Hydrology and Earth System Sciences Discussions
Hydrol. Earth Syst. Sci. Discuss., author comment AC2
https://doi.org/10.5194/hess-2021-600-AC2, 2022

Reply on RC2

Vitali Diaz et al.

Author comment on "Machine-learning approach to crop yield prediction with the spatial extent of drought" by Vitali Diaz et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-600-AC2, 2022

Response to the reviewers

Thank you very much for the review and valuable suggestions. We describe below how we addressed each comment. Our responses begin with the word '[Reply]'. We have numbered each reviewer's comment for ease of reading, and this numbering is indicated between brackets.

Reviewer 2
Thank you for this proposal.
[1] Major comment 1: "Machine-learning approach to crop yield prediction with the spatial extent of drought" doesn't exactly reflect the exact object of your work as drought is mainly considered as agricultural drought through the SPEI index (ie the difference between precipitation and evapotranspiration) used as "drought indicator" in your study. Please modify in accordance the paper title: "...with the spatial extent of agricultural drought" and precize your definition of drought in the introduction Section. Please also explain if soil characteristics (structure, nature) and farmer operational pratices could be taken into account in the ML approach in the futur and how they could influence the "agricultural drought area" calculation? Please discusse it.

[Reply] Thank you very much. Although it is true that drought indices are originally created to analyse a specific type of drought, by considering different aggregation periods it is also possible to identify other types for which they were originally created. In our study this is the case. For this reason, in the title we do not emphasise agricultural drought, because we are not using only aggregation periods usually used for agricultural drought.
Our correlation analysis between crop yield and drought areas derived from different aggregation periods indicates that different types of drought (i.e. meteorological, agricultural, and hydrological) affect (impact) the crop yield to varying degrees throughout the months of the crop period. To build the ML models, this level of affectation could be taken into account by using the different hydro-meteorological variables or selecting different aggregation periods of the meteorological variables, as in this case.
Due to the above, we trust that the title is correct because our analysis is not focused on a type of drought, although it is focused on a specific type of dependent variable, that is, the crop yield.

Regarding the characteristics of the soil to be considered within the ML approach, of course they can be considered. Whether to include them depends largely on the availability and accuracy of this information and the need to do so. In our case, due, in large part, to that the cultivated cropland is high with respect to the area of the region, it was not necessary to include this type of information in our ML approach.
In future studies, we envision different topics in this regard. Here are some examples. The degree of influence of anthropogenic factors, such as farmer operational practices, other types including soil conditions, or the impact of natural phenomena such as drought, could be included in the ML approach presented.
One way to implement the above is as follows. Three or more types of input could be classified, anthropogenic, natural, and different types of combinations. For each set of inputs, the variable selection analysis could be carried out and thereby identify which are the ones that mostly manage agricultural production. Subsequently, the ML models could be built following our proposed approach, i.e. ANN models and equations.
Also, optimisation analysis could be carried out, the best practices (combination) could be found. This type of analysis can be done at different spatial and temporal scales, if the data allows it.
Another line that we see a lot of development in the future is the construction of ML models considering the study area totally discretised in cells. For each cell, the whole approach can be carried out. The availability of spatial data is crucial in this type of analysis, advances in remote sensing and the different earth monitors developed in the last decades could facilitate the implementation of this type of methodologies with the use of more advanced ML approaches.
This research can also be extended into the future to analyse the climate change scenario, either to elucidate the consequences or to find the best crop management practices to face it.
An analysis that is not very complicated to implement and that would, however, include anthropogenic factors or other variables such as the type and conditions of the soil, would be, for example, to weight the drought areas with factors calculated with the additional variables. In this way, the drought areas would be modified to a greater or lesser extent, increasing or attenuating the effects of the drought.

[2] Major comment 2: the correlation analysis seems very tricky and is hard to read. In general the figures are not pleasant and easy to read (Figures 6-7-8-9). There is too much text in the figures and tables.

[Reply] Thank you. We have modified and improved the description of the correlation analysis methodology as well as the presentation of the results. We have also included an additional figure in our manuscript to show the general scheme of how the input and output variables are tied.

[3] Minor comment 1: The proposed approach is a large-scale approach is disconnected from the drivers of the local scales of agricultural management (plot, farm or irrigated schemes) and masking the local variability of soil moisture. How does this complex approach provide added value for these local managers? Please discuss it.

[Reply] Thank you. The use of methodologies that consider the variables you mention, agricultural practices, soil properties and condition, among others, are ideal, however, this is not always possible. Our study presents a methodological alternative for predicting crop yield. In the study area, there are current approaches for crop yield calculation, one based on field visits and a monitoring system based on remote sensing inputs. The drawbacks and advantages are indicated in the Introduction. Our methodology is a complement to these two mentioned tools and provides crop yield predictions that can be compared with the current tools, with the difference that our ML approach produces results before the harvest (i.e. prediction)
Our analysis could be extended further. In subsequent studies we consider that an

analysis of irrigation practices could be made, where the best practices could be identified. Our results indicate that the increase in drought area is highly correlated with the decrease in crop yield. A more detailed analysis will make it possible to identify the best agricultural management practices, identify sub-regions more/less vulnerable to the effects of the different types of drought, and detect various demands on water resources throughout the different farming systems. See also the answer to the comment [1].

[4] Minor comment 2: Up to line 308 there is mention of "drought indicator" without really knowing what it is. Could you move the data section at the beginning of the paper?

[Reply] Thanks. For a better understanding and reading, we have reorganised sections 2 and 3, now Data is presented first in Sect. 2 and then Methodology in Sect. 3. The text in both sections has also been adjusted accordingly.

[5] Additional comment: The impact of a drought on agriculture cannot be measured by assessing crop yields because it depends on farming practices and cropping systems. In this approach, the joint result of the agricultural practices currently practiced during a period of plant water stress is evaluated.

[Reply] Thank you. We have updated our text.

Kind regards,
Vitali Diaz
on behalf of the authors