Integrated versus isolated scenario for prediction dissolved oxygen at progression of water quality monitoring stations

A. Najah, A. El-Shafie, O. A. Karim and O. Jaafar

We would like to thank the referee for her/his valuable comments which clearly contributed to improve both the structure and the content of this manuscript. The pertinence and the constructive dimension of her/his review were greatly appreciated. We have addressed all their comments in the following point-by-point response. We have changed the manuscript accordingly. All changes made to accommodate the reviewer comments are underlined in the revised manuscript.

Referee's comments

The paper is about predicting dissolved oxygen (DO) at Johor River Basin by using Multi Layer Perceptron Neural Network Model (MLP-NN). For the purpose of the authors, two scenarios are proposed. In the first scenario (the isolated one), data about five input parameters including Temperature, Electrical Conductivity, pH, NO3 and NH3-NL and an output dissolved oxygen collected from four different water quality monitoring stations located at Johor River Basin are used for MLP-NN modelling purpose. In the second scenario (the integrated one), in addition to the five input parameters, predicted DO values at upstream stations are also used as a model input. The paper concludes that the predictions made based on the second scenario outperform those made by the first one, in terms of coefficients of correlation between the observed and predicted values.

The paper is clear and well written but there are some major questions that should be answered by the authors:

1. **The paper reports the results of an MLP-NN modelling effort based on collected data. However, the true cause and effect relation between the selected input parameters and the output parameter still requires justification. The comments made on Page 6074 is not convincing in terms of both necessity and sufficiency of the selected input parameters.**

Reply:

The choice of input parameters based on statistical correlation analysis is the most popular analytical technique for selecting input. The drawback of cross-correlation is that only able to capture linear dependence between two variables. Consequently, it can lead to

omission of important inputs that are nonlinearity related to the output. To evaluate the effect of input parameters on the model, two evaluation processes were used. First, a priori knowledge supported by statistical correlation analysis. The second assessment process was based on the prediction accuracy of water quality parameters.

---

2. **In a ms of this type I might expect some discussion of the 'stochastic' nature of the models, and the problems of relating the outcomes to understanding of the real chemical processes involved.**

Reply:

Discussion and information about stochastic nature of the problem under study and the chemical status are now improved in the revised version of the manuscript.

---

3. **Page 6071 Line 8: "Water quality is one of the main characteristics of a river, which purpose is not only for human water supply" It is not clear to me what do authors mean by this sentence?**

Reply:

We totally agreed with the referee, the right definition of Water quality has been rewritten in the revised manuscript. The term water quality is used to describe the condition of water, including its chemical, physical and biological characteristics. Water quality is one of the main characteristics of a river, which purpose is not only for human water supply (Dogan et al., 2009).

---

4. **Page 6073 Line 3: ". water quality parameters in terms of dissolved oxygen (DO), having the dynamic processes hidden in the measured data itself." What do authors mean by phrase "having dynamic processes hidden in the measured data itself". How does a dynamic process can be hidden in the measured data? Please, explain it.**

Reply:

Basically, the proposed neural network model in our study is NOT mainly relying on the physical and/or hydrological behaviour of the system in the study area, it is conceptually prediction model with consideration of the dynamic processes of the data. The major advantage of this method is the ability to predict the behaviour of systems without fully consideration or analytical prediction rules (hydrological/physical).

---

5. **Page 6076 Equation 2: Activation value xi is missing.**

Reply:

The above mentioned mistake has been corrected)

6. **Page 6077 Equation 4: constant multiplier 1/2 is missing.**

Reply:

The above mentioned mistake has been corrected

---

7. **Page 6077 Equations 5 and 6: The all + operators should be replaced by commas.**

Reply:

Owing to the reviewer's feedback, the Equations 5 and 6 have been rewritten and the + replaced by commas

---

8. **Page 6079: The back prop algorithm is well-known. So, it is not necessary to describe it in the paper. Just provide a 1-paragraph description with appropriate references.**

Reply:

The author highly appreciate the reviewer's feedback, BUT the authors are interested to introduce background of the back prop algorithm to give the readers (the hydrologist) intensive information.

---

9. **Page 6079: It better to rewrite the Equation 8 as:**

   <delta> = x^(k+1) - x^k = -[J^TJ + <lambda>I]^(-1)J^TE

   where <delta> defines amount of weight update.

Reply:

Owing to the reviewer's feedback, the Equations 8has been rewritten in revised manuscript

---

10. **Page 6101: Figure 2 can be found in any text, and so is not necessary.**

Reply:

The author highly appreciate the reviewer's feedback, BUT the authors are interested to introduce this figure to give the readers (the hydrologist) brief information about the main architecture of MLP-ANN

---

11. **Page 24: The claim in the last sentence of the page "..... offering a relatively fast algorithm ....." requires justification by giving comparative execution time performance results.**

Reply:

More rigorous justification by giving comparative execution time performance results has been added in the revised manuscript

---

12. **Page 6095: Instead of Table 3, the scatter plots of the inputs vs. DO will be more useful.**

Reply:

The authors highly appreciate the reviewer's feedback, after trying to adopt the scatter figure instead of table BUT we found that the table is more suitable to visualize the affect of weight values that caused the output to match the actual target values. Therefore, the authors request to keep the table in the revised version

---

13. **Page 25: The conclusion section is rather like an introduction section and should be re-written.**

Reply:

The conclusion section has been significantly improved in the revised manuscript

---