Anonymous Referee #1

General comments:

This manuscript proposed a standardized dataset for deep learning-based electromagnetic methods. The database is geophysically constrained, which produces good accuracy performance and satisfactory generalization and consistency. Overall the paper is very well written, and the data shows its high readiness for the community. I would recommend it get published before some of my concerns are addressed.

In response of general comments:

The authors would like to thank you for your valuable time and positive evaluation of the paper under consideration. We agree with the suggestions, and they will improve the manuscript.

Specific comments:

1a. Evaluation (section 4): This study mainly employed the proposed dataset to train a deep-learning (DL) model and surrogate the forward modeling process and demonstrated that this dataset shows its great performance. The assessment method is rational; however, this section needs additional result comparison with previous relevant DL-based studies. By comparing this proposed data with other DL studies that used limited input data, the authors may demonstrate this proposed database can be treated as the benchmark. Otherwise, it is only another DL experiment for improving the computation efficiency.

In response of comment 1a:

In order to make comparison with existing relevant studies, we generate random resistivity models similar to the data set generation approach used in several deep learning studies (Colombo et al., 2021b; Moghadas, 2020; Moghadas et al., 2020; Noh et al., 2020; Puzyrev and Swidinsky, 2021; Qin et al., 2019; Wu et al., 2021b) mentioned in Table 1. The data set generation approach for each of the above-mentioned deep learning studies differ from each other, however, the commonality in all the above-mentioned methods is that the resistivity values for each layer is chosen randomly. For example, Colombo et al. (2021b) generate 5-layer resistivity models with a random combination of resistivity and thickness; Moghadas (2020) generates 12-layer resistivity models with fixed layer thicknesses increasing logarithmically down to 6 meters; Moghadas et al. (2020) generate subsurface models having 12 layers with logarithmically increasing thickness down to 10 m depth; Noh et al. (2020) used 1 to 3 layer models; Puzyrev and Swidinsky (2021) generate resistivity models having 50 layers; and Wu et al. (2021b) generated resistivity models with number of layers randomly set between 1 and 20 with the bottom depth at 1000 m.

Therefore, to keep the network configuration same and to have same level of modelling complexity for a fair comparison; the number of layers, depth discretization, and the number of models will be kept the same as used to train other networks in our study. However, the resistivity values for each layer are picked randomly from a log-uniform distribution to compare with the deep learning studies that employ random resistivity models. The relevant text in the manuscript will be revised and Figure 3 and Figure 4 will also be updated after obtaining the results from random resistivity models.

1b. Besides, please try to find some weaknesses in the training data utilized by previous DL studies and demonstrate your progress on it after comparison. For example, the introduction is well written but those training sets are not involved after then. Benchmark is a strong word that requires more comprehensive assessment and evidence.

In response of comment 1b:

Some of the weaknesses in the training dataset utilized by other methods will be discussed in the revised manuscript. We also discuss the similarity of von Kármán models to the methods that uses pseudorandom resistivity models in some other deep learning studies and give the rationale behind improved performance using the proposed database.

1c. The assessment section (section 4) needs to provide additional quantified comparisons and descriptions for the figures rather than just some summaries.

In response of comment 1c:

The description of the figures and additional quantitative comparison will be included in the revised manuscript.

2. References in Table 1 should be Qin et al. (2019) rather than (Qin et al. 2019), and generally, the table caption is above the table.

In response of comment 2:

References in Table 1 will be formatted appropriately and captions for the tables will be moved above for all the tables in the manuscript.

3. The dataset is formatted as txt, which caused the code reading speed very slow.

In response of comment 3:

We have chosen to use the txt format as it is easily readable. As the data from the txt file needs to be read only once for a particular algorithm, the authors are of the opinion to keep it as it is.