Biogeosciences Discuss.,
https://doi.org/10.5194/bg-2017-503-RC2, 2018
© Author(s) 2018. This work is distributed under
the Creative Commons Attribution 4.0 License.

Biogeosciences
Discussions
Open Access

Interactive comment on "Gap-filling a spatially-explicit plant trait database: comparing imputation methods and different levels of ecological information" *by* Rafael Poyatos et al.

Anonymous Referee #2

Received and published: 12 January 2018

The paper compares different imputation methods for trait databases not only regarding their accuracy in predicting plant traits and trait-trait relationships but also their ability to preserve trait distribution and multi-trait correlation structures. They use a plant trait database that has complete trait observations for 5 traits in 630 plots together with the auxiliary information and the data are georeferenced. Exhaustive test are done to compare the methods in using only species mean data, adding auxiliary data or geographical information with different levels of gaps and 4 different complementary evaluation method i.e. R2, NRMSE, KGE, differences in the correlation matrices. This I believe is done for the first time in such a thorough way. The study is very relevant for biogeographical and ecological studies that need to use trait databases, which are

usually not free of gaps. However, compared to the number of methods and test done, there is not enough explanation and discussion done. I would suggest a re-evaluation of the discussion part. For example try to explain why the selected traits behave differently in the results.

I would suggest to add information on the abbreviations of the methods, auxiliary information and traits for all Figures also the figures in the supplementary. The Figure captions must be self-explanatory.

The paper uses different names like environmental information, ecological information/variables, and auxiliary information/variables etc. for the same thing and this is confusing. I suggest to use only auxiliary information or variable throughout the paper and to be consistent, especially since the data listed in the as ecological information (L90) are not all considered ecological exactly.

ABSTRACT

I suggest that the authors summarize the results in the order of the questions mentioned at the end of the introduction.

L 27: the word globally is misleading.

INTRODUCTION

This part is well written.

L 42: It is not clear what is meant by "primary sources"?

L 91: why is regKrig not in the second test together with MICE and kNN?

L91: perhaps would be better to define what you mean by "optimum level" here- best set of auxiliary variables selected in step ii - and then use the phrase hereafter

METHODS

Chapter 2.3: Please use this part to explain "Mean" and Spmean" technique a bit more.

L121: Please write the full name for MCAR as well.

L173: what type of “Lithology” data is used? Please include it here.

L175-178: Why are some auxiliary information added in a factorial design and some sequentially? What about running a test for each trait independently for realizing the most important auxiliary information that explains their variability best and then using those for the MICE and kNN model (also regKrig)?

Chapter 2.5: I would suggest making clear which statistical method is used for evaluating accuracy, multi-trait correlation structure or bivariate trait relationships.

L232: “as the best method globally” – Why globally?

RESULTS AND DISCUSSION

This part is in general not well written. There are a lot of results that looking at them in the paper and its’ supplementary brings up many questions but the authors did not discuss them well. It’s mostly like a report of results and not much discussion and reasoning of the results. Please discuss the performance discrepancies of the methods for different traits, and for the different evaluation criteria (trait-trait correlation, trait distribution, etc.).

First paragraph of chapter 3.1: Fig. 2 and Fig. S7 shows different results for different traits. How do you interpret this? Please discuss.

L239, L246, L250: Instead of Supplementary S5, please indicate Fig. S number. Please do so for the rest of the paper as well.

L246 Please remove ”however”.

L279: “stand structure” was mentioned before (L25, L90, L146, etc.) as “forest structure”. Please choose one and be consistent.

Chapter 3.2: Why the comparison of adding different levels of auxiliary information

was only used for MICE and not for kNN or RegKrig? In addition, please discuss why mice_ctsp was chosen while according to Fig. 4 mice_sc would be sufficient for Nmass, LMA and WD and mice_st for Hmax and only BL:AS needs mice_cstp (extra p).

L288: Please replace A with S in “Fig. A7 and A12”.

L288-289: what about adding only species and forest structure to RegKrig? Climate and topography should be already accounted for using spatial information in RegKrig.

L292-296: Please discuss and reason the differences of the results for the different traits.

L305: “(data not shown here)”. But why not adding a figure similar to Fig. S8, for the comparison of Δcormat and dataset NRMSE between Spmean, mice_ctsp, kNN_ctsp and RegKrig in the supplementary as well?

L307-315: I would suggest to make this part chapter 3.4 (related to “Imputing traits for the main forest species in Catalonia 2.6)

IMPLICATIONS

This part is well written. I would suggest to add also suggestion for improvement of data collections, e.g. trait collection should be better be accompanied by auxiliary information on coordinates, forest structure, etc.

L322: “Here we deal here” Please remove the extra “here”.

L357: Please change “practices” to “practiced”.

SUPPLEMENTARY

Please refer to the supplementary with their Fig. numbers and not chapter numbers. Please make sure that all supplementary figures are mentioned in the main text. Currently this is not the case.

Fig S7: trait mean (Mean) is missing in the plot.

Fig S2: Please change “palant” to “plant”.

Fig S6: Please define “cor.matrix abs.error” in the figure and “correlation matrix error” in the caption. Do you mean Δcormat? Please be consistent.

Interactive comment on Biogeosciences Discuss., https://doi.org/10.5194/bg-2017-503, 2017.