

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

**A. Bårdossy (Referee)**

bardossy@iws.uni-stuttgart.de

Received and published: 8 January 2013

This paper offers an interesting view on multiobjective evaluation of hydrological models. Instead of the usual identification of a suitable Pareto optimum it intends to compare model structures on the basis of their behaviour with respect to different objectives. It introduces the concept of model consistency – a measure which reflects the simultaneous behaviour of the different evaluation criteria. This is a very good idea.

The paper is not easy to read. I have some experience with PCA and hydrological modelling, but it took me real effort to (hopefully) understand the contents. As I understand the authors:

â€” Consider a model as consistent if a model performs simultaneously well in all objectives.

â€” Evaluate consistence in the two dimensional space of the first two PC-s.

â€” Apply a somewhat ad hoc subjective procedure to compare different model structures.

There are quite a few questions and remarks listed below:

1. The explained variance is an important factor to measure the collinearity of the objectives. Why was this not explicitly used?
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.
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.
4. The NS transformation only ensures that the marginals are normal. The multivariate normality remains unexplored.
5. The transformation used for equation (1) is not symmetrical – overestimation of S are more punished than underestimation. One could use  $F = \ln(\text{abs}(S_m/S_o))$  instead.
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.
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

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

issue.

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  $< 1$ .

In conclusion this is an interesting paper with good new ideas, but I am sure the methodology will be further developed in the future. The presentation requires more efforts to reach a wider audience.

---

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

**HESD**

9, C6189–C6191, 2013

---

Interactive  
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

C6191

