Open Access

Earth System
Science
Data

Discussions

Interactive comment on "Mapping the yields of lignocellulosic bioenergy crops from observations at the global scale" *by* Wei Li et al.

Anonymous Referee #2

Received and published: 28 October 2019

The authors reported 3,963 observations covering five bioenergy crops in the abstract, however, they only used 161 grid cells to train the RF model. The sample size is too limited to map the spatial distribution of global bioenergy crops (over 60,000 grid cells). The comparison of the derived maps with other modeled maps cannot convince me.

1. There were a bunch of variables included in the RF regressions. I suggested to add a diagram to show how random forest algorithm works in your study.

2. At the global scale, there are more than 60,000 grids in $0.5° \times 0.5°$. Here the authors used 161 grid cells for model training, among which you included five types of crop types. I think the training data are not substantial enough to build RF regression models.

3. Section 2.3. I appreciate that the authors compared their derived yield maps with the current three IAMs. However, it still cannot convince me since all these are modeled maps rather than the actual yield data. Is it possible to compare your derived yield maps with the existing inventory? Moreover, the authors assumed the derived maps are in 2010 without no temporal changes. To the best of my knowledge, the technology improvement has led to a significant increase of crop yield during the past several decades. Thus, I think it is not appropriate to compare your yield map with the present day's maps. The long-term average covering the time period of your collected observations is better for comparison. Line 198-199: What do you mean 'actual yield maps'? Is it your derived yield map from RF or other? If yes, I do not think you can consider it as an 'actual yield map'.

4. Figure 3. The spatial distribution of predicted yields seems to highly correlated with MAP. For example, the Amazon basin and Southeast Asia receive a substantial rainfall per year. The spatial distribution of Eucalypt and Miscanthus are so similar, the same as the remaining three crops. Thus, nothing new surprised me.

5. Figure 5. Did you compare your areas with any existing inventory data? It is better to compare yours with them since the total amount of production is also important.

6. Figure 2. You listed the variable importance in the trained RF model. It turns out that MAP is the dominant variable. You provide Figure S8 to show the relationship of bioenergy crop yield with temperature. However, MAT is not quite important compared with other variables. Why did not you show the relationship of each crop with dominant variables, such as MAP, GSL, WAI, etc.

Interactive comment on Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2019-118, 2019.