Comment on nhess-2021-110

Anonymous Referee #2

Referee comment on " Applying machine learning for drought prediction using data from a large ensemble of climate simulations" by Elizaveta Felsche and Ralf Ludwig, Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2021-110-RC2, 2021

Title: Applying machine learning for drought prediction using data from a large ensemble of climate simulations
Author(s): Elizaveta Felsche et al.
MS No.: nhess-2021-110
MS type: Research article: First review
Special Issue: Recent advances in drought and water scarcity monitoring, modelling, and forecasting (EGU2019, session HS4.1.1/NH1.31).

RC2: 'Comment on nhess-2021-110',

This study present a methodology for drought prediction at seasonal scale using machine learning algorithms. The study is treating a highly relevant subject with the usage of a novel methodology based on machine learning for droughts predictions. However, the issue with machine learning and climate is the length of the climate records, which does not allows to build AI models. Therefore, this study proposes a methodology fully based on a down scaled ESM. The study is interesting however, I have major comments listed bellow, in particular, at that stage it is very hard to evaluate properly the manuscript since the data/method is not clear enough:

1) the method (if I understood it correctly) is fully based on model data, therefor it is not really a study about prediction but according to me it is only potential predictability, since this study does not demonstrate any skill in predicting observed past climate in the two regions of interest, but only the ability to forecast the model climate. The paper should be much clearer about this, for example the title and the abstract should use the term "perfect model framework" and/or potential predictability.

Authors thank the reviewer for his/her useful suggestion. The title of the revised version of the manuscript will include it:
"Applying machine learning for drought prediction in a perfect model framework using data from a large ensemble of climate simulations"

2) The method description is very unclear about the prediction aspects. What are the target month analyzed? From which start date? For example, it is really confusing to me to predict SPI1 with one month lead time for different seasons. What do you mean here? Do you mix all together the start date of March (to predict the SPI1 of April), April (to predict the SPI1 of May) and May (to predict the SPI1 of June)? Or do you predict SPI1 integrated over MAM, but in this case, to my understanding it is not SPI1 but SPI3. In any case the methodology should be much clearer about this point, at this stage I cannot evaluate properly the manuscript without this clarification.

The authors apologize for the misleading description and will work on clarifying this in the manuscript. The study predicts SPI1 with a lead time of one month. To predict e.g. SPI1 in April of 2000, the data for twelve months before the event is used as input, this is SPI1 and other variables for the period April 1999 – March 2000. For the calculation of SPI1 in April only precipitation for the month of April is used.

3) "The data from the years 1957 - 1999 was used as training data, the years 2000-2005 were used for the testing purpose." Do you mean that the score calculation is performed only for 6 years from 2000 to 2005? This is a far too short period for any skill assessment. Usually, in seasonal prediction the skill is assessed over the whole hindcast period (1957-2005), using cross validation to construct the prediction.

We are using a large ensemble of 50 members for the study. This means that the period 2000 – 2005 is available 50 times, resulting in 250 model years for the test dataset. The same is true for the training dataset: There the period of 1957-1999 is to be multiplied by 50, resulting inf 43*50=2150 model years. The authors apologize for the misleading description and will work on clarifying this in the manuscript. The authors are aware that cross validation would add value to the study, however given the huge amount of variables and the fact that the overall analysis includes 2500 model years, it would require huge computational resources, that are not available.

4) The discussion does not mention at all the main limitation of this study according to me: at that stage the authors have demonstrated some ability to predict a model using AI, but we don't know how to use such method for real prediction. Would it be possible to apply your model on observation and then verify its skill? If yes, it should be included in the study and if not this should be clearly mentioned.

Authors thank the reviewer for his/her useful suggestion. The authors would argue that the immediate application of the framework on observation is not possible, due to the fact that observational data usually lacks a multitude of variables which were used as input in this study e.g. Heat Fluxes, radiation, etc. The objective of this study was not to develop a framework that can be applied on observation, but to use the large amount of events provided by the large ensemble for prediction. The results obtained by shapely value calculation are of high importance for the choice of variables for a development of a model which could be applied to observational data.

Typos:
This study uses the monthly sea level pressure (pr)

The stong influence of ps/psl and NAO shows the influence of the atmospheric pressure

Typos will all be fixed in the preparation of the revised manuscript version.