Dear Editor and reviewers,

we would like to thank you for your kind letter and for reviewers' constructive comments concerning our manuscript (**A Semisupervised Deep Learning Neural Network Using Pseudolabels for Three-Dimensional Shallow Strata Modelling and Uncertainty Analysis in Urban Areas from Borehole Data**). These comments are all valuable and helpful for improving our article. All the authors have seriously discussed about all these comments. According to the reviewers' comments, we have tried best to modify our manuscript to meet with the requirements of your journal. Some main modifications as follows:

(1) We have added to the reasons why we undertook this work and why we chose the ML approach in the new manuscript.

(2) We have added a table to the new manuscript that includes the network architecture and parameters to ensure reproducibility of our work.

(3) We have revised an ambiguous statement in line 71 of the introduction to the previous version of the manuscript

Detailed modifications and responses are as follows:

Topic editor decision

1) discuss the advantages and limitations of the proposed approach, to clarify explicitly the benefits of this ML approach compared to implicit methods. The purpose is not so much to provide an exhaustive comparison but to address the critical reason for this study in the first place, which is essential.

Thank you for your suggestion. We have included the following statement in lines 64-70 of the revised manuscript to explain why we launched this study and why the problem needs to be solved by the ML approach: "When using the implied surface method to construct a 3D geologic model…model by uncertainty analysis."

2) Ensure that all the details required are provided for reproducibility purposes (network architecture and parameters).

Thank you for your suggestion. We have added a detailed Table 2 to the revised manuscript on line 198 according to your and the other reviewers' suggestions.

referee #1

Line 17: What do you mean by "faces challenges related to uncertainty"? Implicit approaches are more automated than explicit ones, so it should be easier to manage uncertainty. However, they struggle to capture complex structures (e.g., Collon et al., 2016, 10.1190/INT-2015-0178.1).

"Faces challenges related to uncertainty" means that constructing a 3D geological model by implicit approaches may lead to modeling that deviates from the real situation; we have modified this description in the Introduction section on lines 64-67.

Line 17: Kriging is one of the key approaches to implicit modeling, and kriging is the same as Gaussian processes in machine learning. So that opposition between explicit and implicit approaches on one side and machine learning approaches on the other is poorly substantiated. Machine learning is not a new tool for either explicit or implicit modeling. But using deep learning for explicit or implicit modeling is more recent, that would be true.

The main focus of this manuscript is to address the challenges of human labor consumption in explicit modeling and the inability to conduct uncertainty analysis in implicit modeling through machine learning methods. It is not intended to set machine learning modeling against explicit or implicit modeling methods.

Line 18-19: It's not so much the use of machine learning (see my comment above, Gaussian processes work well in data-poor settings), but the use of implicit modeling (see Collon et al., 2016, 10.1190/INT-2015-0178.1, although their case study has considerably less data than this one, so I'm not sure this study can be called "data-poor").

Indeed, this manuscript addresses different modeling data than the article you referenced. The focus of this manuscript is primarily on the challenges encountered in modeling borehole data.

Line 27: What does that mean "better [...] supports uncertainty analysis"?

We want to show that the SDLP algorithm is the best algorithm in the manuscript.

Line 29: This is not a case study with sparse borehole data. Having a borehole every 23 m or so covering most of a 305 by 264 m domain is a high data density for a subsurface project.

The manuscript presents the borehole data obtained from a real engineering project in Shenyang city. The primary objective of this project is to ensure the stability of the building.

Line 37: It is a bit weird to end a list of common data with "other types of data". That list is limited in subsurface projects, so better be exhaustive. Maybe analog data from outcrops could be worth adding.

Thank you for your suggestion. We have modified these problems according to your suggestion.

Line 54: What does "MLS" stand for?

Thank you for your suggestion. We have added this information to the lost section.

Line 70: I don't understand the first point.

We have modified this section.

Line 125: What if the TIN connect two intervals that shouldn't be connected? For instance two sandy intervals, but one is from a sedimentary channel, the other a crevasse splay? What would be the impact?

Incorrectly connecting intervals in geological modeling significantly reduces model accuracy. The connectivity of geological layers is vital for ensuring accurate representations of actual geological conditions. Inaccurate connections can lead to models that fail to reflect the true geological characteristics. Moreover, such errors pose challenges in geological interpretation, especially in boundary areas, because they can affect the accurate understanding of lithology, structural features, and depositional environments. In the modeling process, the sedimentary sequence of the strata is used as the basis for connections between geological layers; therefore, it is believed that there are no issues associated with incorrect connections.

Figure 1: I'm still not sure I fully understand, so on the zoomed section, what are the points H1iPj mean? Are they just to illustrate the balance between different intervals? Or are they new data points created with the TIN? This remains quite confusing.

H1iPj represents unequal-interval sampling performed on deterministic profiles, and its coordinates x, y, and z can be obtained using Formula (2).

Line 170: It would be nice to have a figure describing the network's architecture (this could be done by updating figure 2).

Thank you for your suggestion. We have modified Figure 2.

Line 170: How can a user choose the right number of hidden neurons? How many layers are used? 4 like in figure 2? Why that choice? Is it robust or does it impact predictions significantly?

The structure and hyperparameters of the neural network in this manuscript were primarily referenced from the article "www.zjujournals.com/eng/article/2021/1008-973X/202103021.shtml " and manually adjusted through experimentation. However, the selection of the neural network structure and determination of hyperparameters are not within the scope of this manuscript's research. The primary focus of this manuscript is to demonstrate the effectiveness of using pseudolabels in deep learning networks for constructing 3D geological models.

Line 195: Here a figure would help a lot.

Thank you for your suggestion. We will take your opinion into consideration in our next stage of research.

Line 199: High accuracy based on what criterion?

Thank you for your suggestion. We have added this explanation to the manuscript. A training accuracy of 90% is considered to indicate a high accuracy rate.

Line 200: Is that done on the prisms mentioned before? Actually I'm still confused.

"There might have been a misunderstanding; it is done on the unlabelled data. This statement means that pseudolabelled data are obtained by comparing the predicted results of the unlabelled data with the attributes of the prism where the coordinates of the unlabelled data reside."

Line 234: I'm still not sure what "unlabelled grids" mean. Are those related to the deterministic profiles?

“The unlabelled grids” mean that the grids except from borehole data and pseudolabel data.

Line 240: How was the test set determined?

We set the training set:validation set:test set to be 6:2:2, and we have added this detail to Table.2.

Table 2: What are SAM and DL? And where is HRBF?

“SAM” should be SVM. We apologize, but the HRBF algorithm does not fall under the category of machine learning methods. As a result, it does not provide outputs such as accuracy, precision, recall, or F1 score.

Figure 5 & 6: Is that only for SPDL? What about the other methods?

Because the manuscript focuses on the SDLP algorithm, we only show the results of the SDLP algorithm.

Line 258: Is that the cell size?

Yes, the grid size is the cell size.

Line 266: How was the K-fold set up? By removing individual labels or entire wells?

By removing entire wells.

Line 272: The procedure for the validation is not clear. You're only removing a few wells, so why not simply do a holdout validation (i.e., train only once to predict all the missing wells together directly)? And/or why not do a proper group K-fold cross-validation and predict all the wells?

In fact, the article uses K-fold cross-validation (with K set to 10) and provides the average results of cross-validation.

Line 277: But in practice you won't be able to make such choices. This adds a considerable bias to your validation.

The conditions under which this study was conducted were met when we already had some idea of the results. We will consider the research you mentioned regarding real-life situations as part of our future research plans.

Sections 4.1, 4.2, and 4.4: Those are not discussions, but new results that should be in section 3. The validation should be the same for all methods.

We have reorganized the sections of this manuscript in this way because we want to incrementally explain the effectiveness of our proposed SDLP algorithm.

Line 311: I see no real reason for that. The goal of this paper is to predict geological formations, and the HRBF method should be compared on the same ground.

This manuscript demonstrates the conformity between the drilling data and the established three-dimensional geological model through Fig. 11b-d.

Line 329: A visual comparison is not robust enough for this. You need to add a quantified validation, so include the HRBF method in section 3.

Thank you for your suggestion. Our current research is unable to address the issue you have raised. We will endeavor to conduct further investigations in our next phase of research.

Line 339: That's just untrue, erosive structures such as channels will lead to abrupt changes. You actually cannot say whether the HRBF method or SPDL perform better based on such weak comparison based on unsupported claims.

Thank you for bringing up this concern. We have carefully reviewed the geological reports of the area in question, and there is no evidence of an erosive structure, as you described.

Line 411: That claim only holds is the both algorithms were properly tuned, which is not mentioned in the paper.

The main focus of this manuscript is to highlight the effectiveness of the proposed SDLP algorithm. Throughout the main text, we have provided evidence to support the effectiveness of this algorithm.

Line 411: What about other indicators than accuracy? In imbalanced cases, accuracy is actually a poor indicator of performance since good accuracy can be achieved by predicting the major classes only.

We have considered your previous suggestion and have incorporated precision, recall, and F1 score as our evaluation metrics.

Line 415: However, other approaches to implicit modelling can capture uncertainty, how does you method compare to those?

Currently, our research has not included uncertainty comparisons with other implicit approaches. We will endeavor to supplement our study with this aspect in future work.

referee #2

• L.37 “and other type of data” not needed since listing “included” examples.

Thank you for your suggestion. We have removed “and other types of data”.

• L.71 I don’t understand “is much less than not revealed by borehole data”, please rephrase

We have modified the fuzzy expression.

• L.74 “were mainly” -> do you mean “are usually”?

Thank you for your suggestion; we have modified the tense.

• L.75 it is the distribution of categories that’s imbalanced, not their number

In the manuscript, “the imbalanced number of categories” is shown that is the difficult that these data are used for training data to training deep learning model.

• L.102, “maximum average thickness” -> per formation. The authors might omit that sentence altogether, as the average thickness does not seem to be used or referred to anywhere else.

Thank you for your suggestion; we have removed this term.

• L.124, please specify the (redundant but helpful) 2D horizontal nature of the TIN since talking approximately 3D until then.

Thank you for your suggestion. We have added this explanation to the manuscript.

• L.128, please specify the measure and quantitative threshold used (skewness?) to remove "narrow triangles"

Thank you for your suggestion. We have added this explanation to the manuscript. The threshold for determining whether a triangle is an acute triangle based on the measurement of its smallest angle is set to 20 degrees.

• L.170: ReLU instead of RELU

Thank you for your suggestion. We have modified the spelling.

• L.171: what percentage is used for the dropout?

The dropout percentage is set to 0.1. We aimed to avoid excessive resetting of neurons, which could lead to incorrect predictions and erroneous pseudolabelled data.

• Fig.2 "pridiction" typo

Thank you for your suggestion. We have modified the spelling.

Thank you very much for your consideration.

Best regards!
Yours sincerely,
Jiateng Guo