Hydrol. Earth Syst. Sci. Discuss., 8, C4502–C4509, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C4502/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Evaluating uncertainty estimates in hydrologic models: borrowing measures from the forecast verification community" *by* K. J. Franz and T. S. Hogue

K. J. Franz and T. S. Hogue

kfranz@iastate.edu

Received and published: 19 October 2011

We would like to thank this reviewer for the helpful comments and suggestions. These comments have helped us address significant deficiencies in the original submission.

Reviewer comment: 1. First of all, while reading the introduction, the aim of the paper seems clear: review and demonstrate methods to evaluate probabilistic forecasts. From the results in section 3 it seems that you want to compare and evaluate GLUE, modified GLUE (essentially the same as GLUE) and SCEM, which means comparing uncertainty estimation methods. This doesn't get across at all in the introduction.

REPLY: We recognize that the objectives and message of the paper were not made sufficiently clear in the first submission. Our main objective was to illustrate that forecast verification methods can be applied to evaluated probabilistic model simulations. We believe that the best way to demonstrate this is through a comparison of different parameter estimation methods. However, by presenting different methods, it is difficult not to make comparative statements as the results are discussed. As a result our results and discussion was not in-line with our introduction.

We removed the modified GLUE method from the paper to take emphasis off the parameter uncertainty estimation methods and have refocused the results section to focus on the verification methods. We also removed the discussion and thereby removed several of the evaluation statements regarding the uncertainty estimation methods and the redundancy.

The final paragraph of the revised introduction states our objectives for this paper, it is as follows: The focus of the current study is to provide a succinct overview of a range of available probabilistic verification measures and to demonstrate their application in evaluating and distinguishing model ensemble performance. We utilize two commonly applied parameter estimation methods (Generalized Uncertainty Likelihood Estimator (GLUE; Beven and Binley, 1992) and the Shuffled Complex Evolution Metropolis (SCEM; Vrugt et al., 2003) and an operational rainfall-runoff model (Sacramento Soil Moisture Accounting Model (SAC-SMA; Burnash et al., 1973) for demonstration purposes. We evaluate the uncertainty associated with model ensembles propagated through parameter estimates, although the metrics presented here are readily transferable to evaluate model performance from other probabilistic systems. We are not undertaking explicit evaluation of the "best" parameter estimation method being used, but rather highlighting how the applied metrics can help better inform users on model performance and behavior when different results (ensemble hydrographs) are apparent. We also highlight unique challenges in applying probabilistic verification to hydrologic model ensembles and provide initial guidance on those measures which may be most suitable to the hydrologic community.

Reviewer Comment: 2. In the introduction, a number of methods to probabilistic stream flow prediction are mentioned (ESP, ensemble DA, multi-models), however, one important type of methods is missing, which is particularly useful for practical applications because it is often computationally very efficient: stochastic post-processing methods, such as the Hydrological Uncertainty Processor (Krzysztofowicz and Kelly, 2000), the meta-gaussian approach (Montanari and Brath, 2004), Quantile Regression (Weerts et al., 2011) and the Model Conditional Processor (Coccia and Todini, 2010). These methods should be added to the introduction.

REPLY: We have added post-processing and these references to the introduction. The third paragraph in the revised introduction includes the following sentence: The recent growth of probabilistic streamflow estimates in hydrologic modeling, including ensemble data assimilation methods (Kitanidis and Bras, 1980a, 1980b; Evensen, 1994; Margulis et al., 2002; Seo et al., 2003, 2009), multi-modeling platforms (Ajami et al., 2007; Duan et al., 2007; Vrugt and Robinson, 2007; Franz et al., 2010), Extended Streamflow Prediction (ESP) and other probabilistic forecasting systems (Day, 1985; Krzysztofowicz, 2001; Faber and Stedinger, 2001; Franz et al., 2003; Bradley et al., 2004; Franz et al., 2008; Thirel et al., 2008) and post-processing techniques (Krzysztofowicz and Kelly, 2000; Montanari and Brath, 2004; Coccia and Todini, 2010; Weerts et al., 2011) warrants greater integration of probabilistic model evaluation into the hydrologic community.

Reviewer Comment: 3. The paper claims to demonstrate probabilistic forecast evaluation methods, yet no hindcasts (with e.g. increasing lead times and uncertainties due to state uncertainties, NWP uncertainties etcetera) are used. This makes this claim weak. I suggest to alter this to demonstration of evaluation methods for stochastic simulations.

REPLY: The goal of this paper was to show that metrics used for probabilistic forecast

evaluation can also be used to evaluate model uncertainty estimates. Generating and verifying hindcasts from the parameter uncertainty methods was not our objective. We have restructured the introduction to make our objectives more apparent.

Reviewer Comment: 4. the discussion merely summarizes results again (in a lengthy manner). There are considerable differences found between results using GLUE or SCEM and I would like to see a discussion of these differences and the reasons for these differences rather than the summary of results here.

REPLY: We have removed discussion section and moved all relevant discussion of the verification metrics to the results section. Comparing the GLUE and SCEM was not our main objective. We tried to limit comparison between the two methods except where it facilitated discussion of the verification metrics.

Reviewer Comment: 5. The equations throughout the paper are in one word sloppy. They are often incorrect and abbreviations are used within the equations rather than proper symbols. Symbols are also inconsistently used throughout the text. Some suggestions are mentioned in the remainder of this review.

REPLY: We worked with a statistician at Iowa State to review the notation in our Methods section. We have fixed errors and inconsistencies in the equations throughout Section 2.4

Reviewer Comment: 6. The paper seems lengthy to me. Section 2 can be shortened and description of p. 3089, l. 3, GLUE is the Generalised Likelihood Uncertainty Estimation. l.5, ": : :over a selected set of basins: : :"

REPLY: We have streamlined the paper by removing the modified GLUE method, reduced the discussion of the methods in Section 2, and integrated the discussion and results sections. We have reduced the length of the paper by 25%.

Reviewer Comment: p. 3089, it would help the reader to have a brief overview of the structure of the paper here.

REPLY: We now provide a brief overview of the remaining paper at the last section of the introduction which reads: The study sites, model, parameter estimation methods and verification metrics are presented in Section 2.0. Results from the application of the verification metrics are discussed in Section 3.0. Concluding statements are provided in Section 4.0.

Reviewer Comment: The 'symbols' in the equations are more like abbreviations. Please select proper symbols e.g. LNSE (likelihood NSE), LRMSE etc.

REPLY: We note that we are following previous literature on nomenclature as related to objective functions and do not name our error functions as likelihood functions. For the NSE – we utilize this function twice in the paper – once as a likelihood function for behavioral parameter sets in the GLUE method and later as an objective function in the evalution section. In the latter case, it is more common in the calibration literature to present objective functions without the term "likelihood". In the earlier case, we now note in the paper that the NSE is used as a likelihood function in the GLUE method.

Reviewer Comment: p. 3091. Equation (1) is wrong and should be: LNSE = 1 ôĂĂĂ Pn Pt=1 (xt ôĂĂĂ ot) n t=1 (ot ôĂĂĂ ot) (1) 't' should be formatted as subscripts in all equations. Equation (3) is also incorrect and should be Lbias = Pn i=1 P(xt ôĂĂĂ ot) n i=1 ot (2)

REPLY: Thank you for pointing out these errors. They were incorrectly rewritten during typesetting, we will review the PDFs more carefully in the future before they are sent to reviewers. See equations 5 and 7 in the attached PDF.

Reviewer Comment: p. 3092. l. 4-6. It seems to me that 2 thresholds are being used (LNSE > 0:30 and the 90% behavioural interval, which is not required for GLUE. Why has this been done? W-GLUE is to my mind just GLUE. A different criterium has been applied but the method is still exactly GLUE. I disagree with the use of a new term for this method.

REPLY: We have removed the weighted GLUE (w-GLUE) method to better focus the efforts of our paper. We utlize 90% prediction bounds in both SCEM and GLUE, following the established SCEM approach, in order to remove extremes in both cases and allow for a reasonable comparison between the two methods.

Reviewer Comment: p. 3095, the equations: IQR, MAD and Range are abbreviations and words and should not be used as such in equations (mathematicians would read IQR as I Q R). Please use proper symbols. Eq. 7: Pn i=1 xi (t) ôĂĂĂ q0:50 (t) should be Pn i=1 jxi (t) ôĂĂĂ q0:50 (t) j

REPLY: We recognize that specific fields have other historical terminology. Using abbreviations of these terms, and other equations, is commonly in hydrologic sciences community. IQR and MAD are presented in this manner in our reference Wilks, 2006.

We have fixed equation 7 (now equation 2, see attached PDF).

Reviewer Comment: p. 3095, l. 11, The symbol N should be n. The confusion is because it is not consistent with the annotation in eqs. (1)-(3).

REPLY: We have removed the use of symbol N and fixed inconsistencies in the notation.

Reviewer Comment: p. 3096, l. 20. are used to assess the accuracy of the ensemble mean, if I'm not mistaken.

REPLY: In this statement we are referring to our study specifically. We chose to evaluate the ensemble median.

Reviewer Comment: p. 3111, l. 17-25. Nothing is demonstrated here so either move this statement to section 2 or do not mention it at all.

REPLY: This paragraph was removed.

Reviewer Comment: Fig. 2. Please discuss the consierable differences in the discussion section

REPLY: We have made significant efforts to refocus the paper as a discussion of the use of the evaluation metrics for characterizing the ensembles. Discussing the differences between the two methods beyond what was necessary to present the evaluation metrics would be out of scope for the revised paper.

Reviewer Comment: Fig 5. Why are these CDF's on a double-log scale? This distorts the results very much in disfavour of the low flows.

REPLY: The CDF's were plotted on a log-scale to display the results for the low flows more clearly (i.e. on an arithmetic scale the lines are so close as to be indiscernible). However, as this means of presentation can be confusing, we have changed the scale to arithmetic. See the revised Figure in the attached PDF (now Figure 4).

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 3085, 2011.

$$\overline{MAD} = \frac{1}{n}\sum_{t=1}^{n} median_i \left| x_i(t) - x_{med}(t) \right| \quad (2)$$

$$NSE = 1 - \left(\sum_{t=1}^{n} \left(x_{med}(t) - Q_{obs}(t) \right)^2 \Big/ \sum_{t=1}^{n} \left(Q_{obs}(t) - \overline{Q_{obs}(t)} \right)^2 \right) \quad (5)$$

$$Pbias = \left[\sum_{t=1}^{n} \left(x_{med}(t) - Q_{obs}(t) \right) \Big/ \sum_{t=1}^{n} Q_{obs}(t) \right] \cdot 100\% \quad (7)$$

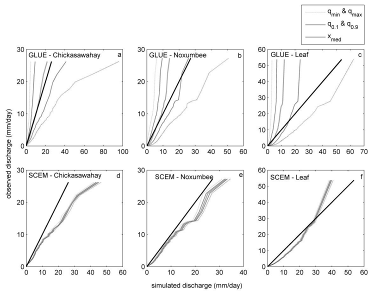

Figure 4: The joint distribution of the lowest (q_{min}), highest (q_{max}), 10th ($q_{0.1}$), and 90th ($q_{0.9}$) quantiles, and the median (x_{med}, or 50th quantile) of the discharge ensembles and the observations from the (a-c) GLUE and (d-f) SCEM parameter estimation methods for select sites. The solid black line in the figures is the 1:1 line and indicates perfect correlation between the simulated and observed discharge.

Fig. 1.