Hydrol. Earth Syst. Sci. Discuss., 12, C4222–C4228, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C4222/2015/

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Technical Note: Application of artificial neural networks in groundwater table forecasting – a case study in Singapore swamp forest" *by* Y. Sun et al.

Y. Sun et al.

tmssy@nus.edu.sg

Received and published: 13 October 2015

Thanks for the valuable comments on our paper.

In correspondence to the 'general comments', please find our reply as follows:

- It's fully acknowledged that application of ANN in hydrology research, more specifically in groundwater table modelling, has been a popular research topic. Two paragraphs are devoted to elaborating the most recent research and findings (please refer to lines 4 to 28 in page 9319 and lines 1 and 2 in page 9320).

The case study of Taormina et al.'s paper is a coastal aquifer, where rainfall and evap-

otranspiration are considered as the major influencing factors (2012). In addition, the approach adopted by Taormina et al. involves two steps with the first step being reconstructing the one-hour-ahead groundwater time series to be used as the inputs for the second step. In contract, we applied ANN to forecast the highly responsive groundwater table in a fresh water swamp forest, and the methodology in our paper is more straightforward, which only uses the rainfall and reservoir levels as the ANN inputs and observed groundwater tables with different leading times as the output. In this context, our paper is the first of this kind and is instructive for similar research in future. We think this paper meets the standard to be published as a 'technical note'.

- ANN is chosen mainly due to its ability in regression analysis and the usage of more accessible variables in mapping the input-output relationships. Its applicability has also been widely verified in many fields of hydrology research. Due to the complicated topography, geological characteristics and hydrological processes, the relationship between the input (reservoir level, rainfall) and the output (groundwater table) is not linear (as exemplified in Fig. 1 and Fig. 2). Hence, linear regression model is not suitable to serve our study purpose (Please refer to lines 14 to 17 in page 9319).

In addition, for our study area – the NSSF, a numerical model has been setup (with Mike SHE) in order to simulate local hydrological conditions. ANN actually outperforms the numerical model in terms of forecasting accuracy at the piezometer locations but with a poorer spatial coverage. In consideration of the limited article length, either the linear regression model or the numerical model is further discussed in the paper.

- A MIMO model with 4 outputs is selected over 4 MISO ANNs mainly for 2 reasons: (1) it's easy to implement as compared to 4 ANNs; and (2) some correlation is observed in the groundwater tables, e.g. the response to dry and wet conditions. Targeting the groundwater table measurements at 4 locations simultaneously, the cross-correlation impact can be captured in the synaptic weights of the trained ANN and hence a better performance is expected.

- The NSSF, as Singapore's only remaining patch of primary freshwater swamp forest, is of utmost importance for conserving a large proportion of the flora and fauna in Singapore. Due to the constraints imposed from setting up monitoring stations in the protected forest, observed evapotranspiration is not available and hence is neglected in the paper. Nevertheless, we expect that 'including evapotranspiration information can in all probability further improve the forecasting accuracy' (please refer to lines 17 and 18 in page 9327).

- The question we try to answer in the paper is: Given the current rainfall and reservoir level measurements, what are the groundwater tables in future going to be? We considered 3 fixed leading times, i.e. 1 day, 3 days and 7 days, which is sufficient for taking intervention actions to maintain favourable hydrological conditions for conserving the ecosystem. Although the input-output correlation is not fully exploited, our simple methodology works well for answering the question, which is supported by the final model performance.

Based on the 'general comments', below revision will be incorporated in the revised manuscript:

- The paper from Taormina et al. (2012) will be properly cited.

- Reason for choosing a MIMO model and reason for neglecting evapotranspiration will be adequately explained.

Our reply to the 'specific comments' is embedded as follows:

- The abstract should clearly state what the novelty of the study is.

R: Well noted. We will further highlight the novelty in the revised version.

- Line 17-18, page 9318. Could you briefly elaborate on these objectives?

R: We will describe briefly on the objectives, e.g. to describe regional groundwater flow patterns, to understand local hydrological processes, etc.

- Line 1, page 9319. ": : : as most of the system forcings are less predictable." This sentence is not very clear.

R: For hydrological numerical models, system forcings include rainfall, evapotranspiration, hydrological variations at the boundaries, etc. Aforementioned forcings, especially the rainfall, are extremely sensitive, variable and unpredictable. We will insert proper explanation in the revised manuscript.

- Lines 21-24, page 9320. I would not use bullet points here; there is no need to emphasize these features of ANNs.

R: Fully agree. The bullet points will be removed.

- Lines 15-16, page 9321. What are the main characteristics of these three categories?

R: The difference of these 3 categories is explained to certain extent in Section 2.3 (lines 6 to 12 in page 9323). As training algorithms are not the research focus and a proper reference has been cited (Haykin, 1999), we would like not to extend the description in our paper.

- Line 9, page 9322. The activation function is used not only "for limiting the amplitude", but also for creating a mapping between input and output variables.

R: Well noted and fully agree. We will revise accordingly.

- Line 16-18, page 9322. This is not correct. The Universal Approximation Theorem (Hornik et al., 1989) states that "every continuous function defined on a closed and bounded set can be approximated arbitrarily closely by a Multi-Layer Perceptron provided that the number of neurons in the hidden layers is sufficiently high and that their activation function belongs to a restricted class of functions with particular properties".

R: Thanks for giving a much more rigorous definition. We will revise accordingly in the manuscript to be in line with this definition.

- Lines 6-9, page 9324. This part should be included in Section 3.1.

R: Fully agree. We will move this part to Section 3.1.

- Lines 9-12, page 9324. Which time lags did you consider?

R: 3 fixed time lags are considered, i.e. 1 day, 3 days and 7 days.

- Line 14, page 9324. It should be stated earlier that the adopted model architecture is MIMO. R: Well noted. We will insert the statement in Section 2.1 to highlight the adopted model architecture.

- Line 19, page 9324. What is the total number of observations?

R: Daily observed data are available for 2 years; hence total observation number is ~730.

- Lines 25-26, page 9325. Is it possible to include the information about the spillway from Upper Seletar reservoir?

R: Not possible at this stage as the information is not available due to confidentiality; it will be interesting to test in future if the data are made available.

- Lines 6-7, page 9326. I would not report the definition of RMSE and r–these metrics are very well known in the modelling community.

R: Despite being well known, we would like to keep these definitions in order to keep the 'technical note' complete and more self-explanatory.

- Table 1. Which period (i.e., training, cross-validation or testing) is being considered here?

R: The ANN performance is evaluated based on the testing data set. We will explain it more explicitly in the revised manuscript.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 9317, 2015.

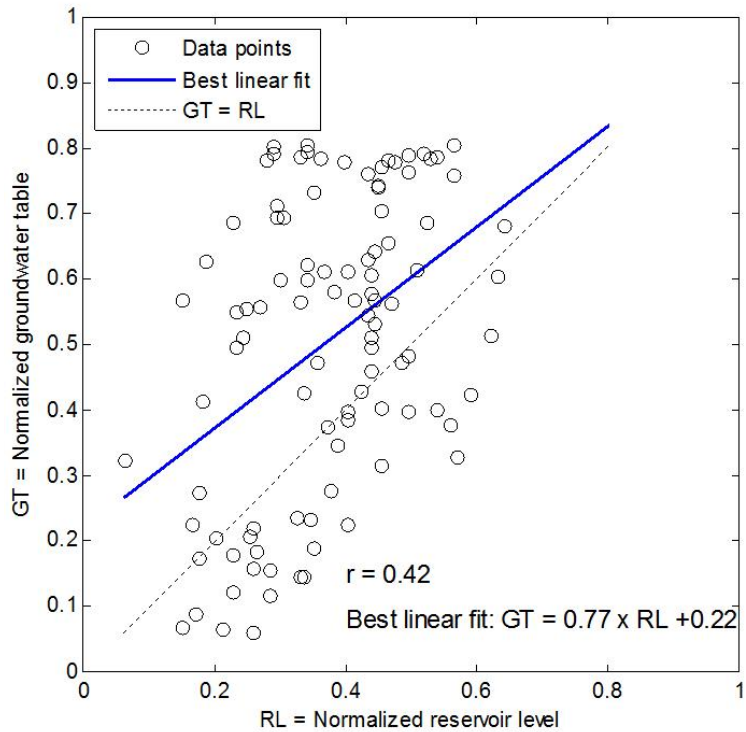

Fig. 1.

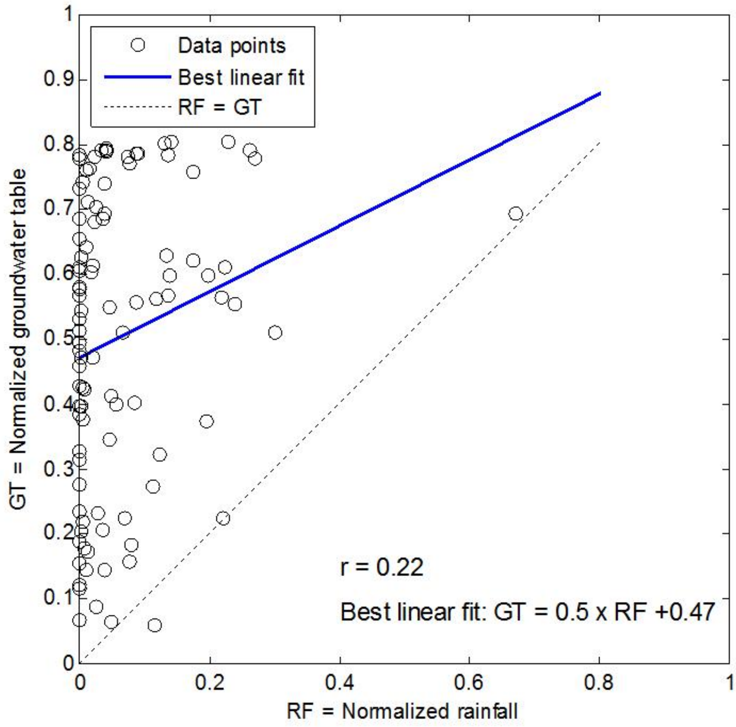
Data points
Best linear fit
RF = GT
r = 0.22
Best linear fit: GT = 0.5 x RF +0.47
GT = Normalized groundwater table
RF = Normalized rainfall
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1

Fig. 2.