General Revision Summary

The study presents a well-executed analysis of uncertainty quantification in deep learning soil spectral models, specifically through the use of Monte Carlo-Conformal Prediction (MC-CP). The paper makes a strong contribution to the field by addressing a crucial gap in soil spectroscopy—reliable uncertainty quantification. The methodological approach is well-documented, and the comparison between MC dropout, Conformal Prediction (CP), and MC-CP is insightful and thorough.

Strengths of the Paper

- Novel Contribution: The paper introduces MC-CP as an method for uncertainty quantification in deep learning soil spectral models. The demonstration of MC-CP's ability to balance expected coverage, computational efficiency, and adaptability to out-of-domain samples is a significant advancement.
- Well-Designed Comparison: The comparison between MC dropout, CP, and MC-CP is informative and shows the trade-offs between these methods.
- Strong Methodological Foundation: The study follows a solid methodological framework.
- Practical Relevance: The application of the proposed method to real-world soil spectral data enhances the practical impact of the study

General Improvements

Terminology and Consistency (Machine Learning vs. Deep Learning)

- The abstract and introduction interchangeably refer to Machine Learning (ML) and Deep Learning (DL). However, the methodology and model used are specifically deep learning-based. Ensure consistency in terminology and explicitly state where ML is a broader category and where DL is specifically applied.
- Incorporate a broader range of examples in the Introduction for Monte Carlo (MC) Dropout and Conformal Prediction (CP), as the current section focuses too narrowly on just two detailed examples.
- There is a lack of clear structure (detailed further in the comments). There are many redundant repetitions, and subordinate clauses with very general information are interspersed throughout, often repeating details that were already mentioned earlier.
- Ensure there is a space between the number and the percentage symbol for proper formatting.

Detailed Comments

Comment No	Lines	*Original*	Review
1	7-10	*"While machine learning has made remarkable strides in predicting various physiochemical properties of soils using spectroscopy, predictions devoid of quantified uncertainty offer limited utility in guiding critical decisions. However,*	The sentence effectively explains that predictions without uncertainty are not useful for decision-making and that uncertainty quantification is rarely used due to limitations. However, the logical connection between these points could be clearer to improve readability and coherence.

		uncertainty quantification remains underutilised in the reporting of soil spectral models, with existing methods facing significant limitations."	
2	10	*" These approaches are either computationally demanding…."*	It is not entirely clear whether this refers to the existing methods mentioned in the previous sentence or to something else, as methods and approaches are not necessarily the same.
3	11-23	-	The structure is confusing in the sense that your method is mentioned without prior explanation, followed by the introduction of two established methods for comparison. Additionally, while introducing these methods, you already include some results. To improve clarity, consider restructuring the section by clearly separating the description of methods, the comparison, and then presenting the results.
4	24-26	*"This breakthrough enhances the real-world applicability of soil spectral models and represents a significant advancement in the field of soil science. [….] further revolutionising decision-making and risk assessment in soil science."*	Shorten this section to two sentences, as the usefulness is stated twice. Avoid redundant explanations to improve clarity.
5	29	*"[...](Padarian et al., 2020; Minasny et al., 2024). These studies are characterised 30 by the use of large soil datasets and require an efficient way of extracting information to predict target attributes."*	The reference is incorrect, as these studies do not discuss what you describe in the following sentence.
6	41-46		There are repetitions in the sentences without adding new content. Shorten them for conciseness.
7	43	*"Despite the significant success of machine learning in predicting soil properties, uncertainty quantification of the prediction remained an underexplored area in soil spectroscopy, and only a few studies have tried to include uncertainty in the model evaluation."*	A reference is needed for the studies mentioned.

8	50-54		I don't see the relevance of explaining the difference between the two types of uncertainty here, as it does not appear to be a topic in the methods section or the discussion.
9	61-66		To my knowledge, bootstrapping is typically used for confidence intervals, not for prediction intervals like MC and CP. Additionally, different methods of quantile regression and Gaussian methods are missing, which would help provide a more complete introduction.
10	68-72		Specify that MC is specifically used for deep learning to avoid ambiguity.
11	96-103	*"In this study, we applied a strategy to increase the PICP of MC dropout while maintaining its advantages in characterising out-of-domain uncertainty. Monte Carlo-Conformal Prediction (MC-CP) was introduced by Bethell et al. (2024). MC-CP integrates the strengths of both MC dropout and CP."*	Clarify that MC-CP is the strategy. Again, avoid repetition to improve clarity and conciseness.
12	113-115		Please specify how many of the removed samples were due to SOC and how many were excluded because of extreme values.
13	116		Clarify why the threshold of 40% clay content was chosen and provide justification for this choice.
14	119		If you are already describing your training and test scheme here, also include the ratio of the splitting mentioned in L203 for consistency and completeness.
15	Chapter 2.2, 2.3, 2.4		For better structure, I suggest organizing the section as follows: 2.2 Methods, with subsections 2.2.1 Monte Carlo Dropout (MC dropout)**,** 2.2.2 Conformal Prediction (CP)**,** and 2.2.3 Monte Carlo-Conformal Prediction (MC-CP)**.**
16	125		Missing abbreviation: MC dropout
17	128	*"In each dropout layer, a certain portion of the neurons is randomly deactivated (weights set to zero) during both training and testing."*	As far as I know, and as stated in the paper by Gal and Ghahramani (2016), neurons are only deactivated during training. While validation can be involved, a specific reason is needed for doing so. Please verify what is happening in your specific use case.
18	137		Check the Mathematical notation and terminology of the journal: https://publications.copernicus.org/for_authors/manuscript_preparation.html#math. I recommend centering the equations for better readability. Additionally, equations should be treated as nouns within the text. So here I would change it to the following: The 90% prediction interval […] of the predictions (Eq. 1): Formula. (Eq. 1)
19	137		When using a formula, ensure that every abbreviation is defined either before or in the sentence following it. In this case, $\mathbf{C}_{MC}$ and $\mathbf{X}_i$ are missing definitions.

20	150	*Table 1*	Stay consistent in using **X** or $\mathbf{X_i}$ throughout the table to maintain clarity and uniformity.
21	161		See comment No. 18
22	170		Stay consistent in the writing of Monte Carlo-conformal prediction**.** Since it is based on Bethell et al. (2024)**,** I recommend following their terminology and formatting.
23	179		See comment No. 18
24	184		See comment No. 18
25	208-209		See comment No. 18 and a reference is missing for the Eq. 5 and 6.
26	210-214		See comment No. 18 a space is missing in Eq. 7 between the fraction and "count".
27	223		I would rephrase it as follows, omitting the word "poor": “A negative R-squared value indicates that the model performs worse than simply using the mean prediction."
28	224-225		Connect the two sentences for example as following: “Such results for out-of-domain samples were expected, as the model did not have any knowledge of soils with clay content larger than 40%, leading most out-of-domain predictions to fall under 40% clay.”
29	238	*“When the evaluation of uncertainty is optimal, the expected coverage of a p% prediction interval is p% (dotted line in Fig. 3)”*	What do you mean by "evaluation of uncertainty"? Please clarify or provide a more precise definition.
30	255		MPIW instead of PIW
31	263	*Table 4*	The PICP value for out-of-domain samples is missing and should be included for completeness.
32	276-281		I do not agree with the strong wording that MC-CP effectively addresses the out-of-domain issue, as the difference in MPIW between in-domain and out-of-domain samples is not significant.
33	299-304		This part should be discussed directly in the uncertainty section rather than in the limitations and future applications section for better coherence.
34	312		Specify the exact deep learning model used.
35	329		The wording should be revised—for an optimal trade-off, the results need to be more significant.