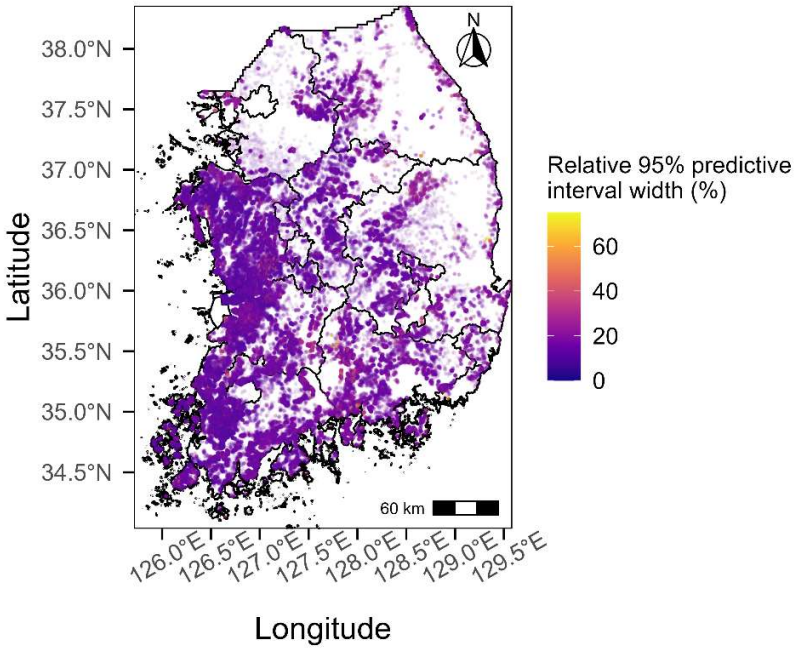
Response to Referee #1

In this manuscript, the authors used Random Forest models to predict SOM content in paddy soils of South Korea. Thanks to an ensemble modeling, they were able to decompose the source of uncertainty between sampling, response variable perturbation and model specification. They used a very large dataset of 283,618 observations. The manuscript is interesting, well written and easy to read, so I have only a few minor remarks to suggest.

Response: We thank Referee #1 for the positive assessment of our manuscript. We appreciate the comments on the presentation of uncertainty, the evaluation of tail weighting, and the clarity of the methods and figures. We have addressed each comment below.

Maybe we need also maps for uncertainty to support what is written in the text (e.g., l. 520 mentioning a spatial variation, l. 523 for a spatial concentration of uncertainty).

Response: To support the discussion of spatial variation in predictive uncertainty, we have added a new map showing empirical 95% predictive interval widths expressed as percentages of the median ensemble prediction at each location (new Figure 5).



I was wondering if the tail weighting of the Random Forest model really worked, considering the big offset at observed SOM > 125 g kg-1 in Figure 5 left.

Response: As pointed out, Figure 5 (now Figure 6) shows substantial underprediction at high SOM concentrations, including observations above 125 g kg^{-1}. Tail weighting was intended to increase the influence of upper-tail observations. We have compared weighted and unweighted

Random Forest models using matched calibration–validation splits and otherwise identical settings. The comparison using 113,444 validation observations shows a modest improvement in overall validation performance and a clearer reduction in negative bias.

	Unweighted	Tail-weighted
RMSE (g kg^{-1})	6.46	6.32
MAE (g kg^{-1})	4.09	4.06
R^2	0.47	0.50
Bias (g kg^{-1})	−0.96	−0.36

However, these overall improvements do not establish improved performance specifically above 125 g kg^{-1}, and substantial underprediction persisted at these extreme concentrations despite tail weighting.

l. 237: “no imputation was applied”. It is unclear.

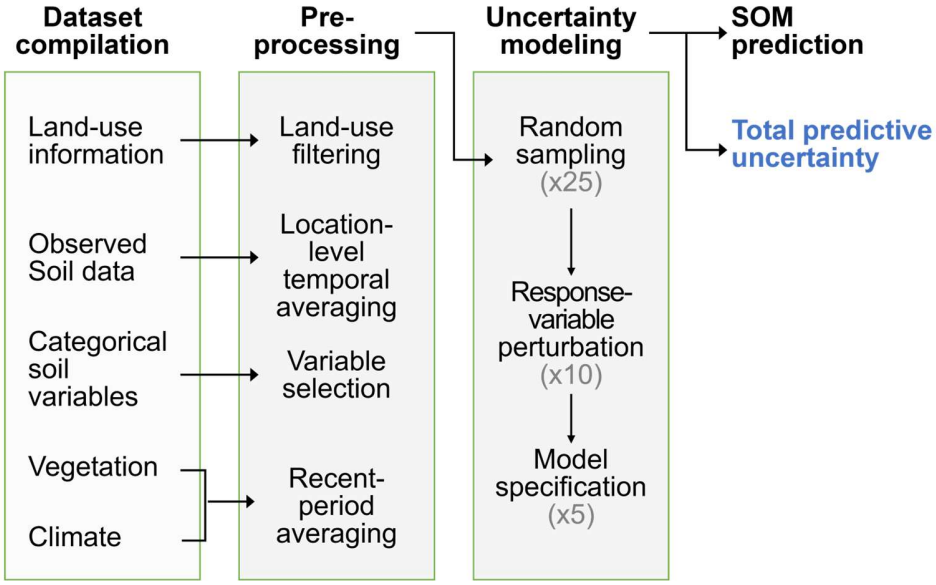
Response: We have replaced “no imputation was applied” with “missing data for the response variable or predictors were not replaced with estimated values” to clarify the meaning [Section 2.2, revised lines 237–238].

Bottom of Table 1: I think “dimensionless” is not the appropriate term.

Response: Thank you for the comment. We have revised the table note to read: “—, CV not reported for near-zero or non-positive means or where its interpretation was not meaningful.”

Figure 2: A better place can be found for the arrow linking “land-use filtering” and “Random sampling”.

Response: We have repositioned the arrow connecting preprocessing to uncertainty modeling to clarify the workflow.



l. 320-332: Try to better organize this with section 2.4.

Response: As suggested, we have moved the description of empirical predictive intervals from Section 2.3 to the beginning of Section 2.4 to present the uncertainty calculations together [Section 2.4, revised lines 418–422].

l. 345: A word is missing.

Response: We have revised "perturbed calibration SOM" to "perturbed calibration SOM concentrations" for clarity [revised line 400].

l. 403: Italics unnecessary for "indexes".

Response: We have revised it [revised line 465].

l. 486: You cannot really refer to Table 2 to illustrate a spatial distribution.

Response: We agree that Table 2 does not demonstrate spatial distributions. We have removed the spatial interpretation and revised the paragraph to describe the SOM groups and associated uncertainty summaries reported in Table 2 [Section 3.2, revised lines 547–578]:

"The first three SOM quantile groups, which encompassed 62.3% of the observations, spanned approximately 1–25 g kg^{-1} (Table 2). The highest SOM group comprised concentrations above approximately 29 g kg^{-1} and had the widest median empirical 95% predictive interval. Although absolute interval widths were generally greater toward the upper end of the SOM distribution, the relative contribution of model-specification uncertainty declined across increasing SOM groups (Table 2). The weaker representation of these high-SOM conditions likely reflects the

greater difficulty of predicting conditions that are less prevalent in the calibration data and/or are associated with environmental combinations that are less well represented across the prediction domain. …"

Table 3: I don't think it is necessary to present methodological aspects below the table. Only the definition of acronyms. Add the word "final" in the caption.

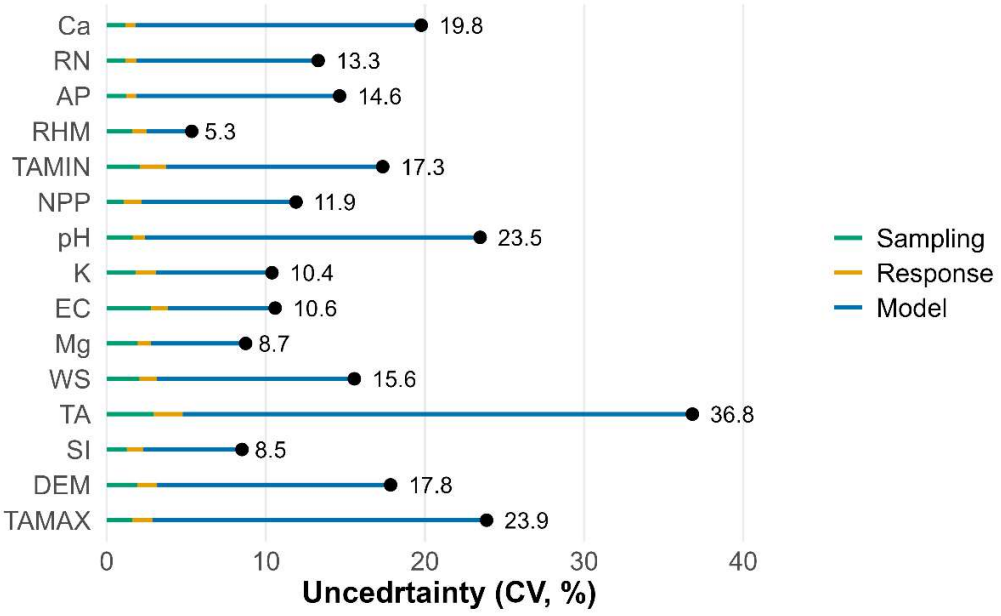
Response: We have removed the methodological details below Table 3, retained only the abbreviation definitions, and added "final" to the caption.

Figure 5: Unit for SOM is missing.

Response: We have added the SOM units (g kg^{-1}) to the caption of Figure 5.

Figure 7: This figure is nice, but I am afraid the colors are not distinct enough given the thickness of the line. You could order variables like in Figure 6.

Response: We have used more distinct colors, increased the line thickness, and ordered the predictors consistently with the original Figure 6 (now Figure 7) to improve readability. The revised figure is now Figure 8.



In the manuscript, please use "concentration" or "content" instead of "values" for SOM.

Response: As suggested, we have replaced "values" with "concentrations" for SOM throughout the manuscript.