Geoscientific Model Development Discussions

Geosci. Model Dev. Discuss., referee comment RC2
https://doi.org/10.5194/gmd-2021-21-RC2, 2021


# Comment on gmd-2021-21

Anonymous Referee #2

Referee comment on "Choosing multiple linear regressions for weather-based crop yield prediction with ABSOLUT v1.1 – Initial tests for the districts of Germany and an over-confidence trap in statistical modelling" by Tobias Conradt, Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2021-21-RC2, 2021

This is a model-description paper. The weather-based crop yield prediction is a critical to yield estimate and climate change impact assessment and adaptation. Focusing on leveraging the limited input features and explanatory power is a key question to be addressed when developing a yield prediction. The major contribution of the study is, according to the author, program coding by which the linear regression with the best explanatory power could be found based on exhaustive combinations of predictors. Such a practice has been widely used in developing empirical models (258 out of 362 studies, lines 78-79), except that few study have developed a formal tool. To this end, the novelty of this study is relatively weak as compared with the standard of GMD.

Major comments:

In the "Introduction" section, author firstly provide the research gap that "the entirety of crop and landscape specific weather effects is hardly captured by the existing models" but the "crop and landscape specific weather effects" is not clear because weather effect consist of extreme adverse weather effect, weather trend and fluctuation effect, and the current linear regression model can capture the effect of annual and seasonal mean weather (Peng *et al* 2018). More specific information about "weather effect" is helpful to emphasize the question aimed to be addressed.

If the key question is to optimize the current linear regression model, the innovation of the study is controversial. As numerous studies have found the linear regression performs not as well as machine learning algorithms in reproducing the spatio-temporal pattern of crop yield (Leng and Hall 2020, Cao *et al* 2021, Cai *et al* 2019, Zhang *et al* 2020), I think optimized current linear regression is not able to improve the yield prediction accuracy significantly. Thus, I suggest to revise the research gap.

After introducing the research gap, nine questions related to yield prediction were provided but the key question aimed to be addressed in this paper is still not clear for me. Also, I think the literature review is not strongly associated with the question this paper intended to address. This study focused to on weather-based yield prediction but there were papers uniting remote sensing data and weather data to predict yield, which may be related to the question 4 but considering two much non-meteorological factors will departure the goal of weather-based yield prediction.

In the “Materials” and “Method” section, it is necessary to provide some information about the weather relevant covariates and their correlated relationship with crop yield, which can help us understand the primary response of crop yield to climate change established in the model.

Reproducing the impact of climate extremes is becoming increasingly important in yield prediction models. As the weather data are monthly, I wonder whether the ABSOLUT model can predict the effect of climate extremes, which is a major source of yield loss in Europe (Trnka *et al* 2014). From the results of Figure 5, the predicted model underestimates the impact of drought in Northern Germany and the district-level performance is not good.

The comparison of other prediction approaches should not be limited to the yield forecast in Germany but some other yield prediction across the major breadbaskets of the world (Cai *et al* 2019, Li *et al* 2021), making the readers understand whether the ABSOLUT is an advanced yield prediction method across the globe. Also, more information about uncertainty of the model should be addressed.

Specific comments:

Line 22: the full name should be added when the ABSOLUT first mentioned.

Line 26: the equation of the ABSOLUT should not be moved to “Method” section.

Figure 2 can demonstrate the model performance at country level but why not provide the results of maize simultaneously.

Line 160: why use some ellipsis?

References

Cai Y, Guan K, Lobell D, Potgieter A B, Wang S, Peng J, Xu T, Asseng S, Zhang Y, You L and Peng B 2019 Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches *Agric. For. Meteorol.* **274** 144–59

Cao J, Zhang Z, Tao F, Zhang L, Luo Y, Zhang J, Han J and Xie J 2021 Integrating Multi-Source Data for Rice Yield Prediction across China using Machine Learning and Deep Learning Approaches *Agric. For. Meteorol.* **297**

Leng G and Hall J W 2020 Predicting spatial and temporal variability in crop yields: An inter-comparison of machine learning, regression and process-based models *Environ. Res. Lett.* **15** 44027

Li L, Wang B, Feng P, Wang H, He Q, Wang Y, Liu D L, Li Y, He J, Feng H, Yang G and Yu Q 2021 Crop yield forecasting and associated optimum lead time analysis based on multi-source environmental data across China *Agric. For. Meteorol.* **308–309** 108558

Peng B, Guan K, Pan M and Li Y 2018 Benefits of Seasonal Climate Prediction and Satellite Data for Forecasting U.S. Maize Yield *Geophys. Res. Lett.* **45** 9662–71 Online: https://doi.org/10.1029/2018GL079291

Trnka M, Rötter R P, Ruiz-Ramos M, Kersebaum K C, Olesen J E, Å½alud Z and Semenov M A 2014 Adverse weather conditions for European wheat production will become more frequent with climate change *Nat. Clim. Chang.* **4** 637–43

Zhang L, Zhang Z, Luo Y, Cao J and Tao F 2020 Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in China using machine learning approaches *Remote Sens.* **12**