

## ***Interactive comment on “Evaluation of Integrated Assessment Model hindcast experiments: A case study of the GCAM 3.0 land use module” by Abigail C. Snyder et al.***

### **Anonymous Referee #1**

Received and published: 28 June 2017

Review Comments on gmd-2017-97; Evaluation of Integrated Assessment Model hindcast experiments: A case study of the GCAM 3.0 land use module

This paper proposes a methodology for evaluating the hindcast results generated by integrated assessment models (IAMs) or land use models. As a case study, GCAM land use results are evaluated. The authors found that global aggregates are not sufficient for evaluating IAMs. Additionally, the deviation measures examined in this work successfully identify parametric and structural changes that may improve land allocation decisions in GCAM. The suggested future work is involving some improvements to the GCAM land allocation system identified by the measures in this work, using the

C1

measures to quantify performance improvement due to these changes, and, ideally, applying these measures to other sectors of GCAM and other land allocation models.

The overall text is well written, and the logic is understandable. However, there are several concerns before publishing. Here, I listed some points that must be modified or improved.

- The way how they use the metrics and to draw the conclusions which argue about the potential model improvement seems not comprehensive and quite naïve. The essence of this hindcast experiment exercise land use is surely determined by the crop demand and trade together with the yield information (either direct observation or extrapolation). At least without the assessment of demand reproducibility, it would be difficult to make conclusions.

- The data chosen to display looks arbitrarily decided and not comprehensive. Starting from global total is fine. Then, the analysis went to specific crop (wheat) and country (USA). Looking from one of the objectives of this study which identifies model improvements, the analysis should be comprehensive. Based on the discussion in the last conclusion part which takes broader issues like FAO data things, the comprehensiveness seems important here also.

- Although the paper says that neither of Fujimori et al.'s techniques are compatible with their goals and methodology, the objective in Fujimori et al. seems quite similar to this paper's method. It is because the method in Fujimori et al. clearly states that “the regression method is focusing on the bias in the discrepancies between the simulation results and statistics by regions and years to identify which regions and years for each variable have large discrepancy.” The authors should discuss what is the advantage and disadvantage of the proposed method in this paper.

- GCAM model description should be more enriched. It is because in the latter part, they discuss about producer price, logit exponent and trade and so on. At least those things should be clearly described.

C2

- The carbon price already exists in the real world around 2010, and is it taken into account? This might have been discussed in their paper Calvin et al 2017 but as far as reading GCAM papers, the land use part is really sensitive to carbon price and sometimes looks unrealistic. It would be better to validate that part.

Other minor points are below.

- Line 7 P1; about the description “this is key in the integrated assessment community, where there often are not multiple models conducting hindcast experiment”, I think the fact that not multiple models conduct hindcast should not be the reason why they need absolute term evaluation. Even if hindcast is carried out many similar models, it should be evaluated independently (for example, macro econometric models like DSGE do validate individually).

- Line 22, P1; It would be better to specify “other model validation exercises”

- Line 3 P2; Are the references all GCAM 3.0?

- Line1 P14; I cannot understand this sentence. are USA producer prices used globally?

- Line 11 P14: The sentence “the scenarios using actual yield information (AY and AYB) lead to GCAM’s land allocation being overly responsive, due to economic agents having more information than their real world counterparts” is strange. From the model point of view, the yield in all four scenarios are given parameters. So the different between (AY, AYB) and (FY, FYB) are not the matter of information quantity difference from real world.

- In conclusion, authors suddenly address about trade and no discussion in results part. It seems strange and would be better to discuss in the results part more and derive some summary in the conclusion part.

---

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2017-97>,

C3

2017.

C4