Reviewer #2

In this work, authors developed a yield prediction system by using a hyper-resolution land surface model HydroBlocks and Random Forest regression. Consequently, authors assessed the sensitivity of predictors towards monitoring droughts. The manuscript is well written with results that can contribute to the ongoing efforts in this area of research. In summary, the manuscript can be accepted for publication once authors address my comments below:

We thank the reviewer for their time and thorough assessment.

1) It would be nice to see some validation of HydroBlocks simulations. I understand getting root zone soil moisture can be difficult. An alternate comparison with SMAP Level 4 root zone soil moisture product can provide an overview of the quality of HydroBlocks simulations. It may also address, to an extent, one of the limitations of input data uncertainty on yield prediction.

We agree with the reviewer that evaluation statistics on the soil moisture itself would be interesting. However, there are no available root zone observations for the study area. Regarding the SMAP L4 suggestion, SMAP L4 root zone is at 1 m depth, while HydroBlocks is 1.5 m, but more importantly, their root zone soil moisture product is the result of a model simulation which is subjected to similar, or even more, data uncertainties as HydroBlocks given its coarser spatial resolution. Also, initial validation results in the Continental United States, with in-situ observations, show that at the surface uncalibrated HydroBlocks can perform better than SMAP L4 (Vergopolan et al., 2020). In our opinion, unfortunately, no suitable validation product exists for this region that would allow us to perform on observation drive independent validation of the simulated root zone soil moisture.

Besides the challenges evaluating the absolute soil moisture values, in our approach, the HydroBlocks soil moisture - as the other predictors - are normalized (considering the period between 2000-2018) prior to training the RF yield model. Thus, comparing the absolute root-zone soil moisture values between (2015-2018) may not be representative of the long-term normalized dynamics that are actually being implemented in the yield prediction.

Vergopolan, N., Chaney, N. W., Beck, H. E., Pan, M., Sheffield, J., Chan, S., & Wood, E. F. (2020). Combining hyper-resolution land surface modeling with SMAP brightness temperatures to obtain 30-m soil moisture estimates. Remote Sensing of Environment, 242, 111740.

2) How is the calibration of HydroBlocks carried out? Authors may have to provide this information.
The model is not calibrated, this will be clarified in the manuscript.

3) Apart from underestimation for high yields, there is also overestimation of yields below ~500 kg/ha. Does this indicate that RF produced yield simulations with lower variability than the observed data? Can authors comment on the overestimation?

Indeed the random forest model resulted in a slight overestimation of low yields, we believe this is a consequence of random forest models being somewhat limited in predicting extreme values, as its final estimates are an average of an ensemble of decision trees. This makes it difficult to correctly estimate the observed yields' outlier values as is the case for many RF applications. The discussion on the random forest limitation will also be expanded to discuss the low yields overestimation.

4) It is surprising to see NDVI not contributing strongly to the prediction of yields. A lot of research on yield prediction depends on NDVI data as a predictor. Can authors throw some more light on this outcome? Is it because there is redundancy in the variance explained by soil moisture and soil temperature compared to that of NDVI towards estimating yields? Since analysis is carried out at monthly resolution, presence of clouds may not be a concern.

In Figure 3 (case 3), our results show that NDVI contributes strongly for yields in the absence of hydrological variables. Certainly soil moisture, surface temperature, and NDVI share some variability. In fact, besides predicting yields, NDVI has been extensively reported in the literature as a predictor for soil moisture as well. What our modeling experiment shows, however, is that soil moisture tended to offer more added value than NDVI. In fact, in the RFE experiment, removing NDVI did not result in loss of model performance, whereas removing hydrological variables did reduce model performance. We agree that cloud coverage was not a major limitation for MODIS, but instead, the limitation of visible sensors in capturing under canopy plant-soil-water dynamics with soil moisture and surface temperature, as discussed in the manuscript. We have added this important interpretation of the revised paper.

5) Spatial scale plays a significant role in driving the soil moisture processes at 30 metre resolution. Can authors assess on the impacts of spatial scales on the yield predictions? Besides, authors may have to describe how various datasets are processed to a consistent spatial resolution.

The yield model is trained based on observations and predictors aggregated at the district scale, and under the assumption that this relationship holds at the fine-scale, the model is applied to predict yields at the fine-scales. I assume the reviewer is wondering about the impact of the hydrological variables' spatial scales on the yield prediction. While coarser-resolution hydrological data would not change the performance of the RF much (as the data is trained aggregated to the district-level), we expect it would result in "smoother" field-scale yield predictions. This is expected because hydrology plays an important role in the spatial patterns of the predicted yields, as soil moisture and soil temperature were the strongest predictors (Figure 3). A discussion on the drivers of the spatial variability in yields is presented in section (4.3). We now suggest further evaluation of the role of spatial scaling for future research in this area in the revised paper.

To obtain a consistent “spatial resolution” for training and predicting, we used a spatial area-weighted average considering only areas where the 30-m land cover is classified as cropland, as described in 3.2.1 section of the manuscript. Thus, despite the model being trained at the district-level, it used the best estimate at the cropland areas for each district. The same approach was used to predict yields at the fine-scales

6) Authors may present time series plots of observed and simulated productivity in addition to the existing results.

A time-series of the observed and predicted soil moisture was not included in the manuscript mainly because PHS training/testing data is not continuous in space and time. Data is obtained for some districts and some years, and thus, time-series comparisons for a single location results in sparse points in time. Consequently, we opted for a scatter plot instead, which allowed us to evaluate the results despite the discontinuities. Nonetheless, in the updated manuscript, we will include a time series of the predicted for some pixels distributed across Zambia.

7) Minor: Figure 8: How did authors do a pairwise comparison of October maximum temperature and April root zone soil moisture? It is not clear as of how Figure 8 is generated from the description.

In Figure 8a, for example, we compared the yield anomaly (of a given season) with the respective October maximum temperature and April root zone soil moisture (for the same season), the hexbins show the mean anomaly yield values when this surface temperature and soil moisture was observed. We will clarify the figure explanation updating the sentences in lines L. 333 to the following: *“To gain further insight into the conditions (e.g., hot and dry vs. hot and wet) associated with yield losses, in Figure 8 we compared how the yield anomalies co-varied with pairwise drought indicators. For example, Figure 8a shows what is the mean yield anomaly associated with its respective October maximum temperature and April root zone soil moisture.”*

8) Minor: Check the units of productivity in the manuscript. At some places, it is given as kg kg/ha instead of kg/ha.

Corrected.