Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-138-RC2, 2016

Geoscientific Model Development Discussions
Open Access
EGU

Interactive comment on "Data-mining analysis of factors affecting the global distribution of soil carbon in observational databases and Earth system models" *by* Shoji Hashimoto et al.

K. Todd-Brown (Referee)

ktoddbrown@gmail.com

Received and published: 17 August 2016

This study used boosted regression tress to identify mismatches in drivers of soil carbon data products and simulated soil carbon in Earth system models. Given how complex Earth system models have become studies like this one which apply statistical methods to both data products and simulation results can provide key insights into what is driving model behavior and contrast that with data driven statistical models. Resulting in better targeted model development.

While I appreciated the authors attempt at brevity I believe that they went a little too far, leaving out key details which are needed to reproduce the results and falling just short of relevant interpretation. The authors need a more detailed overview of the model

Printer-friendly version

Discussion paper

structure of the ESMs examined in this study and whether or not that structure played a role in the attribution. Do you see any effects of model structure on the BRT results? Can these results be tied to particular temperature sensitivity function or number of soil carbon pools represented in the simulation?

Finally this study needs to be placed in context of other studies which have examined the different driving variables in ESMs, particularly the CMIP5 models. While the authors offer some token mentions there is a painful lack of detail on this topic.

Please see below for a line by line reaction:

Please make it clearer that the ESMs are regressed against data products not other ESM output. While the modern ESM NPP and temperature distributions match better with current data products. There are some notable differences in modeled NPP in particular in the CMIP5 models and this could be a source of bias in the analysis.

P1L23-24 C:N ratio and clay content are in most ESMs in the allocation scheme. While it is intractable to investigate each modeling code directly Much of the documentation for these ESMs includes Lignin:N ratio (similar to C:N ratios) and clay content mediating decomposition. CENTURY Parton et al 1988 use Lignin:N ratio and clay content for allocation parameters (IPSL-CM5 Krinner etal 2005 cite CENTURY: Parton et al 1988)

P2L4-8 Should there be a citation here?

P2 L8-11 A more in depth treatment of past attempts to disentangle drivers of data-model differences is called for here. Please expand on each of these treatments with particular attention to the ones that looked at the same models and data products the authors are using in this study. In addition, add something to the discussion to contrast your results with these studies.

Section 2.1 There needs to be some discussion about model structure in the ESM vs data products. These data products are typically constructed using correlation to the local environment (climate + land cover + geology) where the pedon was collected.

Printer-friendly version

Discussion paper

Please summarize the methods used for each specific data product. For ESMs a discussion of their sensitivity functions and pool structure is appropriate (note that BCC was incorrectly stated to have their N-cycle turned on for CMIP5 in Todd-Brown et al 2013).

P2 L33-35 Be more convincing about averaging models from the same center, there is some clustering analysis that is in the supplemental of Todd-Brown et al 2013 that could support this.

P2 L34-35 Todd-Brown et al 2013 averaged all ensembles that were available at the time, this statement is incorrect. Please either provide a different justification for only considering one ensemble or, preferably, go back and re-analyze the data with the multi-ensemble mean (even better if you can incorporate the modeled uncertainty).

Section 2.4 What regridding algorithm did you use? There are several options in CDO, not all are appropriate for soil data, temperature and NPP. Please discussion which algorithm was used and why.

P5 L1 Describe the results here in addition to referencing the figure.

P7 L15 A BRT tutorial is not appropriate to cite under 'Code availability'. Please either link or reference as SI to the actual code used in this analysis (preferred) or remove this section.

Interactive comment on Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-138, 2016.

Printer-friendly version

Discussion paper