**Hydrology and Earth System Sciences Discussions**

# *Interactive comment on* "Legitimising neural network river forecasting models: a new data-driven mechanistic modelling framework" *by* N. J. Mount et al.
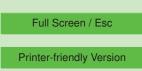
**N. J. Mount et al.**

nick.mount@nottingham.ac.uk

Received and published: 18 March 2013

Response to Reviewer 1 (Holger Maier).

We are grateful to the Reviewer for recognition of the importance of our paper in addressing one of the key issues facing neural network models in hydrological research; namely the credibility of their application. We are also pleased to note his view that the paper is 'well written and organised', and that the 'proposed approach and illustrative case study are clear and useful'. The Reviewer makes four constructive suggestions that, if addressed, could improve the quality and contribution of our paper. Adhering to the numbering system used in his review, the following revisions will be implemented

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

in our revised submission.

1. Different levels of generality.

The Reviewer is correct in noting that different aspects of our paper may be applied at differing levels of generality. Specifically, we present a framework that can be applied at a general level, and a more specific method (partial derivative sensitivity analysis of MLP ANNs) that is used to implement the framework in an example application of NNRFs. The present order and structure of the material in our introduction would clearly benefit from revision. We propose to move the material that develops the general DDMM framework to the beginning of that section, and follow this with more specific issues associated with our NNRF exemplar. New subheadings will be introduced to reinforce the different levels of generality associated with the framework and the exemplar.

We also note that, within Point 2, the Reviewer has questioned why our framework is restricted to the forecasting of hydrological variables in rivers? As we have stated in the paper (page 150, line 6) our framework is 'generic'. However, the Reviewer's comments indicate that we could make this point more clearly. Therefore, in clarifying the generality of the framework (see above comments) we also propose to highlight its wider applicability beyond river forecasting.

2. Terminology – River Forecasting.

The Reviewer questions what is meant by the term Neural Network River Forecasting (NNRF). We did not define the term here but instead referenced Abrahart et al., (2012) in which the term is defined:

"NN rainfall-runoff and streamflow modelling are collectively termed NN river forecasting (NNRF) in which 'The basic jobs of a river forecast model are to estimate the amount of runoff a rain event will generate, to compute the routing, how the water will move downstream from one forecast point to the next, and to predict the flow of

water at a given forecast point throughout the forecast period' (NOAA, 2011)."

As the terminology would appear unfamiliar to some readers of HESS, we will include a direct quotation of this definition in the revised paper.

3. Purpose of the proposed framework.

The Reviewer suggests that the purpose of our framework is to "assess how well underlying physical processes have been captured by a calibrated ANN model". This is not an accurate reflection of the purpose of the framework as presented in our paper. We note that on Page 147, Line 22 we highlight the view taken in data-based mechanistic methods: that a model should be sufficiently 'real'. This is typically taken to mean physical process representation. However, in the DDMM framework that we present we are careful to express a model's legitimacy in terms of the stability, coherency, continuity and magnitude of its mechanisms (Figure 2); thereby avoiding the need to conflate legitimacy with physical process representation. This view of legitimacy is more closely aligned with model verification (in which a model's mechanisms are assessed against a conceptual blueprint (AIAA, 1998; Balci, 1998; Davis, 1992; Sargent, 1998; 2010)) than validation (in which a model's performance is assessed against a set of defined performance criteria (Carson, 1986; Curry et al., 1989; Beven and Binley, 1992, Rykiel, 1996)). Indeed, a model's legitimacy may be assessed intrinsically (e.g. do the behaviours indicate unacceptable local overfitting?) as well as extrinsically (e.g. do the behaviours mirror physical processes?).

The Reviewer questions our use of the term validation in the paper and suggests that our framework is not validation, but rather part of model selection, which is a pre-validation activity. We are not fully in agreement with the Reviewer as we recognise validation as an overarching term that is commonly used to encompass all stages and components in model assessment (c.f. Anderson and Bates, 2001). However, we do recognise that the paper would be improved by more clearly positioning our framework in the terminology of model verification and validation (as presented above). This

will be done briefly in the revised Introduction, but we recognise the need for a fuller discussion of such issues and that will be the topic of a follow-up, standalone paper.

4. Input selection and model structure.

The Reviewer requests greater discussion about how the information delivered by the framework, and the sensitivity analyses, could be used. He suggests it may be useful for improved input selection and the determination of optimal model structure – both central issues associated with the question of how to develop ANN models with appropriate complexity. He notes that our method could be useful in this context and requests the inclusion of an appropriate discussion and certain relevant literature. We note that sensitivity analysis is a common approach for optimising input selection and/or model complexity. Whilst we do not use our framework to deliver such optimisation, we do agree that it could be useful in such respects. We will, therefore, be pleased to include additional text to discuss this aspect.

5. Physical plausibility of ANN mechanisms.

We would be pleased to include some additional literature in relation to elucidating the internal workings of ANN models at the start of Section 2 so that partial derivative sensitivity analysis is more broadly contextualised.

6. Comparison to different methods.

The suggestion to include a comparison of our partial derivative sensitivity analysis method with other methods for elucidating ANN internal mechanisms (e.g. connection weights) is a good one. As suggested, we will develop a qualitative, tabulated comparison.

Response to Reviewer 2 (Anonymous).

We are grateful to this Reviewer for recognising the interest associated with our research and its potential contribution. The Reviewer makes nine numbered suggestions that we respond to in the same order.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

1. Emphasis on black-box model weaknesses.

The Reviewer suggests that we reduce our emphasis on the black-box characteristics of ANNs and its associated limitations. They also suggest that we highlight the success of ANNs in various hydrosystem problems.

We accept that ANNs have been widely used in hydrosystems research and that they have been shown to be effective in replicative and predictive problems. Indeed, we clearly state on Page 147, Lines 12-16 that:

"the main benefit of NNRFs over statistical models is that they have been found to deliver enhanced levels of model fit when assessed against calibration and validation data sets. Consequently, it has been suggested that NNRFs can deliver forecasts with reduced error and can be used to extend the horizon over which forecasts can reliably be made."

We are, therefore, surprised that the Reviewer feels our attitude towards ANNs is overly negative. The issue here is what one determines to be a 'successful and satisfactory application'. The purpose of this paper is to recognise that, whilst ANNs are effective at optimising complex, multidimensional hydrological relationships, the extent to which their optimised outputs are demonstrably 'right for the right reasons' is far less clear. Indeed, the lack of mechanistic knowledge surrounding ANN models remains a core criticism of them in the literature as well as one that has been identified as preventing their wider uptake and acceptance by the hydrological community (Abrahart et al., 2010; Abrahart et al., 2012). Addressing this critical issue is the focus of our work and a pressing matter that we should not be afraid to emphasise. Nonetheless, we will ensure that the successes of ANNs are properly stated in our revised Introduction.

2. Simplicity and linearity of cases.

The Reviewer is correct in recognising the simplicity of the modelling cases we present. This contrasts with the complexity of the arguments that the paper necessarily ex-

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

amines. In this context, we would suggest that using relatively simple exemplars to illustrate the application of a new, novel framework, is an ideal starting point. We respectfully highlight the fact that this is a paper about how we model, rather than what we model, and it is this fact that justifies the complexity of the arguments and simplicity of the examples presented.

With respect to the linearity of the cases we present, we do accept that they are not highly non-linear. However, inspection of Figures 8 and 9 coupled with the metrics presented in Tables 5 and 6 does demonstrate modest non-linearity in the modelling problem. That said, we again stress that this paper is not about what is being modelled, but about how it is being modelled, and in this context we would argue that the degree of linearity in the examples is of limited relevance.

3. The potential of other ANNs.

This comment parallels aspects of the suggestions raised by Reviewer 1 in Point 6 (above). We agree that other ANN techniques can elucidate aspects of internal mechanisms and we will review these in a new comparative table.

4. Length of introduction.

We agree that the introductory sections of this paper are substantial and note that Reviewer 1 recognises the need for even greater length and detail – not less. Greater clarity will be provided by the reordering and restructuring of the Introduction that will be made in response to Reviewer 1 (Point 1). However a reduction in the size of the Introduction would limit the extent to which vital arguments about model legitimacy can be rehearsed. As stated under Reviewer 2, Point 1 (above), we will ensure that the value of ANNs is properly portrayed.

5. Model inputs.

The Reviewer is correct to note the potential usefulness of sensitivity analysis in supporting model input selection and structure. We note that this view parallels that made

by Reviewer 1 in Point 4. The implications will be discussed as outlined in our response to Reviewer 1.

6. Explanation of Equation 2.

The request to provide the relative sensitivity of each input in Equation 2 is not possible because computed relative sensitivity will vary continuously across the input space. For this reason, each data point has a separate relative sensitivity calculated for its inputs – resulting in the cloud of points that are observed in Figures 8 and 9. Therefore, there is no single relativity sensitivity value that can be defined and explained for each input. Instead we adopt the brute-force, computational approach that we explain on Page 153 – an approach that extends the idea of relative sensitivity values being local rather than global. We will include a sentence alongside Equation 2 that clarifies this point.

7. Similarity of DDMM framework to pre-processing in ANNs.

As ANNs are a form of data-driven modelling, the correspondence between elements of Figure 1 and neural network model building is to be expected. However, we draw the Reviewer's attention to box B2 and B3 and note that efforts to elucidate the internal mechanisms of ANN models and provide assessments of their legitimacy are not standard pre-processing operations. These additions are the core foci of our paper.

8. Global-local relative sensitivity plots.

Yes. The data can be separated into calibration and validation plots and this will be done in the revised paper.

9. Ordering of Figures 8 and 9.

We are grateful to the Reviewer for spotting this error in the Figure sequencing and this will be rectified.

References.

Abrahart, R.J., Anctil, F., Coulibaly, P., Dawson, C.W., Elshorbagy, A., Mount, N.J., See,

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

L.M., Shamseldin, A.Y., Solomatine, D.P., Toth, E., Wilby, R.L. 2012. Two decades
of anarchy? Emerging themes and outstanding challenges for neural network river
forecasting. Progress in Physical Geography, 36(4), 480-513.

Abrahart, R.J., See, L.M., Dawson, C.W., et al. 2010. Nearly two decades of neural
network hydrologic modeling. In: Sivakumar, B. Berndtsson, R. (Eds.) Advances in
Data-Based Approaches for Hydrologic Modeling and Forecasting. Hackensack, NJ,
World Scientific Publishing. 267–346.

American Institute of Aeronautics and Astronautics. 1998. Guide for the Verification
and Validation of Computational Fluid Dynamics Simulations, AIAA-G-077-1998, Re-
ston, Virigina, USA.

Anderson, M.G., Bates, P.D. 2001. Model Validation: Perspectives in Hydrological
Science. John Wiley and Sons, Chichester, UK.

Balci, O. 1998. Verification, validation and testing. In: Banks, J. (Ed.), Handbook of
Simulation, John Wiley and Sons, Chichester, UK. 335-396.

Beven, K.J., Binley, A.M. 1992. The future of distributed models: model calibration and
uncertainty prediction. Hydrological Processes, 6, 279-298.

Carson, J.S. 1986. Convincing users of a model's validity is a challenging aspect of a
modeler's job. Industrial Engineering, 18 (6), 74-85.

Caswell, H. 1976. The validation problem. In: Patten, B. (Ed.), Systems Analysis and
Simulation in Ecology, Vol. IV. Academic Press, New York. 313-325.

Davis, P.K. 1992. Generalizing concepts of verification, validation and accreditation for
military simulation. R-4249-ACQ, October 1992, RAND, Santa Monica, CA.

National Oceanic and Atmospheric Administration (NOAA) 2011. National Weather
Service Middle Atlantic River Forecast Center: The models and the final product. Avail-
able at: http://www.erh.noaa.gov/marfc/Science/models.html

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Rykiel, E.J. 1996. Testing ecological models: the meaning of validation. Ecological Modelling, 90, 229-244.

Sargent, R.G. 2010. Verification and validation of simulation models. In: Johansson, B., Jain, S., Montoya-Torres, J., Hugan, J., Yucesan, E. (Eds.), Proceedings of the 2010 Winter Simulation Conference, Baltimore, Maryland, USA. 166-183.

Sargent, R.G. 1998. Verification and validation of simulation models. In Carson, J.S., Manivannan, M.S., Medeiros, D.J., Watson, E.F. (Eds.) Proceedings of the Winter Simulation Conference 1998, Washington DC, USA. 121-130.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 145, 2013.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper