Response to comments

Paper #: *essd-2019-118*
Title: *Mapping the yields of lignocellulosic bioenergy crops from observations at the global scale*
Journal: *Earth System Science Data*

Reviewer #1:

Comment #1

The authors use a machine learning technique (random forest-RF) to develop an upscaled global (0.5 x 0.5 degrees) yield data set for five bioenergy crops. To justify how realistic this empirically-derived global bioenergy yield map, the authors further compare their product with the yield map used by the Integrated Assessment Models (IAM). In general, I agree with the authors that this dataset can become potentially a useful product for either benchmarking the global crop models (e.g. LPJ alike models) or being as input to IAMs. However, I think the method and results of this manuscript suffer from the following major weaknesses, which cannot make me convinced that this is a reliable product.

Response #1

We thank the reviewer for the comments and suggestions. Please see the detailed point-by-point responses below.

Comment #2

1. The authors disregard the details of temporal resolution and coverage of training data sets.

Response #2

We agree that the temporal resolution and coverage of the training dataset are important for training the machine learning model given the temporal variations of climate conditions. Therefore, as suggested, we analyzed the sampling time in the training dataset. There are ~30% of the yield observations without reported sampling year in the original dataset and also ~30% in the aggregated 0.5-degree data used for random forest training. We thus arbitrarily set the 2 years before the publication year as the sampling year for the yield observations without reported sampling years (e.g. set 1997 as the sampling year if the reference paper was published in 1999). The frequency of the sampling years in the 0.5-degree data used for random forest training is shown in **Fig. R1**. The sampling years range from 1969 to 2016 with a median year of 1999.

We then derived temperature (T), precipitation (P) and short-wave radiation (SW, from CRUNCEP because BESS SW starts from 2001) and soil water availability index (WAI) at the sampling year for each grid cell and re-trained the random forest (RF). However, the OOB R^2 is **<u>0.54</u>**, **<u>lower than</u>** the original value of **<u>0.63</u>**. Possible reasons may include: 1) RF training may largely respond to the spatial gradients of climate and soil conditions, and thus the contribution of temporal variation may be low; 2) Climate conditions at the sampling year may be a good predictor of yields for annually harvested herbaceous crops, but yields of woody crops like eucalypt, poplar and willow may also be impacted by the previous years in the whole growing cycle. Unfortunately, there are only about 18% observations with both reported harvest year and age, impeding the derivation of the mean climate conditions during the whole growing cycle.

In addition, using the climate conditions at the sampling years also changed the variable importance (**Fig. R2**) compared to the original one (**Fig. 2a**). Precipitation is no longer an important contributing variable while contributions of the other variables are more or less similar to those in the original trained RF.

We will add this test as a sensitivity test and discuss accordingly in the revised manuscript.

Figure R1 Histogram of sampling years in the yield observation grid data used for random forest training.

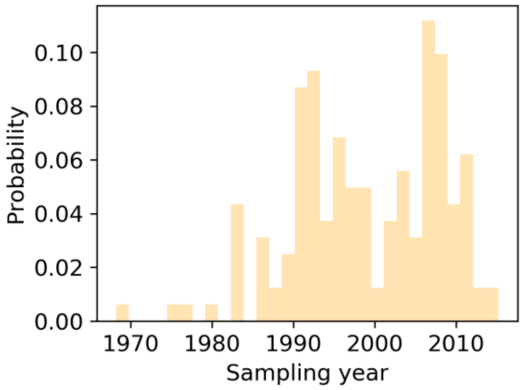
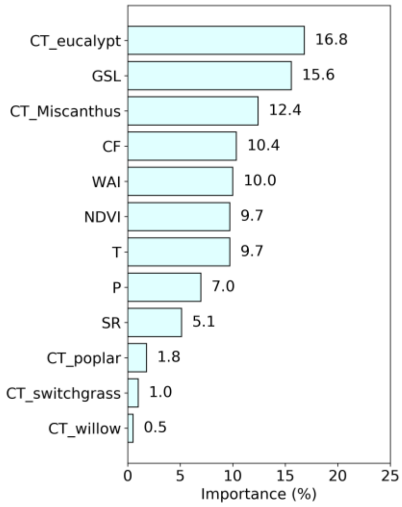


Figure R2 Variable importance in the trained RF model using climate conditions at the sampling years.



Comment #3

2. The authors haven't provided good reasoning for how they decided the training data sets. The temperature dataset in CRUNCEP is similar to the CRU data set, which is based on observations, but precipitation has less good reliability. Also, why do authors choose satellite-based short-wave radiation? Does the median value in the high-resolution dataset have any advantages over the 0.5-degree data set (e.g. the CRU sunshine hours)? The water available index is a model-derived data set, but actually, there should be some satellite-based dataset to indicate soil moisture. In a word, I think the authors should give strong reasoning on why they have chosen their training data sets.

Response #3

We understand the reviewers' concerns, and we will add the reasons as well as more sensitivity tests to explain why we choose these climate forcing data in the revised manuscript (see below).

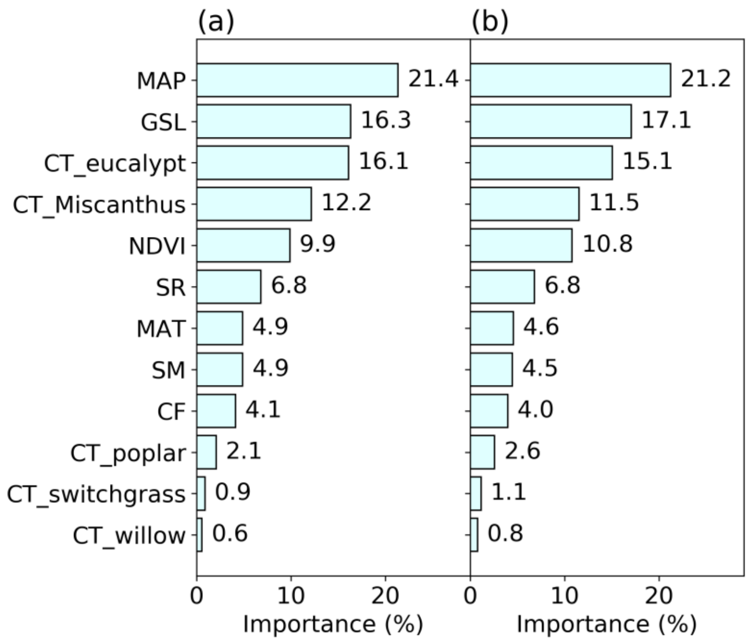
1) The CRUNCEP data is based on CRU climatology but only used NCEP to generate the diurnal and daily variability (*Viovy, 2017*; ftp://nacp.ornl.gov/synthesis/2009/frescati/model_driver/cru_ncep/analysis/readme.htm). We used **annual** precipitation in the random forest regression, and thus it should be the same as that from CRU.
2) In fact, the high-resolution datasets didn't help much in improving the Random Forest training. As discussed on **L313-325**, we tried higher-resolution (0.01 degree) MAP and MAT data from

WorldClim and trained the RF at higher resolution (0.01 degree) but the OOB R^2 didn't improve. We will further emphasize this point in the revised manuscript

As for the radiation data, shortwave radiation (SR) from CRUNCEP was simply converted from the cloudiness provided by CRU based on the calculation of clear sky incoming solar radiation as a function of date and latitude of each pixel (*Viovy, 2017*). By contrast, SR data from BESS was computed based on a series of forcing data from Terra & Aqua/MODIS Atmosphere and Land products, including "solar zenith angle from MODIS Atmospheric Profile product (MOD/MYD07_L2), dark target and deep blue combined aerosol optical depth at 500 nm from MODIS Aerosol product (MOD04_L2), cloud optical thickness, cloud top pressure, cloud top temperature, surface pressure and surface temperature from MODIS Cloud product (MOD06_L2), total column precipitable water vapor and total ozone burden from MODIS Atmospheric Profiles product (MOD/MYD07_L2), and land surface shortwave albedo from MODIS Albedo product (MCD43D61)" (*Ryu et al., 2018*). The SR data from BESS was also highly consistent with the observational field data (R^2=0.95, *Fig. 2* in *Ryu et al., 2018*). Therefore, we would expect SR from BESS is more reliable and accurate than SR from CRUNCEP. Still, we tested the RF performance using SR from CRUNCEP and the OOB R^2 remained unchanged (0.63), possibly due to the relatively low contribution of SR in the random forest training (7%, **Fig. 2a**) and the high spatial correlation between SR from BESS and from CRUNCEP. This will be added in the revised manuscript.

3) As suggested, we replaced the model-derived WAI with satellite-based surface soil moisture (SM) data, including the mean annual soil moisture data from Soil Moisture and Ocean Salinity (SMOS) during 2010-2018 (*Li et al., 2020*) and Soil Moisture Active Passive (SMAP) during 2015-2018, *O'Neill et al., 2019*). The OOB R^2 for SMOS and SMAP are 0.60 and 0.59 respectively, compared to the original value of 0.63. The lower performance may be caused by the fact that satellite-based soil moisture data only accounted for soil water status in the top centimeters whereas productivity is influenced by root-zone soil moisture. In addition, the importance ranking changed from #4 for WAI (**Fig. 2a**) to #8 for SM_SMOS and SM_SMAP (**Fig. R3**). The order of other variables remains unchanged. This will be added in the revised manuscript.

Figure R3 Variable importance in the trained RF model using soil moisture (SM) data from SMOS (a) and SMAP (b).



Reference:

Li, X., Al-Yaari, A., Schwank, M., Fan, L., Frappart, F., Swenson, J., & Wigneron, J. P. Compared performances of SMOS-IC soil moisture and vegetation optical depth retrievals based on Tau-Omega and Two-Stream microwave emission models. Remote Sensing of Environment, 236, 111502. 2020.

O'Neill, P. E., S. Chan, E. G. Njoku, T. Jackson, and R. Bindlish. SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 6. [Indicate subset used]. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center. doi: https://doi.org/10.5067/EVYDQ32FNWTH. 2019.

Ryu, Y., Jiang, C., Kobayashi, H. and Detto, M.: MODIS-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000, Remote Sens. Environ., 204, 812–825, doi:10.1016/j.rse.2017.09.021, 2018.

Viovy, N. CRUNCEP dataset, description available at: ftp://nacp.ornl.gov/synthesis/2009/frescati/temp/land_use_change/original/readme.htm. 2017.

Comment #4

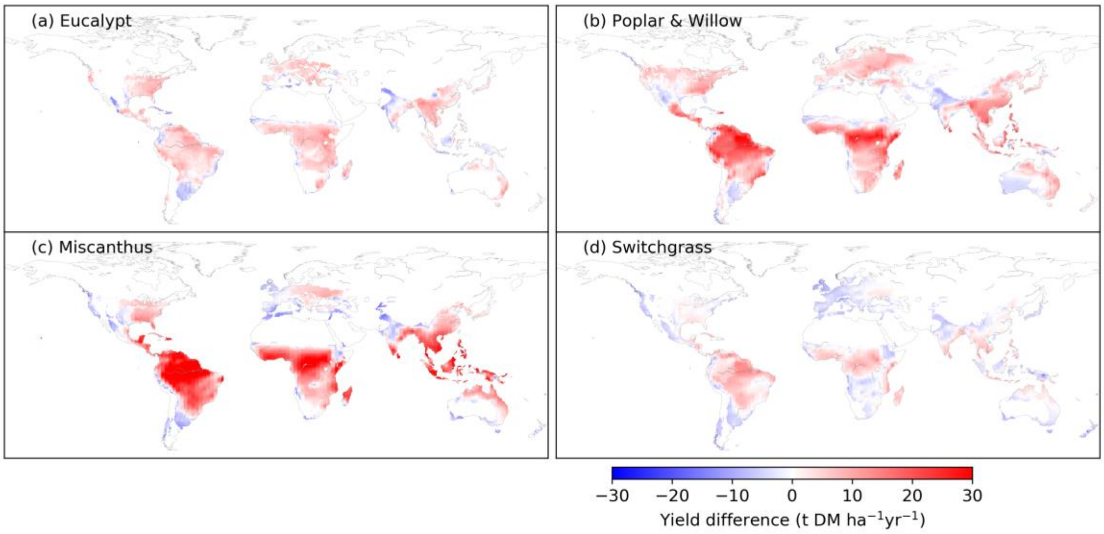
3. Given the big deviation shown between the yield map used by IAMs and the yield map derived by the authors, it is difficult to convince me of the reliability of the yield map generated by the random forest approach. I also wonder why the authors don't compare this product with their model estimates (Li et al., 2018b). Because the ORCHIDEE model has also been calibrated based on the same global bioenergy crop yield data set in Li et al. (2018a), it would be more logical to compare the derived product with the ORCHIDEE model estimate in the spatial scale.

Response #4

As suggested, we compared the yield map derived from random forest with the yields simulated by the land surface model — ORCHIDEE (**Fig. R4**). Because poplar and willow were taken as one plant functional type (PFT) in ORCHIDEE, the average yields of poplar and willow from random forest were used for comparison (**Fig. R4b**). The yields simulated by ORCHIDEE are generally higher than those from random forest, especially for Miscanthus and Poplar&willow. This could be largely expected because in this version of ORCHIDEE, there are no nutrient limitations on plant growth, no effect of pests and disease on crops, and the management practices were implicitly included when adjusting the productivity parameters in the model to match the site observations with management like irrigation, fertilization or specific high-productive genotype. There could be a similar case in LPJml (*Heck et al., 2016*), and that is why the IAMs calibrated the LPJml yields based on currently observed yields to get the potential yield maps (see details **on L199-215**).

On the other hand, the predictions from random forest are largely constrained by the yield range of observations, representing the yields that can be achieved (or were achieved during the period when yield data were reported) under current (optimal) technology. This is exactly the purpose of producing this data product in our study, which is observation-based and can be used to benchmark the yields simulated by land surface models or IAMs.

Figure R4 Comparison of bioenergy crop yields between the RF map and maps simulated by ORCHIDEE (ORCHIDEE yields minus RF yields where yields are available in both paired maps).


(a) Eucalypt
(b) Poplar & Willow
(c) Miscanthus
(d) Switchgrass
−30 −20 −10 0 10 20 30
Yield difference (t DM ha^{-1}yr^{-1})