Geophysical Research Letters

Supporting Information for

Instantaneous Bayesian imaging for large-scale transient electromagnetic data using probabilistic neural networks

Jian Chen[1], Stefano Galli[1], Alessandro Signora[1], Nicole Anna Lidia Sullivan[2], Bo Zhang[3], and Gianluca Fiandaca[1,2]

[1]The EEM Team for Hydro & eXploration, Department of Earth Sciences "Ardito Desio", University of Milano, Via Botticelli 23, Milano (Italy).

[2]The EEM Team Spin-Off company, Via Botticelli 23, Milano (Italy).

[3]Institute of Earth exploration Science and Technology,Jilin University, Changchun (China).

Contents of this file

Text S1

Data generation and definition of STD

For data generation, this study's training dataset comprises a forward modeling dataset (405,000 sets of data) and a field-measured dataset (15,000 sets of data). To calculate electromagnetic response data, constructing a theoretical resistivity model is necessary. To simulate the actual geological conditions as much as possible, this study refers to the generation approach proposed by Wu et al. (2021a). Firstly, 2-4 anomaly layer positions are randomly selected within the 120 m underground space, with a minimum interval of 15 m between anomaly layer positions. Secondly, random values within the logarithmic resistivity range of -1 to 4 are assigned to these anomaly positions. Lastly, we determined the resistivity values for the 30-layer earth model's positions using cubic spline interpolation.

EEMverter (Fiandaca et al., 2023) was utilized for forward calculations on the resistivity model to obtain electromagnetic response data. The input data of this study consists of electromagnetic response data and data STD. For field-measured data, STD is known, whereas additional calculations are required for forward modeling data (Auken et al., 2008):

$$\mathrm{STD} = \left[\mathrm{STD}_{\mathrm{uni}}^2 + \left(V_{\mathrm{noise}} / V \right)^2 \right]^{1/2}. \tag{S1}$$

where $\mathrm{STD}_{\mathrm{uni}}$ is uniform noise standard deviation and its value is set to 0.03, V is the forward modeling noiseless data, and Vnoise is the background noise contribution:

$$V_{\mathrm{noise}} = b \cdot \left(t / 1 \cdot 10^{-3} \right)^{-1/2}. \tag{S2}$$

where t is gate time, b is the noise level at 1 ms and its value is set to 1 $\mathrm{nV/m^2}$.

After obtaining the STD of the forward modeling data, we applied the EEMverter software for resistivity inversion of both simulated and field measured data. During the inversion process, the STD of the data was taken into consideration, and a relatively loose longitudinal constraint was applied (with a constraint factor set to 3, meaning that a penalty would be imposed if the resistivity change between adjacent layers exceeded threefold). This approach ensures that the inversion results rare relatively consistent with the actual geological conditions. The inversion objective function is as follows:

$$Q_{\mathrm{inv}} = \left[\frac{1}{N_d + N_m - 1} \left(\sum_{i=1}^{N_d} \frac{\left(d_{obs,i} - d_{forward,i} \right)^2}{\left(STD_i \cdot d_{obs,i} \right)^2} \right) + \sum_{i=1}^{N_m - 1} \frac{\left(m_i - m_{i+1} \right)^2}{\sigma_m^2} \right]. \quad \text{(S3)}$$

where N_d is number of time gates, N_m is munber of layers, d_{obs} is the target data to be inverted, $d_{forward}$ is the forward modeling data, m is the inversion parameter, and σ_m is the constraint standard deviation in the vertical direction of the model.

Ultimately, the TEM data, the data STD, and the resistivity models obtained from inversion are each subjected to logarithmic transformation and normalization. The normalized data are combined with the STD to form the input dataset, while the inverted models are utilized as training labels, thus completing the construction of the training set. The settings of training parameters are shown in Table S1.

Table S1. cPNN-Net training parameters

Training parameters	
Number of layers	30
Number of training dataset	420,000 (Contains 15,000 sets of field measurement data collected from different geological areas)
training strategy	Adam optimizer + Early stopping
Number of time gates	19 (Low moment 3 + High moment 16)
Time gate (s)	6.340E-06~1.247E-04
Train time	8.4 hours
Platform	Windows laptop (Intel(R) Core(TM) i7-12700H CPU @ 2.30 GHz, 16.0 GB RAM), PyCharm, Pytorch

The cPNN-Net in this study runs on a Windows laptop (Intel(R) Core(TM) i7-12700H CPU @ 2.30 GHz, 16.0 GB RAM) using the PyCharm platform. During the network training, a total of 420,000 data pairs are used as the training set, with a batch size of 400. The Adam optimizer (Kingma et al., 2015) is employed to perform gradient descent on the mini-batches of data, and the early stopping (Prechelt 1998) is implemented to determine the optimal number of training iterations. The training process takes a total of 8.4 hours.

Text S2

TEM resistivity distribution

In this study, the introduced cPNN structure incorporates an MDN-Net, capitalizing on the utility of Gaussian Mixture Models (GMMs) to effectively simulate arbitrary distribution functions. This approach is particularly advantageous for the inversion of geophysical parameters, where the complexity is often unknown and unpredictable. Utilizing GMMs as the foundational model for simulating parameter distribution functions offers robustness and adaptability in capturing the heterogeneity of geophysical data. However, it's noteworthy that augmenting the number of Gaussian kernels within the model may significantly increase computational demands. This consideration prompts a critical evaluation of the trade-off between model complexity and computational efficiency. To this end, we introduce a 'toy model' approach designed to optimize the configuration of Gaussian kernels by simulating the posterior resistivity distribution specific to TEM data.

As shown in Figure S1, with simplification for clarity and focus, our toy model assumes a subsurface composed of seven distinct earth layers, employing a configuration of five Gaussian kernels to model the resistivity distribution. The dataset for training encompasses 400,000 data derived from forward modeling, with theoretical resistivity models serving as training labels. This methodological simplification is strategic and helps to focus on exploring the characteristics of the theoretical resistivity distribution function without increasing the complexity of the real inversion models.

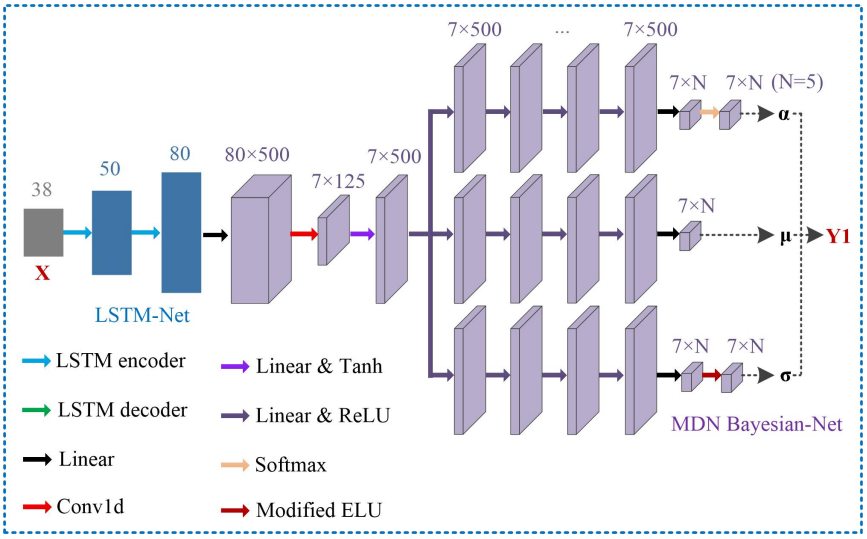

Figure S1. MDN-Net structure with a mixture of five Gaussian kernels. The input is TEM data and corresponding STD value, and the label data is corresponding theoretical resistivity model.

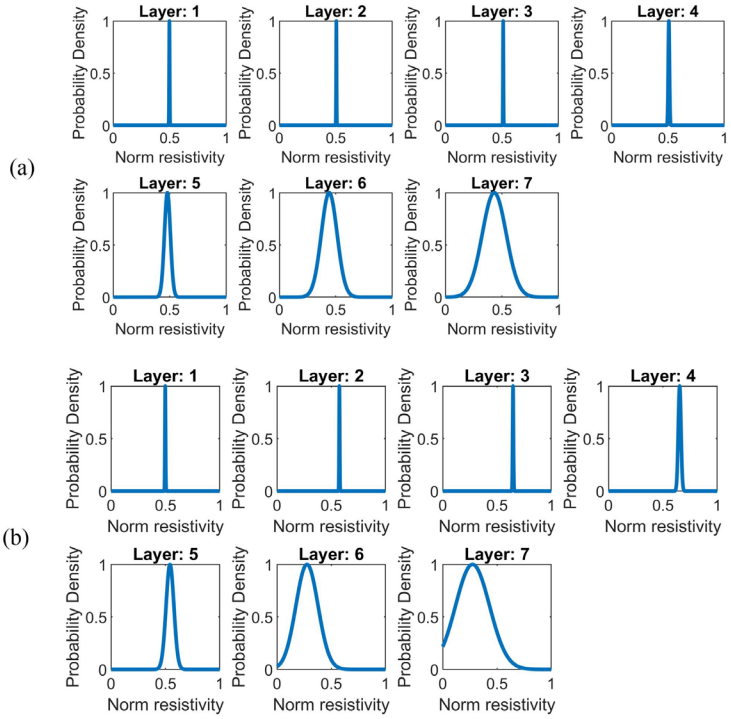

Figure S2. Application of mixture Gaussian distribution to model posterior resistivity in a seven-layer geological formation at Lake Iseo. (a) The first set of data collected at Lake Iseo (b) The 100th set of data collected at Lake Iseo.

After the completion of the training phase for the ‘toy model’ network, an evaluation was conducted using 100 randomly selected sets of TEM sounding data to test the predictive capability concerning resistivity distribution across different subsurface layers. The outcomes from these tests revealed that the simulated posterior resistivity distribution functions for each layer closely align with the characteristics of a standard Gaussian distribution. Figure S2 provides a visual representation of the resistivity distribution for two randomly selected data sets.

Given these outcomes, the decision was made to streamline the model’s complexity by adopting a single Gaussian distribution model (setting the number of mixture Gaussian kernels, N, to 1) for simulating the resistivity distribution functions for each layer. This step effectively reduces the computational demand, enabling the reallocation of resources towards enhancing the model’s depth resolution. Specifically, we extend the model to 30

distinct earth layers, significantly improving the resolution of the inversion results. This methodological adjustment underscores a strategic trade-off between model complexity and computational efficiency, demonstrating that a simplified model can still capture the essential distribution characteristics needed for high-resolution inversion, provided it is carefully calibrated and validated against empirical data

Text S3

MDN principle and loss function design

MDN can directly output an estimate of the Bayesian posterior distribution $p(\mathbf{m}|\ \mathbf{d})$, which can be parameterized under GMMs as:

$$p(\mathbf{m}|\ \mathbf{d}) = \sum_{k=1}^{M} \boldsymbol{\alpha}_k(\mathbf{d})\phi_k(\mathbf{m}|\ \mathbf{d}) \tag{S4}$$

where M is the number of Gaussian kernel functions, $\boldsymbol{\alpha}_k(\mathbf{d})$ is the weight coefficient of each Gaussian kernel, and the sum of the weight coefficients is 1. $\Phi_k(\mathbf{m}|\ \mathbf{d})$ is a Gaussian density function that satisfies:

$$\phi_k(\mathbf{m}\,|\,\mathbf{d}) = \frac{1}{(2\pi)^{c/2}\boldsymbol{\sigma}_k(\mathbf{d})^c}\exp\left\{-\frac{\|\,\mathbf{m}-\boldsymbol{\mu}_k(\mathbf{d})\,\|^2}{2\boldsymbol{\sigma}_k(\mathbf{d})^2}\right\}. \tag{S5}$$

where $\boldsymbol{\mu}_k(\mathbf{d})$ is the mean value of the kth kernel function, $\boldsymbol{\sigma}_k(\mathbf{d})$ is the standard deviation, and c is the dimension of $\mathbf{m}$. The mixture weight coefficient $\boldsymbol{\alpha}_k(\mathbf{d})$, mean $\boldsymbol{\mu}_k(\mathbf{d})$, and standard deviation $\boldsymbol{\sigma}_k(\mathbf{d})$ together form the output of the MDN. By utilizing these parameters, the approximate estimation of the Bayesian posterior distribution can be realized, from which the maximum a posteriori (MAP) model and random sampling model can be extracted (Wu et al. 2021b). In the cPNN structure designed in this paper, the DNN imaging-Net provides an approximate mean solution model of the Bayesian posterior distribution, which forms an effective complement to the Bayesian posterior distribution estimation provided by the MDN. In actual training, the DNN imaging-Net use the mean squared error (MSE) loss function, while MDN uses a loss function based on the logarithmic maximum likelihood principle, which satisfies:

$$Loss = \lambda_1 \mathrm{MSELoss}_{DNN} + \lambda_2 \mathrm{Loss}_{MDN} \tag{S6}$$

where

$$\mathrm{Loss}_{MDN} = -\log\left[\, p(\boldsymbol{m}|\ \boldsymbol{d})\right] = -\log\left[\sum_{k=1}^{M}\boldsymbol{\alpha}_k(\boldsymbol{d})\phi_k(\boldsymbol{m}|\ \boldsymbol{d})\right]. \quad \text{(S7)}$$

In this study, the chosen configuration for the Mixture Density Network (MDN) employs a singular Gaussian kernel in the mixture model, effectively simplifying the setup by setting the number of mixture Gaussian kernels to 1. Additionally, the weighting parameters λ_1 and λ_2 are selected as λ_1=1 and λ_2=10, respectively. This parameter configuration ensures that the overall objective function focuses on the training quality of both DNN and MDN.

Text S4

Lake water depth extraction strategy

In this study, conducted on Lake Iseo, the FloaTEM boat was equipped with Sonar and GPS devices to obtain accurate bathymetric information. To assess the accuracy of the cPNN network inversion results, a comparison with actual Sonar Bathymetry data was considered most convincing. Due to the battery capacity limits of the equipment, Sonar data covered approximately 70% of the survey lines, with the depth detection range between 1-75 m. Therefore, the following procedures were implemented to extrapolate lake depth from the inversion data:

1). Identification of the clay layer's position within the inversion results. Given that the electrical resistivity range of the lake water is approximately 30-55 ohm-m, and the layer beneath the lake water is typically a clay layer with even lower resistivity, usually below 30 ohm-m, we began by examining the resistivity values from the top layer downwards. When the resistivity of a layer dropped below 30 ohm-m, the depth at that layer interface was recorded as the estimated lake depth.

2). Matching the inversion-derived depths with corresponding bathymetry data by locating the nearest coordinate point in the Sonar data for each survey point based on the coordinates. The distance between the two sets of coordinates was calculated, and any pair with a distance greater than two meters was excluded from the analysis, along with their corresponding lake depth data derived from the inversion.

3. Further processing the inversion-derived water depth data. Any water depth less than one meter or greater than 75 meters was set to zero to ensure consistency with the Sonar's detection range. Finally, all data with an extracted water depth of zero meters are deleted because these data are either beyond the sonar detection range or mean that the presence of the clay layer is not detected.

The MDN outputs the distribution function of resistivity, so the lake water depth extracted from the resistivity values in the distribution will exhibit uncertainty. For the estimation of uncertainty in lake water depth shown in Figure 3g in the main text, the calculation method is as follows:

1) Within the one-sigma range, calculate the maximum and minimum resistivity values for each layer output by the MDN.

2) Extract the lake water depth parameter from the maximum resistivity value. Starting from shallow to deeper layers, when the maximum resistivity value is less than 35 ohm·m, the corresponding layer is likely to be a clay layer. The depth of this layer is considered one extreme value of the lake water depth.

3) Extract the lake water depth parameter from the minimum resistivity value. Starting from shallow to deeper layers, when the minimum resistivity value is less than 30 ohm·m, the corresponding layer is considered potentially a clay layer. The depth of this layer is considered another extreme value of the lake water depth.

4) Using these two extreme values, the errorbar function is applied to plot the uncertainty range of the lake water depth distribution. We have selected different thresholds for the maximum and minimum values here, which will make the estimation of the lake depth range more accurate.

It should be noted that this method of estimating lake depth from inversion has errors, due to variations in the resistivity of the clay layer beneath the lake and the potential discrepancy in the nearest coordinate matching. Nevertheless, it is enough to serve as additional reference information to evaluate the overall inversion quality.

Text S5

DOI cumulative sensitivity

The MDN-Net within the cPNN structure can output the standard deviation values, σ, corresponding to the Gaussian distribution functions of the resistivity for each layer of the model. Essentially, σ offers insights into how sensitive the inversion process is to changes or variations in resistivity at different depths. First, we calculate the thickness-normalized sensitivity, S, to eliminate the influence of layer thickness variations:

$$S_j = \frac{1}{h_j \cdot \log\left(\sigma_j\right)^2}. \tag{S8}$$

where, σ_j is the standard deviation value for the jth layer, and h_j is the thickness of the jth layer. The calculation of S_j does not include the last last semi-infinite layer. This adjustment is crucial because it accounts for the fact that the influence of a layer's thickness could otherwise mask the true sensitivity to resistivity changes during the inversion. By normalizing the standard deviation to the layer thickness, the derived sensitivity measure, S, furnishes a more accurate reflection of the relative sensitivity across layers. Then, the cumulative sensitivity S^* at different depth is calculated by summing S from deeper layers to shallower ones, satisfying the following:

$$S_j^* = \sum_{i=M,-1}^{j} S_i. \tag{S9}$$

After obtaining the cumulative sensitivity, it is necessary to define a sensitivity threshold. The imaging depth corresponding to a cumulative sensitivity exceeding this threshold is deemed the Depth of Investigation (DOI) for the depth probability neural network. Currently, the threshold setting is empirical. Our testing has revealed that positioning the threshold within the range of 0.008 to 0.01 generates outcomes closely aligned with DOIs defined by traditional algorithms. Increasing the threshold beyond this range leads to a progressive reduction in DOI depth.

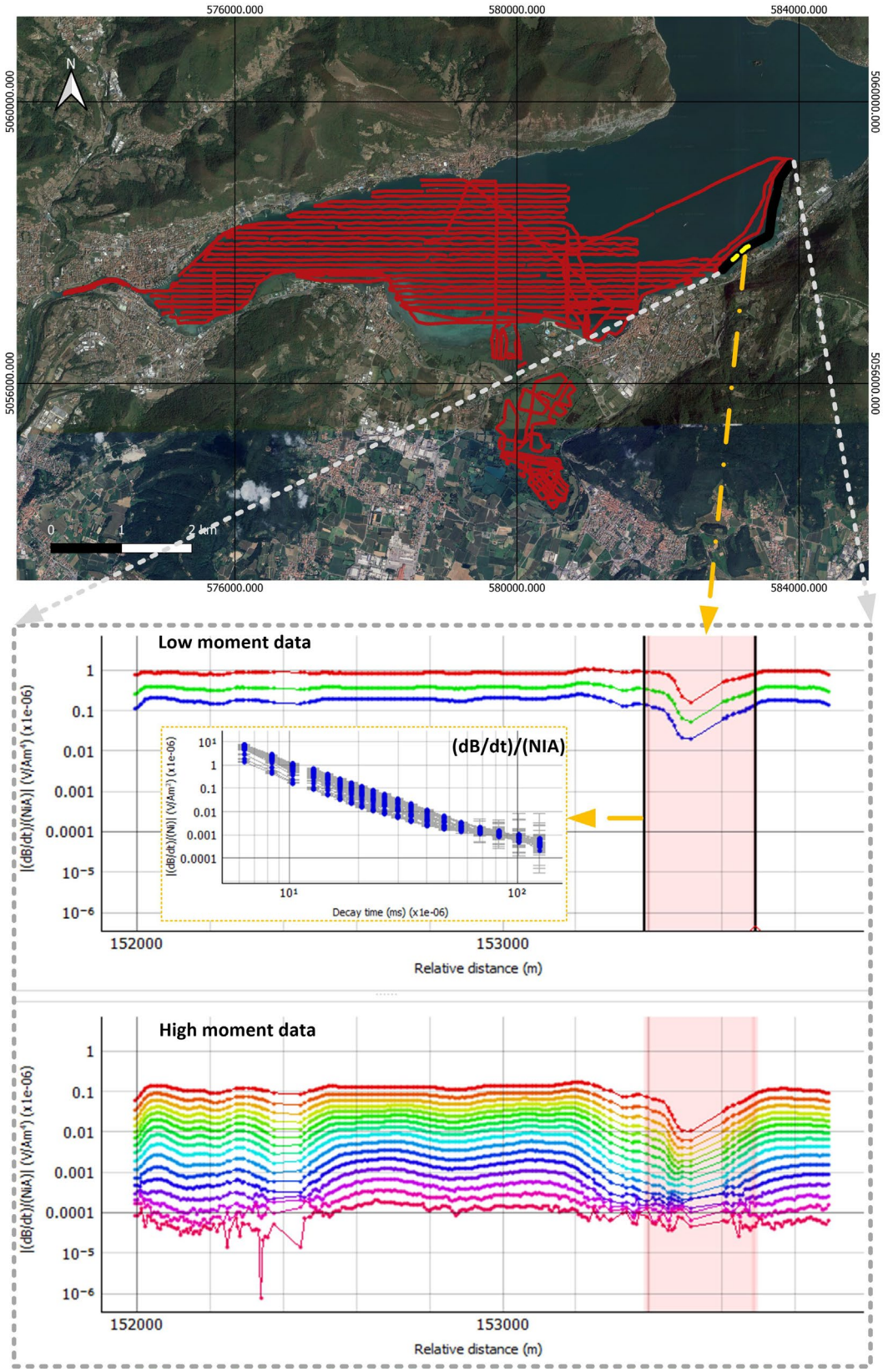

Figure S3. Quality visualization of FloaTEM survey data along a lakeshore Line.

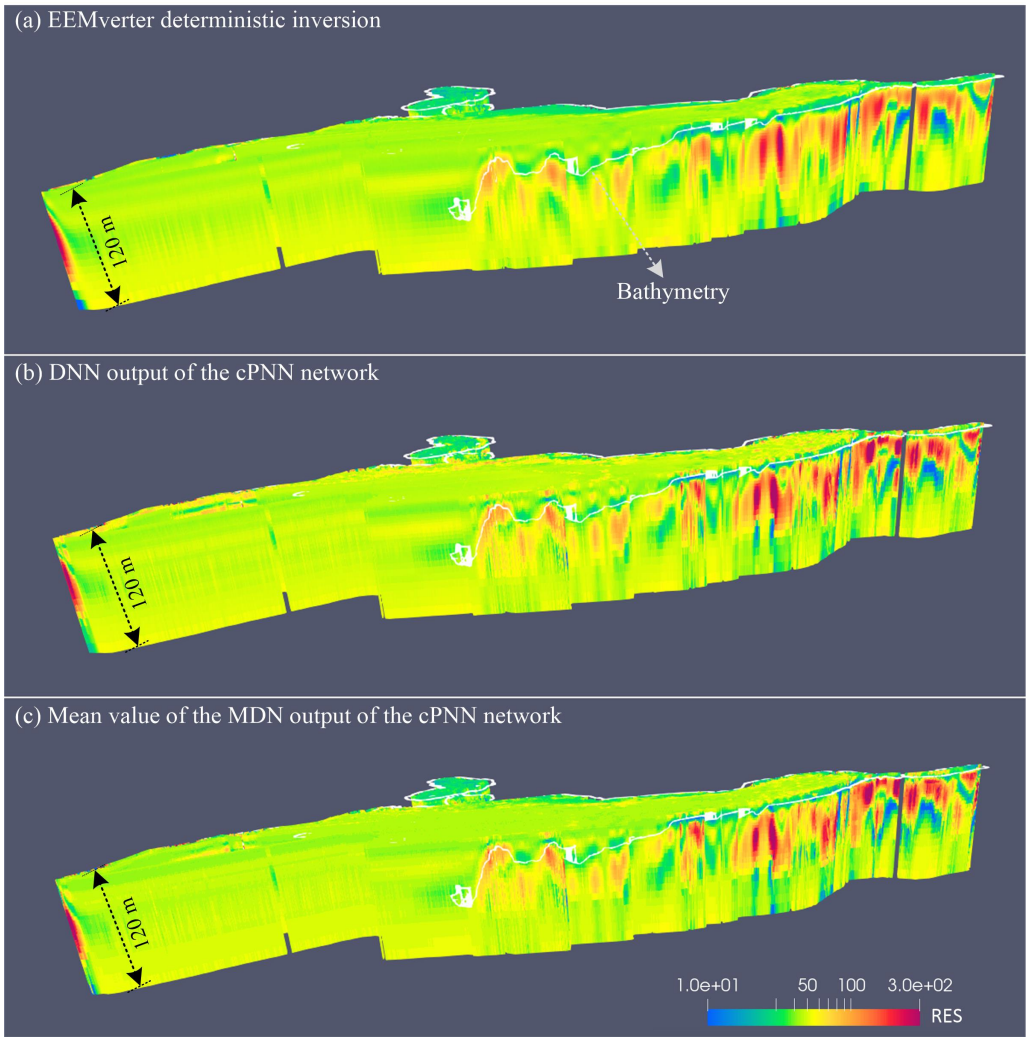

Figure S4. Lake Iseo tTEM data inversion results. From top to bottom: a) EEMverter deterministic inversion; b) DNN output of the cPNN network; c) mean value of the MDN output of the cPNN network. For enhanced clarity in visual representation, a), b), and c) are presented with consistent X and Y scales, while the Z scale is expanded tenfold.

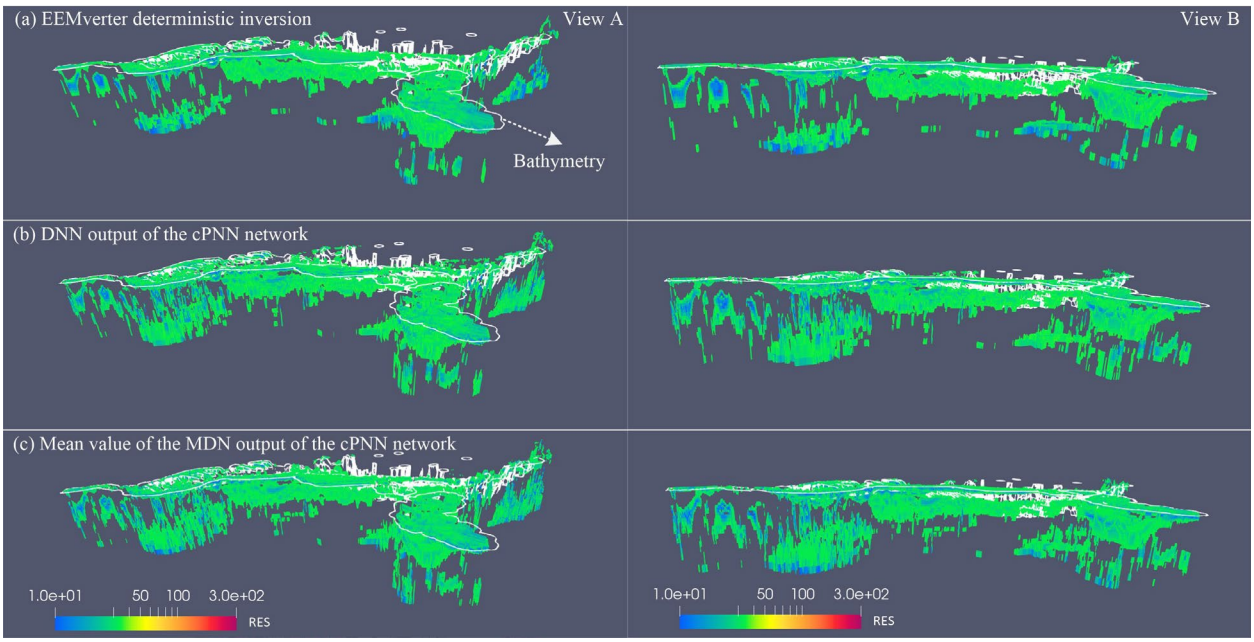

Figure S5. Distribution of the subaqueous clay layers extracted from the inversion results. From top to bottom: a) EEMverter deterministic inversion; b) DNN output of the cPNN network; c) mean value of the MDN output of the cPNN network. For enhanced clarity in visual representation, a), b), and c) are presented with consistent X and Y scales, while the Z scale is expanded tenfold.

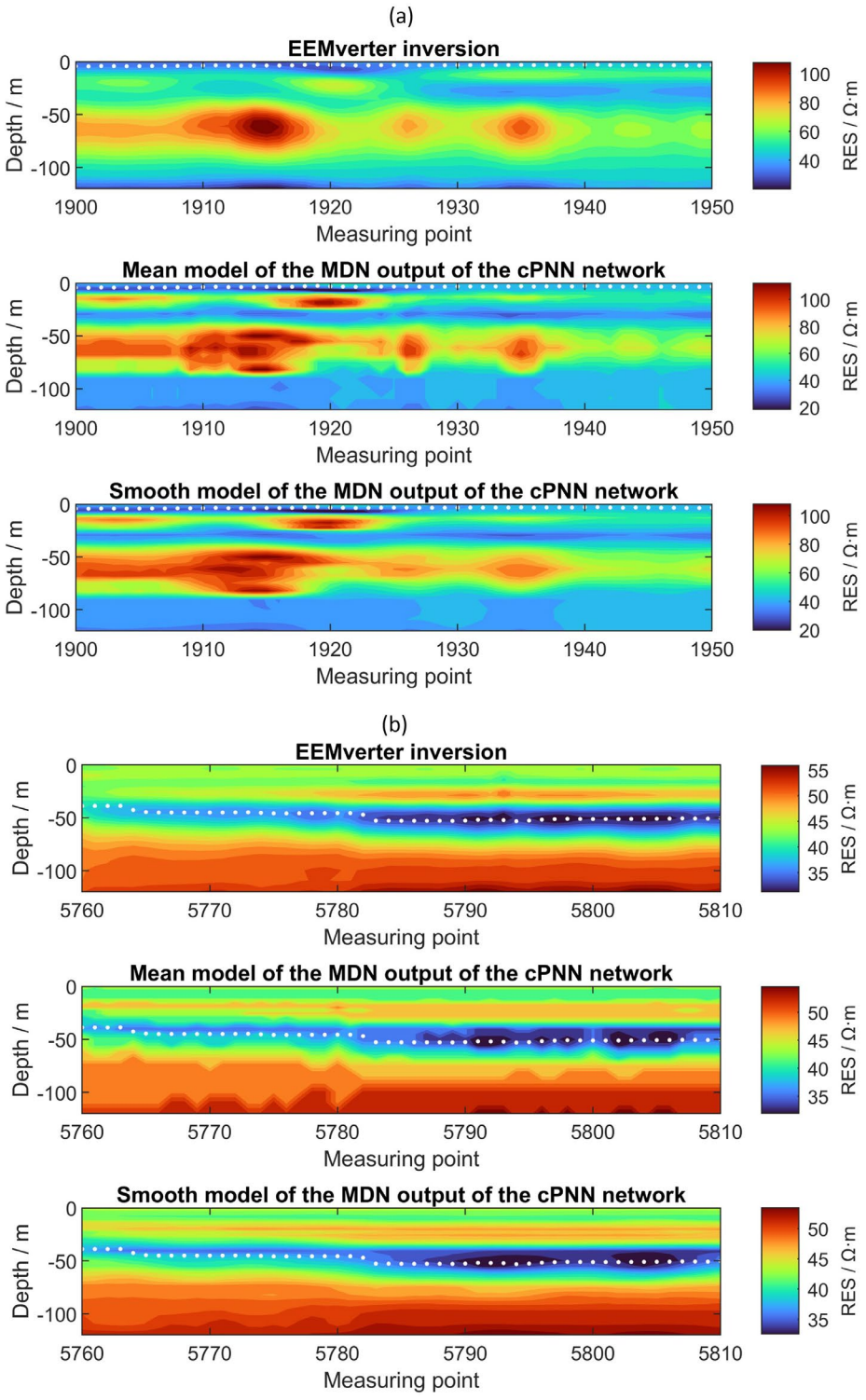

Figure S6. Comparison of smooth model extracted from MDN-Net resistivity distribution function with EEMverter inversion model and MDN-Net mean model. (a) The 1900-1950 survey points data of lake Iseo. (b) The 5760-5810 survey points data of lake Iseo. The white dots in the figure represent sonar bathymetry data.