Geosci. Model Dev. Discuss.,
doi:10.5194/gmd-2016-138-AC1, 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.

Shoji Hashimoto et al.

shojih@ffpri.affrc.go.jp

Received and published: 7 October 2016

Referee #1

We greatly appreciate your thoughtful and constructive comments. We have revised the manuscript on the basis of your comments, and our responses to the Major and Specific comments can be found below. According to the editorial instructions, our response is structured as follows: (1) comments from Referees; (2) author's response; and (3) author's changes to the manuscript. Thank you very much.

###########################

Major comments

Printer-friendly version

Discussion paper

###########################

Comment 1: Need for improved texts and figures.

Response: We have revised the text and figures on the basis of your specific comments. Please see our responses to your specific comments.

Changes to the manuscript: Please see our responses to your specific comments.

#######

Comment 2: More information for reproducibility

Response: We have added more details about our methodology (e.g., software details and procedural details) along with the code that we used. We cannot attach the CMIP5 data because of their terms of use (http://cmip-pcmdi.llnl.gov/cmip5/terms.html), but we have attached the other data (data for observational databases) in the Supplement.

Changes to the manuscript: Please see our responses to your specific comments 6. Additionally, please see the Supplement.

#######

Comment 3: "The authors need to expand on the genuinely interesting parts, and offer more interesting, novel insights on how their results apply, and will be useful, to future work. The lack of reproducibility (above) means that it's also not clear how any of this would inform or be useful for modelers seeking to improve their software and science."

Response: This type of the manuscript was "Methods for assessment of models" (please see http://www.geoscientific-model-development.net/about/manuscript_types.html), and the main purpose of the paper was to demonstrate the possibility of using a machine learning algorithm to compare model outputs with observational databases; it was not aimed at identifying new mechanisms of the soil carbon cycle. We have improved the reproducibility of our analyses by adding more information about the methods along with the code and

Printer-friendly version

Discussion paper

applied data in the Supplement. We have modified the discussion to clearly convey our message.

Changes to the manuscript: Please see our responses to your specific comments, and Discussion in the revised manuscript.

##########################

Specific comments

##########################

Comment 1. Page 1, line 1: I'd suggest either "Data-mining analysis of the global..." or "Factors affecting the global..."

Response: We have changed the title to "Data-mining analysis of the global..."

Changes to the manuscript: "Data-mining analysis of the global distribution of soil carbon in observational databases and Earth system models"

#######

Comment 2. P. 1, lines 9, 20, and 26-27: these three short sentences could be deleted with no real loss

Response: We have deleted the first two sentences, but we prefer to retain the third sentence (lines 26-27) because, as mentioned above, we would like to emphasize this point.

Changes to the manuscript: We have deleted the first two sentences.

#######

Comment 3. P. 1, l. 25: "elucidate the nature" of the databases? Confusing

Response: We have rewritten the sentence.

Changes to the manuscript: (Page 1, line 24–25) "The results of this study should aid in identifying the causes of mismatches between observational SOC databases and ESM outputs and improve the modelling of terrestrial carbon dynamics in ESMs."

#######

Comment 4. P. 2, l. 4: what recent study?

Response: This is Todd-Brown et al. 2013. We have added "(Todd-Brown et al., 2013)".

Changes to the manuscript: (Page 2, line7– 8) "a recent study (Todd-Brown et al., 2013) has found that. . ."

#######

Comment 5. 5. P. 3, l. 8-9: divided over what spatial scale? Some more detail in this entire para- graph would be useful

Response: A resolution was added.

Changes to the manuscript: (Page 3, line 30– 31) "The wetland ratio was calculated by dividing the number of wetland grids at 30 seconds by the total grids at 1°."

#######

Comment 6. Methods: need to give version numbers CDO, R, and all packages used. Also, I'm shocked at the complete lack of any mention of data or code availability (no, that one sentence on p. 7 doesn't count). It's 2016, and I expect all code and data (at least that backing the main results) to be included as Supplementary info, or posted in a repository. It's not acceptable to produce results from a black box; see also http://www.geoscientific-model-development.net/about/code_and_data_policy.html

Response: We agree with the importance of openness. We have added details about the software used in this study and about the main code that we used. We have also attached the data used in this study. We cannot attach the CMIP5 output (please see the terms of use of CMIP5: http://cmip-pcmdi.llnl.gov/cmip5/terms.html), but our results are easily reproduced by pasting CMIP5 data into the attached data and applying the data codes.

Changes to the manuscript: (Page 3, line 38–Page 4, line 2) “All global databases, except for the databases with a spatial resolution of 1° by default, including observational and ESM model outputs, were regridded to a spatial resolution of 1° for the analyses. Regridding of data in the NetCDF format was performed using the Climate Data Operators (CDO) software, version 1.6.9, provided by the Max Plank Institute for Meteorology (https://code.zmaw.de/projects/cdo). A bilinear interpolation, which is one of the most widely used algorithms, was used (remapbil in CDO).”

(Page 4, line 15– 23) “We used the open-source BRT package (brt.functions.R) in R software version 3.2.1 and 3.2.2 (R Core team, 2013) developed by Elith et al. (2008). The gbm package was used (version 2.1.1) to run the BRT package. The calculations were performed in Mac OS X (version 10.9.5 and version 10.10.5). To do so, the “windows” function in the “brt.functions.R” needed to be replaced with the “quartz” function in R. In practice, three parameters in the BRT packageâĂŤ the learning rate (lr), tree complexity (tc), and bag fraction (bg)âĂŤcontrol the BRT performance. The lr determines the contribution of each tree, the tc controls the number of splits, and the bg is the proportion of data selected at each step. The number of trees was determined using the cross-validation method in the R package. The maximum number of trees was set to 15,000. The tc value was set to 5. We tested different lr (0.001, 0.005, 0.01, 0.05, 0.1) and bg values (0.5, 0.6, 0.7) and used the best parameter set for each database, but the changes in parameter values had little effect on the model performance.”

Please see the Supplement.

#######

Comment 7. P. 4, l. 17: “Relationships with a mean annual temperature were relatively close to each other” – what does this mean? Clarify

Response: We have rephrased the sentence.

Changes to the manuscript: (Page 5, line 1) "Relationships with the mean annual temperature were similar."

#######

Comment 8. P. 4, l. 34: "The contribution of each variable varied between ESMs" ?

Response: We have rephrased the sentence.

Changes to the manuscript: (Page 5, line 19) "The contributions of some variables varied among ESMs"

#######

Comment 9. P. 4, l. 36: "large inconsistencies. . .demonstrated low contributions" – what?

Response: We have modified the sentence.

Changes to the manuscript: (Page 5, line 20–22) "Large inconsistencies between the observational databases and ESMs were found in the low contributions of clay content and the CN ratio and in the high contributions of NPP in ESMs (Figs. 5a and 5b)"

#######

Comment 10. P. 5, l. 23-24: this is an interesting point, and should be expanded upon. What are the implications, if the seemingly wide variety of CMIP5 models in fact uses a much smaller number of fundamental assumptions or modeling approaches? I'm pretty sure that Kathe Todd-Brown made this point in one of her papers; see also Alexander et al. (2015), 10.5194/gmd-8-1221-2015

Response: We have expanded this part.

Changes to the manuscript: (Page 6, line 4–21) "Analyses of the ESM outputs showed large variability, but the influential factors were predominantly similar among the ESMs

Printer-friendly version

Discussion paper

(Fig. 5). This similarity most probably indicates that the structures of the models that describe SOC dynamics in the ESMs are similar. One reason for the similarity is probably because some ESMs share common code (Alexander and Easterbrook, 2015). Another reason may be rooted in the basic structure of the soil carbon model: SOC is calculated as the balance between dead organic matter input to soil and carbon emissions from the decomposition of organic matter in soil, and these processes are influenced by temperature and water conditions. The SOC pool is characterized by its turnover time (decomposition constant). In general, decomposition exhibits an exponential response to temperature, which is more severe than its response to water. As a result, modelled SOC is strongly influenced by NPP (litter input), temperature, and turnover time, which have been demonstrated by previous studies (Exbrayat et al., 2014; Todd-Brown et al., 2013) and were also confirmed in our analyses. As shown in Table 2, SOC submodels in ESMs differ in the number of SOC pools and function types of temperature and moisture. Todd-Brown et al. (2013) have reported the absence of any pattern of agreement between ESM outputs and observational SOC databases with soil carbon pools, temperature and moisture sensitivity functions, and Exbrayat et al. (2014) have found that turnover times of SOC in ESM outputs are not affected by the number of SOC pools. Our analyses also indicated that a match or mismatch of major contributing factor between ESM outputs and observational databases are not strongly related to these properties of SOC submodels. Thus, it is likely that the spatial pattern of SOC from ESMS are more strongly affected by the basic structure, driving variables (NPP and temperature) and parameterisations (turnover time and influential parameters of temperature and moisture sensitivity) than by the number of pools and the function types of temperature and moisture sensitivity."

#######

Comment 11. P. 6, l. 13-14: "The use of temperature sensitivity. . ." - ?

Response: We have deleted the sentence.

Changes to the manuscript: Deleted.

#######

Comment 12. P. 7, l. 2-3: would such model-data fusion ever be possible, given the extremely long running time of modern ESMs?

Response: I agree. As you point out, model-data fusion of the whole ESM system is very difficult because of the long running time. In practice, I think that applying model-data fusion to a limited part of ESMs (e.g., ecosystem carbon cycle models) would be realistic. We have added the above points to the main text.

Changes to the manuscript: (Page 8, line 4– 7) "Constraining model parameters with observational databases through data assimilation, such as a Bayesian approach, would improve the performance of ESMs. Applying such model-data fusion to whole ESMs, however, would require a very long running time; therefore, model-data fusion to a part of an ESM (e.g., ecosystem carbon cycle model) would be realistic."

#######

Comment 13. Table A1: an URL or reference for each model would be useful

Response: We have added a URL for each model.

Changes to the manuscript: Please see Table 2 in the revised manuscript.

#######

Comment 14. Table A2: this classification was applied to. . .? Where is it from?

Response: We used soil texture data in the ISLSCPII database (Table 1), and a classification in the database was used (Table A2). Because the contribution of soil texture was not high, the relationships between the SOC and soil texture are not shown. To clarify that this soil texture classification is for the soil texture shown in Table 1, we have modified the caption of Table A1 and A2 in the revised manuscript.

Changes to the manuscript: (Page 15, Table A1) “Classification of soil texture in ISLSCPII (see Table 1).”

#######

Comment 15. Figures 1 and 2: these are so tiny I'm not sure they convey any information, really

Response: The maps have been redrawn.

Changes to the manuscript: Please see Fig. 1 and Fig. 2 in the revised manuscript.

#######

Comment 16. Figure 6 should be the central, most important figure of the entire paper–showing how variable importance compares between observational databases and ESMs–but it's very difficult to see what's going on. I'd suggest re-thinking this, and carefully con- sidering the most effective way to show this

Response: We have redrawn Fig. 6 using a box plot, and we have added a new figure to clearly show the results from each ESM. We have redrawn other figures, too.

Changes to the manuscript: Please see Fig. 5 in the revised manuscript.

Please also note the supplement to this comment:
http://www.geosci-model-dev-discuss.net/gmd-2016-138/gmd-2016-138-AC1-supplement.zip

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

Printer-friendly version

Discussion paper

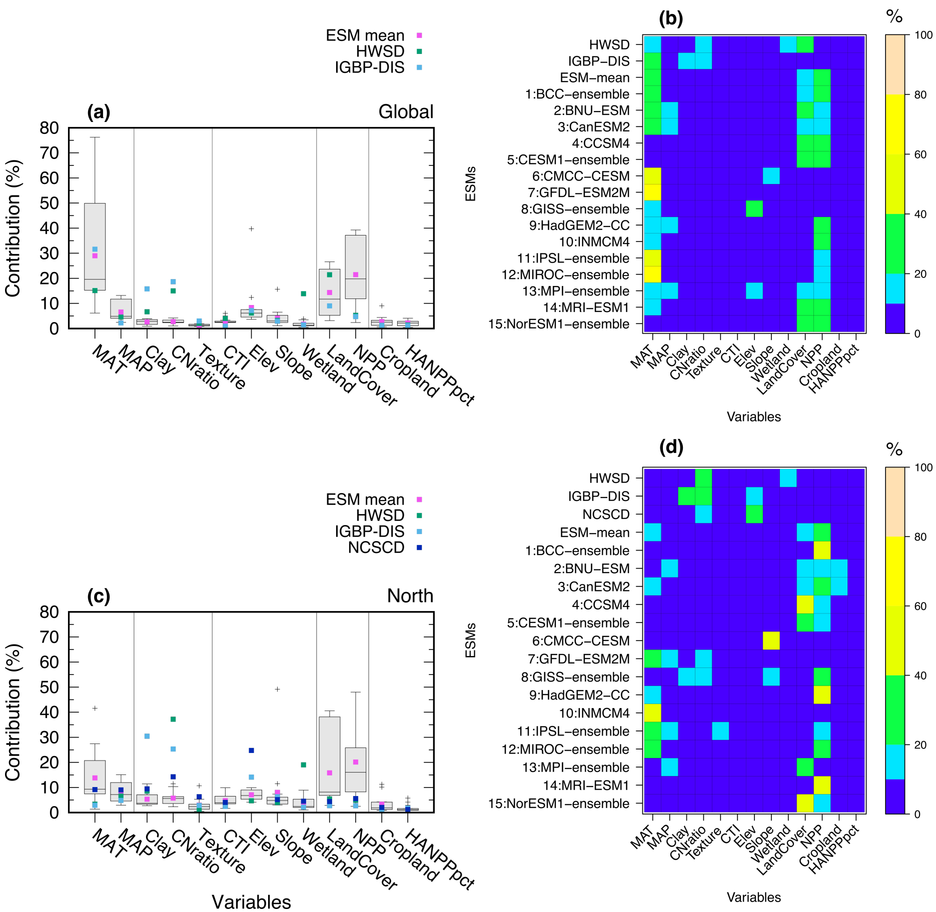

Fig. 1. New Fig. 6 (Fig. 5 in the revised manuscript)

Printer-friendly version

Discussion paper