bgd-11-C3642-2014

We would like to thank the referee for his thorough remarks and criticism of our study. At first we would like to address the general criticism. Referee comments are presented in *italics*.

We are aware about the limitations of crop and climate modeling and that the uncertainty of our results is increased simply by coupling both approaches. This is a problem all modelers do face and must address. We tried to improve our explanation on this matter, especially in regard to several of the following points.

"However, downscaling climate projections to a higher resolution should not be seen as increasing confidence in data…"
It was not our intent to imply this. We did revise the text for clarification. The downscaling of data can never improve the confidence in it and was not supposed to. However, regional climate effects (e.g. due to relief) will exist in the future, thus the downscaling of the data will only aid in pointing out regional differences within the scope of a large-scale climate projection.

"…(ensemble modelling)…"
It is certainly a valid point to refer to the advantages of ensemble modeling. WETTREG was indeed part of such ensemble runs (Kreienkamp. et al. 2011) and was thus tested against results from other models. Model bias and results are well documented by the developers. The use of ensembles is further not unanimously preferred, as there are also doubts if the validity of results is improved. To name just one: many models use the same initial parametrization. Thus they might all carry the same inherent flaws and thus a multi-model mean would not compensate for this (Tebaldi und Knutti 2007). Also, extreme years would be further reduced in multi-model means. As discussed this is already a problem, as the climate series is relatively smooth.

Furthermore, the downscaled climate model data was provided through the *State Authority for Mining, Energy and Geology,* who use it for several local assessments in Lower Saxony. Thus we had no direct involvement in the decision to use this specific climate model.

* Kreienkamp, F., Baumgart, S., Spekat, A. und Enke, W. (2011) *Climate Signals on the Regional Scale Derived with a Statistical Method: Relevance of the Driving Model's Resolution*. Atmosphere 2(2) S. 129–145
* Tebaldi, C. und Knutti, R. (2007) *The use of the multi-model ensemble in probabilistic climate projections*. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 365(1857) S. 2053–2075

"Ultimately, a wide range of impacts and adaptation responses is possible …"
This is connected to the remark about ensembles and we do agree. We would like our results to be read as "under the assumptions of this model the results will look like this". That other scenarios are probable is not disputed. We tried to evaluate the main responsible factors for the given yields, to give an idea how these results might change if input variables would be altered.

"…the use of WETTREG climate model (alone), but also of BioSTAR crop model (alone) is poorly justified in this study…"
The one main reason is computational power. We could have either used several models on a coarse scale or one model respectively on a rather local level (i.e. at a higher resolution). We deliberately chose the local assessment but agree that a comparison to other models would add value to this study. More so, our aim was less to give an outline of general maize development throughout the century, but of the different varieties relative to each other.

"This is also more relevant given the simplified assumptions behind this crop model (section 2.2), referred to as being in beta stage development (page 8, line 5)"

We agree that beta stage might be at least misleading and we clarified this in our text. The BioSTAR model is developed by a group that did not necessarily follow defined stages of *alpha*, *beta* or *release candidates* as might be known from professional (commercial) software developers. Thus "it might be referred to as being in beta stage" does only mean that changes have been made in software layout since. The now published model version (0.6.0) has some added capabilities compared to the version used in this study, however no official model version was used back then (it might be around 0.4.0 to 0.5.0). This might be a more adequate representation than *beta stage*.
Still, the results of the used model were validated as described by the model's main author (Bauböck 2013) with the same version as used by us.

"At least the authors should expand the exercise to other crop models simulating maize production. They may refer to benchmarking actions run at international level, e.g., for maize, Bassu et al. (2014)."

At the time we conducted the study we have not been aware of the effort of Bassu et al. (2014) and thankfully include this information. We added a small section to our report discussing this matter using this reference.

We would also like to address the minor issues raised by the referee.

In section 2.4, it is not clear the time step (daily, hourly …) of the weather data used to input the crop model.

The data are monthly means that are broken down by the model into daily values. Problems with intra-annual variability are referred to in the discussion.

For Penman/Monteith ("Penman-Monteith", with a hyphen not a solidus) evapotranspiration (page 10, line 6), for instance, the FAO-56 approach for daily calculation was probably adopted, but please specify

Yes, the FAO-56 approach was adopted. It should however be noted that this was purely done to calculate backwards air humidity from available ETpot data, as only this variable was provided. The crop model then calculates its own evapotranspiration.

"At any rate the few other existing studies are hinting towards the same result". Please cite such published studies.

This does refer to the studies of Southworth et. al 2010 and Liu et al. (2013) from the introduction / discussion part. As advised we referenced them again in the conclusion.