

Interactive comment on “A framework to assess the realism of model structures using hydrological signatures” by T. Euser et al.

T. Euser et al.

t.euser@tudelft.nl

Received and published: 28 January 2013

Dear referee,

Thank you for your positive and constructive review. The method presented in this paper will indeed be developed further. Currently we are working on the development of a more quantitative assessment of the framework and on the influence of more and different signatures on the effectiveness of the framework. These additions would go beyond the scope of this manuscript, which is rather meant as a proof-of-concept paper, and will be the subject of a future publication.

C6506

We would like to comment on your remarks and questions:

1. *The explained variance is an important factor to measure the collinearity of the objectives. Why was this not explicitly used?*

We agree that collinearity would give an overall measure of the spread between the different vectors. However, in our understanding, two objectives, while showing a high degree of collinearity, can be directly OR inversely correlated, which is of crucial importance for the interpretation of signatures in FARM. As a thought experiment (see also Figure 1), let us consider two highly correlated and thus also highly collinear signatures. If the PCA shows nearly identical vectors (Figure 1a), this means that the model is capable of almost equally well reproducing these two signatures with the same parameter set, resulting in a consistent model. HOWEVER, if the very same signatures are now highly inversely correlated (i.e. $r \approx -1$) (Figure 1b), they would still be collinear, but the vectors would point into opposite directions in the PCA. This would mean that the model is not capable of reproducing the two signatures with the same parameter set. Rather, while improving performance according to one signature, the performance with respect to the other one decreases, resulting in an inconsistent model. In addition, it can be seen in figures 8 and 9 of the manuscript that the variance explained is comparable for different model structures, so this criterion would not discriminate well between different model structures used for this study.

2. *Contradictions between the different evaluation criteria can be identified directly from the correlation matrix. The PCA procedure is also based the correlation (or covariance) matrix. It would be interesting to provide at least for one of the models the correlation matrix of the evaluation criteria.*

C6507

The main reason to use a PCA is to reduce the amount of dimensions, this cannot be done by investigating the correlation matrix. With a smaller amount of dimensions, you lose some information, but the remaining results are usually easier and more intuitively to interpret. Thus, we think that the correlation or covariance matrix is less suitable to use as a comparison tool, as it does not give a visual output. We will add the correlation matrix of the evaluation criteria for one of the model structures to show what the results are based on.

3. *The normal score (NS) transformation of the evaluation criteria might be dangerous. If the distribution of the evaluation criteria is skewed the NS transform might bring very different performances close together.*

We agree that this risk exists. However, in a preliminary stage, we tested the effect of applying this transformation on the evaluation criteria. It appeared that the effect of this transformation on the data was minimal. We will note that in the manuscript.

4. *The NS transformation only ensures that the marginals are normal. The multivariate normality remains unexplored.*

We agree that the multivariate normality remains unexplored. However, according to Jolliffe (1986, page 199¹) the requirement of multivariate normality is of less importance if the PCA is used for dimension reduction, and thus as a mere descriptive tool as is the case with FARM. We will address this shortly in the paper.

5. *The transformation used for equation (1) is not symmetrical - overestimation of S are more punished than underestimation. One could use $F = \ln(|S_m/S_o|)$*

¹Jolliffe, I.T., 1986. Principal Component Analysis, Springer-Verlag New York Inc.

C6508

instead.

This transformation would indeed not be symmetrical if both S_m and S_o would change for different parameter values. However, S_o represents the observed hydrograph, so within a catchment the value of a certain signature of the observed hydrograph is constant. Due to the absolute value overestimation and underestimation are punished to the same extent. We will address this shortly in the paper.

6. *The prior distribution used to generate the sample might be of large importance. It would be interesting to test how the results change if a non-informative prior is used for example.*

We agree that a prior distribution can have a large influence on the results. However, we are not sure what you are referring to. For the model evaluation we used uniform sampling between certain bounds. However, due to your remark, we realize that we did not make this clear in the paper. The values presented in table 2 of the paper are the bounds for each parameter, used for the uniform sampling. We will clarify this in the table and the paper.

7. *The paper puts all blame on the model - but we know that inputs and even outputs are often erroneous. They might lead to contradicting performance for the different evaluation criteria. It would be interesting to have a comment of the authors on this issue.*

We agree that errors in the input and output data can influence the model evaluation. However, due to the relative small catchment areas, the influence of heterogeneity in the catchment, which is not represented in the data, is assumed to be of limited importance. So, the main influence of data errors can be found in case

C6509

of extreme events, which are not adequately sampled by the instruments. These errors will probably have the largest influence on the signature representing the peak distribution. On the other hand, by using signatures, mainly the dynamics of the measured and observed hydrograph are taken into account. These dynamics are more likely to represent catchment behaviour and to be less sensitive to small measurement errors than evaluation criteria that compare each point of the hydrograph individually. We shall comment on this in the revised version of the manuscript as well.

8. *A minor remark : What are the units of the axes on the PC figures? Normally the evaluation criteria represent unit vectors thus their 2D projection on PC1 and PC2 should have a length < 1.*

The Principal Components are dimensionless, because the ratios between the same signatures of different hydrographs are taken into account and these ratios are dimensionless. To increase the readability of the graphs, all loadings are multiplied by 15. We will add this in the figures which presents the results of the PCA.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 12989, 2012.

C6510

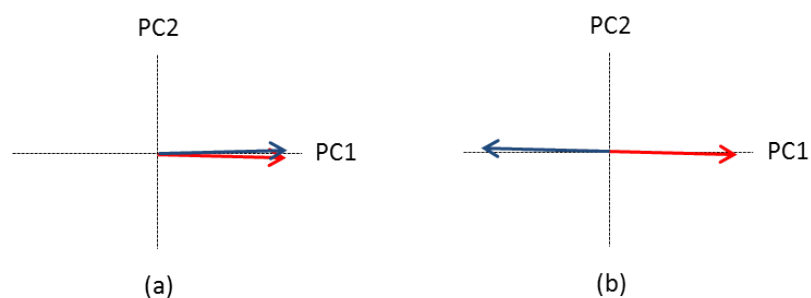


Fig. 1. Thought experiment for using collinearity

C6511