Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-364-RC1, 2020


Hydrology and Earth System Sciences
Discussions



# *Interactive comment on* "Field-scale soil moisture bridges the spatial-scale gap between drought monitoring and agricultural yields" *by* Noemi Vergopolan et al.

**Hannah Kerner (Referee)**

hkerner@umd.edu

Received and published: 7 September 2020

The manuscript presents maize yield prediction results based on a model that combines hydrological, meteorological, and remote sensing features in a random forest regression. The authors performed feature importance and sensitivity analyses to determine which features influenced maize yield predictions the most and which types of features contributed most to yield prediction accuracy. Overall, the paper is well written and presents useful findings for future studies.

My main criticism for this manuscript is that the random 80%/20% train/test split of yield observations and use of the test set in model optimization are likely overestimating the



Printer-friendly version

Discussion paper

performance of the test set compared to the performance that could be expected in practice. The 80%/20% random split across all available yield observations across 70 districts and 8 years does not ensure the test set is independent from the training set. For example, observations of the same district in different years probably have high correlation, as are observations in the same year but different districts. The authors are also optimizing the model for their test set by performing RFE using the test set (they should instead use a third validation set or cross-validation with the training set as was done in 2.3.2). The goal (I assume) of this study is to present a method that can be used to predict maize yields in future years – for the test set to be representative of the performance in this setting, there should be no overlap in years present in the training and test sets (e.g., the training set could include observations from 6 of the 8 years and the test set include observations from the other two years).

Additional comments:

- Why did the authors use 250m MODIS instead of 30m Landsat-resolution NDVI? The latter is much closer to the field scales observed in Zambia and would have the same resolution as the HydroBlocks simulations.

- What computational, time, or cost resources would be required to use the HydroBlocks model operationally to predict maize yields in all of Zambia in future years? Is this feasible to do operationally? (Also, note that sometimes the authors write “Hydroblocks” and sometimes “HydroBlocks”.)

- The ESA-CCI 2016 land cover map is used as a cropland mask, and the authors assume all cropland is maize. How valid is this assumption (i.e., what percentage of crops grown in Zambia are typically maize?)? What is the accuracy of this land cover map across Zambia? (I have not seen promising results for this map in Africa.) This could affect the authors’ interpretation of shrubland percentage as an important indicator of maize yields.

- The colormap in Figure 9 (right) is missing a title.

- The Figure number is not shown on line 434.

- “Unknowing” should be “Not knowing” on line 60.

- "its" should be "their" on line 2.

- Lines 251-252: the second set of MAE and R-squared values should say they are for the training set (only the test set is mentioned).

---

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-364, 2020.



Printer-friendly version

Discussion paper