Fairbairn *et. al.*, (2016):
http://www.hydrol-earth-syst-sci-discuss.net/hess-2016-195/
Response to comments from Dr Jean-Philippe Vidal

July 26, 2016

Firstly, we would like to thank Dr Jean-Philippe Vidal for his comments and useful suggestions.

1 Major comment

Dr Jean-Philippe Vidal was right to point out that many of the stations included in the calculations are influenced by anthropogenic water management, which is not simulated by the MODCOU hydrogeological model. He was concerned that the results might be interpreted as being closer to anthropogenically influenced streamflow. He suggested the following:

1. To consider only catchments with low anthropogenic influence in order not to compare apples and oranges and avoid drawing conclusions on the ability of SIM (with or without data assimilation) to simulate anthropogenically influenced streamflow,
2. To show scatter plots of NSEs instead of distributions (possibly with marginal distributions) to reduce the potential spatial bias effect mentioned above.

Response:

We have decided to follow Dr Jean-Phillipe Vidal's suggestions by showing the results for the stations with low-anthropogenic influence. We have used the suggested reference networks of Giuntoli *et al.* (2012, 2013) to extract a subset of 67 river gauges with low-anthropogenic influence from the original 546 stations, valid for both low and high flows. A map of these stations is shown in Fig. R3.1 (at the end of the document). A scatter plot is shown of the Nash efficiency of these stations (labeled as "Low anth. influence") and all the other stations (labeled as "High anth. influence") in Fig. R3.2.

For the sake of clarity, in Fig. R3.2 only the stations are shown in the range of Nash scores -1.0 to 1.0. The 'low anth. influence' stations follow a similar pattern to the 'high anth. influence' stations. Furthermore, we calculated the Median Nash efficiency scores for the 67 stations in Table R3.1 (at the end of the document). Note that the median is calculated rather than the mean because the majority of stations ($> 80\%$) have positive Nash efficiency scores, but a few outliers have scores near to -100. The median is a more appropriate metric as it is less sensitive to extreme outliers and is a better indicator for highly skewed distributions (Moriasi *et al.*, 2007). The scores for this subset are improved relative to the 546 stations in Table R3.1, as expected. In particular, the percentage of stations with good scores (Nash efficiency > 0.6) is increased significantly. The discharge bias is also slightly less for the stations with low anthropogenic influence relative to the 546 stations. This supports Jean-Phillipe Vidal's suggestion that part of the positive bias in the discharge ratio of the NIT simulation for the 546 stations could be attributed to abstractions not being accounted for. However, the majority of the discharge bias in the NIT simulation is still present with the 67 stations with low anthropogenic influence. Moreover, the relative performance of the experiments is very similar. Therefore, the conclusions of the experiments are not affected by the ability of SIM (with or without data assimilation) to simulate anthropogenically influenced streamflow.

We will explain these results in Section 3.4 of the paper. Figures R3.1 and R3.2 will be included in a supplement.

2 Minor comment

Their interpretation of the NSE detailed P8L8-9 is incorrect:a negative NSE value means that the model performs worse than a constant model with a value equal to the average of all observations.

Response: We agree. We will replace "The Nash efficiency can range from $-\infty$ to 1, with 1 corresponding to a perfect match of the model to the observed data and scores less than zero implying that the model mean is a worse predictor than the observations." with "The Nash efficiency can range from $-\infty$ to 1, with 1 corresponding to a perfect match of the model to the observed data and a negative value implying that the model performs worse than a constant model with a value equal to the average of all the observations."

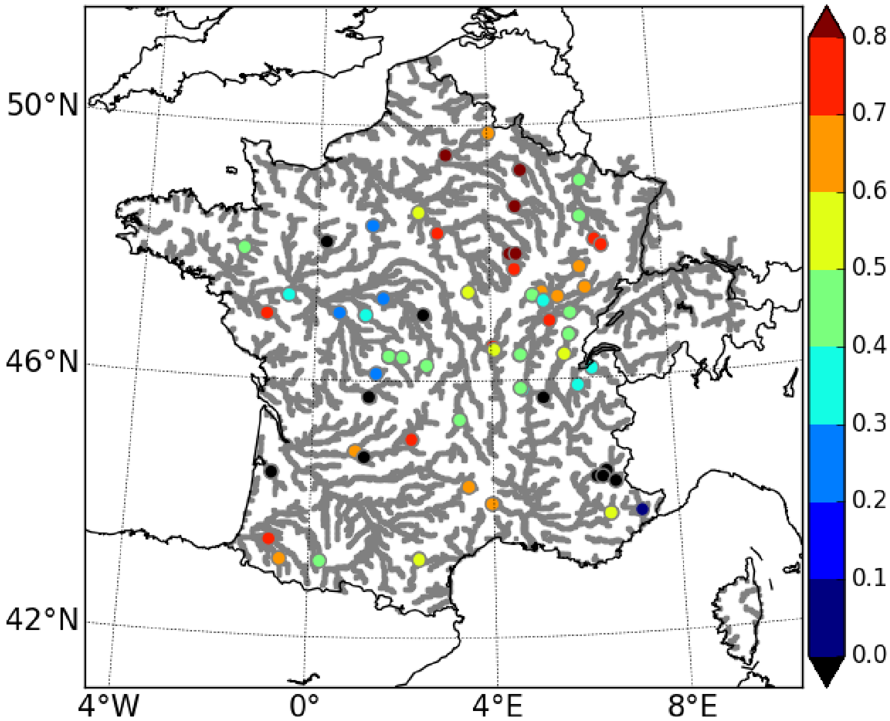

Figure R3.1: Map of the SIM discharge Nash efficiency scores for the 67 stations with low-anthropogenic influence over France for the NIT simulation, calculated over the period 2007-2014. The river network is also shown.

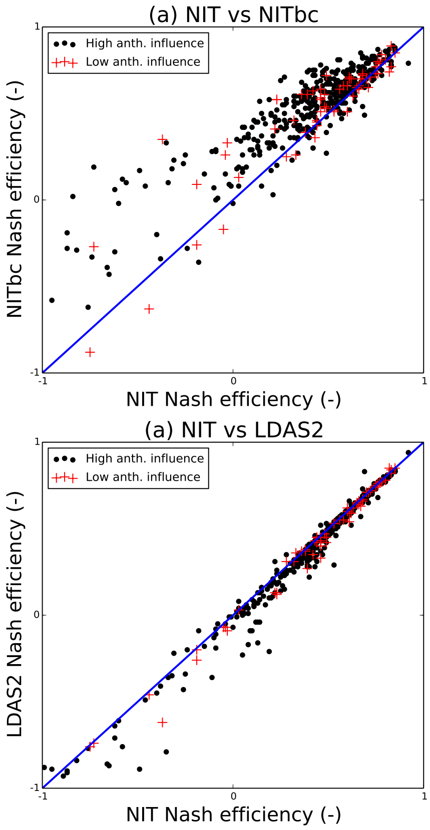

Figure R3.2: Scatter plots of the SIM discharge Nash efficiency scores for the 546 stations over France for (a) NIT vs NITbc and (b) for NIT vs LDAS2. The stations are classified with either low (67 stations) or high anthropogenic influence (479 stations). For the sake of clarity, the Nash scores are shown between -1.0 and 1.0. The scores are calculated over 2007-2014.

Table R3.1: Median Nash efficiency (NE) and discharge ratio (Qs/Qo) scores over the 546 river gauges over France and for the subset of 67 gauges with low anthropogenic influence, calculated over 2007-2014. Also shown are the % of stations with a Nash score above 0.6. The best scores are shown in bold font.

Experiment	NE for 546/67 stations	Discharge ratio for 546/67 stations	% stations with NE > 0.6 for 546/67 stations
NIT	0.44/0.48	1.19/1.16	26%/44%
NIT_m	0.48/0.54	1.15/1.12	30%/48%
NIT_{bc}	**0.56/0.60**	**1.02/0.99**	**42**%/**59**%
LDAS1	0.44/0.48	1.18/1.15	27%/44%
LDAS2	0.41/0.45	1.21/1.18	23%/40%
$LDAS1_{bc}$	**0.56/0.60**	**1.02/1.00**	**42**%/**57**%
$LDAS2_{bc}$	0.53/0.54	1.08/1.06	38%/53%
$LDAS2_{QC}$	0.40/0.45	1.21/1.18	21%/39%

References

I. Giuntoli, B. Renard, and M. Lang. *Changes in Flood risk in Europe*, pages 199–211. Kundzewicz, Z. K., Ed., 2012. ISBN 9780415621892. IAHS Special Publication 10.

I. Giuntoli, B. Renard, J.-P. Vidal, and A. Bard. Low flows in france and their relationship to large-scale climate indices. *Journal of Hydrology*, 482:105–118, 2013.

D.N. Moriasi, J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, and T.L. Veith. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *American Society of Agricultural and Biological Engineers*, 50:885–900, 2007.