of input forcing (\mathbf{x}_0) are allowed. A deficient model structure infers quantile models at quantiles $\tau = 0.2$ and $\tau = 0.3$ with biases $\lambda^{\tau=0.2}$ and $\lambda^{\tau=0.3}$ such that $\lambda^{\tau=0.2} > \lambda^{\tau=0.3}$. The inferred quantile models then “at best” model quantiles $0.2 + \lambda^{\tau=0.2}$ and $0.3 + \lambda^{\tau=0.3}$ instead of 0.2 and 0.3. The order of (probable) predictions at quantiles $\tau = 0.2$ and $\tau = 0.3$ thus ends up being reverse of the order of quantile values $\{0.2, 0.3\}$ as shown by the projections of $0.2 + \lambda^{\tau=0.2}$ and $0.3 + \lambda^{\tau=0.3}$ quantiles on the $y(1)$ – $y(2)$ plane. This reverse ordering implies a crossing of quantile predictions of a given model structure.

with a set of assumptions that in general hold for hydrological model estimation problems.

[64] First, it is shown that a minimum exists. Then the necessary conditions that a minimum should satisfy are shown based on a formal definition of model structure deficiency. The necessary conditions show how structure deficiencies affect optimal model selection. These conditions also expose how certain properties of hydrological model predictive equations can lead to an interplay between model structure deficiencies and predicting quantiles based on hydrological models.

[65] One fundamental property of quantile predictions is that that it should not cross, i.e., no two quantile predictions should cross each other at any time t . Thus the extension of more commonly used quantile regression on a class of linear functions to arbitrarily nonlinear hydrological models is nontrivial since nonlinearities may lead to quantile crossing. This essentially motivates a formal study of conditions under which quantile hydrological predictions do not cross and conditions under which the quantile predictions may cross. The analysis reveals that model structure deficiencies play a crucial role. Since model structure deficiencies constrain quantile model estimation and prediction, structure deficiencies can also be implicitly inferred from the performance of model structures across different structures.

[66] The formal analysis imparts objectivity to the assessment of model structure deficiencies without subscribing to several simulations. It is also testable, since the statements on the interplay between model structure deficiency assessment, quantile prediction noncrossing and model structure deficiencies that are proposed in the following hold as long as the stated assumptions hold. These propositions are falsified and thus are not applicable in cases that contradict any of these assumptions (though in many cases, propositions may not be sensitive to some of the assumptions).

[67] The outline of the following analysis is as follows. Proposition 1 first states that quantile model predictions do not cross in the simplest case when model structure deficiency is absent. Its proof indicates that model structure deficiencies introduce an effect that makes estimated quantile predictions to cross. Building upon Proposition 1 it is further argued that model structure deficiencies introduce bias in hydrological modeling of observed quantiles. Proposition 2 states the necessary conditions observed by optimal model parameters in presence of model structure deficiencies. It is used to show that monotonicity of the model predictive equation in atleast one of the parameters is sufficient to ensure that the bias that is introduced due to model structure deficiency is independent of parameter dimensionality and that it is time invariant. This is done by first showing that when a model structure is deficient, it leads to a representation of deficiency in terms of Lagrange multipliers of the constrained minimization problem of quantile model estimation.

[68] These are powerful results based on a formal and rigorous treatment of structural deficiencies. This becomes evident in Proposition 3, which states that the asymmetric loss function values can order model in terms of their biases. This proposition is subsequently used in corollary 4 to propose that the optimal value of asymmetric loss function orders corresponding model structures in terms of its deficiencies.

[69] Note that the formal analysis contributes to several questions. It studies the effect of structure deficiency on quantile predictions by formalizing an otherwise abstract notion of structure deficiency, nontrivial extension of quantile regression to hydrological modeling, and comparative assessment of model structures in terms of its deficiencies through the ordering of optimal asymmetric loss function values.

[70] To begin with, the definitions of model structure as a constraint set and the objective function of quantile model selection problem are provided. It is followed by the assumptions made.

[71] **Definition 1:** A model structure is defined as a constraint set

$$C(\mathbf{x}) = \{ \{ \mathbf{S}, \mathbf{k} \} | h_t(\mathbf{S}, \mathbf{k}; \mathbf{x}) = 0, t = 1, \dots, N-1, \mathbf{S} \in \mathbb{R}_+^N, \mathbf{k} \in \mathbf{K}, \mathbf{x} \in \mathbb{N} \subset \mathbb{R}_+^M \},$$

where

$$h_t(\mathbf{S}, \mathbf{k}; \mathbf{x}) = S_{t+1} - S_t - x_t + f(S_t, \mathbf{k}),$$

$\mathbf{S} = \{S_t\}_{t=1, \dots, N}$ is a N dimensional vector of state variables (without loss of generality it is assumed that there is only one storage variable per time step), \mathbb{N} is a compact subset of M -dimensional positive real space and \mathbf{K} is the parameter set of the model structure. As is shown later, the results of the paper are insensitive to the dimensionality of S_t . Thus, we retain it as such with any loss of generality.

[72] **Definition 2:** The objective function of QE2 is (earlier referred to as ρ_τ),

$$q(\mathbf{S}, \mathbf{k}; \tau, \mathbf{x}) = \int \sum_{t=1}^N [(1-\tau)|y_t - f(S_t, \mathbf{k})| + \tau|y_t - f(S_t, \mathbf{k})|_+].$$

$$p(\mathbf{y}|\mathbf{x})d\mathbf{y}/N.$$

[73] Note the expectation operation in the definition above. It implies that the analysis is for large sample sizes. Some simplifying assumptions are made below in order to elucidate key points on the existence of a minimum, on the necessary conditions for a minimum, on model structure deficiency and on the asymmetric loss function as a measure of model structure deficiency.

[74] **Assumption 1:** The parameter set \mathbf{K} that defines the model structure for a given model equation is compact.

[75] **Assumption 2:** The model equation $f(S_t; \mathbf{k})$ is differentiable, is monotonic in at least one element of \mathbf{k} and increasing in S_t . Further $\frac{\partial f}{\partial k_j}$ is independent of $\frac{\partial f}{\partial k_{j'}}$ where $k_j, k_{j'}$ are two distinct elements of \mathbf{k} .

[76] **Assumption 3:** Input forcing vector is nonzero, i.e., $\mathbf{x} = \{x_1, \dots, x_t, \dots, x_N\} \neq \mathbf{0}$.

[77] **Assumption 4:** Initial model storage is sufficiently greater than 0, i.e., $S_0 \gg 0$.

[78] **Assumption 5:** The observed variable of interest is bounded, i.e., $\underline{y} \leq y \leq \bar{y}$.

[79] **Assumption 6:** The cumulative probability density $F(\mathbf{y}|\mathbf{x})$ is differentiable and

$$F(\underline{y}|\mathbf{y}^{-t}, \mathbf{x}) = 0, \forall \underline{y}^{-t} \leq \mathbf{y}^{-t} \leq \bar{\mathbf{y}}^{-t}$$

where $\mathbf{y}^{-t} = \{y_1, \dots, y_{t-1}, y_{t+1}, \dots, y_N\}$.

[80] **Assumption 7.** We avail of a global optimizer that can identify a minimum of a quantile model selection problem.

[81] We first prove that the minimum of QE2 exists under the above stated assumptions and definitions. It is followed by a formal definition of model structure deficiency. Then the necessary conditions that a minimum of QE2 should obey when a model structure is not binding (deficient) are derived.

3.1. Existence of a Minimum of QE2

[82] The program QE2 is a form of a finite horizon optimal control problem [Lyon and Pande, 2006]. Existence of solution to QE2 can be proved based on a variant of Weierstrass theorem [Mordukhovich, 1976], which suggests that a minimum of a continuous function is attainable on a compact set.

[83] We note that the constraint set $C(\mathbf{x})$ defined as below is compact due to the compactness of \mathbb{N} , the continuity of $f(S_t, \mathbf{k})$ in S_t and \mathbf{k} and the compactness of \mathbf{K} (assumptions 1 and 2). Note that the assumptions on \mathbb{N} , $f(\cdot, \cdot)$ and \mathbf{K} ensure that S_t is bounded for any t leading to the compactness of $C(\mathbf{x})$. Further, let the objective function of QE2 be redefined by the substitution mentioned previously and consider its expectation with respect to the conditional probability of \mathbf{y} on \mathbf{x} ($p(\mathbf{y}|\mathbf{x})$) with a differentiable cumulative density function $F(\mathbf{y}|\mathbf{x})$ (Assumption 6), where $\mathbf{y} = \{y_t\}_{t=1, \dots, N}$:

$$q(\mathbf{S}, \mathbf{k}; \tau, \mathbf{x}) = \int \sum_{t=1}^N [(1-\tau)|y_t - f(S_t, \mathbf{k})|_- + \tau|y_t - f(S_t, \mathbf{k})|_+] p(\mathbf{y}|\mathbf{x}) d\mathbf{y}/N,$$

which is continuous in (\mathbf{S}, \mathbf{k}) .

[84] Thus, QE2 has a solution.

3.2. Necessary Conditions for a Minimum and Model Structure Deficiency

[85] Let $(\mathbf{S}^*, \mathbf{k}^*)$ be a global minimum (Assumption 7) given that such a minimum exists. This optimum will be nonzero for nonzero \mathbf{x} and for sufficiently large (Assumptions 3 and 4) initial storage. The necessary condition for $(\mathbf{S}^*, \mathbf{k}^*)$ to be a global minimum (Lemma 3.7 of Avriel [1976] and Bertsekas and Ozdaglar [2002]) is:

$$\mathbf{w}'\mathbf{z} \geq 0, \forall \mathbf{z} \in T_C(\mathbf{S}^*, \mathbf{k}^*; \mathbf{x}), \mathbf{w} \in \partial q(\mathbf{S}^*, \mathbf{k}^*; \tau), \quad (3.1)$$

where $\partial q(\mathbf{S}^*, \mathbf{k}^*; \tau)$ is the set of all subgradients of q at $\mathbf{S}^*, \mathbf{k}^*$ (see Avriel [1976] for definitions) and $T_C(\mathbf{S}^*, \mathbf{k}^*; \mathbf{x})$ is the tangent cone of the constraint set C at $\mathbf{S}^*, \mathbf{k}^*$.

[86] Condition (3.1) provides us with a sufficient condition for quantile noncrossing, i.e., when two quantile predictions do not intersect for any \mathbf{x} . It also links model structure deficiency to quantile crossing under certain conditions. We note that when $(\mathbf{S}^*, \mathbf{k}^*)$ is in the interior of the constraint set $C(\mathbf{x})$, which is the case when the constraint set (the model structure) is not binding (i.e., the model structure does not constrain hydrological modeling of an observed quantile),

$$T_C(\mathbf{S}^*, \mathbf{k}^*; \mathbf{x}) = \mathbb{R}^J, \quad (3.2)$$

where J is the sum of the sizes of vectors \mathbf{S}^* and \mathbf{k}^* [Bertsekas and Ozdaglar, 2002]. Condition (3.2) leads to the case stated in Proposition 1. Condition (3.2) along with (3.1) requires the set of subgradients $\partial q(\mathbf{S}^*, \mathbf{k}^*; \tau)$ to contain a zero. Thus when $\partial q(\mathbf{S}^*, \mathbf{k}^*; \tau)$ contains a 0, the constraints (represented by the tangent cone T_C) are not binding and do not affect the optimum.

[87] **Definition 3:** A model structure represented by the set \mathbf{K} that obeys condition (3.2) is not deficient.

[88] Definition 3 implies that the model space or structure is flexible enough not to constrain the attainment of the minimum of the objective function. A model structure is deficient when condition (3.2) does not hold. i.e., $T_C(\mathbf{S}^*, \mathbf{k}^*; \mathbf{x}) \neq \mathbb{R}^J$.

[89] We first examine the case when a model structure is not binding (i.e., not deficient) before elaborating on a more general case of model structure deficiency. The latter is the case where the possibility of quantile crossing appears. We then provide necessary conditions for quantile noncrossing for deficient model structures such as in section 2.3. It provides insights into how model structure deficiency introduces biases in quantile predictions and as a result into how optimal values of the asymmetric loss function can order model structures in terms of its deficiencies.

[90] **Proposition 1:** Let Definitions 1–3 and Assumptions 1–4 hold. Let $f(S_t^*; \mathbf{k}^*; \tau)$ be a τ quantile model, with $(S_t^*; \mathbf{k}^*; \tau)$ solution to QE2. Then quantile model predictions do not cross, i.e., $f(S_t^*; \mathbf{k}^*; \tau_1) \leq f(S_t^*; \mathbf{k}^*; \tau_2)$ for any $0 < \tau_1 \leq \tau_2 < 1$.

[91] The proof is provided in Appendix B.

[92] We now analyze the case when model structures do not obey condition (3.2), i.e., the case of deficient model structures.

[93] **Proposition 2:** Let definitions and assumptions hold. The model structure is deficient in the sense that $T_C(\mathbf{S}^*, \mathbf{k}^*; \mathbf{x}) \neq \mathbb{R}^J$. Let $f(S_t^*; \mathbf{k}^*)$ be a τ quantile model, with $(S_t^*; \mathbf{k}^*)$ the solution to QE2. Then $(S_t^*; \mathbf{k}^*)$ obeys the following necessary conditions:

$$\sum_{t=1}^N [\Pr(y_t \leq f(S_t^*; \mathbf{k}^*)|\mathbf{x}) - \tau + \lambda_t] \nabla_{\mathbf{k}} f(S_t^*; \mathbf{k}^*) = 0 \quad (3.3)$$

$$[\Pr(y_t \leq f(S_t^*; \mathbf{k}^*)|\mathbf{x}) - \tau + \lambda_t] \frac{\partial f(S_t^*; \mathbf{k}^*)}{\partial S_t} - \lambda_t + \lambda_{t-1} = 0, \forall t = 1, \dots, N \quad (3.4)$$

where $\nabla_{\mathbf{k}}$ is the gradient operator with respect to \mathbf{k} (i.e., $\nabla_{\mathbf{k}} = \left[\frac{\partial}{\partial k_1}, \frac{\partial}{\partial k_2}, \dots, \frac{\partial}{\partial k_{|\mathbf{k}|}} \right]$) and $\lambda_t \neq 0$ are the Lagrange multipliers corresponding to the constraints h_t .

[94] The proof is provided in the appendix. We here note that if the prediction equation is a function of more than one states, as is generally the case, equation (3.4) can be restated as,

$$[\Pr(y_t \leq f(S_t^*; \mathbf{k}^*)|\mathbf{x}) - \tau + \lambda_t] \nabla_{\vec{S}_t} f(\vec{S}_t^*; \mathbf{k}^*) - \lambda_t + \lambda_{t-1} = 0, \forall t = 1, \dots, N \quad (3.4')$$

where \vec{S}_t^* is a vector of states that influences the prediction variable of interest at time t . However as shown later, such a modification does not affect the results. Thus, the equation of form (3.4) is retained throughout the paper without any loss of generality.

[95] Let $\lambda_t^{\tau*}$ be the Lagrange multipliers that obey (3.3) and (3.4). For $\lambda_t^{\tau*} = 0, \forall t = 1, \dots, N; \forall \tau$, (3.3)–(3.4) collapse to the case of Proposition 1. In case when the constraints set is binding, there is some t and τ such that $\lambda_t^{\tau*} \neq 0$. Also the magnitude of $\lambda_t^{\tau*}$ indicates the strength with which the model structure is binding at a quantile τ and time t as it is equal to the bias in estimating the τ -th quantile of $F(\mathbf{y}|\mathbf{x})$ (as to be shown in equation (3.3a)). The Lagrange multipliers therefore define the degree of model structure deficiency. As is also shown later, the optimized objective function, i.e., the asymmetric loss function, encapsulates the total effect of model deficiency at all time steps at a given quantile level. A comparison of the asymmetric loss functions of two model structures at a given quantile thus measures the degree of flexibility that one model structure offers relative to the other (relative deficiency of one with respect to the other).

3.3. Necessary Conditions for Noncrossing Quantile Models in Presence of Model Structure Deficiency

[96] The Lagrange multipliers that measure model structure deficiencies can lead to crossing of quantile predictions. Model structure deficiencies at different quantiles can thus lead to the violation of a desirable property that quantile predictions should not cross at any point in time.

[97] For a given model structure, the predictions of a τ -quantile specific estimated model may not be ordered by τ . In this case the estimated quantile models cross and do not correspond to the respective quantiles of $F(\mathbf{y}|\mathbf{x})$. A deficient model structure is therefore a necessary condition for quantile predictions to cross. Thus crossing of quantile predictions is a diagnostic of model structure deficiency. This becomes evident upon further inspection of (3.3) and (3.4).

[98] Equations (3.3) and (3.4) under Assumption 2 that f is monotonic in atleast one element of \mathbf{k} and increasing in S_t lead to, $\forall t = 1, \dots, N$:

$$[\Pr(y_t \leq f(S_t^*; \mathbf{k}^*)|\mathbf{x}) - \tau + \lambda_t] = 0, \quad (3.3a)$$

$$\begin{aligned} \lambda_t = & \left(1 - \frac{\partial f(S_{t+1}^*; \mathbf{k}^*)}{\partial S_{t+1}}\right) \lambda_{t+1} \\ & - (\Pr(y_{t+1} \leq f(S_{t+1}^*; \mathbf{k}^*)|\mathbf{x}) - \tau) \frac{\partial f(S_{t+1}^*; \mathbf{k}^*)}{\partial S_{t+1}} \end{aligned} \quad (3.4a)$$

[99] Substituting (3.3a) into (3.4a) yields

$$\lambda_t = \lambda_{t+1} = \lambda^*. \quad (3.5)$$

[100] Equation (3.3a) along with (3.5) suggest that the λ s act as bias in predicting observed quantiles. Further if for two quantiles $\tau_1 \leq \tau_2$, $\lambda_t^{\tau_1*}$ and $\lambda_t^{\tau_2*}$ are such that $\tau_1 - \lambda_t^{\tau_1*} > \tau_2 - \lambda_t^{\tau_2*}$ then from equation (3.3a) $\Pr(y_t \leq f(S_t^{\tau_1*}; \mathbf{k}^{\tau_1*})|\mathbf{x}) > \Pr(y_t \leq f(S_t^{\tau_2*}; \mathbf{k}^{\tau_2*})|\mathbf{x})$ or that the quantile predictions cross. Here $(S_t^{\tau_2*}; \mathbf{k}^{\tau_2*})$ solves $[\Pr(y_t \leq f(S_t^{\tau*}; \mathbf{k}^{\tau*})|\mathbf{x}) - \tau + \lambda_t^{\tau*}] = 0$. Thus structure deficiencies can lead quantile predictions to cross. However if quantile predictions do cross, it is surely due to nonzero

λ s or the bias due to model structure deficiencies. Thus a necessary and sufficient condition for two quantile models not to cross at quantiles τ_1, τ_2 such that $\tau_1 \leq \tau_2$ is $\tau_1 - \lambda_t^{\tau_1*} \leq \tau_2 - \lambda_t^{\tau_2*}$ for all $t=1, \dots, N$.

[101] Note that if the equation of form (3.4') rather than (3.4) is used, allowing for more than one state variable to influence the prediction variable of interest, condition (3.4a) still holds except that it now holds for all the elements of \vec{S}_{t+1}^* . Let the vector \vec{S}_{t+1}^* have R element, i.e., $\vec{S}_{t+1}^* = \{S_{t+1}^{1*}, S_{t+1}^{2*}, \dots, S_{t+1}^{R*}\}$. Then equation (3.4a) transforms into, $\forall r = 1, \dots, R$:

$$\begin{aligned} \lambda_t = & \left(1 - \frac{\partial f(\vec{S}_{t+1}^*; \mathbf{k}^*)}{\partial S_{t+1}^r}\right) \lambda_{t+1} \\ & - (\Pr(y_{t+1} \leq f(\vec{S}_{t+1}^*; \mathbf{k}^*)|\mathbf{x}) - \tau) \frac{\partial f(\vec{S}_{t+1}^*; \mathbf{k}^*)}{\partial S_{t+1}^r} \end{aligned} \quad (3.4a')$$

[102] Substituting (3.3a) into (3.4a') still yields (3.5). Thus, we retain the formulation for a single state variable without any loss of generality as consequent results depend solely on (3.3a) and (3.5).

[103] An inspection of ((3.3)–(3.4)) and equations (3.3a), (3.4a), and (3.5) reveal that the bias is independent of parameter dimensionality and it is time invariant. These are two powerful properties that are revealed by the formal analysis. This is because the monotonicity assumption (monotonicity of the predictive equation in atleast one parameter) in Assumption 2 forces the quantile prediction to match the observed quantile upto a constant (bias) independently of parameter dimensionality. However, parameter dimensionality may effect quantile specific model estimation on finite sample due to model complexity (through arguments related to rates of convergence of estimators [see, e.g., Pande *et al.*, 2009, 2012b]). The theoretical development here is only for large sample sizes.

[104] The following proposition reveals the main property of the asymmetric loss function (the objective function of QE2 at the optimum) used in this paper, showing that an ordering of its optimal values orders different model structures in terms of its deficiency.

[105] **Proposition 3:** Let Definitions 1–2 and Assumptions 1–6 hold. Let $q(\mathbf{S}, \mathbf{k}; \tau, \mathbf{x})$ be defined as in Definition 2 for the two τ -quantile estimators as q_1 and q_2 respectively such that $q_2 - q_1 \geq 0$. Let $f_1 = f_1(\hat{S}_t; \hat{\mathbf{k}})$ and $f_2 = f_2(\tilde{S}_t; \tilde{\mathbf{k}})$ be two arbitrary quantile estimators such that (a) bias in f_2 or f_1 ($\Pr(y_t \leq f_2|\mathbf{x}) - \tau$ or $\Pr(y_t \leq f_1|\mathbf{x}) - \tau$) has the same value for all t , (b) bias in f_2 has the same sign as bias in f_1 , i.e., $\text{sgn}(\Pr(y_t \leq f_2|\mathbf{x}) - \tau) = \text{sgn}(\Pr(y_t \leq f_1|\mathbf{x}) - \tau)$ at any time t . Then the magnitude of bias in $f_2 >$ the magnitude of bias in f_1 , i.e., $|\Pr(y_t \leq f_2|\mathbf{x}) - \tau| > |\Pr(y_t \leq f_1|\mathbf{x}) - \tau|$ at any time t . Or,

$$q_2 - q_1 \geq 0 \Rightarrow |\Pr(y_t \leq f_2|\mathbf{x}) - \tau| \geq |\Pr(y_t \leq f_1|\mathbf{x}) - \tau|.$$

[106] The proof of the proposition is provided in Appendix B. A corollary of Proposition 3 is provided in the following that suggests that deficiency of a model structure in optimally modeling a quantile τ can be measured relative to the deficiency of another using the asymmetric loss function. The bias in optimally modeling the τ -th quantile of

$F(\mathbf{y}|\mathbf{x})$ is represented by λ^* that is a Lagrange multiplier (due to (3.5) and (3.3a)) corresponding to the model structure constraints defined as $C(\mathbf{x})$ in Definition 1. In effect, the Lagrange multipliers measure this bias, which has been defined as model structure deficiency. Thus, note that Proposition 3 allows an ordering of model structures at a particular quantile in terms of its respective bias in optimally estimating the τ -th quantile of $F(\mathbf{y}|\mathbf{x})$. Corollary 4 is an application of proposition when definitions of and assumptions hold.

[107] **Corollary 4:** Let definitions and assumptions hold. Let model structure deficiency, or the bias in optimally estimating the τ -th quantile of $F(\mathbf{y}|\mathbf{x})$ by a model structure (a model structure represented by \mathbf{K}), be defined by the Lagrange multiplier $\lambda^{\mathbf{K}*}$ given by equations (3.3a), (3.4a), and (3.5). Let $q(\mathbf{S}^*, \mathbf{k}^*; \tau, \mathbf{x}, \mathbf{K})$ be the asymmetric loss function (objective function in Definition at the optimum). Further let \mathbf{K}_1 and \mathbf{K}_2 represent two model structures such that $\text{sgn}(\lambda^{\mathbf{K}_1*}) = \text{sgn}(\lambda^{\mathbf{K}_2*})$ and let q_1^* and q_2^* represent the value of its asymmetric loss function at the optimum. Then

$$[108] \quad 1) q_1^* - q_2^* \leq 0 \Rightarrow |\lambda^{\mathbf{K}_1*}| \leq |\lambda^{\mathbf{K}_2*}|.$$

$$[109] \quad 2) q_1^* - q_2^* = 0 \Rightarrow |\lambda^{\mathbf{K}_1*}| = |\lambda^{\mathbf{K}_2*}|.$$

[110] **Proof:** 1) Let f_1 and f_2 in Proposition represent the optimal estimators from the two model structures \mathbf{K}_1 and \mathbf{K}_2 , respectively. It then follows from equations (3.3a) and (3.5) that the bias in f_2 or f_1 ($\Pr(y_t \leq f_2|\mathbf{x}) - \tau = \lambda^{\mathbf{K}_2*}$ or $\Pr(y_t \leq f_1|\mathbf{x}) - \tau = \lambda^{\mathbf{K}_1*}$) has the same value for all t (time invariance). Since it is given that $\text{sgn}(\lambda^{\mathbf{K}_1*}) = \text{sgn}(\lambda^{\mathbf{K}_2*})$, it follows from Proposition that $q_1 - q_2 \leq 0 \Rightarrow |\lambda^{\mathbf{K}_1*}| \leq |\lambda^{\mathbf{K}_2*}|$.

[111] 2) Follows from proposition for the case when proposed inequality holds with equality. ■

[112] Corollary 4 suggests that differences in biases of optimal quantile estimators, due to differences in their structural deficiency, are nonlinearly measured by the differences in optimal q (the asymmetric loss function). Thus an ordering of model structures by optimal q (the asymmetric loss function) provides an ordering of model structures in terms of their deficiencies. Note that, for any two flexible model structures, i.e., when condition (3.2) holds for two model structures, the difference between the asymmetric loss functions is also 0 (in addition to case 2 within corollary 4). The difference in the asymmetric loss functions will be nonzero only in the situation when two model structures are deficient to different degrees at a given quantile. It is nonnegative, as corollary 4 suggests, when model structure 2 is more deficient than model structure 1. The difference in the asymmetric loss function however can be positive for one quantile while negative for another since the ordering of model structures due to corollary 4 holds for one quantile at a time. The asymmetric loss function thus identifies relative deficiencies of model structures at different quantiles of a variable of interest.

4. Discussion

[113] Model structure deficiency (rigidity) is encapsulated by the asymmetric loss function. Thus explicit quantification of model structure rigidity (deficiency) in terms of bias in estimating the τ -th quantile of $F(\mathbf{y}|\mathbf{x})$ as represented by the Lagrange multipliers can be relaxed as long as one

can obtain a global minimum. A difference in the asymmetric loss functions of two model structures is a function of the underlying data generating process that is reflected in the conditional cumulative density function $F(\mathbf{y}|\mathbf{x})$. The cumulative density function in turn is a function of the underlying processes and measurement errors which for given input forcings generate a time series of variable of interest such as streamflow, evaporation or storage. Its decomposition into respective components requires assumptions on the structure of processes or measurement errors [Renard et al., 2010], which we refrain from in this study. We however remark that an assumption on the structure of measurement errors is sufficient, as it reveals the distribution of measurement-error-corrected conditional distribution of \mathbf{y} on \mathbf{x} given $F(\mathbf{y}|\mathbf{x})$ (uncorrected conditional distribution).

[114] No assumption is made on the structure of the cumulative distribution of observed variable of interest \mathbf{y} conditioned on input forcing \mathbf{x} , $F(\mathbf{y}|\mathbf{x})$ (except continuity of F , boundedness of \mathbf{y} and near zero probability of occurrence of \mathbf{y} at its lower bound). This distribution is a function of underlying processes as well as of measurement errors present in the data set. Acceptable assumptions on the latter allow resolution of these two sources, allowing inference of model structure uncertainty. Measurement error vectors can be generated to create multiple instantiations of measurement-error-corrected data sets. An ordering based on deficiency of model structures can then be an ordering based on model structure uncertainty when the objective function is an average of asymmetric loss function over these multiple instantiation of measurement-error-corrected data sets. However, it remains a relative assessment of one model structure with respect to others. We intend to pursue this in future work.

4.1. Personal Belief in Defining Model Structure Deficiency?

[115] The only “personal belief” that has been invoked in the analysis is the existence of “truth.” It however is inconsequential to the presented analysis if the truth “cannot” be realized by our knowledge in the form of variety of hydrological model structures. But if a model structure can replicate the truth, the truth lies in the N dimensional (output) space spanned by this model structure. Here N represents the number of sample points and can even be infinity.

[116] It is not a matter of personal belief to know that when a solution lies in the interior of a constraint set, the constraint set is not binding on the problem of finding the solution. A constraint set in context of the paper is a model structure and the solution represents a model that replicates the “truth.” The model structure is thus defined not to be deficient when it contains the “truth.”

[117] The complement of this definition then defines a deficient model structure, which does not require any set of beliefs. The definition of a deficient model structure as stated in the paper then follows after a few more steps. Hence the definition of a deficient model structure is subject to a personal belief on the existence of truth (such as the underlying physical laws), absence of which would, in my opinion, nullify the essence of any model structure improvement exercise.

[118] It has been duly noted in the paper at several places that quantile model selection at present does not distinguish

between structural errors and measurement errors. A solution to distinguish between the two (and thus topic for future research) has also been proposed in the paper. Therefore a perfect hydrological model (one that is an exact representation of the complete system) driven with corrupted input data from potentially unknown initial states may constrain the estimation of quantiles of the observed (corrupted) output. The “perfect” hydrological model structure is deficient if the “perfect” hydrological model structure on its own cannot see the truth given the uncertainties.

[119] We use observations as a representation of the “truth.” If the representation of truth is corrupted, one needs models for measurement errors to “reveal” the truth so that the “perfect” model structure can be put to use in an unconstrained manner. It is however possible that the “perfect” model structure is sufficiently complex and can model measurement errors alongside the processes even though we may not be able identify how it does so. Why deficiency may reduce with increasing complexity is because known measurement error models can often be represented by hydrologic model concepts. Then the truth lies within the output space of the “perfect” model structure and it is no longer deemed deficient even though we cannot isolate process representation from measurement errors.

[120] Consider an example (similar to the one in the paper) to demonstrate two cases (i) when the “perfect” model is not sufficiently complex and (ii) when it is sufficiently complex to contain the observations. The truth corrupted by noise is what we observe.

[121] Let the truth be represented by $y^* = \beta_o^* + \beta_1^* x$. Let it be corrupted by additive noise $\varepsilon = \eta x^2$ (defining measurement errors) where x is iid random variable and η is normally distributed with mean 0 and variance 1. The observations are then given by $y = y^* + \varepsilon$. Let $q_{y|x}(\tau)$ for $0 < \tau < 1$ represent the τ th quantile of observations y conditioned on a given x . Conceptually, quantile hydrological model selection attempts to model a quantile of observations conditioned on a given data set. It is straightforward to note that,

$$\begin{aligned} q_{y|x}(\tau) &= q_{y^*|x}(\tau) + q_{\varepsilon|x}(\tau) = \beta_o^* + \beta_1^* x + q_{\eta|x}(\tau)x^2 \\ &= \beta_o^* + \beta_1^* x + q_{\eta}(\tau)x^2. \end{aligned}$$

[122] Now consider the first case when we attempt to model $q_{y|x}(\tau)$ by using a class of linear functions. Note that the class of linear functions is “perfect” in the sense that the “truth” (observations treated for their corrupted selves) comes from this class of functions. However the class of linear functions is unable to see the “truth” within the observations due to measurement errors. Thus, using the class of linear regressors without treating for measurement errors leads to bias of $q_{\eta}(\tau)x^2$ in estimating the τ th quantile.

[123] Now consider the second case in two parts. First consider a class of quadratic functions. If a class of linear functions is deemed “perfect,” so is the class of quadratic functions since it contains the class of linear functions. If we again attempt to model the τ th quantile of the observations, the bias in the case of the class of linear regressors is now absorbed by the coefficient of the quadratic term in the case of the class of quadratic functions. The model structure is then deemed sufficiently complex and not deficient.

[124] Naturally, the above construct does not allow us to distinguish between the truth and corruption. Let some benchmark study on measurement errors suggest that the corruption is additive and distributed as $\tilde{\eta}x^2$ with $\tilde{\eta}$ distributed as Gaussian with unknown mean and a positive standard deviation, the truth can be revealed and model structure deficiency can be isolated (if any) from the measurement corruption. The case of using the “perfect” model structure of the class of linear regressors is straightforward and is similar to the class of quadratic functions case presented previously. Instead consider the class of constants. We use the class of functions of type $\beta_0 + \beta_2 x^2$ with the first addendum representing our class of models (of constants) to represent the “truth” while the second addendum is to reveal the “truth” (which we know represents the corruption based on certain benchmarking). Then, when modeling the τ th quantile of the observations, the bias in representing the truth is $\beta_1^* x$ while the specification of corruption is recovered in the estimate of β_2 . In particular, we recover $q_{\eta}(\tau)$ (the cdf of η) through τ th quantile estimate of β_2 when τ is varied between 0 and 1. We can then estimate the mean and the variance of corruption accordingly.

[125] We here note that an improvement in a model structure, even without first decorruping observations (to reveal the truth), such that deficiency reduces, in itself ensures that we are improving the conceptualization of both the truth and the measurement error jointly. What we do not know is how to distinguish between the two. We can distinguish between the two if we have benchmarking studies that can define the type of measurement errors that we have. For example, consider the case above. If we move from a model structure of a class of constant functions via the class of linear functions to a class of quadratic functions, the bias (deficiency) in modeling a τ th quantile of observations reduces from $\beta_1^* x + q_{\eta}(\tau)x^2$ via $q_{\eta}(\tau)x^2$ to 0 for $x > 0$. This would suggest to a modeler, who is unaware of the type of measurement errors, that she is reducing deficiency even though she is unable to distinguish between the sources. Once the modeler becomes aware of the type of measurement errors (through certain benchmarking studies), she would be able to distinguish the effect of structure deficiency from the effect of measurement errors.

4.2. Convergence of Quantile Model Selection: Finite Sample Performance

[126] One may then argue that increasing complexity leads to overfitting. Indeed it may if the sample size is small. This has also been noted in the paper. The theory is presented for large sample sizes (note the expectation operator). The theory is either applicable for large samples or for the cases in which the complexity is regularized so that the effect of complexity on model estimation is minimum. All the examples and the case studies either are for large samples (all flexible model case studies span 6 years) or the complexity of the problem has been controlled for (the case of dryland model where several constraints common to the two structures are considered to regularize the model selection problem). The case study in section 2.3.1 is a cyclic one, extending the data to infinity does not affect the result. The same applies for the parsimonious case study of

western India (in *Pande* [2013]) since we are modeling a seasonal cycle.

[127] The simulation results can be sensitive to small data sizes just as in the case of any other calibration/model selection problem (including Bayesian estimators). For a given small amount of data, it may be more sensitive to boundary quantiles (that are close to either 0 or 1). This has also been stated in the paper. However, the choice of absolute error deviations when selecting a particular quantile model is a robust estimator that is insensitive to outliers. Since the sensitivity of estimation to small sample sizes is often (but not always) due to outliers, it is relatively (relative to other measures such as square of residuals) robust.

[128] A question remains whether the estimator converges as the sample size goes to infinity. By convergence I mean the convergence of a quantile model estimated on finite sample to a quantile model estimated on infinite sample. The complexity of a model structures affects its rate of convergence [see, e.g., *Pande et al.*, 2009, 2012; L. Arkesteijn and S. Pande, On hydrological model complexity, its geometrical interpretations and prediction uncertainty, submitted to *Water Resources Research*, 2013]. Structures with large complexity converge slower than the structures with low complexity. However, the convergence is ensured as long as the complexity is finite. The complexities of most hydrological model structures, including the ones used in this paper, are finite based on recent results for a class of hydrological models [*Pande et al.*, 2012b], for k-nearest neighbor hydrological models [*Pande et al.*, 2009] and for the flexible model structures used in this study (Arkesteijn and Pande, submitted manuscript, 2013). Thus, the convergence at large sample size is almost ensured.

[129] Since the complexity of model structures is the same for any quantile, one may argue that the rate of convergences should also be the same. However, the rate of convergence of an estimation depends both on the performance measure and the model structure used. The performance measure plays the role of transforming the effect of model structure complexity on model estimation [see, e.g., *Pande et al.*, 2012b; Arkesteijn and Pande, submitted manuscript, 2013]. Since the performance measure in quantile model selection problem is an asymmetric loss function that is composed of absolute deviations, the complexity of the estimation problem remains finite. However since the asymmetry of the loss function depends on the quantile under study, the complexity of the estimation problem may differ for different quantile model selection problems. To summarize, convergence is ensured for quantiles, when using conceptual model structures as used in this study though the rate of convergence may be different for different quantiles.

4.3. Consistency of Bayesian Estimators

[130] One of the arguments of the paper has been that unless a likelihood function is fully specified it does not lead to an appropriate posterior distribution. A model selection based on inappropriate posterior distribution may be inappropriate as well. However distributions that satisfy the axioms of probability are still distributions. Nonetheless, distributions that are not fully specified are “personal degrees of belief” [*Bernardo and Smith*, 1994, p. 35] and thus are subjective. This is in spite of a framework within

which it may not make sense to use adjectives such as “objective” [*Bernardo and Smith*, 1994, p. 35]. This is because one arrives at such a framework by assuming that the notion of “rational beliefs” based on accumulated information cannot be separated from the notion of “rational actions” [*Bernardo and Smith*, 1994, p. 15].

[131] Based on the assumption that a model selection problem is no different from a choice problem under uncertainty, one can indeed invoke *von Neumann-Morgenstern* [1947]’s expected utility theory. Within such or other associated frameworks, a distribution need not even be additive. See, for example, *Gilboa* [1987]’s seminal article for a discussion on the generalization of expected utility maximization under “personal belief” of uncertainty.

[132] However, whether a choice of a model under (personal beliefs of) uncertainty is asymptotically consistent still remains a question. The proof of convergence of prior belief to the “true” posterior necessarily requires that the latter is in the support of the former [see, e.g., *Freedman*, 1963, 1965; *Barron*, 1988; *Feldman*, 1991]. A belief is a distribution defined on the set of distributions that possibly generated the observations. If the problem is misspecified, i.e., none in the set of distributions (which a modeler assumes based on her specification of the likelihood function) generated the observations, the convergence of posterior beliefs (for any given prior belief) to the true posterior distribution is impossible (see, for example, the discussion of the assumptions and results of *Feldman* [1991]; also see Theorem 1 of *Freedman* [1963], in particular regarding the topological carrier of a prior belief). The case studies presented in sections 2.3.1 and 2.3.2 showcase the examples of model misspecification, where no prior specification on the Generalized Likelihood function may result in a consistent estimation of the posterior.

[133] Indeed the role of prior specification is well recognized in cases when model misspecification is absent. However even in these cases of “strong a priori belief,” the convergence of Bayesian beliefs to the true posterior is not ensured without additional conditions [*Feldman*, 1991].

5. Conclusions

[134] A theory for quantile model selection and model structure deficiency assessment was presented in this paper. The case studies and the formal analysis of quantile model selection problem suggested that degree of model structure deficiencies (or rigidities) as measured by the Lagrange multipliers corresponding to the model structure constraints are embedded in the asymmetric loss function.

[135] The unique contribution of this paper was the mathematical formulation of quantile model selection problem in the Lagrangian form. It elucidated why the asymmetric loss function can be used to assess model structure deficiency. The degree of model structure deficiency was reflected in the Lagrange multipliers of the constraints posed by a model structure on a quantile model selection problem. This also led to the formal definition of model structure deficiency (or rigidity) and the formulation of a sufficient condition for model structure flexibility. This formal analysis was presented for a predictive equation with reasonable assumptions that hold in general for

hydrological models. The case studies further supported that the formal analysis holds for hydrological models in general.

[136] One main insight from the formal analysis of quantile model selection problem was that the asymmetric loss function at any quantile can order model structures by its structure deficiencies. Further it was shown that crossing of two quantile predictions is necessarily due to model structure deficiency. This was also revealed by various case studies that were undertaken. It also revealed that the bias due to structure deficiency is independent of model parameter dimensionality and is time invariant. The analysis crucially depended on two assumptions that a model prediction of a variable of interest is nondecreasing in state variables and is monotonic in one of the parameters, such as recession parameters. Neither of these two assumptions is unrealistic when fluxes such as evaporation or streamflow are considered. Another assumption of differentiability of a model structure predictive equation may seem restrictive but it can be relaxed.

Appendix A: Description of Kernel Density Independence Sampling-Based Monte Carlo Scheme (KISMCS)

[137] KISMCS is a Markov Chain Monte Carlo method with Metropolis-Hastings updates [Kuczera and Parent, 1998] using an independence sampler [Brooks, 1998]. The independence sampler ensures that candidate observations are drawn independently of the current state of a chain, thereby ensuring efficient exploration of the target distribution [Pasarica and Gelman, 2010]. The M-H acceptance-rejection criteria are also used to sample across n chains, which ensures that the chains are well mixed. Kernel density estimation [Haerdle, 2004] on last m samples in a chain is used to calculate standardized importance weights [Kuczera and Parent, 1998; Givens and Raftery, 1996] within the independence sampler to ensure fast convergence of sampled points to the target distribution. An overdispersed distribution, a multivariate t -distribution, is used as the kernel [Gelman and Rubin, 1992] to ensure exhaustive exploration of the parameter space. The convergence proof of the algorithm is standard [Roberts and Smith, 1994] but is beyond the scope of this paper. Exhaustive tests of the algorithm and its comparison with other MCMC algorithms is a topic for future study.

[138] The algorithm is implemented with $m=600$, a Gaussian reference bandwidth for multivariate t distribution, sampled parameter covariance matrix updates every 600 function evaluations and the number of chains, n =parameter dimensionality of a model under consideration.

Appendix B

[139] **Proposition 1:** Let Definitions 1–3 and Assumptions 1–4 hold. Let $f(S_t^*; \mathbf{k}^*)$ be a τ quantile model. Then quantile models selected are noncrossing, i.e., $f(S_t^{\tau_1*}; \mathbf{k}^{\tau_1*}) \leq f(S_t^{\tau_2*}; \mathbf{k}^{\tau_2*})$ for $\tau_1 \leq \tau_2$.

[140] **Proof:** Definition 3 along with condition (3.1) yields necessary conditions for an unconstrained minimum,

$$0 \in \partial q(\mathbf{S}^*, \mathbf{k}^*; \tau) \quad (\text{B1})$$

[141] Thus, the following holds for the 0 element of $\partial q(\mathbf{S}^*, \mathbf{k}^*; \tau)$ based on condition (B1):

$$\int \sum_{t=1}^N ([I(y_t - f(S_t^*; \mathbf{k}^*)) - \tau] \nabla f(S_t^*; \mathbf{k}^*)) p(\mathbf{y}|\mathbf{x}) d\mathbf{y} = 0,$$

or

$$\sum_{t=1}^N [\Pr(y_t \leq f(S_t^*; \mathbf{k}^*)|\mathbf{x}) - \tau] \nabla f(S_t^*; \mathbf{k}^*) = 0 \quad (\text{B2})$$

where $I(v)$ is an indicator function, which takes a value of 1 when $v \leq 0$, else 0.

[142] Finally, condition (B2) yields (under Assumption 2 that $f(S_t; \mathbf{k})$ is monotonic in at least one element of \mathbf{k} or $f(S_t; \mathbf{k})$ is increasing in S_t),

$$\Pr(y_t \leq f(S_t^*; \mathbf{k}^*)|\mathbf{x}) - \tau = 0, \quad \forall t = 1, \dots, N, \quad (\text{B3})$$

[143] Note for example, that under the assumption of $\frac{\partial f(S_t; \mathbf{k})}{\partial S_t} \geq 0, \forall t = 1, \dots, N$, Left Hand Side of condition (B2) can only be 0 when condition (B3) holds.

[144] Let τ -quantile model estimates be indexed by τ . Let $(S_t^{\tau_1*}, \mathbf{k}^{\tau_1*})$ and $(S_t^{\tau_2*}, \mathbf{k}^{\tau_2*})$ satisfy condition (3.4a) for $\tau_1 < \tau_2$. From equation (B3) and under Assumption 6 that $\Pr(y_t \leq \mu|\mathbf{x})$ is continuous nondecreasing function in μ ,

$$\Pr(y_t \leq f(S_t^{\tau_1*}; \mathbf{k}^{\tau_1*})|\mathbf{x}) \leq \Pr(y_t \leq f(S_t^{\tau_2*}; \mathbf{k}^{\tau_2*})|\mathbf{x}), \quad \forall t = 1, \dots, N \\ \Rightarrow f(S_t^{\tau_1*}; \mathbf{k}^{\tau_1*}) \leq f(S_t^{\tau_2*}; \mathbf{k}^{\tau_2*}), \quad \forall t = 1, \dots, N.$$

[145] Thus conditioned on \mathbf{x} quantile models selected are noncrossing for the case when the model structure is not deficient [obeys equation (3.2)],

$$f(S_t^{\tau_1*}; \mathbf{k}^{\tau_1*}) < f(S_t^{\tau_2*}; \mathbf{k}^{\tau_2*}), \quad \forall t = 1, \dots, N.$$

[146] **Proposition 2:** Let Definitions 1–2 and Assumptions 1–4 hold. The model structure is deficient in the sense that $T_C(\mathbf{S}^*, \mathbf{k}^*; \mathbf{x}) \neq \mathcal{H}$. Let $f(S_t^*; \mathbf{k}^*)$ be a τ quantile model, with $(S_t^*; \mathbf{k}^*)$ solution to QE2. Then $(S_t^*; \mathbf{k}^*)$ obeys the following necessary conditions:

$$\sum_{t=1}^N [\Pr(y_t \leq f(S_t^*; \mathbf{k}^*)|\mathbf{x}) - \tau + \lambda_t] \nabla_{\mathbf{k}} f(S_t^*; \mathbf{k}^*) = 0 \quad (3.3)$$

$$[\Pr(y_t \leq f(S_t^*; \mathbf{k}^*)|\mathbf{x}) - \tau + \lambda_t] \frac{\partial f(S_t^*; \mathbf{k}^*)}{\partial S_t} - \lambda_t + \lambda_{t-1} = 0, \quad \forall t = 1, \dots, N \quad (3.4)$$

where $\nabla_{\mathbf{k}}$ is the gradient operator with respect to \mathbf{k} and $\lambda_t \neq 0$ are the Lagrange multipliers corresponding to constraints h_t .

[147] **Proof:** Gradients of h_t , i.e., ∇h_t , exist (due to differentiability of f under Assumption 2 though this assumption can be relaxed without affecting the conclusions in what

follows). For a nonzero minimum, columns of a matrix H with rows $\nabla h_t(\mathbf{S}^*, \mathbf{k}^*)$, $t = 1, \dots, N$, will be linearly independent where $\nabla(\tilde{\mathbf{n}}) = [\frac{\partial}{\partial S_1}, \dots, \frac{\partial}{\partial S_t}, \dots, \frac{\partial}{\partial S_N}, \frac{\partial}{\partial k_1}, \dots, \frac{\partial}{\partial k_t}, \dots, \frac{\partial}{\partial k_N}]^T(\tilde{\mathbf{n}})$ given Assumption 2. Note that $H_{tt'} = 0$ (t -th row and t' -th column of H), for $t, t' \leq N$, $t \neq \{t', t' + 1\}$ thereby ensuring linear independence of first N columns. Linear independence amongst columns $j = N+1, \dots, J$ is due to Assumption 2.

[148] Thus, QE2 satisfies Mangasarian-Fromovitz type constraint qualification [Mangasarian and Fromovitz, 1967], thereby admitting Lagrange multipliers (see CQ1 of Bertsekas and Ozdaglar [2002]).

[149] Any solution $(\mathbf{S}^*, \mathbf{k}^*)$ of QE2 then obeys the following [Bertsekas and Ozdaglar, 2002],

$$\sum_{t=1}^N [\Pr(y_t \leq f(S_t^*; \mathbf{k}^*) | \mathbf{x}) - \tau] \nabla f(S_t^*; \mathbf{k}^*) + \sum_{t=1}^N \lambda_t \nabla h_t(\mathbf{S}^*, \mathbf{k}^*) = 0,$$

where $\lambda_t \neq 0$ are the Lagrange multipliers corresponding to constraints h_t . The above necessary conditions for S_t and \mathbf{k} read as:

$$\sum_{t=1}^N [\Pr(y_t \leq f(S_t^*; \mathbf{k}^*) | \mathbf{x}) - \tau + \lambda_t] \nabla_k f(S_t^*; \mathbf{k}^*) = 0$$

$$[\Pr(y_t \leq f(S_t^*; \mathbf{k}^*) | \mathbf{x}) - \tau + \lambda_t] \frac{\partial f(S_t^*; \mathbf{k}^*)}{\partial S_t} - \lambda_t + \lambda_{t-1} = 0, \forall t = 1, \dots, N$$

where ∇_k is the gradient operator with respect to \mathbf{k} .

[150] **Proposition 3:** Let Definitions 1–2 and Assumptions 1–6 hold. Let $q(\mathbf{S}, \mathbf{k}; \tau, \mathbf{x})$ be defined as in Definition 2 for the two τ -quantile estimators as q_1 and q_2 respectively such that $q_2 - q_1 \geq 0$. Let $f_1 = f_1(\hat{S}_t; \hat{\mathbf{k}})$ and $f_2 = f_2(\hat{S}_t; \hat{\mathbf{k}})$ be two arbitrary quantile estimators such that (a) bias in f_2 or f_1 ($\Pr(y_t \leq f_2 | \mathbf{x}) - \tau$ or $\Pr(y_t \leq f_1 | \mathbf{x}) - \tau$) has the same value for all t , (b) bias in f_2 has the same sign as bias in f_1 , i.e., $\text{sgn}(\Pr(y_t \leq f_2 | \mathbf{x}) - \tau) = \text{sgn}(\Pr(y_t \leq f_1 | \mathbf{x}) - \tau)$ at any time t . Then the magnitude of bias in $f_2 >$ the magnitude of bias in f_1 , i.e., $|\Pr(y_t \leq f_2 | \mathbf{x}) - \tau| > |\Pr(y_t \leq f_1 | \mathbf{x}) - \tau|$ at any time t . Or,

$$q_2 - q_1 \geq 0 \Rightarrow |\Pr(y_t \leq f_2 | \mathbf{x}) - \tau| \geq |\Pr(y_t \leq f_1 | \mathbf{x}) - \tau|.$$

[151] **Proof:** Consider the objective function,

$$q(\mathbf{S}^*, \mathbf{k}^*; \tau, \mathbf{x}) = \int_{\underline{\mathbf{y}}}^{\bar{\mathbf{y}}} \sum_{t=1}^N [(1-\tau)|y_t - f(S_t^*, \mathbf{k}^*)|_- + \tau|y_t - f(S_t^*, \mathbf{k}^*)|_+] p(\mathbf{y} | \mathbf{x}) d\mathbf{y} / N$$

$$= \frac{1}{N} \sum_{t=1}^N \int_{\underline{\mathbf{y}}}^{\bar{\mathbf{y}}} [(1-\tau)|y_t - f(S_t^*, \mathbf{k}^*)|_- + \tau|y_t - f(S_t^*, \mathbf{k}^*)|_+] p(\mathbf{y} | \mathbf{x}) d\mathbf{y}$$

[152] Following is a decomposition of objective function for an arbitrary estimate of τ -quantile (not a result of minimization):

$$q = \frac{1}{N} \sum_{t=1}^N \int_{\underline{\mathbf{y}}^{-t}}^{\bar{\mathbf{y}}^{-t}} I_t p(\mathbf{y}^{-t} | \mathbf{x}) d\mathbf{y}^{-t} \quad (\text{B4})$$

where $\bar{\mathbf{y}}$ and $\underline{\mathbf{y}}$ is the upper and lower bound N -vectors of \mathbf{y} and \mathbf{y}^{-t} is \mathbf{y} vector without y_t (Assumption 5), i.e., $\mathbf{y}^{-t} = \{y_1, \dots, y_{t-1}, y_{t+1}, \dots, y_N\}$ and

$$I_t = \int_{\underline{\mathbf{y}}}^{\bar{\mathbf{y}}} [(1-\tau)|y_t - f(S_t, \mathbf{k})|_- + \tau|y_t - f(S_t, \mathbf{k})|_+] p(y_t | \mathbf{y}^{-t}, \mathbf{x}) dy_t$$

[153] The form as above is the most generic form to preserve any intertemporal dependence.

[154] Further, q is linearly additive in I_t .

[155] Using integration by parts I_t can be decomposed as (further details in section B1),

$$I_t = \left\{ \int_{F^{-1}(\tau)}^f [\hat{F}(y_t) - \tau] dy_t \right\} + \left\{ \int_{\underline{\mathbf{y}}}^{F^{-1}(\tau)} [\hat{F}(y_t) - \tau] dy_t + \tau \int_{\underline{\mathbf{y}}}^{\bar{\mathbf{y}}} (1 - \hat{F}(y_t)) dy_t \right\} \quad (\text{B5})$$

[156] Equation (B5) can then be further simplified as,

$$I_t = I_t^1 + I_t^2, \quad (\text{B6})$$

where

$$I_t^1 = \int_{\underline{\mathbf{y}}}^{F^{-1}(\tau)} [\hat{F}(y_t) - \tau] dy_t + \tau \int_{\underline{\mathbf{y}}}^{\bar{\mathbf{y}}} (1 - \hat{F}(y_t)) dy_t \quad (\text{B7})$$

$$I_t^2 = \int_{F^{-1}(\tau)}^f [\hat{F}(y_t) - \tau] dy_t \quad (\text{B8})$$

[157] The above shows that while I_t^1 is independent of $f(S_t, \mathbf{k})$, I_t^2 is not. Rather, $f(S_t, \mathbf{k})$ appears in its upper bound. The integral I_t^2 is nondecreasing in quantile prediction made by $f(S_t, \mathbf{k})$. It would also be nonnegative in case when f is positively biased. Since I_t^1 is independent of $f(S_t, \mathbf{k})$, I_t is nondecreasing in prediction made by $f(S_t, \mathbf{k})$.

[158] Let two estimators (may not be optimal) f_1 and f_2 (predictions at time t with reference to time index suppressed) be such that a) bias in f_2 or f_1 ($\Pr(y_t \leq f_2 | \mathbf{x}) - \tau$ or $\Pr(y_t \leq f_1 | \mathbf{x}) - \tau$) has the same value for all t , b) bias in f_2 has the same sign as bias in f_1 , i.e., $\text{sgn}(\Pr(y_t \leq f_2 | \mathbf{x}) - \tau) = \text{sgn}(\Pr(y_t \leq f_1 | \mathbf{x}) - \tau)$ at any time t . Then a nonnegative difference in the corresponding I_t (note that $\hat{F}(y_t) = F(y_t | \mathbf{y}^{-t}, \mathbf{x})$) implies that the magnitude of bias in $f_2 \geq$ the magnitude of bias in f_1 , i.e.,

$|\Pr(y_t \leq f_2 | \mathbf{x}) - \tau| \geq |\Pr(y_t \leq f_1 | \mathbf{x}) - \tau|$ at any time t . This is now shown to hold.

[159] Let $I_{t,1}$ be I_t in (B6) for $f=f_1$. Similarly let $I_{t,2}$ be I_t in (B6) for $f=f_2$. Then the following needs to be proved:

$$\begin{aligned} I_{t,2} - I_{t,1} &= \int_{f_1}^{f_2} [F(y_t | \mathbf{y}^{-t}, \mathbf{x}) - \tau] dy_t \geq 0 \Rightarrow |\Pr(y_t \leq f_2 | \mathbf{x}) - \tau| \\ &\geq |\Pr(y_t \leq f_1 | \mathbf{x}) - \tau| \end{aligned}$$

[160] Let $\Pr(y_t \leq f_2 | \mathbf{x}) - \tau = \lambda_2$ and $\Pr(y_t \leq f_1 | \mathbf{x}) - \tau = \lambda_1$. Consider the only two cases below that obey conditions a) and b) above.

[161] 1) Case 1: $\lambda_2, \lambda_1 \geq 0$ for all t .

[162] This implies that $[F(y_t | \mathbf{y}^{-t}, \mathbf{x}) - \tau] \geq 0$ for $y_t = f_2$ and $y_t = f_1$. Further since $[F(y_t | \mathbf{y}^{-t}, \mathbf{x}) - \tau]$ is nondecreasing in y_t , $[F(y_t | \mathbf{y}^{-t}, \mathbf{x}) - \tau] \geq 0$ for all y_t between f_1 and f_2 . Thus,

$$\int_{f_1}^{f_2} [F(y_t | \mathbf{y}^{-t}, \mathbf{x}) - \tau] dy_t \geq 0$$

is only possible when

$$[F(f_1 | \mathbf{y}^{-t}, \mathbf{x}) - \tau] = \lambda_1 \leq [F(f_2 | \mathbf{y}^{-t}, \mathbf{x}) - \tau] = \lambda_2 \Rightarrow |\lambda_1| \leq |\lambda_2|$$

[163] Thus,

$$\begin{aligned} I_{t,2} - I_{t,1} &= \int_{f_1}^{f_2} [F(y_t | \mathbf{y}^{-t}, \mathbf{x}) - \tau] dy_t \geq 0 \iff \\ |\Pr(y_t \leq f_2 | \mathbf{x}) - \tau| &\geq |\Pr(y_t \leq f_1 | \mathbf{x}) - \tau| \end{aligned} \quad (\text{B9})$$

for any t .

[164] 2) Case 2: $\lambda_2, \lambda_1 \leq 0$

[165] Just as in case 1, we can conclude that $[F(y_t | \mathbf{y}^{-t}, \mathbf{x}) - \tau] \leq 0$ for all y_t between f_1 and f_2 . Thus,

$$\int_{f_1}^{f_2} [F(y_t | \mathbf{y}^{-t}, \mathbf{x}) - \tau] dy_t \geq 0$$

is only possible when

$$[F(f_1 | \mathbf{y}^{-t}, \mathbf{x}) - \tau] = \lambda_1 \geq [F(f_2 | \mathbf{y}^{-t}, \mathbf{x}) - \tau] = \lambda_2.$$

[166] Since,

$$\lambda_1 \geq \lambda_2 \Rightarrow -|\lambda_1| \geq -|\lambda_2| \Rightarrow |\lambda_1| \leq |\lambda_2|$$

[167] Thus,

$$\begin{aligned} I_{t,2} - I_{t,1} &= \int_{f_1}^{f_2} [F(y_t | \mathbf{y}^{-t}, \mathbf{x}) - \tau] dy_t \geq 0 \iff \\ |\Pr(y_t \leq f_2 | \mathbf{x}) - \tau| &\geq |\Pr(y_t \leq f_1 | \mathbf{x}) - \tau| \end{aligned} \quad (\text{B10})$$

for any t .

[168] Finally, note that since q is linearly additive in I_t and since the bias, $|\Pr(y_t \leq f_2 | \mathbf{x}) - \tau|$ or $|\Pr(y_t \leq f_1 | \mathbf{x}) - \tau|$, is time invariant,

$$q_2 - q_1 \geq 0 \iff I_{t,2} - I_{t,1} \geq 0$$

for all t .

[169] This because $q_2 - q_1 \geq 0$ requires $I_{t,2} - I_{t,1} \geq 0$ for atleast one t . However, if $I_{t,2} - I_{t,1} \geq 0$ for atleast one t , $I_{t,2} - I_{t,1} \geq 0$ for all t since bias is time invariant.

[170] From (B9) and (B10), we thus have:

$$q_2 - q_1 \geq 0 \Rightarrow |\Pr(y_t \leq f_2 | \mathbf{x}) - \tau| \geq |\Pr(y_t \leq f_1 | \mathbf{x}) - \tau|.$$

B1. Integration by Parts

[171] Integration by parts:

$$\int u \frac{\partial v}{\partial x} dx = uv - \int v \frac{\partial u}{\partial x} dx$$

[172] Consider I_t , where $\hat{p}(y_t) = p(y_t | \mathbf{y}^{-t}, \mathbf{x})$ and $f = f(S_t; \mathbf{k})$:

$$I_t = \int_{\underline{y}}^{\bar{y}} (1 - \tau) |y_t - f|_{-} \hat{p}(y_t) dy_t + \int_{\underline{y}}^{\bar{y}} \tau |y_t - f|_{+} \hat{p}(y_t) dy_t = A + B$$

[173] When $y_t \leq f$, $B = 0$ and

$$A = \int_{\underline{y}}^f (1 - \tau) |y_t - f| \hat{p}(y_t) dy_t = (1 - \tau) \bar{A}.$$

[174] For $u = |y_t - f|$, $v = \hat{F}(y_t) = \int_{\underline{y}}^{y_t} \hat{p}(z) dz$, we have from integration by parts:

$$\bar{A} = [y_t - f | \hat{F}(y_t)]_{\underline{y}}^f - \int_{\underline{y}}^f \hat{F}(y_t) (-1) dy_t = \int_{\underline{y}}^f \hat{F}(y_t) dy_t$$

[175] Thus,

$$A = (1 - \tau) \int_{\underline{y}}^f \hat{F}(y_t) dy_t.$$

[176] When $y_t > f$, $A = 0$ and

$$B = \int_{\underline{y}}^{\bar{y}} \tau |y_t - f| \hat{p}(y_t) dy_t = \tau \bar{B}.$$

[177] For $u = |y_t - f|$, $v = \hat{F}(y_t) = \int_{\underline{y}}^{y_t} \hat{p}(z) dz$, we have from integration by parts:

$$\begin{aligned}\bar{B} &= [\bar{y}_t - f|\hat{F}]_{\underline{y}}^{\bar{y}} - \int_{\underline{y}}^{\bar{y}} \hat{F}(y_t)(+1)dy_t = [\bar{y} - f|\hat{F}(\bar{y}) - \int_{\underline{y}}^{\bar{y}} \hat{F}(y_t)dy_t \\ &= (\bar{y} - f) - \int_{\underline{y}}^{\bar{y}} \hat{F}(y_t)dy_t.\end{aligned}$$

[178] Thus,

$$B = \tau\bar{B} = \tau(\bar{y} - f) - \tau \int_{\underline{y}}^{\bar{y}} \hat{F}(y_t)dy_t.$$

[179] Finally,

$$\begin{aligned}I_t &= A + B \\ &= \int_{\underline{y}}^f \hat{F}(y_t)dy_t + \tau(\bar{y} - f) - \tau \int_{\underline{y}}^{\bar{y}} \hat{F}(y_t)dy_t \\ &= \int_{\underline{y}}^f \hat{F}(y_t)dy_t - \tau(f - \underline{y}) + \tau(\bar{y} - \underline{y}) - \tau \int_{\underline{y}}^{\bar{y}} \hat{F}(y_t)dy_t \\ &= \int_{\underline{y}}^f [\hat{F}(y_t) - \tau]dy_t + \tau \int_{\underline{y}}^{\bar{y}} (1 - \hat{F}(y_t))dy_t \\ &= \left\{ \int_{F^{-1}(\tau)}^f [\hat{F}(y_t) - \tau]dy_t \right\} \\ &\quad + \left\{ \int_{\underline{y}}^{F^{-1}(\tau)} [\hat{F}(y_t) - \tau]dy_t + \tau \int_{\underline{y}}^{\bar{y}} (1 - \hat{F}(y_t))dy_t \right\} \\ &= I_t^2 + I_t^1\end{aligned}$$

Appendix C: Model Structures Specifications

[180] Figure C1 illustrates the model structures used for the synthetic case study (section 2.3.2) and the real world case study (section 2.3.3). Both the studies use daily precipitation, potential evaporation and streamflow data set of the French Broad river basin, USA [Duan *et al.*, 2006] from 1970 to 1975.

[181] The linear reservoir without a threshold model structure (model structure 1) determines flow $Q(t)$ as a linear function of the reservoir storage $S(t)$, i.e., $Q(t) = S(t)/K_s$. The linear reservoir with a threshold model structure, i.e., model structure 2, has two flows: slow flow $Q_s = \min(S(t), S_{\max})/K_s$ and the fast flow $Q_f = \max(0, S(t) - S_{\max})/K_f$. Both the model structures assume that $E(t) = \min(P(t), E_p(t))$. Therefore they are effectively forced by $\max(0, P(t) - E_p(t))$.

[182] The model structure 3 is composed of reservoirs to model the unsaturated zone, the saturated zone and river routing. Precipitation $P(t)$ contributes to the unsaturated zone. Evaporation, $E(t)$, overland flow, $R(t)$, and percolation to the saturated zone are generated from the unsaturated zone as nonlinear functions of $S_u(t)/S_{u\max}$, where $S_u(t)$ is the storage in the unsaturated zone and $S_{u\max}$ is its stor-

age capacity. Evaporation and overland flow are modeled as:

$$\begin{aligned}E(t) &= E_p[1 - \exp(-\alpha_E S_u(t)/S_{u\max})]/[1 - \exp(-\alpha_E)] \\ R(t) &= P(t)[1 - \exp(-\alpha_F S_u(t)/S_{u\max})]/[1 - \exp(-\alpha_F)]\end{aligned}$$

where the parameters α_E and α_F are nonlinear controls and E_p is the potential rate of evaporation. Percolation ($Q_P(t)$) is linearly related to $S_u(t)/S_{u\max}$ as,

$$Q_P(t) = Q_{P\max} S_u(t)/S_{u\max}.$$

[183] The slow flow, $Q_s(t)$, is a linear function of saturated zone storage, $S_s(t)$,

$$Q_s(t) = S_s(t)/K_s$$

where K_s is the slow flow time constant.

[184] Finally, overland flow $R(t)$ and slow flow $Q_s(t)$ are routed through two (fast) linear reservoirs each with time constant K_f .

[185] Table C1 summarizes relevant model structure quantities.

[186] The synthetic case study (section 2.3.2) uses model structure 3 to generate synthetic streamflow data. The model structure 3 is forced by 1970–1975 daily precipitation and potential evapotranspiration. It assumes $E(t) = \min(P(t), E_p(t))$ and the following values for its parameters: $S_{u\max} = 10$ mm, $Q_{P\max} = 2$ mm/day, $\alpha_E = 100$, $\alpha_F = -15$, $\alpha_S = 1E-6$, $K_f = 4$ days, $K_S = 25$ days. Thus all the model structures in the synthetic case study (section 2.3.2) are forced by the same time series of $\max(0, P(t) - E_p(t))$.

[187] For the real world case study (section 2.3.3), model structure 3 with suppressed evaporation scheme, i.e., $E(t) = \min(P(t), E_p(t))$, is called structure 3 while model structure 3 without the suppressed evaporation scheme, i.e., $E(t) = E_p[1 - \exp(-\alpha_E S_u(t)/S_{u\max})]/[1 - \exp(-\alpha_E)]$, is called model structure 4. Thus the evaporation scheme distinguishes model structures 1 to 3 from the model structure 4 and serves as a major source of deficiency for model structures 1 to 3.

Appendix D: First-Order Approximation of Bias That Measures Model Structure Deficiency

[188] Let $F(\bar{y}) = \Pr(y \leq \bar{y})$, where conditioning variables have been suppressed.

[189] At the optimal, by (3.3a) and (3.5), the following equality holds for a τ -quantile predictor, y_τ^{pred} :

$$F(y_\tau^{pred}) - \tau + \lambda = 0.$$

[190] Here τ is the quantile to be modeled and λ is the bias in estimating the τ^{th} observed quantile. The inverse of F exists since $F(\omega)$ is differentiable w.r.t. ω from Assumption 6. It is also differentiable. By first order approximation, we have

$$F^{-1}(\tau - \lambda) \approx F^{-1}(\tau) - \lambda \frac{\partial F^{-1}(\tau)}{\partial \tau}.$$

[191] Thus, the bias in predicting the τ th quantile can be estimated as:

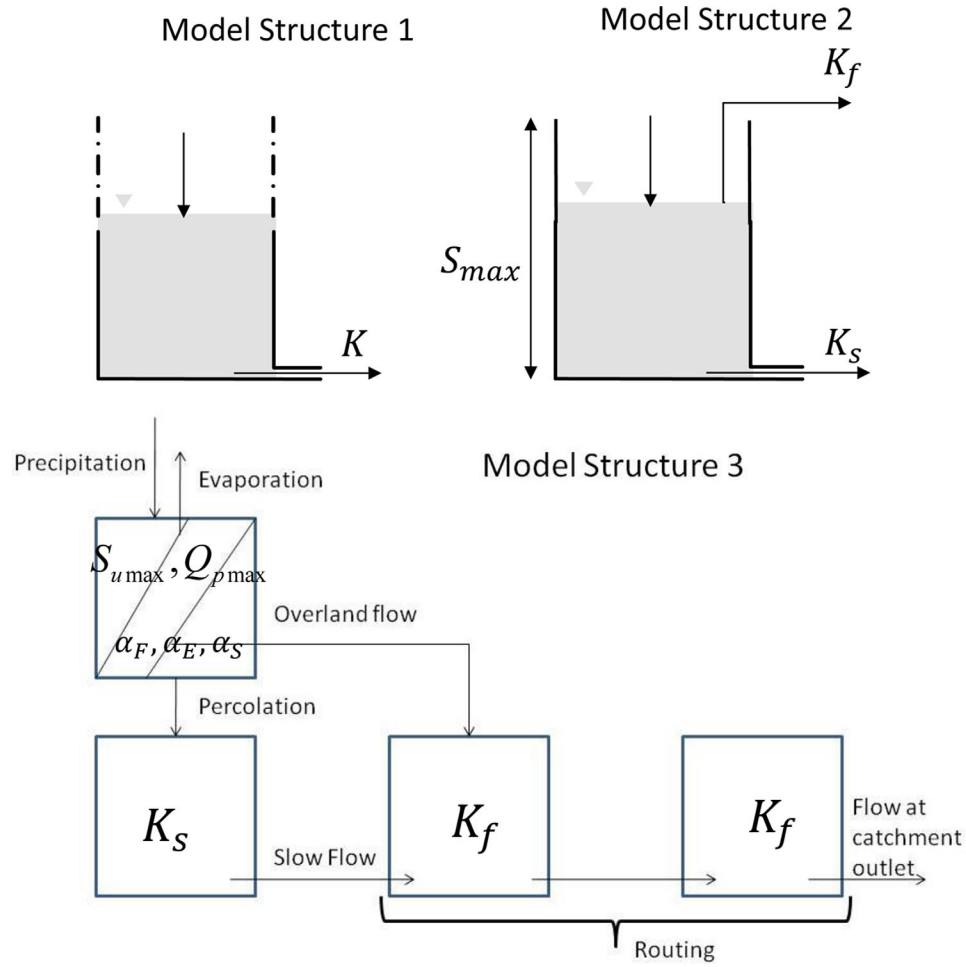


Figure C1. The three model structures considered in the synthetic case study and the French Broad river basin case study.

Table C1. Description of Parameters (to Estimate), Variables, Coefficients, and Indices Used in the Model Structures 1–3

Symbol (Units)	Description	Min	Max
<i>Model Structure 1 (Linear Reservoir Without a Threshold)</i>			
Parameters			
K_s (day)	Recession parameter	1	150
Variables			
Q (mm/d)	Flow		
<i>Model Structure 2 (Linear Reservoir With a Threshold)</i>			
Parameters			
K_s (day)	Slow flow recession parameter	1	150
K_f (day)	overthreshold recession parameter	1	10
S_{max} (mm)	Storage capacity (threshold)	0	1000
Variables			
Q_f (mm/d)	Fast flow		
Q_s (mm/d)	Slow flow		
<i>Model Structure 3</i>			
Parameters			
$S_{u \max}$ (mm)	Top layer/unsaturated zone moisture parameter	0	1000
$Q_{p \max}$ (mm/d)	Maximum percolation rate	0	100
α_E (—)	Curvature parameter for evaporation	0	100
α_F (—)	Curvature parameter for overland flow	–100	0
α_S (—)	Curvature parameter for percolation	–10	10
K_s (day)	Baseflow time constant	1	150
K_f (day)	Routing time constant	1	10
Variables			
$S_u(t)$ (mm)	Upper layer/unsaturated zone soil water storage		
$S_s(t)$ (mm)	Lower layer/saturated zone soil water storage		
$E(t)$ (mm/d)	Evaporation		
$R(t)$ (mm/month)	Overland flow		
$Q_p(t)$ (mm/d)	Percolation		
$Q_s(t)$ (mm/d)	Baseflow		
Others			
$P(t)$ (mm/d)	Precipitation		
$E_p(t)$ (mm/d)	Potential evapotranspiration		
t	day index, $\{1, \dots, T\}$		

$$\lambda \approx - \frac{[F^{-1}(\tau - \lambda) - F^{-1}(\tau)]}{\frac{\partial F^{-1}(\tau)}{\partial \tau}}.$$

[192] The quantity $F^{-1}(\tau)$ is estimated as the τ th quantile of the observed time series, $F^{-1}(\tau - \lambda)$ is estimated by taking the average of τ th quantile “prediction” at the indices that correspond to $F^{-1}(\tau)$ on the observed time series. $\frac{\partial F^{-1}(\tau)}{\partial \tau}$ is numerically estimated on the observed time series.

Appendix E: Bayesian Criteria Used

[193] Three Bayesian criteria are used to approximate the marginal log likelihood of a model structure.

[194] 1. Bayesian Information Criteria (BIC) [Kass and Raftery, 1995]:

$$\log m(y|M) \approx \log f(y|M, \theta^*) - 0.5 \log(N) |\theta|.$$

[195] 2. Harmonic mean of the log-likelihood values of the posterior distribution (HM1) [Kass and Raftery, 1995]:

$$\log m(y|M) \approx \frac{1}{m} \left(\sum_{i=1}^m (\log f(y|M, \theta^*))^{-1} \right)^{-1}.$$

[196] 3. A variant of Chib and Jeliazkov [2001] (HM2):

$$\log m(y|M) \approx \log f(y|M, \theta^*) + \log \pi(\theta^*|M) - \log \pi(\theta^*|y, M).$$

[197] Here $m(y|M)$ is the marginal likelihood that data y are from a model structure M , $f(y|M, \theta^*)$ is the likelihood that the data y are from a model that is from a structure M and parameterized by θ^* , θ^* represents the maximum likelihood parameter estimate (MLE) for a given model structure M , $|\theta^*|$ is the dimensionality of the parameter set, $\pi(\theta^*|M)$ is the prior probability of the MLE θ^* , $\pi(\theta^*|y, M)$ is the posterior probability of θ^* , N is the sample size and m is the size of parameter sets sampled from the posterior distribution $\pi(\theta^*|y, M)$. The General Likelihood function is used (see section 2.3.1) for $f(y|M, \theta)$.

[198] For the Bayesian criteria HM2, $\pi(\theta|y, M)$ is non-parametrically estimated using multivariate kernel density estimation. For the case studies $N=2192$ days and $m=600$.

[199] **Acknowledgments.** The author is grateful to Michiel A. Keyzer for his critical review and suggestions, to Gerrit Schoups for providing the MATLAB code for the Generalized Likelihood function and several discussions on the applications and to Mojtaba Shafiei, Huub Savenije and Ashvani K. Gosain for discussions on an early version of the manuscript. Thanks are due to several referees including Nataliya Bulygina and Jasper A. Vrugt for their critical review of the manuscript. The author also thanks the AE and the Editor for their patience with previous versions of the manuscript.

References

- Avriel, M. (1976). *Nonlinear Programming*, Engelwood Cliffs, Prentice Hall, N. J.
- Bernardo, J. M., and A. F. M. Smith, (1994). *Bayesian Theory*. Chichester: Wiley.
- Barron, A. R. (1988), The exponential convergence of posterior probabilities with implications for Bayes estimators of density functions, Tech. Rep. 7, Dep. of Stat., Univ. of Ill. at Urbana-Champaign, Champaign, Ill.
- Berk, R. H. (1966), Limiting behavior of posterior distributions when the model is incorrect, *Ann. Math. Stat.*, 37, 51–58.

- Bertsekas, D. P., and A. E. Ozdaglar (2000), Psuedonormality and a Lagrange multiplier theory for constrained optimization, *J. Optim. Theory Appl.*, 114(2), 287–343.
- Beven, K. J., P. J. Smith, and J. Freer (2008), So just why would a modeller choose to be incoherent?, *J. Hydrol.*, 354(1), 15–32.
- Brooks, S. P. (1998), Markov Chain Monte Carlo Method and its applications, *J. R. Stat. Soc., Ser. D*, 47(1), 69–100.
- Chib, S., and I. Jeliazkov (2001), Marginal likelihood from the Metropolis-Hastings output, *J. Am. Stat. Assoc.*, 96(543), 270–281.
- Davidson, R., and J. G. MacKinnon (2004), *Econometric Theory and Methods*, pp. xviii + 750, Oxford Univ. Press, New York.
- DiCiccio, T. J., and A. C. Monti (2004), Inferential aspects of the skew exponential power distribution, *J. Am. Stat. Assoc.*, 99(466), 439–450.
- Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28(4), 1015–1031, doi:10.1029/91WR02985.
- Duan, Q., et al. (2006), The Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, *J. Hydrol.*, 320, 3–17, doi:10.1016/j.jhydrol.2005.07.031.
- Feldman, M. (1991), On the generic non-convergence of Bayesian actions and beliefs, *Econ. Theory*, 1, 301–321.
- Fisher, R. A. (1922), On the mathematical foundations of theoretical statistics, *Philos. Trans. R. Soc. London A*, 222, 309–368.
- Freedman, D. A. (1963), On the asymptotic behavior of Bayes estimates in the discrete case, *Ann. Math. Stat.*, 34, 1386–1403.
- Freedman, D. A. (1965), On the asymptotic behavior of Bayes estimates in the discrete case II, *Ann. Math. Stat.*, 36(2), 454–456.
- Gelman, A., and D. A. Rubin, (1992), Inference from iterative simulations using multiple sequences, *Stat. Sci.*, 7(4), 457–511.
- Gilboa, I. (1987), Expected utility with purely subjective non-additive probabilities, *J. Math. Econ.*, 16, 65–88.
- Givens, G. H., and A. E. Raftery (1996), Local adaptive importance sampling for multivariate densities with strong nonlinear relationships, *J. Am. Stat. Assoc.*, 91(433), 132–141.
- Haerdle, W., et al. (2004), *Nonparametric and Semi-Parametric Methods*, pp. 305, Springer, Berlin.
- Heckman, J. J. (2005), The scientific model of causality, *Sociol. Methodol.*, 35, 1–97, doi:10.1111/j.0081-1750.2006.00164.x.
- Kass, R. E., and A. E. Raftery (1995), Bayes factors, *J. Am. Stat. Assoc.*, 90, No. 430, pp. 773–795.
- Kavetski, D., G. Kuczera, and S. W. Franks (2006), Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, *Water Resour. Res.*, 42, W03407, doi:10.1029/2005WR004368.
- Keyzer, M. A., and S. Pande (2009), Instrumentalization using quantiles in semiparametric support vector regression, WP 09-04, 61 pp., Cent. for World Food Stud., Amsterdam.
- Koenker, R., and G. Basset (1978), Regression quantiles, *Econometrica*, 46(1), 33–50.
- Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *J. Hydrol.*, 211, 69–85.
- Lyon, K., and S. Pande (2006), The costate variable in a stochastic renewable resource model, *Natural Resource Modeling*, 19(1), 45–66.
- Marshall, L., A. Sharma, and D. J. Nott (2006), Modelling the catchment via mixtures: Issues of model specification and validation, *Water Resour. Res.*, 42, W11409, doi:10.1029/2005WR004613.
- Mangasarian, O. L., and S. Fromovitz (1967), The Fritz John necessary optimality conditions in the presence of equality and inequality constraints, *J. Math. Anal. Appl.*, 17, 37–47.
- McMillan, H., T. Krueger, and J. Freer (2012), Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality, *Hydrol. Processes*, 26, 4078–4111, doi:10.1002/hyp.9384.
- Mordukhovich, B. S. (1976), Existence of optimal controls, translated from Itogi i Nauki i Tekhniki, *Sovremennye Probl. Math.*, 6, 207–261.
- Pande, S., M. McKee, and L. A. Bastidas (2009), Complexity-based robust hydrologic prediction, *Water Resour. Res.*, 45, W10406, doi:10.1029/2008WR007524.
- Pande, S., H. H. G. Savenije, L. A. Bastidas, A. K. Gosain, (2010), A parsimonious modeling approach to water management in dryland areas, in *Hydrocomplexity: New Tools for Solving Wicked Water Problems*, vol. 338, edited by S. Khan, H. H. G. Savenije, S. Demuth, and P. Hubert, IAHS Press, Wallingford, U. K.

- Pande, S., et al. (2011), Water valuation at basin scale with application to western India, *Ecol. Econ.*, 70, 2416–2428, doi:10.1016/j.ecolecon.2011.07.025.
- Pande, S., H. H. G. Savenije, L. A. Bastidas, and A. K. Gosain (2012a), A parsimonious hydrological model for a data scarce dryland region, *Water Resour. Manage.*, 24, 909–926, doi:10.1007/s11269-011-9816-z.
- Pande, S., L. A. Bastidas, S. Bhulai, and M. McKee (2012b), Parameter dependent convergence bounds and complexity measure for a class of conceptual hydrological models, *J. Hydroinformatics*, 14(2), 443–463.
- Pande, S. (2013), Quantile hydrologic model selection and model structure deficiency assessment 2: Applications. *Water Resources Research*, DOI: 10.1002/wrcr.20422.
- Pasarica, C., and A. Gelman (2010), Adaptively scaling the Metropolis Algorithm using expected squared jumped distance, *Stat. Sinica*, 20, 343–364.
- 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.
- Roberts, G. O., and A. F. M. Smith (1994), Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms, *Stochastic Processes Appl.*, 49, 207–216.
- 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.
- Smith, T., A. Sharma, L. Marshall, R. Mehrotra, and S. Sisson (2010), Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments, *Water Resour. Res.*, 46, W12551, doi:10.1029/2010WR009514.
- 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.
- von Neumann, J., and O. Morgenstern (1947), *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press, Second Edition.
- Vapnik, V. (2002), *The Nature of Statistical Learning Theory*, 2nd ed., Springer, New York.
- White, H. (1982), Maximum likelihood estimation of misspecified models, *Econometrica*, 50, 1–25.