

Interactive comment on “Importance of spatial and depth-dependent drivers in groundwater level modeling through machine learning” by Pragnaditya Malakar et al.

Anonymous Referee #2

Received and published: 26 September 2020

Comments on Manuscript Hess 2020-208-manuscript ‘Importance of spatial and depth-dependent drivers in groundwater level modeling through machine learning’

Groundwater is an important source of water, in particular for the transboundary areas of IGBM Rivers. This study used a liner regression approach based on dominance analysis and machine learning methods to identify the spatial and depth-wise drivers based on a large network of ground-based observations. Some interesting conclusions are found by the authors, including e.g. the groundwater level change is primarily influenced by abstraction and population in most of the IGBM; the machine learning methods can well simulate the groundwater level and the performance decreases from

C1

shallow to deep observation wells. The conclusions can be useful for groundwater management in the IGBM areas. However, the quality of this manuscript is not good enough for publication in HESS. The detailed comments are shown as the following.

Detailed comments:

1) Machine learning methods are popular over the years. The authors gave an introduction to machine learning methods used in GWL. I expect that more prevailing methods should be mentioned in the introduction instead of ANN and SVR. I also expect a comparison of these prevailing methods. 2) From the manuscript, it is difficult to see the originality of this study. For me, the only originality might be the use of a large network of monitoring wells to identify the spatial and depth-wise drivers. 3) Line 120: Although a large network of monitoring wells was used, the time resolution is rather coarse. Also can the authors show us the time series of monitored water levels? 4) For the dominance analysis, the independent variables seem dependent, such as groundwater withdrawals and population, temperature and potential evapotranspiration. Will this affect the results of dominance analysis? 5) Section 2.5: I am curious why the authors used two somewhat old-fashioned models including ANN and SVM. It is very easy to over-train these two types of models. I suggest the authors to use other models including LSTM. 6) Line 251: replace ‘has’ with ‘have’ 7) If the ML methods used in the study have some weakness regarding the low generalizability of the methods, risk of overtraining, why did the authors choose other machine learning methods? 8) Line 268: it seems to me that only half of the observation wells having correlations greater than 0.6 is not much. 9) Line 328: it is expected that the ANN and SVM models have limitations in areas with higher groundwater abstraction. 10) Figure 4: why large deviations in Indus? 11) Figure 6: how were the relative contributions calculated? Based on coefficient of determination?

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2020-208>, 2020.

C2