This paper presents an interesting and comprehensive study on the use of machine learning (ML) algorithms for predicting surface soil moisture content at a global scale. The authors use eight different machine learning algorithms trained with in situ data to predict soil moisture at the top 5 cm of the soil across the globe at a spatial resolution of 1 km and a temporal coverage of 2000-2018. The performance of various ML algorithms was evaluated. They also created 10 ensemble models from five optimized base models with the highest performance metrics and propose KNR_RFR_XB as the best ensemble for soil moisture prediction.

The paper provides a thorough description of the model building and validation process. However, prior to publication, the following questions should be addressed:

Abstract

Please specify which soil layer, period (coverage), temporal and spatial resolution were considered in the generation of the machine learning models.

Line 15-20: Instead of using the term "best data product", it would be more appropriate to use an adjective such as reliable or well-validated, as the data product from machine learning may be produced empirically without explicit knowledge of the physical processes involved. This may introduce additional uncertainties into the final output. (O., S., Orth, R. Global soil moisture data derived through machine learning trained with in-situ measurements. Sci Data 8, 170 (2021). https://doi.org/10.1038/s41597-021-00964-1)

Line 24-34: Please clarify what it means to have the best performance on "test_random" set, "test_temporal" and "test_independent-stations" and reword this sentence to focus on the meaning rather than the jargon terms. Also, please use full wording before notation (XB = extreme gradient boosting).

Introduction

Please provide citations for all claims or statements that require them (e. g., Line 42:45, Line 47:49, Line 52:54)

Line 65-68: The statement "Another advantage of using ML techniques is that they can help to reduce the uncertainty…" may not be accurate, as machine learning may introduce additional uncertainties into the final output (see response to line 15:20). It would be helpful to discuss this issue in more detail and provide references to support your argument.

Line 75: Please specify which soil layer, period (coverage), temporal and spatial resolution were considered in the generation of the machine learning models.

Data

Line 91:92: Did you filter out NAN values from the predictor data based on the in-situ soil moisture data? If not, what did you do? Please make the statement clear.

Line 104: Why is "Year/ DOY " a predictor?

Line 140: To improve readability, could you please ensure that all temporal coverages are consistent (2000-2018). I noticed that line 117 mentions the years 2000-2019, but I believe only the years 2000-2018 were used for analysis.

Line 232: " The full data contained a total of 735,475 registries (each sample contained nineteen predictor variables) and was differentiated in the training set and three different test sets, based on the following strategy (also described in Table 2)"

I am finding this a bit confusing and would appreciate some clarification. Is it correct that the combined data of all predictors and in-situ data is 735,475, and that this data was split into train and test samples as shown in Table 2?

Additionally, I noticed that Table 1 shows fewer than 19 predictors. Could you please explain how you arrived at the number 19 predictors? Is it necessary to use all 19 predictors, or could feature or predictor importance be performed to determine the various influences of predictors and select important features if necessary?

Methodology

Line 279: Introduce the full wording e.g., coefficient of determination before using the associated notation $\mathbf{r^2}$.

Line 350: Could you please explain the reason for choosing both **r** and $\mathbf{r^2}$ as evaluation metrics? What is the significance of using both in this context?

Line 384: Arranging Table 3, as well as any other relevant tables, in ascending or descending order based on performance can improve readability.

Line 430: 470: Could you please explain why the Kruskal-Wallis test is necessary when you can already compare models using various performance metrics such as RMSE and $\mathbf{r^2}$?

Line 480: Please move Figure 5 into supplementary information.

In addition to addressing these questions, it would also be helpful to compare the performance metrics and uncertainties of your proposed ensemble model with other existing soil moisture products or methods. This would provide more context for your results and help readers evaluate their significance. It would also be interesting to discuss the limitations and assumptions of your machine learning algorithms and how they affect the reliability and applicability of your soil moisture predictions. Finally, it would be useful to suggest future work or research directions to improve or extend your study.