Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2017-571-RC1, 2017

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Modelling salinity in river systems using hybrid process and data-driven models" *by* Jason M. Hunter et al.

Anonymous Referee #1

Received and published: 25 October 2017

This study developed hybrid process and data-driven model to improve single-driven model performance for modelling salinity in river systems. Despite the paper is well organized and interesting to read, the manuscript in its present form has some weaknesses (mainly lack novelty and scientific findings). General and special comments: (1) The introduction and methodology (13 pages) are too long. Please make it concise and shorter and emphasize the novelty of the study. (2) The results and discussion (1 page excluding tables and figures) are too shorter. Please enrich it and offer more valuable analyses and scientific findings. (3) In Figure 5, the descriptions of "Below 30,000"ãĂĄ"Below 50,000"etc, are imprecise. Please replace it with 'Below 30,000 and above 15,000' etc. Besides, the symbol "sigma" easily causes readers' misunderstanding that the results of Model 1 are equal to the sum of Model 2-5. Please make

major revision for Figure 5. (4) In Eq. (3), so many researchers suggested that it needs use the index Gbench (or Coefficient of Persistence) by replacement of NSE to judge the good-of-fit of the model, when you applied a data-driven model, such as ANN on the basis of benchmark series. Please refer to the reference "Seibert, J. (2001). On the need for benchmarks in hydrological modelling. Hydrological Processes, 15(6), 1063-1064". (5) In Figure 8, the inputs are so important for data-driven model. Why you ANN model just has the exogenous inputs (ex, Lock 5: electrical conductivity with 5-day lag, Lyrup pump station: water level with 3-day lag, Lock 5: flow rate 5-day lag, and Lock 5: water level with 5-day lag), but hasn't the autoregressive input (Lock 4: salinity with 1-day lag). As known, the contributions of autoregressive input for model performance are higher than 80%-90%, however, the contributions of the exogenous inputs for model performance are only 10%-20%. Please explain it. (6) In Figure 8, how do you identify the time-lags of inputs? Please add your methods and results to demonstrate their suitability. (7) Section 2.3, the methodology for identification of most suitable model types is not scientific and imprecise. From the results of Table 2, the most suitable model types are identified based on the degree of data availability and process understanding. How do you quantify the degree of data availability and process understanding? Please make major revision of section 2.3 for enhancing the reliability of this method. (8) In Page 20, lines 438-440: Please add the results of trials with one to four hidden nodes to demonstrate that ANN with three hidden nodes preforms best on the calibration data. Providing the results of RMSE and NSE. (9) In conclusion: you stated the limitation of your methodology "While the approach has been developed specifically for the modelling of salinity in rivers, ….". In fact, this methodology just developed specially for modelling salinity in Murray River of South Australia. Hence, the title of this paper might be changed as follow. Modelling salinity in Murray River of South Australia using hybrid process and data-driven models.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2017-571, 2017.