

Replies to Referee #2's comments on "Insights into the prediction uncertainty of machine-learning-based digital soil mapping through a local attribution approach" (egusphere-2024-323)

We would like to thank Referee #2 for the positive analysis and the constructive comments. We agree with most of the suggestions and, therefore, we have modified the manuscript to take on board their comments. We recall the reviews and we reply to each of the comments in turn.

Referee #2:

This manuscript is well written, clear and relevant, and presents methods that could provide stakeholders with valuable insights into where the uncertainty comes from: this has the potential to make uncertainty more concrete for them.

I appreciate the use of a synthetic test case, which makes the whole procedure a lot easier to understand.

We thank Referee #2 for this positive feedback.

I don't have any major criticisms. I would be pleased to see this manuscript published after attention to the following minor details :

Line 44: However, at a local scale, these methods don't (?) provide any information for a prediction at a certain spatial location.

We thank Referee #2 for noticing this problem. We have reformulated as follows: "However, these methods do not allow to measure the influence of the covariates for a prediction at a certain spatial location."

Line 157: pushes the prediction uncertainty?

We agree that this term is confusing. The sentence has been replaced as follows: "i.e. whether the considered covariate influences the prediction upwards or downwards in relation to the base value"

Line 442: I don't see any circular pattern on the bottom middle panel of Figure 13 (in the bottom right one however, they are really clear).

We thank Referee #2 for noticing this problem. We have corrected the text by referring to Figure 13, bottom right.

Synthetic test case: isn't the fact that in Z1, the biggest contributor to uncertainty is Tmean-Tmax (and that respectively in Z2, the biggest contributor is Pwetest) be linked to the fact that these covariates have uniquely high (respectively low) values there, that are not represented in the dataset? If you agree, this in my opinion would be interesting to put in the discussion.

We thank Referee #2 for the analysis. These additional elements are helpful for a better understating of the synthetic case and have been added to Sect. 4.1.

Orleans,
April 29th, 2024

J. Rohmer¹ on behalf of the co-authors

¹ BRGM, 3 av. C. Guillemin - 45060 Orléans Cedex 2 – France