Hydrol. Earth Syst. Sci. Discuss., 12, C831–C841, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C831/2015/


Hydrology and Earth System Sciences
Discussions

# *Interactive comment on* "Inverse modelling of in situ soil water dynamics: accounting for heteroscedastic, autocorrelated, and non-Gaussian distributed residuals" *by* B. Scharnagl et al.

**Th. Wöhling (Referee)**

thomas.woehling@uni-tuebingen.de

Received and published: 1 April 2015

Summary

In this study, Bayesian inference is applied to estimate in-situ soil hydraulic properties using field observations of soil moisture and MCMC simulation. Three different error residual models were implemented in the likelihood function and posterior parameter distributions resulting from corresponding MCMC runs were analysed with respect to prediction quality, parameter uncertainty, and the fit to multivariate Gaussian priors.



The authors demonstrate the importance of an adequate error model choice and its effect on posterior parameter estimates.

Major Comments

1. Scope and novelty: The topic of the paper is well within the scope of HESS and interesting beyond the application presented here. In a previous paper by the authors (Scharnagel et al. 2011), which used the same data set, the influence of the prior information on posterior parameter distributions was investigated. As a follow-up, this paper addresses the impact of residual error models. In a paper very similar to the present study, Schoups & Vrugt (2010) investigated "a formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic and non-Gaussian errors". Wöhling and Vrugt (2011) applied the Bayesian framework and MCMC simulation to infer soil hydraulic properties from a comprehensive field experiment and tested two residual error models, including the autoregressive AR(1) model. Both the aforementioned studies used uniform priors. Despite the progress being made in these (and related) studies, the present manuscript provides two novel aspects that are interesting beyond the presented case study: 1) How does the setup of the parameter inference scheme influence robust and parsimonious posterior parameter distributions that are in agreement with prior belief? 2) Why does the autoregressive AR(1) model inflate uncertainty when used in the likelihood function although time series of observations used for calibration are correlated?

2. Methodology: The Bayesian framework used here is well suited to investigate these questions. The techniques for the implementation of prior parameter knowledge are state-of-the-art. As a benchmark, a MCMC run was conducted using a commonly applied likelihood function that assumes Gaussian, homoscedastic, independent errors, which is closely related to the weighted sum-squared error (SSE). In a second MCMC run, a heteroscedastic, non-Gaussian error model with autocorrelation was used. Interestingly, the resulting parameter uncertainty bounds are inflated and unrealistic, similar to what was reported in Wöhling and Vrugt (2011). The authors hypothesize, that the

AR(1) scheme introduces a systematic bias to the inference scheme. In a third run, the authors try to address this problem by adding a term in the AR(1) scheme. This term represents the mean standardized residuals multiplied by the autocorrelation coefficient.

3. The chosen methodology has several important effects on the parameter inference that need further explanation and analysis. Examining the AR(1) scheme as expressed in Eq. (12), it can easily be seen that residuals are minimized for a correlation coefficient, $p \rightarrow 1$. Since the inference scheme is based on minimization of residuals (by the likelihood function), it can be expected that values close to unity are obtained for p if it is a free parameter in the inference scheme. This is confirmed by the results of the MCMC run using likelihood 2, that yields p = 0.996 (Table 1). Note that p = 1 results in division by zero in Eq. (14). The additional term in the AR(1) scheme introduced in Eq. (20) does not alleviate this fundamental problem. Instead, it can be shown that it substitutes the AR(1) scheme by the mean (standardized) residuals because for p = 1 in Eq. (20), only the newly introduced mean residual term remains. Indeed, the MCMC run using likelihood 3 results also in a correlation coefficient close to unity (p = 0.96). Since the mean residuals are (by definition of the objective function) close to zero, the AR(1) scheme becomes essentially not effective, and error residual model of likelihood function 3 becomes a SSE formulation for heteroscedastic, non-Gaussian residuals.

4. The results of the present study support these arguments. The performance of the maximum likelihood estimate for the runs with likelihoods 1 and 3 are similar. To analyse this further, the likelihood 3 could be compared with a modified likelihood 2 that does not use the AR(1) scheme (only heteroscedastic, non-Gaussian errors). The results are expected to be similar to those of using likelihood 3 and it would be interesting to see whether the posterior densities plot in a similar region as likelihood 3 (Figure 7).

5. The problem with the AR(1) scheme is therefore not solved. In fact it can be argued that it is not even desirable to use the AR(1) scheme. Firstly, because all residual pairs in the time series are treated with the same correlation coefficient, which is certainly not



a correct assumption given that external forcing creates singular events (e.g. re-wetting spikes). Secondly, there is the fundamental problem in the implementation in the likelihood function that leads to $p \rightarrow 1$, which essentially prevents the inference scheme from accessing the information content of the time series which lies in its dynamics. Consequently, the argumentation that the modified AR(1) scheme "removes the bias in the model predictions" (p 2170, line 13ff and others) appears to be misleading or even wrong and the study should be revised in accordance to the comments above.

6. Regarding the second question formulated in comment 1), the authors argue throughout the manuscript, that the likelihood 3 results in statistically correct posterior parameter density function because the remaining residuals are closer to normality. This argument can be challenged for several reasons. Firstly, the AR(1) scheme does not provide the correct statistical framework in the context of Bayesian inference and forces the correlation coefficient to be close to unity as described above. Second, and equally important, it must be argued that the physical model (here Hydrus) and the error residual model cannot be separated. Since models are always imperfect representations of reality, structural errors are always present and will be compensated in one form or another by the choice of the error residual model. It should be noted that the residual error model does not "treat" the error statistically, it only compensates for the imperfectness of the model. Therefore it can be argued that the physical model and the error model are in fact one unit and not independent from each other. Otherwise, true measurement errors, for example, should be obtained from experiments and assigned directly to the likelihood function instead of being estimated as part of the inference scheme as proposed here. Thirdly, the degrees of freedom provided by the setup of the physical model and by the statistical error model are extremely large. A single homogeneous soil layer and a single water content time series suggest a rather simple problem. Yet, 8 error model parameters are estimated in addition to 6 physical model parameters. This seems to be a rather uneven balance that has an impact on posterior parameter estimates. Even the lower boundary condition of the physical model is estimated. One could wonder, what introduces more uncertainty: the assumption of

a fix (but unknown) lower boundary, or the assumption of Gaussian, heteroscedastic, uncorrelated error residuals. This is actually an interesting question to be addressed and would help to put the results of this study into a general context of the relative importance of uncertainty sources. For these arguments, it is therefore recommended, to delete all references related to “statistically correct” or “invalid” parameter estimates from the manuscript (instances were found in the abstract, introduction, discussion, conclusions).

7. Further to the previous comment, the authors state that the posterior parameter densities for likelihood 3 agree better with the prior (Figure 3). However, the actual fit to the data (Table 2) is very similar to the fit when using likelihood 1. This means that the difference between these runs lies perhaps in the relative weighting between data likelihood and the prior that is in turn determined by the weighting of the residuals. The residuals are weighted by a constant variance in likelihood 1 and a parameterized variance (PCHIP) for likelihood 3. Both the constant variance and the PCHIP parameters are variables in the parameter estimation scheme. It can be expected that this approach underestimates parameter uncertainty. As seen in Figure 7, the posteriors resulting from utilizing likelihoods 1 and 3 cover separate, isolated, and rather small areas in the parameter space, although the performance of the maximum likelihood estimates are very similar. From a Bayesian point of view, the solutions of likelihood 1 AND 3 should be contained in the true posterior, because it seems that they can’t be rejected on the grounds of the (mis)fit to the calibration data. This can be tested by setting and fixing the variance in likelihood 1 (and correspondingly likelihood functions 2 and 3) to a value that is realistic for both measurement and model structural error, i.e. much larger than the values obtained from the inference. This should lead to more realistic, but not necessarily smaller uncertainty bounds.

8. It doesn’t resolve the parameter uniqueness problem but would perhaps lead to more realistic parameter posteriors (and thus to more realistic uncertainty bounds). The use of prior information does only partly resolve the problem of constraining the parameter



space to realistic values. The soil hydraulic properties stored in the ROSETTA data base are largely derived from laboratory analysis of small-scale soil samples, which have been shown to be unreliable for estimating field-scale properties (e.g. Wöhling et al. 2008, 2009, and others). Therefore, it must be considered that i) the prior used here might be biased and ii) solutions that are outside of that prior are not necessarily bad or even “statistically incorrect”.

9. Generally on the discussion of adequate error models, it would be more instructive to analyse the structure of the (unprocessed) residuals and to develop/use diagnostics that help to improve the structure of the physical model rather than the error model.

10. In the assessment of hydraulic parameters at the field (or column) scale, it cannot be expected that “true” parameters can be estimated. The heterogeneity of soil stratigraphy is never resolved in the model. Therefore, only effective parameters for a certain strata can be inferred. This, of course is not independent from the effective parameters of other (model) layers (Wöhling et al. 2009). This is another reason why the preference to constrain the parameter posterior to an area within the prior must be challenged. Rather than applying complex residual error models with high parameterization effort (likelihood 3 requires 8 error model parameters compared in addition to the 6 physical model parameters and 1 boundary condition), the inclusion of other data types is more efficient and reliable to constrain effective parameter values and to build better predictive models. It has been shown in the cited study (Wöhling et al. 2011) that water content data alone is not sufficient to realistically constrain the water retention function of a multi-layer field soil and that this approach leads to biased prediction of pressure head. It is therefore recommended to test the predictive model presented here on additional experimental data of different type. Large errors in predicting other data types when fitting to only water contents (or other single data types) were also shown for another field data set for soil-plant systems by Wöhling et al. (2013). This study demonstrates that biased model predictions can be overcome by using multiple data sources. It should be noted that model structural and parameter uncertainty

can be treated explicitly and formally in the Bayesian framework also in the context of soil-plant models (Wöhling et al. 2015). It is suggested that the authors consider these findings and revise the manuscript with regard to the relative importance of different uncertainty sources, the discussion about realistic uncertainty bounds, and how to derive realistic parameter posteriors.

11. Model Structure: The methodology part is could be slightly restructured. It is more intuitive to start with the Bayesian Theorem (2.4), then explain the likelihood functions (2.3.), then the prior and the solution scheme. Please revise accordingly. The introduction in subsection 2.3 (up to likelihood 1) can be shortened substantially. It basically states that assumptions have to be made regarding bias, variance, and correlation of the residuals.

12. Aims of the study: These are actually never made explicit! Please revise the introduction section with regard to comment 9 and close this section with a clear identification of the research gaps and corresponding objectives of this study. Suggestions along these lines can be found under comment 1 above.

13. Discussion and Conclusions: The statement that AR(1) introduces bias that can be resolved by likelihood 3, and the discussion of "statistically invalid" parameter estimates when not considering autocorrelation (amongst others) dominates the discussion and conclusion sections. Please revise this argumentation considering the comments above. These statements are not generally valid and, by the way, also not necessary because the analysis of error model choice alone is an interesting topic. However, the position taken by the authors that assumptions of Gaussianity, and homoscedastic, independent error residuals necessarily leads to biased parameter estimates appears to be not warranted in the presence of the other, potential larger uncertainty sources, that are not considered and treated here.

Other comments

1. P2160, line 7ff: The results from this study are cited slightly incorrect. A simultane-



ous fit to both water content and pressure heads was found when both data types were used in the parameter inference. In contrast, the inferred parameters and corresponding predictions were strongly biased when only water content or tension data was used in the calibration.

2. P2160, Line 27, The deficit of the AR(1) scheme is not obvious at this point. Moreover, it is not the AR(1) scheme itself, but the use in the Bayesian inference as explained in the major comments 3 – 5.

3. P2162, Line 24, How realistic is a surface pressure of -1km? Does this ever occur in the model? What consequence has it for the upper boundary condition? Please explain, whether flux or head is assumed for pressure heads exceeding this threshold.

4. How deep is the groundwater table at the experimental site? Can the lower boundary condition really be assumed as a constant head? Typically, pressure heads still vary at 1m depth. Did you try different boundary conditions, such as free drainage?

5. I would suggest to replace the term "likelihood models" with "likelihood function" to avoid confusion. Bayesian inference requires a likelihood function. In contrast, error residuals are described by error models.

6. P2163, Line 18, "Formal statistical inference requires a likelihood model that describes the statistical features of the time series of residuals as closely and consistently as possible." Residuals and model structure choice act in unity and are not independent. True statistical independence cannot be achieved in this context. Therefore I would suggest removing this sentence.

7. Bias is one of the most common errors in hydrology. Among the many different error model components analysed in this study, why was bias not included? Particularly since it is stated explicitly as a systematic error of the physical model: P2177, lines 11ff.

8. Why is heteroscedasticity expressed by PCHIP and not by a simple relative error

model? This would reduce the unnecessarily large dimensionality of the inverse problem. Why were the parameters of PCHIP (and the variance in likelihood 1) fitted and not taken from the data? This leads to an underestimation of uncertainty and potentially to bias. Does the data support such a highly parameterized variance model like PCHIP? How transferrable is the entire set of inferred error parameter values? Did you test the obtained parameter posterior on independent observations, i.e., other locations of the experimental field?

9. P2170, line 12: Should this be Eq. (12)?

10. The first sentence in 2.4 is odd, please revise.

11. P2173, lines 12ff. If the original DREAM code was used, the prior is only implemented for sampling the initial parameter population. The correct consideration of the prior can (and has been), however, be implemented by adapting the Metropolis step of the original code.

12. P2175, line 3ff. when referring to the fit of likelihoods, please be specific in that you refer to the fit of the maximum likelihood estimate. It is used in a rather loose descriptive way in many instances throughout the manuscript.

13. P2175, line 21. The diagnostic plot for likelihood 1 shows that residuals differ only slightly from Gaussianity (Figure 3c). In fact the use of likelihood 3 seems to result in an even stronger deviation from Gaussianity (Figure 6f)?!

14. P2175, line 29: It's the predictions that are meaningless.

15. P2176, line 5 & 10: The AR(1) model has not a bias per se, as explained above. Please remove. Same on P2179, Line 18 and other instances.

16. P2176, line 23. The correlation coefficient for the MLE of likelihood 3 is still very close to 1! It can't be exactly unity because of the division by zero.

17. P2177, line 26ff. "statistically invalid" Please remove sentence and other similar



statements.

18. P2178, line 25ff: Tight parameter uncertainty bounds do not guarantee precision of the process model. Instead, the setup of the inference scheme with a large number of statistical parameters created an artificially overconfident model. The reality check would be the application to independent data, on different sites, or even different data types.

19. P2179, Line 11: "using a likelihood model which neglects autocorrelation essentially implies to treat the soil hydrological model as if it was perfect". This is not the case, please remove this statement.

20. P2182, line 22ff: The conclusion that likelihood 3 parameter estimates are "statistically valid" and superior to likelihood 1 is not supported by the analysis. The reasons for that are discussed in the previous comments. It is more instructive to aim at realistic parameter and predictive uncertainty estimates.

21. Figure 8: The uncertainty bounds particularly for the water retention curve are VERY small. Please plot the corresponding field data in order to assess the quality of these estimates.

Despite the substantial comments, I believe the topic of the study is important and I am confident that the paper could become a valuable contribution to hydrological science. I hope the authors find the comments insightful and useful to improve their manuscript.

Additional References

Schoups, G. & Vrugt, J. A. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic and non-Gaussian errors. Water Resources Research, 2010, 46, W10531.

Wöhling, Th., Barkle, G.F., Vrugt, J.A. (2008). Comparison of three multiobjective optimization algorithms for inverse modeling of vadose zone hydraulic properties. Soil Science Society of America Journal, 72(2), 305-319.

Wöhling, Th., Schütze, N., Heinrich, B., Šimunek, J. and Barkle, G.F. (2009) Three-dimensional modeling of multiple Automated Equilibrium Tension Lysimeters to measure vadose zone fluxes. Vadose Zone Journal, 8(4), 1051–1063.

Wöhling, Th., Gayler, S., Priesack, E., Ingwersen, J., Wizemann, H.-D., Högy, P., Cuntz, M., Attinger, S., Wulfmeyer, V., Streck, T. (2013). Multiresponse, multiobjective calibration as a diagnostic tool to compare accuracy and structural limitations of five coupled soil-plant models and CLM3.5. Water Resources Research, 49(12), 8200-8221, doi: 10.1002/2013WR014536.

Wöhling, Th., Schöniger, A., Gayler, S., Nowak, W. (2015). Bayesian model averaging to explore the worth of data for maximum-confidence soil-plant model selection and prediction. Water Resources Research, (available online), doi.org/10.1002/2014WR016292

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 2155, 2015.