Reviews for hess-2023-192

Manuscript title: On the challenges of global entity-aware deep learning models for groundwater level prediction

Overview

The authors develop an entity-aware deep learning model for spatially and temporally continuous groundwater level prediction using a combined Long Short-Term Memory (LSTM) and Multi-Layer Perceptron (MLP) network. They rely on ground observations from 108 wells in Germany and other dynamic and static predictor data obtained from multiple sources to train, validate, and test the model. The authors also perform some interesting comparisons of four model variants, namely, the time series feature-driven model (TSFeat), environmental feature-driven model (ENVfeat), random static features (RNDfeat), and dynamic inputs only features (DYNonlyfeat). While there are some issues with spatial generalizability in the out-of-sample setting, the model shows satisfactory performance with the Nash-Sutcliffe Efficiency (NSE) > 0.8 in an in-sample setting.

Overall, the manuscript is generally well-structured, with detailed explanations of the methodologies and data. However, the following comments should be addressed before this manuscript is published.

Major Comments

- The authors should discuss relevant literature that incorporates process-based, machine-learning, or hybrid models and remote sensing data for groundwater level monitoring. The Introduction section should highlight the relevance of this topic more and refer to some of the negative impacts of groundwater depletion and why groundwater level monitoring is essential.

Hasan, M.F., Smith, R., Vajedian, S. et al. Global land subsidence mapping reveals widespread loss of aquifer storage capacity. *Nat Commun* 14, 6180 (2023). https://doi.org/10.1038/s41467-023-41933-z

Herrera-García, G. et al. Mapping the global threat of land subsidence. Science 371, 34–36 (2021). https://doi.org/10.1126/science.abb8549

Famiglietti, J. The global groundwater crisis. Nature Clim Change 4, 945–948 (2014). https://doi.org/10.1038/nclimate2425

Wada, Y. et al. Global depletion of groundwater resources. Geophys. Res. Lett. 37, 1–5 (2010). https://doi.org/10.1029/2010GL044571

Faunt, C.C., ed., 2009, Groundwater Availability of the Central Valley Aquifer, California: U.S. Geological Survey Professional Paper 1766, 225 p.

Li, B., et al. (2019). Global GRACE Data Assimilation for Groundwater and Drought Monitoring: Advances and Challenges. Water Resources Research, 55(9), 7564–7586. https://doi.org/10.1029/2018WR024618

Ahamed, A., Knight, R., Alam, S., Pauloo, R., & Melton, F. (2022). Assessing the utility of remote sensing data to accurately estimate changes in groundwater storage. *Science of The Total Environment*, *807*, 150635. https://doi.org/10.1016/j.scitotenv.2021.150635

Bierkens, M. F. P., & Wada, Y. (2019). Non-renewable groundwater use and groundwater depletion: a review. *Environmental Research Letters*, *14*(6), 063002. https://doi.org/10.1088/1748-9326/ab1a5f

- In addition to the predictor data summary, the authors should include a description of the predictor data listed in Tables 1 and 2 and the associated uncertainty. In Table 2, what does self-derived mean? Would snow water equivalent and soil moisture be helpful as additional predictors to capture the groundwater dynamics better? The authors should make a stronger case for selecting HYRAS 3.0 than other globally available land-surface models like the Global Land Data Assimilation System (GLDAS), which provides spatially and temporally continuous estimates of various hydrological processes acting as critical drivers of groundwater dynamics.

Rodell, M., et al. (2004). The Global Land Data Assimilation System, Bull. Amer. Meteor. Soc., 85(3), 381-394. https://doi.org/10.1175/BAMS-85-3-381

Razafimaharo, C., Krähenmann, S., Höpp, S. et al. New high-resolution gridded dataset of daily mean, minimum, and maximum temperature and relative humidity for Central Europe (HYRAS). Theor Appl Climatol 142, 1531–1553 (2020). https://doi.org/10.1007/s00704-020-03388-w

- The authors should include the model forecasts beyond January 2016. While it may be challenging to obtain in-situ groundwater levels between 2016-present, it would be interesting to observe how the model predictions compare to the GRACE- and GRACE Follow-On (GRACE-FO)-based total water storage changes (https://grace.jpl.nasa.gov/data/data-analysis-tool/) at a regional or national scale. This comparison would serve as an additional model validation and strengthen the manuscript.

- There should be an additional section (or a subsection within the Introduction) describing the study area and related studies on groundwater level changes. Also, the spatial distribution of the 108 well locations should be shown on a map.

- What are the 11 land cover classes in the CLC data? How are these used in the model? Can the categories be reduced by aggregating to a base class? E.g., crops aggregated to 'Agriculture,' urban/industry to 'Urban,' and so on? Is there no significant change in built-up or irrigated areas within the temporal domain of the model? The potential effects of land use changes on the model performance should be discussed. Also, the percentage of land use classes should be described in the Study Area section.

- The corresponding time series of the dynamic predictors for the two wells in Figure 5 should be added and tied up with the discussion related to the permutation feature importance.

- Evapotranspiration (ET) is the second largest component of the water cycle after precipitation (https://openetdata.org/what-is-evapotranspiration) and is a critical driver of groundwater use, which, in turn, is correlated to groundwater levels (Majumdar et al., 2020; 2022; Brookfield et al., 2023; Melton et al., 2021; Senay et al., 2022). Why didn't the authors include it as a dynamic predictor and instead rely on the potential ET (Table 2)? While the OpenET and the Landsat-derived actual ET products are currently available only over the conterminous United States (CONUS), the globally available 500 m MOD16 actual ET is available within the temporal domain of the model. Thus, the authors should justify the choice of their predictors.

Majumdar, S., Smith, R., Butler, J. J., & Lakshmi, V. (2020). Groundwater withdrawal prediction using integrated multitemporal remote sensing data sets and machine learning. *Water Resources Research*, 56(11), e2020WR028059. https://doi.org/10.1029/2020WR028059

Majumdar, S., Smith, R., Conway, B. D., & Lakshmi, V. (2022). Advancing remote sensing and machine learning-driven frameworks for groundwater withdrawal estimation in Arizona: Linking land subsidence to groundwater withdrawals. *Hydrological Processes*, *36*(11), e14757. https://doi.org/10.1002/hyp.14757

Brookfield, A. E., Zipper, S., Kendall, A. D., Ajami, H., & Deines, J. M. (2023). Estimating Groundwater Pumping for Irrigation: A Method Comparison. *Groundwater*. https://doi.org/10.1111/gwat.13336

Melton, F., et al. (2021). OpenET: Filling a Critical Data Gap in Water Management for the Western United States. *JAWRA Journal of the American Water Resources Association*. https://doi.org/10.1111/1752-1688.12956

Senay, G. B., et al. (2022). Mapping actual evapotranspiration using Landsat for the conterminous United States: Google Earth Engine implementation and assessment of the SSEBop model. *Remote Sensing of Environment*, 275, 113011. https://doi.org/10.1016/j.rse.2022.113011

Running, S., Mu, Q., Zhao, M. (2021). MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V061 [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. Accessed 2023-10-17 from https://doi.org/10.5067/MODIS/MOD16A2.061.

- Lines 50-60: While the proposed machine learning-based method of using multiple wells to develop an entity-aware global groundwater level prediction model is new, earlier studies have integrated remote sensing, climate, and hydrogeologic data in a machine learning framework for estimating annual groundwater withdrawals (Majumdar et al., 2020; 2021; 2022; Wei et al., 2022) and land subsidence (Smith & Majumdar, 2020; Hasan et al., 2023). For the studies on groundwater withdrawal estimation, a single machine learning model was trained and validated using in-situ pumping measurements from multiple wells across vast geographical areas (states of Kansas and Arizona in the U.S.). Thus, the authors should clearly convey that the novelty lies in groundwater level monitoring rather than the entire hydrogeology domain.

Majumdar, S., Smith, R., Butler, J. J., & Lakshmi, V. (2020). Groundwater withdrawal prediction using integrated multitemporal remote sensing data sets and machine learning. *Water Resources Research*, 56(11), e2020WR028059. https://doi.org/10.1029/2020WR028059

Majumdar, S., Smith, R., Conway, B. D., Butler, J. J., Lakshmi, V., & Dagli, C. H. (2021). Estimating Local-Scale Groundwater Withdrawals Using Integrated Remote Sensing Products and Deep Learning. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 4304–4307. https://doi.org/10.1109/IGARSS47720.2021.9554784

Majumdar, S., Smith, R., Conway, B. D., & Lakshmi, V. (2022). Advancing remote sensing and machine learning-driven frameworks for groundwater withdrawal estimation in Arizona: Linking land subsidence to groundwater withdrawals. *Hydrological Processes*, *36*(11), e14757. https://doi.org/10.1002/hyp.14757

Wei, S., Xu, T., Niu, G.-Y., & Zeng, R. (2022). Estimating Irrigation Water Consumption Using Machine Learning and Remote Sensing Data in Kansas High Plains. *Remote Sensing*, *14*(13), 3004. https://doi.org/10.3390/rs14133004

Smith, R., & Majumdar, S. (2020). Groundwater storage loss associated with land subsidence in Western United States mapped using machine learning. *Water Resources Research*, *56*(7), e2019WR026621. https://doi.org/10.1029/2019WR026621

Hasan, M.F., Smith, R., Vajedian, S. et al. Global land subsidence mapping reveals widespread loss of aquifer storage capacity. *Nat Commun* 14, 6180 (2023). https://doi.org/10.1038/s41467-023-41933-z

Minor Comments

- Line 94- Fix typo: followed by a brief *introduction.*
- What are the spatial resolutions of the predictor data listed in Table 2? How do the authors map the groundwater wells to these gridded raster datasets?
- Report other error metrics like the coefficient of determination (R^2), root mean square error (RMSE), and the mean absolute error (MAE) in Table 3.
- The CLC acronym is not defined.

- Do the authors scale all the features? What scaling is applied?
- Lines 235-240: For the out-of-sample setting, are the scores only calculated for the testing period of a well that has been left out of model training? Why not calculate the score for the entire period?