Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-406-RC1, 2018

Hydrology and Earth System Sciences Discussions

Interactive comment on "A framework for likelihood functions of deterministic hydrological models" *by* Lorenz Ammann et al.

J. Vrugt (Referee)

jasper@uci.edu

Received and published: 28 September 2018

Review of HESS-submission "A framework for likelihood functions of deterministic hydrological models" by Ammann et al.

Summary: In this paper the authors introduce a parametric framework to residual analysis. This approach leads to formulation of a likelihood function which, with a suitable prior distribution, helps to evaluate the posterior density of nontraditional residual time series, e.g. truncated and subject to various degrees of skew, kurtosis and serial correlation. The framework allows for the use of transient nuisance variables (hyper parameters) to help accommodate so-called non-stationary residual patterns. The framework presented herein differs a bit from the standard likelihood paradigm in that the starting

point is some parametric family of distributions which describes the likelihood of observing the data, Q, given current model output, Qdet. Authors claim that the proposed likelihood function improves probabilistic inference of hydrologic models via MCMC machinery – with a more realistic description of parameter and predictive uncertainty. I enjoyed reading this paper as it combines theory development with practical application. The paper is well written and should be of interest to the readership of HESS. I hope the authors consider the following comments – I believe those will help to further improve the quality of this manuscript. Note, comments appear in order of my reading of the paper.

1. Page 5, Line 9-11. Authors state that most (many) modelers will have an intuitive idea about the probability distributions of the observations for a given model output. I disagree with this assertion. For the sake of my argument, lets follow the hydrologic example as presented in this work. Let's assume that the model simulates a discharge of 20 mm/day. What would be a reasonable expectation of the actual (observed) discharge at that time? 15? 30? I cannot confidently claim that I would know what probability distribution to assume for the observed discharge at that time. Of course, if 20 mm/day is among the largest simulated values, then I would generally expect the dispersion of this supposed distribution to be larger than for a simulated value of 5 mm/day. Yet, this is only the dispersion – I would not really have an idea about the underlying distribution – would I center this distribution on 20 mm/day? Or is my model systematically under or overestimating the data so that I should shift the distribution to higher or lower values, respectively. Of course, for low discharge values I know that the distribution is truncated at zero – and probably has a tail to the right. But then again do I center the distribution on the model simulated value? Or do we shift it up or downward? In other words, I do not agree with the assertion that many modelers will have an intuitive idea what the distribution of the observed discharge would be if the model output is known.

2. Page 5, Line 22-23. The authors refer to Eq. (3) before presenting Eq. (2). Do

not understand why this is done – would think that text can be presented so that Eq. (3) follows first – then followed by Eq. (2). Note, is Eq. (3) needed after all? The right-hand-side of Eq. (3) can be placed at end of Eq. (2) – then the index needs to be fixed.

3. Page 5, Line 27-29: I do not understand the statement that truncation at zero would lead to lighter tails on the lower end. Yes, truncation would move the probability of negative streamflow values to streamflow values larger than zero. In essence, one could then argue that the tail at the right-hand-side may become larger – as the pdf has to integrate to unity. Yet, because of truncation the left tail is essentially gone if simulated streamflow values are close to zero. The wording "lighter tails" may be a bit confusing as the tail is truncated. It is no longer there.

4. Page 5, Eq. (2) – (3) – thus, eta is the normally transformed counterpart of Q – with truncation accounted for?

5. Equation (4) – authors may consider for normal distribution, N, instead \mathcal(N)(a,b), where "a" (mean) is the first term between brackets in Eq. (5) and "b" is the second term in Eq. (4). In text below Eq. (4) authors could then explain that "a" is the mean of the distribution and b is the variance.

6. Eq. (6) – reference should be given.

7. Page 6, Line 12-14. Maybe I am missing something here, but with any other likelihood function one can ignore missing data as well? One simply does not include this particular observation in the likelihood function. The authors may have a point if serial correlation is considered – then this removal is not straightforward as it breaks the AR-error model.

8. Eq. (7) – top line of curly brace may fit on one line if authors define rho = (ti+1 – ti)/tau, and then use rho in the equation – maybe etatrans written as etaT.

9. Then notation – not sure about the guidelines of HESS, but should theta (param-

eter vector) not be upright-bold instead of italic-bold? Same holds for the nuisance variables, psi.

10. Is notation DQ required or would fQ suffice instead? Then, the text would talk about a distribution of Q – instead of DQ.

11. A limitation of Eq. (4) is that serial correlation at higher-order lags cannot be modelled, right? Unless you specify different "rho's" in Eq. (6) – but this then leads to multiple likelihoods. This limitation should be stated in the text as residuals may exhibit/show residual correlation beyond lag-1.

12. In Eq. (8) how do we compute the first term on the right-hand-side – that is – the likelihood of the zeroth discharge observation (at t0)? Do we assume normality with dispersion of variance/(1-rhoˆ2)?

13. Page 7, Line 12-13: The statement "the likelihood function can be evaluated analytically" is a bit confusing to me. What does the word "analytical" mean in this context? Most other commonly used likelihood functions in the applied (hydrologic) literature are simple to evaluate in practice, right? That means numerically. All that is needed are the model output and the data? What is different in the present context?

14. The authors use the affine invariant ensemble sampler of Foreman and Mackay et al. (2013) to sample the posterior parameter and nuisance variable distribution. The article would benefit from some more background information – that is – algorithmic settings (number of walkers, the types of moves that are considered, etc.). Note, that this ensemble sampler has many elements in common with the DREAM family of MCMC algorithms – which uses parallel direction and snooker moves. For later work it may be interesting to compare both methods in terms of efficiency – and to evaluate the power and usefulness of the walk, stretch and replacement move. Note, that the ensemble sampler has two important shortcomings; 1) detailed balance requires the use of a relatively large number of walkers (chains) – this is a significant disadvantage for higher dimensional problems as each chain needs burn-in before reaching the

target distribution, and 2) the walkers require stepwise updating – this guarantees reversibility but does not make the sampler amenable to distributed computing, wherein each chain is evolved on a different core/node.

15. Equation (10) – the subscript "F" in the flashiness index, should this not be regular font – that is – upright? As "F" is an abbreviation for "flashiness" and not a variable. Same holds for some of the other summary metrics used in this paper, for example the Nash-Sutcliffe efficiency (subscript "N" should be regular = upright font). Note, that on Page, 8, Line 25 correct notation is used for the flashiness index of the deterministic model output.

16. Page 5, Line 24: "maximum posterior parameter values" – this is rather awkward wording as it literally means – the largest posterior parameter values. And it is not clear what this means either as each dimension of the target distribution will have a maximum posterior value – but all these maxima combined are unlikely to make up an actual posterior sample. Instead, what the authors should use is "maximum a-posteriori density (MAP) parameter values" – that is – the parameter values that maximize the product of the prior density and the likelihood.

17. Eq. (15) and (16) list the flux and water balance equations used by the hydrologic model – but equally important what numerical solution method is used to solve these equations? I assume that the authors have used an implicit solution with time-variable integration step? Solution maintains mass balance?

18. Page 12, Line 5: Why are these model parameters held constant? Why are they not part of the inference – this would be much stronger in my view. If held constant, then how does one know the assumed values are reasonable for the catchment of interest? Note, if I look at the equations then m, alpha and beta must have a large impact on the simulated model output. Hence, unless these parameters have a strong physical underpinning I do not see why one would keep them fixed in the present work. Certainly, the values of m, alpha and beta will affect the residual analysis.

19. The authors do not consider highly relevant work by Scharnagl et al. (2015) published in HESS: Inverse modeling of in situ soil water dynamics: accounting for heteroscedastic, autocorrelated, and non-Gaussian distributed residuals. This work also used a Student distribution for the conditional density of the residuals – and combined this with the template function of Fernandez and Steel (1998) to enable treatment of skewed residual distributions. Given the similarities with the work presented in this paper I think it is important for the authors to consider the listed work of Scharnagl et al.

20. Eq. (18) – does this function satisfy the laws of total expectation and total variance? This is a concern not typically addressed in the hydrologic literature – but the paper by Hernandez-Lopez in HESS (2017) makes some important points regarding preservation of expectation and variance of the error model.

21. I am wondering whether readability of the paper would improve if the section on error models is placed directly after the likelihood section. Indeed, the likelihood contains tau – which is then defined (among others) in the error model section.

22. Page 11, Line 16: What has happened to the index time in the formulation of Qdet? It appears on the left-hand side but does not appear on the right-hand side. Also, what are Qs and Qf? These entities are introduced but they are not discussed nor do they appear elsewhere in the paper?

23. At this point I am wondering why the authors are not using the more common terminology of P(.) for prior distribution and L(.|.) for likelihood function.

24. Figure 6 – the values of eta show a strong temporal correlation for error model E2 and E3. Would it be possible to plot, in some way, the decorrelated eta values (with serial correlation removed).

25. In general, it may be useful if the authors include a plot of the marginal posterior distributions of the model parameters and nuisance variables. As it stands it is difficult

to determine which parameters are well defined and which variables are not well defined by inference against the measured data (for one or more error models). In fact, the authors could compute the KL divergence of the prior and posterior distributions for each error model. In any case, it would be good to have insights on how well the parameters and nuisance variables are defined. Do their posterior distributions extend over the entire prior ranges, or are they limit to a small region inside the prior distribution? Note, Figure 6 goes a long way but is difficult to interpret as the matrix plot is rather small and the x-ranges are scaled according to the posterior uncertainty.

26. Figures 3 and 4: I find these results a bit difficult to interpret. The color/symbol coding is not necessarily clear – making it difficult to interpret the findings. I am sure the authors can find a way of plotting from which the main results are directly visible. Then, again, other readers may like to digest this plot.

27. Figure 5: Difficult to see the differences between the three panels. Would it be possible to enlarge the horizontal length of each of the subplots? Right now, the measured data interacts too much with the grey region, particularly when the posterior prediction/simulation uncertainty is small.

28. Note, the authors use the wording “prediction” – one could argue though that what is presented are simulations as the rainfall for the next is assumed known when simulating streamflow values.

29. Page 24, Line 9 – 12: Is this not due in large part because of ignoring the laws of total expectation and total variance? Per my previous comment on this topic.

30. I think a weakness of this paper is that the authors do not compare their findings against another likelihood function. In the introduction section, the authors discuss strength and limitations of previously used/developed likelihood functions – they use this as justification for their own approach. Yet, my own practical experience suggests that a simple AR-1 likelihood would already do quite a reasonable job. This likelihood is easy to include in the present paper. What is more, the authors should consider

the generalized likelihood function – it is argued that this likelihood has a limitation because of the treatment of serial correlation on non-standardized residuals – this is easy to remedy in practice. Then, the argument of analytic tractability I do not really follow (Page 3, Line 22).

31. Would the inference not lead to more realistic results if the authors augment their likelihood with an error model for the rainfall data? This would carry another set of nuisance variables / hyper parameters (depending in large part on the choice of rainfall prior) but make the inference more robust.

32. Just a thought – but is nonstationary the right wording in the present application of the likelihood function? If tau does vary between rainfall and dry periods – but these two values of tau repeat themselves in the future (e.g. are constant) – then one may argue that overall the residual time series is a stationary time series. Tau just differs between rainfall and non-rainfall days.

33. Overall, I think the author should better recognize the highly related work of Scharnagl (2015) published in the same journal (HESS). Indeed, this paper used the Student distribution with the Fernandez and Steel template function for skew.

I hope these comments are useful to further improve the paper, Jasper Vrugt
jasper@uci.edu

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2018-406, 2018.