Hydrol. Earth Syst. Sci. Discuss., 11, C6568–C6577, 2015
www.hydrol-earth-syst-sci-discuss.net/11/C6568/2015/

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Virtual laboratories: new opportunities for collaborative water science" *by* S. Ceola et al.

S. Ceola et al.

serena.ceola@unibo.it

Received and published: 27 February 2015

article

Response to Hessel Winsemius (Referee)

The authors gratefully acknowledge Hessel Winsemius for his positive and constructive review. In what follows in *italics* are the comments provided by the Referee, and in **bold** fonts the authors' response, inclusive of the indication on how the text will be modified within the next days to comply with the Referee' recommendations and comments.

In this paper, a much-needed proposition for development of virtual laboratories for collaborative research in hydrology, ensuring reproducibility and repeatability of experiments, is made. The paper's focus seems rather trivial, but I agree with the authors that it is not! In fact many hydrological studies to date lack reproducibility due to lack of data sharing, limited metadata, poorly shared and documented experiment protocols and experiment outputs.

The authors wish to thank Hessel Winsemius for his recognition of our main goal.

I do have a number of comments of which the last one is the most important. I hope the authors find them useful for improving the manuscript. The most important comments are given below. I have also provided an annotated manuscript with some smaller remarks that should be treated.

Please see our detailed replies below. The revised manuscript including the suggested changes will be uploaded in the next few days.

No reference is made to past or ongoing global model intercomparison studies (some also including social interaction with the natural system) that may also benefit from the methods presented, e.g. ISLSCP, ISIMIP, EU-WATCH. I suggest to add some of these including references. The fact that these studies are global scale rather than local definitely sets this study apart from them as many of the issues raised by the authors (e.g. differences in preprocessing procedures, parameter selection, state handling) are difficult to resolve or less important at global scale, but can be tackled more appropriately at local scales.

HESSD
11, C6568–C6577, 2015

In response to this comment we have added the following sentence, in which we present and cite the suggested global model inter-comparison studies. Thanks for this.

"Model inter-comparison studies at a global scale, including social interactions with the natural system, like e.g. ISLSCP (http:/daac.ornl.gov/ISLSCP_II/islscpii.shtml), EU-WATCH (http:/www.eu-watch.org/) and ISI-MIP (https:/www.pik-potsdam.de/research/climate-impacts-and-vulnerabilities/research/rd2-cross-cutting-activities/isi-mip), but also comparative model inter-comparison experiments in hydrology (i.e. performed by different and independent research groups) such as MOPEX (Duan et al., 2006, Andreassian et al., 2006), DMIP (Reed et al., 2004) or LUCHEM (Breuer et al., 2009), though successful with respect to data sharing, have contributed little to disentangle the causes of performance differences between different models and to increase our understanding of underlying hydrological processes. "

The 7 stages mentioned do not seem to be very specific for hydrology (although their implementation in the virtual laboratory of course is hydrology specific) but instead could be applied on any scientific model intercomparison experiment. This raises the question if you are here proposing a general framework for virtual laboratories, showing an application in hydrology, or that you are proposing a hydrology-specific framework. From the remainder of the paper, we can conclude that it is probably the latter. It would be stronger if you can emphasize how these stages are specific to hydrology compared to other scientific fields or change them so that they are hydrology specific.

We have now better contextualised the proposed workflow stages as hydrology specific steps. Please refer for instance at STAGE 3: Collect input data, where

HESSD

11, C6568–C6577, 2015

we included new reference to specific hydrologic data (i.e., precipitation, temperature and river discharge). As kindly suggested by Hessel Winsemius in one of the following comments, we decided to add a new workflow stage as follows. Immediately after the definition of a scientific question/problem (STAGE 1), the involved partners need to clearly outline a protocol (i.e. experiment guideline) in order to ensure experiment reproducibility and therefore reduce and control the degree of freedom of single modellers (STAGE 2: Set up protocols).

Stage 3: the reworking of data into model specific inputs. Stage 3 suggests that any modeller can do any preprocessing he/she deems fit. In this transformation process however, much of the intercomparability of the experiments may be polluted by the fact that one modeller does something else with the data than another. I would propose that the degree of freedom is controlled through the proposed protocols and that you clarify this in the description of stage 3.

Please refer to our previous reply.

An important comment is that the connection between the 7 steps and the description of the collaborative experiment, performed in the SWITCH-ON project is not very clear. Please refer to the steps in the description of the 2 experiments performed so that the reader can make this connection more easily.

We acknowledge this lack of information in the original version of our manuscript. Accordingly to this useful comment, we have now added some reference to the workflow stages directly in the experiment description.

Moreover, in the experiment description, a lot of focus is on the protocol design (which

HESSD

11, C6568–C6577, 2015

I agree is very important) while the protocol design receives very limited attention in the general 7-step description. It is somewhere hidden in step 4. I would emphasize more on the protocol design and describe in the 7-steps more accurately what the protocols should embody. In fact, you could argue that the protocol design should be a concrete separate step. Please consider this option.

Thanks for this remarkable advice. Please refer to our previous reply, where we introduce a new workflow step (STAGE 2: Set up protocols).

In page 13463, l. 24-25, the authors state that "with different model implementations, the main purpose of the modelling exercise needs to be clearly defined". Whilst I fully agree with this, strictly speaking, the second experiment design did not adhere fully to this statement. The authors indicate this also in page 13462 l. 17 "we did not specify what model improvement meant a priori". I can imagine that this observation in fact led to the statement above, however this is not clear from the discussion. Please add a sentence that explains whether the lack of specification of the meaning of model improvement indeed led to the conclusion that the purpose of the experiment needs to be very clearly defined.

The original version of our manuscript was not totally clear on this subject. Indeed, the purpose of the second experiment was to profit from researchers personal experience in order to improve model performances. The added value of this second experiment relied on the scientific knowledge of researchers, being capable of exploring alternative modelling options which will be helpful for future hydrological experiments in the VWSL. Therefore, we did not deliberately specify, as in Protocol 1, how to run the experiment. In Protocol 2, researchers could freely choose to improve model performance by either reaching a higher statistical metric, less equifinality among parameters or a more reliable model

HESSD

11, C6568–C6577, 2015

in terms of realistic internal variables. According to this, in the new version of the manuscript, which will be uploaded soon, we added a couple of sentences providing a thorough explanation.

My most important comment: the experiments performed are rather simple (same model structure, same spatial representation, same data, data handling) and perhaps not very representative for the type of collaborative model experiments that the hydrological community would like to perform in the forthcoming decade. Whilst including a more complex experiment is perhaps beyond the scope of this paper, it would make the paper a lot stronger, if the authors can demonstrate that the suggested procedure for protocol establishment indeed applies even when a completely different (more complex) experiment would be performed. For instance, the suggested protocols for the experiments performed amongst the research groups would not yield a satisfactory intercomparability when the science question would be related to differences between model structures, where all groups would use different hydrological models and/or different levels of process or input distribution in their models throughout the 15 catchments, or when e.g. different ways to include man-made interactions in models would be studied. In these examples, models may have very different states and fluxes, and may even have different spatial representations of states and fluxes making their intercomparison a lot more difficult. It would make the framework a lot more convincing if the authors can perform a thought experiment in the discussion that demonstrates the validity of the proposed framework, even in more complex cases that will become important in this decade of Panta Rhei such as mentioned above. I hope these comments prove useful and I am looking very much forward to an improved manuscript.

We totally agree with this comment, since this paper presents the results of a relatively simple hydrological exercise run in a collaborative framework. However, the experiments discussed here show that it is important to revisit experiments

HESSD

11, C6568–C6577, 2015

that are seemingly simpler than existing inter-group model comparisons to understand how small differences affect model performance. What is clear is that it is fundamental to control for different factors that may affect the outcomes of more complex experiments, such as modeller choice and calibration strategy. In more complex situations the virtual experiments could be conducted through comparisons at different levels of detail. For example, if models with different structures were to be compared there will be no one-to-one mapping of the state variables and model parameters and the comparison would applied to a higher level of conceptualizations. There are a number of examples in the literature where comparisons at different levels of conceptualization have been demonstrated to provide useful results. One such example is Chicken Creek model inter comparison (Holländer et al., 2009, 2014) where the modellers were given an increasing amount of information about the catchment in steps, and in each step the model output in terms of water fluxes were compared. The Chicken Creek inter comparison involved models of vastly different complexities, yet provided interesting insights in the way models made assumptions about the hydrological processes in the catchment and the associated model parameters. Another example is the Predictions in Ungauged Basins (PUB) comparative assessment (Blöschl et al., 2013) where a two step process was adopted. In a first step (Level 1 assessment), a literature survey was performed and publications in the international refereed literature were scrutinised for results of the predictive performance of runoff, i.e. a meta-analysis of prior studies performed by the hydrological community. In a second step (Level 2 assessment) some of the authors of the publications from Level 1 were approached with a request to provide data on their runoff predictions for individual ungauged basins. At Level 2 the overall number of catchments involved was smaller than in the Level 1 assessment but much more detailed information on individual catchments was available. Level 1 and Level 2 were therefore complementary steps. In a similar fashion, virtual experiments could be conducted using the protocol proposed

HESSD

11, C6568–C6577, 2015

in this paper at different, complementary levels of complexity. The procedure for protocol development (Figure 5), which notably checks on independent model choices between partners and feedbacks to earlier stages in protocol development, will help in developing protocols for more complex collaborative experiments, addressing real science questions on floods, droughts, water quality and changing environments. As such more elaborated experiments are part of ongoing collaborative work, we are not able to present their outcomes in the manuscript, but we will certainly upload their description in the Virtual Water-Science Lab space of the SWITCH-ON website (http:/www.water-switch-on.eu/lab.html), accessible to all HESS Readers, in the next months, as soon as they are completed. As this is a learning process, the adequacy of the protocol development procedure itself will be evaluated during these experiments. The modelling study presented in this paper therefore represents a relatively simple, yet no less important first step towards collaborative research in the Virtual Water-Science Laboratory. Finally, we thank Hessel Winsemius for his detailed and helpful review.

Please also note the supplement to this comment: http:/www.hydrol-earth-syst-sci-discuss.net/11/C6069/2015/hessd-11-C6069-2015-supplement.pdf

We modified the manuscript according to the proposed suggestions. Thank you. Please also see below our reply to one of Hessel Winsemius's comments.

Hessel Winsemius's comment at page 13454: *This would in terms of calibration/validation yield the same as NSE, why did you use both?*

The metrics we selected to measure the goodness-of-fit of the TUWmodel are

HESSD
11, C6568–C6577, 2015

widely employed indicators that separately focus on high, low and ordinary flows. Even though we are well aware that NSE and RMSE provide comparable information on model performance, we deemed important to compute both.

REFERENCE

Andreassian, V., Hall, A., Chahinian, N., and Schaake, J. (2006) Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment - MOPEX - IAHS Proceedings & Reports No. 307, IAHS Press.

Blöschl, G., M. Sivapalan, T. Wagener, A. Viglione and H. H. G. Savenije (Eds.) (2013) Runoff Prediction in Ungauged Basins - Synthesis across Processes, Places and Scales. Cambridge University Press, Cambridge, UK, 465 pp.

Breuer, L., Huisman, J. A., Willems, P., Bormann, H., Bronstert, A., Croke, B. F. W., Frede, H. G., Graeff, T., Hubrechts, L., Jakeman, A. J., Kite, G., Lanini, J., Leavesley, G., Lettenmaier, D. P., Lindstroem, G., Seibert, J., Sivapalan, M., and Viney, N. R. (2009) Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM). I: Model intercomparison with current land use, Advances in Water Resources, 32, 129–146, doi:10.1016/j.advwatres.2008.10.003.

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H., Gusev, Y., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E. (2006) Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, Journal of Hydrology, 320, 3 – 17, doi:http:/dx.doi.org/10.1016/j.jhydrol.2005.07.031,

http:/www.sciencedirect.com/science/article/pii/S002216940500329X.

Holländer, H.M., T. Blume, H. Bormann, W. Buytaert, G. Chirico, J.-F. Exbrayat, D. Gustafsson, H. Hölzel, P. Kraft, C. Stamm, S. Stoll, G. Blöschl, H. Flühler (2009) Comparative predictions of discharge from an artificial catchment (Chicken Creek) using sparse data, Hydrology and Earth System Sciences, 13, 2069-2094.

Holländer, H. M., H. Bormann, T. Blume, W. Buytaert, G. B. Chirico, J.-F. Exbrayat, D. Gustafsson, H. Hölzel, T. Kraußе, P. Kraft, S. Stoll, G. Blöschl, and H. Flühler (2014) Impact of modellers' decisions on hydrological a priori predictions. Hydrology and Earth System Sciences, 18, 2065-2085, doi:10.5194/hess-18-2065-2014.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D., and Participants, D. (2004) Overall distributed model intercomparison project results, Journal of Hydrology, 298, 27–60, doi:10.1016/j.jhydrol.2004.03.031.