Dear Dr. Dimitri Solomatine,

My co-authors and I wish to express our sincere appreciation for your attention and comments on our paper.

Most of the referees' comments are agreeable, while some remain debatable. Our reply corresponding to the comments is detailed in the 'Revision Notes' table below.

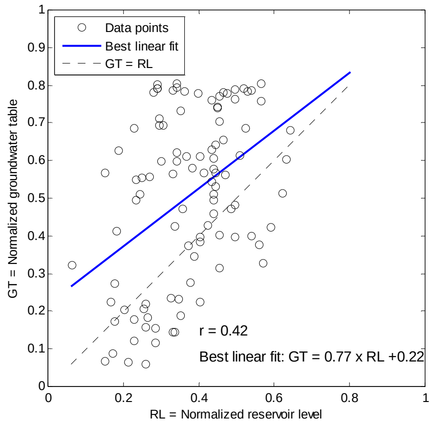
The comments have been incorporated in the revised paper, which is believed to have significantly improved. We herein submit it again for a possible publication.

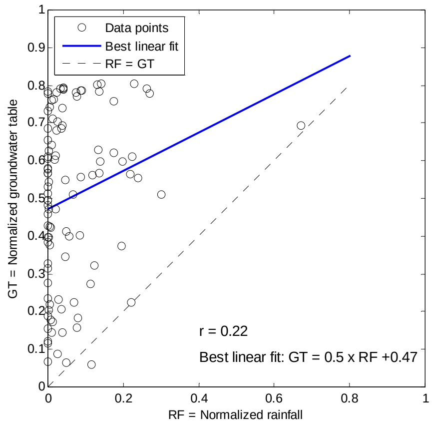
Thank you for your kind consideration.
We look forward to hearing from you.

Yours sincerely,
Yabin Sun

Revision Notes (Manuscript Number: hess-2015-294)

Anonymous Referee # 1		
No.	**Comments**	**Authors' Response**
01	Methods and tools. The use of ANNs for time series modelling has been a very popular research topic, and several advances have been proposed in the past decade – e.g., multi-objective calibration, modelling of the prediction uncertainty, input variable selection, improved calibration schemes, etc. (see Maier and Dandy, 2010). The methodology here adopted is an application of some well-known, existing tools, so it does not represent a methodological advancement.	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 42 to 66 in page 2). Our study, for the first time to the best of our knowledge, applied ANN to forecast the highly responsive groundwater table in a freshwater swamp forest; the methodology in our paper is straightforward and easy to implement. 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 in HESS as a 'technical note'.
02	Modelling problem. According to the authors, the main novelty stands in (1) the adoption of a short prediction horizon – justified by the fast dynamics of the water table, and (2) the use of exogenous variables (rainfall and reservoir water level data), instead of historical groundwater tables, as input to the ANN. A similar approach is adopted by Taormina et al. (2012), who modelled hourly fluctuations of the groundwater using rainfall and evapotranspiration data.	The case study of Taormina et al.'s paper is a coastal aquifer, where rainfall and evapotranspiration are considered as the major influencing factors (2012). The approach adopted by Taormina et al. is complicated, which 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. The paper from Taormina et al. (2012) is properly cited in the revised manuscript (please refer to lines 60 to 62 in page 2).
03	First, the results obtained with ANNs are not benchmarked, so it is hard to say whether they could be improved or whether the adoption of a non-linear model is needed. How do ANNs compare with a simple linear regression, for instance? Why using a MIMO model instead of four MISO ANNs? Second, I do not fully understand why evapotranspiration has been neglected – it should be influential in such a forested area. Third, I am not too convinced by the use of the input	ANN is chosen mainly due to its ability in regression analysis and the usage of more accessible variables in mapping the input-output relationships. 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 figures below). Hence, linear regression model is not suitable to serve our study purpose (please refer to lines 50 to 52 in page 2).

<table>
<tr>
<td></td>
<td>variables. Which time lags have been considered? Why not using multiple time lags or (temporal) aggregation of the input variables to fully exploit the input-output correlation?</td>
<td>




A MIMO model with 4 outputs is selected over 4 MISO ANNs mainly for 2 reasons: (1) it’s easy to implement; and (2) cross-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. This is added in the revised manuscript (please refer to lines 153 to 163 in pages 4 and 5).
The NSSF is of utmost importance for conserving a large proportion of the freshwater 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 it is excluded in the paper. Nevertheless, we expect that ‘including evapotranspiration information can in all probability further improve the forecasting accuracy’ (please refer to lines 227 and 228 in page 6).
</td>
</tr>
</table>

		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 predicting the groundwater tables in future given the current rainfall and reservoir level measurements. This is supported by the final model performance.
04	The abstract should clearly state what the novelty of the study is.	The novelty is further highlighted in the revised manuscript.
05	Line 17-18, page 9318. Could you briefly elaborate on these objectives?	The objectives are briefly described in the revised manuscript (please refer to lines 29 and 30 in page 1).
06	Line 1, page 9319. " as most of the system forcings are less predictable." This sentence is not very clear.	System forcings in hydrological models, such as rainfall, evapotranspiration and hydrological variations at the boundaries, are extremely sensitive, variable and unpredictable. This is further explained in the revised manuscript (please refer to lines 38 and 39 in page 2).
07	Lines 21-24, page 9320. I would not use bullet points here; there is no need to emphasize these features of ANNs.	Agree. The bullet points are removed in the revised manuscript.
08	Lines 15-16, page 9321. What are the main characteristics of these three categories?	The difference of these 3 categories is explained in Section 2.3 (please refer to lines 120 to 125 in page 4).
09	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.	Agree. It is revised accordingly (please refer to lines 107 and 108 in page 3).
10	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".	Thanks for giving a much more rigorous definition. It is revised in the manuscript to be in line with this definition (please refer to lines 109 to 112 in page 3).
11	Lines 6-9, page 9324. This part should be included in Section	This part explains the selection for ANN inputs, and we think it is more appropriate to be included

	3.1.	in Section 3.2.
12	Lines 9-12, page 9324. Which time lags did you consider?	3 fixed time lags are considered, i.e. 1 day, 3 days and 7 days.
13	Line 14, page 9324. It should be stated earlier that the adopted model architecture is MIMO.	Agree. This is inserted in Section 2.1 to highlight the adopted model architecture (please refer to lines 153 to 158 in pages 4 and 5).
14	Line 19, page 9324. What is the total number of observations?	Daily observed data are available for 2 years; hence total observation number is ~730.
15	Lines 25-26, page 9325. Is it possible to include the information about the spillway from Upper Seletar reservoir?	Not possible at this stage as the information is not available due to confidentiality; it will be interesting to test in future when the data are made available.
16	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.	Agree. It is removed in the revised manuscript.
17	Table 1. Which period (i.e., training, cross-validation or testing) is being considered here?	The ANN performance is evaluated based on the testing data. This is explicitly explained in the revised manuscript (please refer to lines 199 and 200 in page 6).

Anonymous Referee # 2		
No.	**Comments**	**Authors' Response**
01	The Authors say the used reservoir levels and rainfall as input to the ANN. It is not clear if they used lagged data or data at the step before the output.	The reservoir levels and rainfall are fed to the networks as input, while the output is the future observed groundwater tables after 1 day, 3 days and 7 days. It is revised accordingly in the manuscript (please refer to lines 149 to 152 in page 4).
02	It is not clear how the Authors assumed the architecture of the network and how they chose the input.	The reservoir levels and rainfall are chosen as the inputs as they are the major water source and driving force for the regional groundwater (please refer to lines 143 to 152 in page 4). A single hidden layer MLP is selected due to the universal approximation theorem (please refer to lines 109 to 112 in page 3), whereas the number of hidden neurons is determined by trial and error (please refer to line 60 in page 5).
03	Training data set seems to be too limited. I wonder if this may cause overfitting problems, as it seems to be.	An entire year's data are selected as the training data, which covers a complete annual cycle and is considered to be sufficient to train the network in a robust manner (please refer to lines 167 to 169 in page 5).
04	Looking at figures 3 and 5 as well as to table 1, it seems that there is an immediate decay of fitness, when the prediction is pushed at 3 and 7 days ahead. This may be related to overfitting problems or to bad selections of	Judging from Table 1, when move from 1 day to 3 and 7 days prediction, the performance of ANN does decay, but not to a drastic extent, e.g. at P1 from 5.4 to 8.2 and 9.9, at P3 from 5.2 to 6.6 and 8.6. Therefore, the overfitting problems may not be dominating in our study case and we've also used cross validation data trying to avoid overfitting

	the input.	(please refer to lines 171 to 174 in page 5).
05	The ANNs fail at reproducing peaks and dry periods, in particular for 3 and 7 days ahead prediction. Again, this seems to be related to an improper choice of the input or to a lack of information content of the input.	The peaks, especially at P4, are not perfectly captured because of the missing information of spillway discharge (please refer to lines 194 to 198 in pages 5 and 6). The dry periods are not well predicted because such a drought condition does not exist in the training data (please refer to lines 188 to 191 in page 5).
06	It is not clear if the Authors compared their model with a simpler one, i.e. linear models, like ARX or ARMAX. Maybe, these models may have similar performances with the proposed ANNs.	This is answered in No. 03 from Referee # 1.