IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 10 August 2023, accepted 30 August 2023, date of publication 11 September 2023, date of current version 18 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3314335

APPLIED RESEARCH

Optimal Inversion Method for Composite Layered Soil Model Considering Outlier Dispersion

XIAOBING XIAO[1], YONGXIANG CAI[1], XIAOMENG HE[1], HUAPENG LI[1], YUE LI[1], XINYI HE[1], TAO YUAN[2], AND QIAN CHEN[2]

[1]Guizhou Power Grid Company Ltd., Guiyang 550000, China
[2]School of Electrical Engineering, Chongqing University, Chongqing 400044, China

Corresponding author: Qian Chen (1462571881@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 51777020 and in part by the Science and Technology Project of Country Electric Power Research Institute of Guizhou Power Grid Corporation "Research on Rapid Extinction of Lightning Arc and Cooperative Lightning Protection Technology of Distribution Line in Mountainous Area" under Grant 0666002021030101ZD00004.

ABSTRACT Accurate soil structure models are crucial references for substation grounding system design. Typically, inversion algorithms are employed to obtain uniform or horizontal layered soil models based on measured apparent resistivity. However, soil resistivity outlier distribution areas can affect the accuracy of these inversion algorithms, particularly when these areas are near the surface. Traditional algorithms do not account for the outlier distribution of soil resistivity, leading to significant discrepancies between the calculated results of the design scheme and actual operation, thereby impacting the safety and economy of the grounding system. Therefore, this paper proposes an inversion method for soil structures with outlier distribution characteristics based on deep belief networks (DBNs). Firstly, we introduce a statistical criterion for identifying the outlier distribution characteristics of soil resistivity. Subsequently, we construct a database of soil models with outlier distribution characteristics to train the DBN. Finally, we verify the inversion accuracy of the optimal DBN using apparent resistivity data measured in a 220 kV substation and the Qinghai-Tibet Railway. The results demonstrate that the inversion accuracy of the method proposed in this paper is comparable to that of the traditional method for horizontally layered soil but exhibits a remarkable improvement of approximately 40% when dealing with soil apparent resistivity exhibiting outlier distribution characteristics.

INDEX TERMS Grounding, soil model, outlier dispersion, deep belief network.

ABBREVIATION LIST

DBN	Deep belief networks.
CDEGS	Current Distribution, Electromagnetic Interference, Grounding, and Soil Structure Analysis Package.
GA	Genetic algorithm.
ANN	Artificial neural network.
RMSE	Root mean square error.
ML	Machine learning.
RBM	Restricted Boltzmann Machine.
MSE	Mean squared error.
M-MSE	Mini-batch training MSE.
F-FMSE	Full-batch train MSE.
V-MSE	Validation MSE.
SPSS	Statistical Package for the Social Sciences.
ROC	Receiver operating characteristic curve.
AUC	The area under the ROC curve.
SD	Steepest descent method.
SA	Simulated Annealing.
BFGS	Broyden-Fletcher-Goldfarb-Shanno algorithm.
PSO	Particle swarm optimization.
ρ_i	The resistivity of the i th layer of soil.
h_i	The thickness of the i th layer of soil.
$S(x_0, y_0, z_0)$	The coordinates of the center of the outlier dispersion area.

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

d_1, d_2, h	The length, width and height of the outlier distribution area.
ρ_y	The resistivity of the outlier area.
$\rho_{f1}, \rho_{f2}, \rho_{f3}$	The apparent resistivity calculated in three conditions.
a	The average value of a group apparent resistivity.
b	The standard deviation of a group apparent resistivity.
H	The ratio of the a to b.
$G_{0.9}(12)$	The value of the outlier confidence of 0.9 in the Grubbs table corresponding to 12 data, which is 2.134.

I. INTRODUCTION

A reliable grounding system is crucial for ensuring the safe operation of a power system and the protection of personnel [1]. Therefore, an accurate soil structure model is essential for designing a grounding system. In the past, the uniform soil model was commonly used for grounding design, derived by averaging the apparent resistivity values obtained from the Wenner Four-point method at different spacing intervals [2]. However, this model can lead to significant errors if there are substantial variations in soil resistivity values along the vertical direction. As a result, it is less commonly applied nowadays, and the horizontally layered soil model has gained widespread use instead. This model determines the soil resistivity and thickness of each layer by inverting the apparent resistivity data obtained through the Wenner Four-point method. The inversion process involves numerically calculating the electromagnetic forward modeling of the soil model and iteratively optimizing the inversion function [3]. The calculation of Green's function of potential distribution under horizontal multi-layer soil structures is commonly performed using the complex image method [4], [5]. The objective function for inversion adopts the root mean square error (RMSE) function of apparent resistivity. Over time, the iterative inversion optimization process has evolved from the earliest gradient-based solution methods to intelligent algorithm–based search methods.

Dawalibi utilized the steepest descent method to invert parameters for the horizontal two-layer soil model [3]. This marked the emergence of analytical methods and optimization algorithms based on gradient operations for soil model inversion. Alamo improved convergence speed by using the Newton-Raphson method for inverting the two-layer soil model [6]. In the same year, Dawalibi integrated various results of soil model inversion and grounding parameter calculations into the grounding design software CDEGS (Current Distribution, Electromagnetic Interference, Grounding, and Soil Structure Analysis Package) [7]. Pan and Yang further contributed to the optimization calculation of the two-layer soil model by adopting the Powell and simplex algorithms, respectively [8], [9]. Although these direct search methods circumvent complex and tedious Green's function derivations and are not heavily dependent on the selection of initial values, they do not apply to complex soil models and thus did not see widespread use.

Zeng summarized the optimization process of multi-layer soil models using the Gauss-Newton and Marquardt methods and conducted engineering tests [10]. Later, Zou and Wen proposed a two-stage inversion method and a quasi-Newton method, respectively, improving convergence speed and solution complexity [11], [12]. In 2005, Gonos introduced artificial intelligence algorithms into the soil model inversion process for the first time, significantly simplifying the calculation process for two-layer and three-layer soil model inversions [13]. The inversion results show that the method based on genetic algorithm (GA) is greatly improved in accuracy compared with the traditional method based on theoretical derivation and drawing ρ-a curve. In the same year, Lee employed the artificial neural network (ANN) to invert parameters for the two-layer soil model [14], marking the earliest inversion method based on big data. The inversion results of 40 sets of measured data show that the error between calculated apparent resistivity and the measured value is less than 2%. However, this method was limited in its application due to the overly simplistic soil model. Zhang introduced particle swarm optimization to invert the horizontal multi-layer soil parameters, and compared with other literatures, it shows that the accuracy of this method is higher than that of genetic algorithm and simulated annealing method [15]. Subsequently, researchers have continuously applied and enhanced artificial intelligence algorithms for direct search, bypassing the need for a rigorous selection of initial values and derivations. Examples include particle swarm optimization, variable precision genetic algorithm, differential evolution algorithm, and various heuristic algorithms. In [16], GA technique with 40th degree polynomial is used to decompose the integral of kernel function and apparent resistivity estimation. Complex image method along with Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm is used to optimise the complex soil parameters [5]. Multilayer soil parameters are evaluated using Chebyshev polynomials with differential evolution technique in [17]. Improved particle swarm optimisation technique in [18] and differential evolution in [19] are used to estimate the parameters of complex soil structure. The inversion of one-dimensional soil parameters in the frequency domain with considering multilayered soil model based on simulated annealing algorithm is proposed in [20]. In recent years, researchers have persistently explored the horizontally layered soil model and made advancements in its measurement process, forward modeling, and inversion optimization [21], [22], [23], [24], [25], [26], [27].

The horizontally layered soil model considers the distribution of soil resistivity in the vertical direction, significantly improving the inversion accuracy compared to the uniform soil model. However, practical soil often contains blocks or lenses with extremely high resistivity, such as the cement foundation of nearby buildings around an underground substation or large rocks in the soil near a tower grounding

device in mountainous areas. These areas have resistivity several times or even tens of times higher than the surrounding soil, significantly impacting the accuracy of the horizontally layered soil model obtained through inversion. To achieve an apparent resistivity RMSE lower than the recommended 10% from reference [3], [15], [16], these high resistivity lumps cannot be disregarded, necessitating the establishment of a soil model with apparent resistivity outlier distribution characteristics. Some researchers have obtained the three-dimensional spatial distribution of earth resistivity by inverting magnetotelluric measurement data [24], [25], but this method faces challenges in engineering applications. Firstly, the electromagnetic forward calculation for this complex three-dimensional soil structure is not yet fully mature, and it can only be accurately calculated for the grounding electrode, not the grounding grid parameters. Additionally, the magnetotelluric method's measurement range spans several to hundreds of kilometers, resulting in poor exploration accuracy for shallow resistivity. In practice, the grounding device's performance is primarily influenced by its low resistivity, and the four-point method exhibits high measurement accuracy within a range of several hundred meters [27]. Hence, the Four-point method becomes the preferred measurement approach for grounding system design. However, obtaining accurate outlier dispersion areas through traditional inversion optimization algorithms becomes challenging when dealing with complex initial models and numerous parameters. So far, no method exists for directly obtaining the composite layered soil model with an outlier distribution area based on apparent resistivity measured by the Four-point method. Thus, there is an urgent need to develop a new algorithm, different from previous inversion approaches, that can efficiently and directly derive the composite layered soil structure model from apparent resistivity data acquired through the Four-point method.

The development of machine learning (ML) approaches has opened new possibilities for soil prediction and geotechnical assessments, including soil erosion susceptibility modeling and soil model inversion [37], [38]. These approaches feature a non-linear structure and aim to find robust relationships between input and output parameters readily available in data. Unlike physically based models, ML techniques do not focus on explaining the underlying physical processes or mathematical reasoning for behavior. Instead, they recognize expected and unexpected patterns within the data. The key advantage of ML approaches lies in their ability to predict some difficult or expensive parameters, such as soil model parameters, using other readily available factors. Lee used ANN to invert the parameters of two-layer horizontal layered soil, and used 40 sets of data to verify the accuracy of the method [14]. Niyogi used Support Vector Machine, Random Forest and Deep Neural Network to predict the shear strength of soil [39]. Moghadas used ANN to study the relationship between soil water content and electrical conductivity. The results show that this method is superior to the Rhoades model in some aspects [40]. Zhang used U-net to invert the anomalous body in the soil, including the contaminated area and the degree of mineralization of the mountain [41]. Liu used a cellular neural network to separate gravity anomalies. The results show that the method has strong lateral resolution and can be used for soil survey and mine resource exploration [42].

This paper proposes an inversion method based on deep belief networks (DBNs), capable of obtaining a soil structure model with an outlier distribution area based on the measured apparent resistivity. The method overcomes the limitations of current soil model inversion algorithms with low accuracy when faced with outlier distribution areas in the soil, providing an effective reference for the inversion idea of the composite layered soil model.

II. ESTABLISHMENT OF COMPOSITE SOIL MODEL WITH DISCRETE DISTRIBUTION CHARACTERISTIC

This paper presents the establishment of a soil model exhibiting an outlier distribution of apparent resistivity. As depicted in Fig.1, the surface of the layered soil structure features an outlier distribution area characterized by exceptionally high soil resistivity, thereby forming a composite layered soil model that accounts for outlier dispersion conditions. The Schlumberger Four-point method was selected for several advantages: less frequent probe movements, faster measurements, easier discrimination of lateral and depth changes in geology, and the ability to measure deeper soil depths compared to the Wenner arrangement [16].

A. DEFINITION OF THE OUTLIER DISPERSAL AREA

The parameters characterizing the outlier dispersal area are as follows:

1) The location of the outlier dispersal area is determined using a three-dimensional coordinate system. The X-axis coincides with the arranged probe line, the Y-axis is parallel to the surface and perpendicular to the probe line, and the Z-axis is perpendicular to the surface plane. The coordinate origin is positioned at the center of the surface probe line. Represented as a cuboid, the outlier dispersal area has center coordinates denoted as $S(x_0,y_0,z_0)$, reflecting the position change of the whole outlier dispersal area.

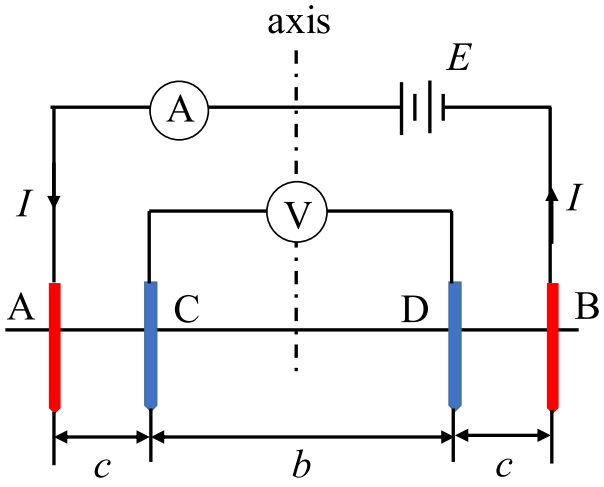
2) The length and width of the upper and lower surfaces in the outlier dispersal area are d_1 and d_2, respectively, and the height is h, representing the size variation of the outlier dispersal area.

3) The resistivity in the outlier dispersal area is denoted by ρ_y.

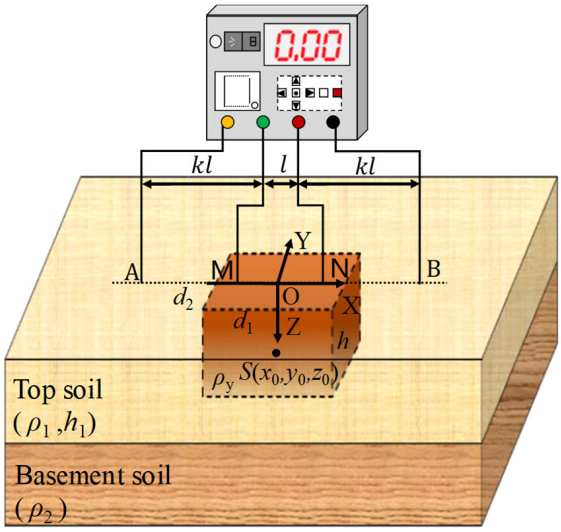
Hence, the parameters of the outlier dispersal area model encompass $S(x_0,y_0,z_0)$, h, d_1, d_2, and ρ_y.

B. INFLUENCE OF OUTLIER DISTRIBUTION AREA ON APPARENT RESISTIVITY OF THE SOIL MODEL

To investigate the influence of the outlier distribution area on the apparent resistivity of the soil model, we employed the



(a) Schlumberger arrangement



(b) Composite stratified soil model considering the outlier dispersion characteristics of apparent resistivity

FIGURE 1. Illustration of (a) Schlumberger arrangement and (b) composite stratified soil model considering apparent resistivity's outlier dispersion characteristic.

composite two-layer soil model and established a method to compare the apparent resistivity error between the traditional soil model and the model presented in this paper. First, the Schlumberger Four-point method measurement simulation platform is established using CDEGS, as depicted in Fig.1.

Fig.1(a) illustrates the Schlumberger Four-point method, recommended by IEEE standard 81-2012 [29], for soil apparent resistivity measurement. The apparent resistivity can be calculated as follows:

$$\rho = \frac{\pi c(b+c)U}{bI} \tag{1}$$

where b denotes the distance between the internal probes, c denotes the distance between the internal probe and the adjacent external probe, U denotes the potential difference of internal probe spacing, and I denotes excitation current.

TABLE 1. Outlier dispersion areas of soil resistivity under three conditions.

Model parameters	Condition 1	Condition 2	Condition 3
S	(0,0,5)	(0,0,30)	
d_1	20	20	without high
d_2	20	20	resistivity
h	10	10	area
ρ_y	10000	10000	

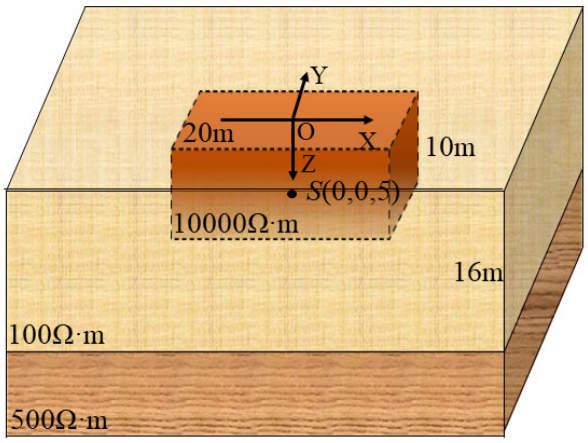
To explore the influence of outlier distribution area, 12 different probe distances were set up; b was set to 1 m, and c was set to kb, where $k = 0.9, 1.5, 2.3, 3.5, 5.0, 7.0, 9.5, 13.5, 19.5, 27.5, 39.5, 54.5$, respectively. The next step involved defining the composite layered soil model, incorporating the outlier distribution characteristics of soil resistivity. Subsequently, the apparent resistivity was calculated for 12 different probe spacings using the Equation (2):

$$\rho = \frac{k(k+1)\pi U}{I} \tag{2}$$

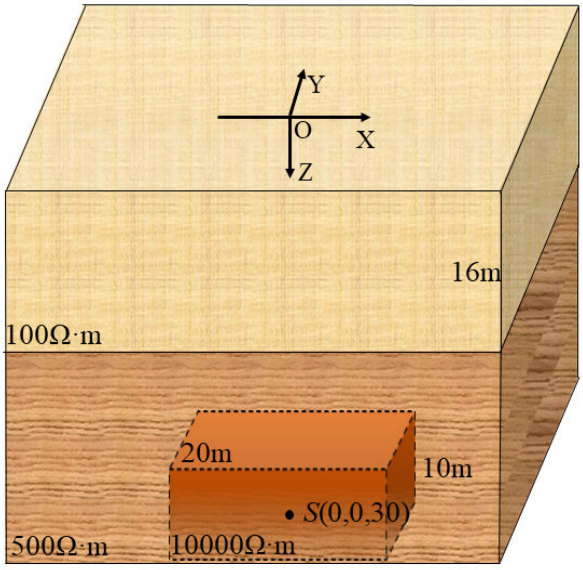
The apparent resistivity calculated using Equation (2) serves as the actual measured value, enabling the inversion of either a uniform or layered soil model using CDEGS. The RMSE of apparent resistivity can also be determined. By adjusting the parameters of the outlier distribution area, multiple sets of apparent resistivity values and traditional horizontal layered soil models can be obtained. If the RMSE of these models is less than the 10% threshold recommended in the literature, it indicates equivalence between the traditional and proposed soil models. On the other hand, if the RMSE exceeds the threshold, it suggests the need to consider the influence of the outlier distribution area in the soil.

This method is applied to analyze the changes in the apparent resistivity RMSE when setting different parameters for the outlier distribution area in a typical horizontal two-layer soil model.

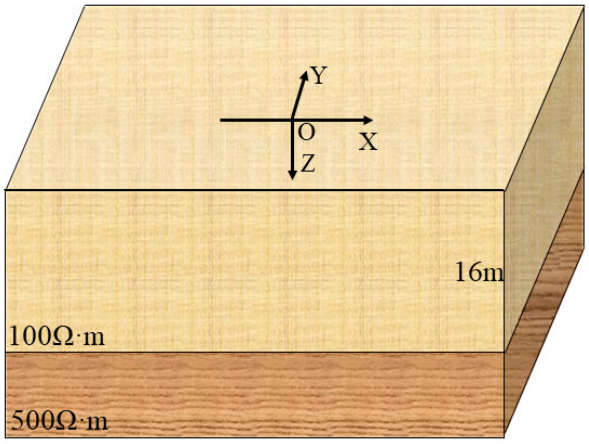
To investigate the influence of the depth of the outlier distribution area on the results of soil apparent resistivity measurement and soil model inversion, we placed the outlier distribution area at the top and bottom of the horizontal two-layer soil. Subsequently, we compared the measurement and inversion results with a horizontal two-layer soil model lacking an outlier distribution area. The resistivity of the first soil layer was set to 100 Ω·m, with a thickness of 16 m. The resistivity of the second soil layer was set to 500 Ω·m. The parameters of the soil resistivity outlier distribution area are detailed in Tab.1. By contrast, the soil models for the three cases are depicted in Fig. 2. The measured apparent resistivity of the three cases is ρ_{f1}, $\rho_{f2,}$ and ρ_{f3}, respectively, and the results are shown in Tab.2. Based on these apparent resistivity inversion sets, we obtained the traditional horizontal layered soil model with the smallest RMSE. The RMSE values of the model parameters and their apparent resistivity are shown in Tab.3.



(a) Soil model of condition 1



(b) Soil model of condition 2



(c) Soil model of condition 3

FIGURE 2. Typical soil models for three conditions.

C. ESTABLISHMENT OF SOIL MODEL LIBRARY

In the context of a two-layer soil structure, high-resistivity soil blocks in the upper and lower layers can significantly influence the measurement of soil apparent resistivity. Consequently, this affects the final inversion of the obtained soil structure. To explore this scenario, let's assume that the surface soil resistivity is 100 Ω·m with a thickness of 16 m, while the bottom soil resistivity is 500 Ω·m. The parameters characterizing the soil resistivity dispersion area are presented in Tab.1, and a schematic diagram depicting three typical soil models can be found in Fig.2. Using CDEGS software, the apparent resistivity data under various probe distances are calculated, and these values are then utilized as the measured data for inverting the soil structure. This process enables the derivation of the traditional horizontal layered soil model that yields the minimum RMSE. The RMSE values for the model parameters and their apparent resistivity are listed in Tab.2.

TABLE 2. The apparent resistivity of two-layer horizontal soils considering different outlier dispersion areas.

AB/2/m	ρ_{f1}/Ω·m	ρ_{f2}/Ω·m	ρ_{f3}/Ω·m
1.4	9983	100	99
2.0	9950	100	100
2.8	9858	100	100
4.0	9571	100	100
5.5	8818	100	100
7.5	6696	102	100
10.0	4900	105	103
14.0	206	110	109
20.0	216	123	123
28.0	254	146	145
40.0	306	183	181
55.0	368	224	224

TABLE 3. The layered soil model obtained by inversion.

Model parameters	Condition1	Condition2	Condition3
ρ_1/Ω·m	12412	99	100
h_1/m	2.7	12.6	15
ρ_2/Ω·m	70	327	506
h_2/m	6.5	/	/
ρ_3/Ω·m	837	/	/
RMSE	34.16%	2.48%	0.19%

As shown in Tab.2 and Tab.3, ρ_{f2} and ρ_{f3} exhibit nearly identical values, and the RMSE of the horizontal layered soil model inverted using these values remains within 10%. This implies that the influence of the dispersion area on the horizontally layered soil model is negligible when the area is distant from the surface. Hence, we can conclude that the soil model considering the outlier distribution area is equivalent to the traditional layered soil model in such cases. However, when the outlier distribution area is close to the surface, the calculated apparent resistivity ρ_{f1} and ρ_{f3} differ significantly, and the apparent resistivity displays characteristics of outlier distribution. Under these circumstances, transforming the outlier distribution area into the layered soil

model introduces substantial errors, resulting in an RMSE of 34.16%. In this scenario, the soil model considering the outlier distribution area is not equivalent to the traditional soil model. Consequently, when the outlier distribution area is located on the surface, it becomes essential to account for the outlier distribution conditions in the soil model.

III. CRITERION OF APPARENT RESISTIVITY DISPERSION IN SOIL MODEL

In statistical science and computer science application engineering, outlier points refer to data points that deviate greatly from the rest of the data in a given dataset. They are defined as observation points inconsistent with the overall data pattern or significantly deviating from the overall trend. The objective of identifying outliers is to eliminate noise or identify potentially meaningful knowledge [28].

Various rules currently exist for determining outliers, including the Layda method, the Shawville method, the Grubbs method, and the Dixon test method, among others [30]. However, the criteria for determining outliers in a dataset cannot be universally generalized for different practical engineering problems. In some cases, outliers with profound significance within a specific professional field may not be identified, significantly affecting the accuracy of certain calculations. For the inversion problem of the composite layered soil model addressed in this paper, the resistivity of the surface block area defined in Section II is several times higher than that of the surrounding soil. Consequently, certain measured apparent resistivity values may deviate from the mean value, distorting the overall apparent resistivity and profoundly affecting the model's accuracy. The aforementioned rules are not sufficient for determining whether a specific apparent resistivity value is an outlier or possesses outlier distribution characteristics. Therefore, whether a complex soil model considering the outlier distribution area is required should be determined through the statistical analysis of apparent resistivity.

We modified the Grubbs method to establish a statistical criterion suitable for identifying the presence of outlier distribution in apparent resistivity for the soil model. We constructed a composite layered soil model similar to that described in Section II using CDEGS to develop this criterion. Subsequently, we introduced block areas on the surface with resistivity values more than ten times higher than the environmental resistivity. Through extensive calculations and debugging, we formulated the following conditions to judge outliers in the apparent resistivity of actual soil structure measurements:

1) H is defined as the ratio of the average value to the standard deviation of a group of soil apparent resistivity, $0.75 < H < 1.35$;

2) For a measured value in this group of apparent resistivity, $2.5G_i > G_{0.9}(12)$. G_i is defined as follows [29]:

$$G_i = \left| \frac{X_i - a}{b} \right| \tag{3}$$

TABLE 4. Test results of outliers of apparent resistivity in Table 2 according to the outlier dispersion criterion in this paper.

ρ_{f1}/Ω·m	H	G_i	ρ_{f2}/Ω·m	H	G_i	ρ_{f3}/Ω·m	H	G_i
9983		2.714	100		1.514	99		1.530
9950		2.696	100		1.514	100		1.468
9858		2.645	100		1.514	100		1.468
9571		2.486	100		1.514	100		1.468
8818		2.068	100		1.514	100		1.468
6696	1.13	0.890	102	3.09	1.390	100	3.07	1.468
4900		0.108	105		1.204	103		1.282
206		2.714	110		0.894	109		0.910
216		2.708	123		0.088	123		0.041
254		2.687	146		1.338	145		1.324
306		2.658	183		3.632	181		3.557
368		2.624	224		6.174	224		6.225

where X_i represents the ith measured value in the apparent resistivity of this group, a denotes the average value of the apparent resistivity of this group, b denotes the standard deviation of the apparent resistivity of this group, $G_{0.9}(12)$ represents the value of the outlier confidence of 0.9 in the Grubbs table corresponding to 12 data, which is 2.134.

If the above two conditions are simultaneously met, a certain measured value within this apparent resistivity group is considered an outlier point, indicating that this group exhibits the characteristics of an outlier distribution.

By employing the outlier distribution criterion for the apparent resistivity of the proposed soil model, we can differentiate the three sets of apparent resistivity values presented in Tab.2. The results are summarized in Tab.4.

From the calculation results presented in Tab.4, it is evident that the first group of apparent resistivity data satisfies the first condition of the outlier dispersion criterion formulated in this paper, and the G values for the 1st, 2nd, 3rd, 4th, 8th, 9th, 10th, 11th, and 12th measured values meet the second condition. As a result, these measured values are considered outliers, indicating that the apparent resistivity of the first group of soil models exhibits characteristics of outlier dispersion. Therefore, it becomes necessary to account for the block outlier dispersion area on the surface during the inversion of the soil model. In contrast, the G values for the second and third groups satisfy the conditions, but the H values lie outside the interval defined by the outlier dispersion criterion in this paper. Consequently, these two data sets lack outlier distribution characteristics, and the traditional horizontal layered soil model can achieve higher accuracy when utilized.

IV. SOIL MODEL INVERSION METHOD BASED ON DEEP BELIEF NETWORKS

A DBN is a multi-hidden layer neural network that combines unsupervised and supervised training methods [31]. By initializing unsupervised network parameters and employing supervised optimization, the DBN enables

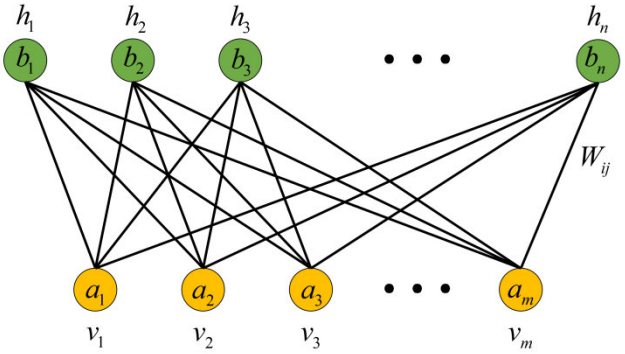


FIGURE 3. Structure of the restricted boltzmann machine.

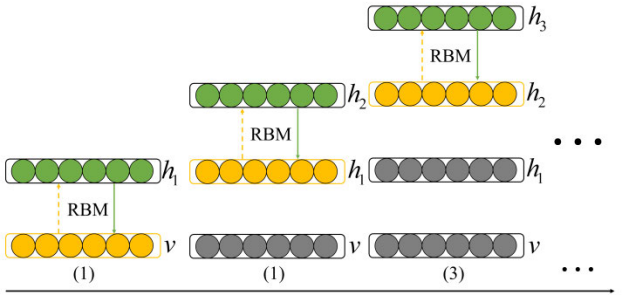


FIGURE 4. Unsupervised training process for DBN.

more straightforward avoidance of the optimization problems associated with nonlinear mapping compared with traditional methods. For the soil model inversion problem, this translates complex numerical and iterative calculations into a large data sample training process for the optimal network. Upon completion of network training, researchers can quickly obtain refined and accurate soil structure models for similar types of soil areas within a certain range without the need to adjust the network's super parameters. This significantly enhances the efficiency of soil model parameter inversion. Most notably, due to its distinct big data operation mechanism, the DBN can also learn the discrete distribution characteristics of apparent resistivity in shallow local areas. As a result, the DBN can be utilized to predict this composite layered soil model, thereby showcasing its greatest advantage.

A. TRAINING METHODS AND COMPOSITE EVALUATION METHODS OF DBN

The fundamental unit of the DBN is the restricted Boltzmann machine (RBM), which functions as a random neural network [31]. By learning the probability distribution of the sample, the network parameters automatically adjust to achieve specific functions. The basic structure of the DBN, as illustrated in Fig.3, consists of the visual layer (v) and the hidden layer (h). The visual layer comprises n neurons, while the hidden layer includes n neurons. In this context, v_i represents the i-th visual layer neuron, and h_j represents the j-th hidden layer neuron. The connection weight matrix between the i-th visual layer neurons and the j-th hidden layer neurons is denoted as W_{ij}, whereas a_i and b_j represent the biases of the i-th visual layer neurons and the j-th hidden layer neurons, respectively. The RBM is used to identify the mapping relationship between v and h, transforming the input v into the output h through feature representation.

The conversion between v and h is represented by a joint probability distribution between them, defined by the following energy function:

$$E(v,h) = -\sum_{i=1}^{m} a_i v_i - \sum_{j=1}^{n} b_j h_j - \sum_{i=1}^{m}\sum_{j=1}^{n} h_j W_{ij} v_i \quad (4)$$

The probability that neurons are activated is [32]:

$$P(v_i = 1|h) = \sigma(a_i + \sum_{j=1}^{n} W_{ij} h_j) \quad (5)$$

$$P(h_j = 1|v) = \sigma(b_j + \sum_{i=1}^{m} W_{ij} v_i) \quad (6)$$

The learning process of RBM is to calculate the parameters a_i, b_j, and W_{ij} to maximize the likelihood function. The expression is as follows:

$$L = \log \prod_{t=1}^{T} P_\theta(v^{(t)}) = \sum_{t=1}^{T} \log P_\theta(v^{(t)}) \quad (7)$$

where L represents the log-likelihood function of the training sample set; T is the number of training sample sets; t is the sample number; P is the joint probability distribution; and θ is the model parameter set, including a_i, b_j, and W_{ij}. The expression for solving the optimal model parameter θ_m is as follows:

$$\theta_m = \arg\max_{\theta} \sum_{t=1}^{T} \log P_\theta(v^{(t)}) \quad (8)$$

θ is solved iteratively according to the derivative of the likelihood function L:

$$\theta_{i+1} = \theta_i + \eta \frac{\partial L}{\partial \theta_i} \quad (9)$$

where θ_i denotes the value of θ at i times of iteration i, θ_{i+1} denotes the value of θ at $i+1$ times of iteration, and η denotes the learning rate.

The above presents the training process for a single restricted Boltzmann machine, and the DBN consists of multiple stacked restricted Boltzmann machines. The unsupervised training process is depicted in Fig.4. First, the sample is employed to train the bottom RBM [33]. Once the parameters of a single RBM achieve their optimal solution, they remain unchanged. Subsequently, the subsequent RBMs are trained in a layered manner, ultimately optimizing the parameters of each RBM.

Subsequently, the RBM network is expanded to a BP neural network, as illustrated in Fig.5. The BP neural network

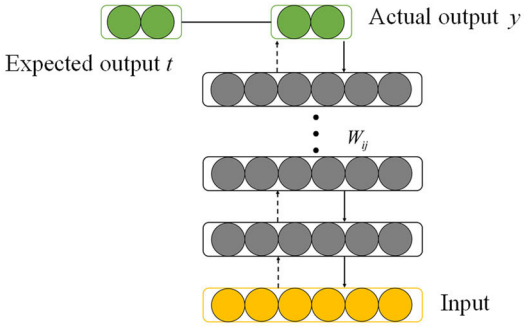


FIGURE 5. BP neural network.

undergoes supervised optimization using the backpropagation algorithm.

During the forward transfer process of the BP neural network, the actual output is represented as y, while the expected output is denoted as t. Accordingly, the error can be defined as follows:

$$\delta = t - y \tag{10}$$

The error δ is inversely propagated, and W_{ij} is adjusted to the opposite direction of its derivative to δ:

$$W_{ij}^{(k+1)} = W_{ij}^{(k)} - \eta(d\delta/dW_{ij}) \tag{11}$$

where η is the learning rate, k is the current number of training, and then one training is completed. The network's weights are continually updated by utilizing many samples, enabling the final output to approximate the expected output accurately and achieve any nonlinear mapping from input to output.

During the unsupervised training process, the basic structure of DBN - the Restricted Boltzmann Machine (RBM), requires adjustments of hyperparameters, including the number of network layers and the number of neurons in each layer. The training process adopts the method of mini-batch sample training (Mini-batch), wherein each iteration involves selecting a mini-batch sample from the sample set and computing it using the network. Therefore, the number of mini-batch training samples (Batchsize) and the number of training epochs for all samples (Epochs) were adjusted based on the experience in this study. The learning rate and momentum are crucial parameters that were fine-tuned during the learning process. They jointly control the size of each adjustment of network weights, and appropriate parameter settings can optimize network performance. The evaluation index of network performance is the mean squared error (MSE) between the actual output of the network and the expected output of the sample. The calculation equation is as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - a_i)^2 \tag{12}$$

where n denotes the number of calculated samples, y_i represents the expected output of the soil model parameters

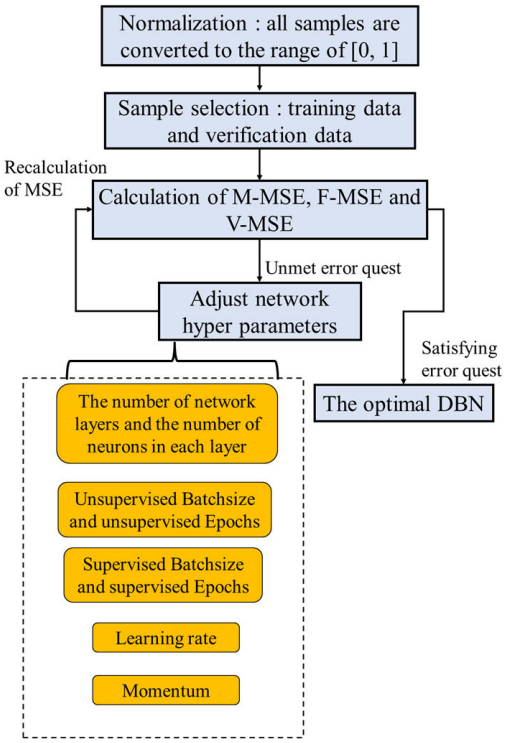


FIGURE 6. The training process of DBN.

corresponding to a measurement sample, and a_i represents the actual output of the soil model parameters corresponding to a measurement sample in the network operation process. We evaluated the performance of the DBN using three types of mean square errors. The Mini-batch training mean square error (M-MSE) represents the cumulative mean square error of the training samples as the network parameters are gradually adjusted in the training process. Full-batch train MSE (F-MSE) represents the mean square error between the network output and the expected output after completing the training, with no further parameter adjustments. Validation MSE (V-MSE) represents the mean square error between the output of the untrained input network and the expected output of all validation samples after completing the network training with no parameter adjustments. The network training process is illustrated in Fig.6.

B. DATABASE CONSTRUCTION AND TRAINING OF OPTIMAL NETWORK

The hardware environment for DBN training in this paper is as follows: the processor model was an Intel (R) Core (TM) i5-10400 CPU, with 6 cores and 12 threads and a main frequency of 2.90 GHz. The RAM parameter was 16.0 GB, and the computing platform is a 64-bit operating system based on an x64 processor with Windows 10 as the operating system version. The programming software environment was Matlab R2020a, 64-bit (Win64).

Because of the scarcity of measured data with soil apparent resistivity exhibiting outlier distribution characteristics

TABLE 5. The parameters' range of soil model considering apparent resistivity's discrete distribution characteristic.

Model parameters	Value	Model parameters	Value
$\rho_1/\Omega\cdot$m	[50,500]	h/m	[5,10]
h_1/m	[0,20]	x_0	[-10,10]
$\rho_2/\Omega\cdot$m	[50,500]	y_0	[-10,10]
d_1/m	[5,50]	z_0	h /2
d_2/m	[5,50]	$\rho_y/\Omega\cdot$m	[10000,15000]

for DBN training, we constructed a soil apparent resistivity database through CDEGS. We created uniform, two-layer, and three-layer soil models, each containing outlier distribution areas, with 2000 instances in each category. Of the 2000 instances, 70% were used for network training, and the remaining 30% were used for model validation.Given limited space, we used the double-layer composite soil model as an example to illustrate.

The model parameters are randomly set within the range specified in Tab.5, and the soil model samples are established with careful consideration of factors such as the number of training samples [14], [35], [36], parameter ranges, training complexity, and model parameter accuracy. The apparent resistivity is calculated for 12 different probe distances, as described in Section II-A, using CDEGS. The results indicate that the calculated apparent resistivity values meet the statistical criterion for apparent resistivity dispersion in the soil model, as outlined in this paper. After constructing the database, the DBN is trained following the steps shown in Fig.6. The training data and validation data are normalized, and the number of input neurons in the network is set to 12, corresponding to the apparent resistivity values at 12 probe spacings. The number of output neurons in the network is set to 9 (ρ_1, h_1, ρ_2, d_1, d_2, h, x_0, y_0, ρ_y), corresponding to the number of model parameters of the composite two-layer soil model.

Next, the network is trained using the samples to obtain three mean square errors: M-MSE, F-MSE, and V-MSE. The network hyperparameters are manually adjusted, including the number of hidden layers, the number of neurons in the hidden layer, the number of unsupervised mini-batch training cycles (unsupervised Epochs), the number of unsupervised mini-batch training samples (unsupervised Batchsize), the number of supervised mini-batch training cycles (supervised Epochs), the number of supervised mini-batch training samples (supervised Batchsize), the learning rate, and the learning momentum. This optimization minimizes the mean square error between the network output and the expected output, optimizing the network weight and bias. The network mean square error varies with the number of samples and hyperparameters, as shown in Fig.7, and the final adjusted network optimal hyperparameters are presented in Tab.6. The three mean square error results of the network are shown in Tab.7.

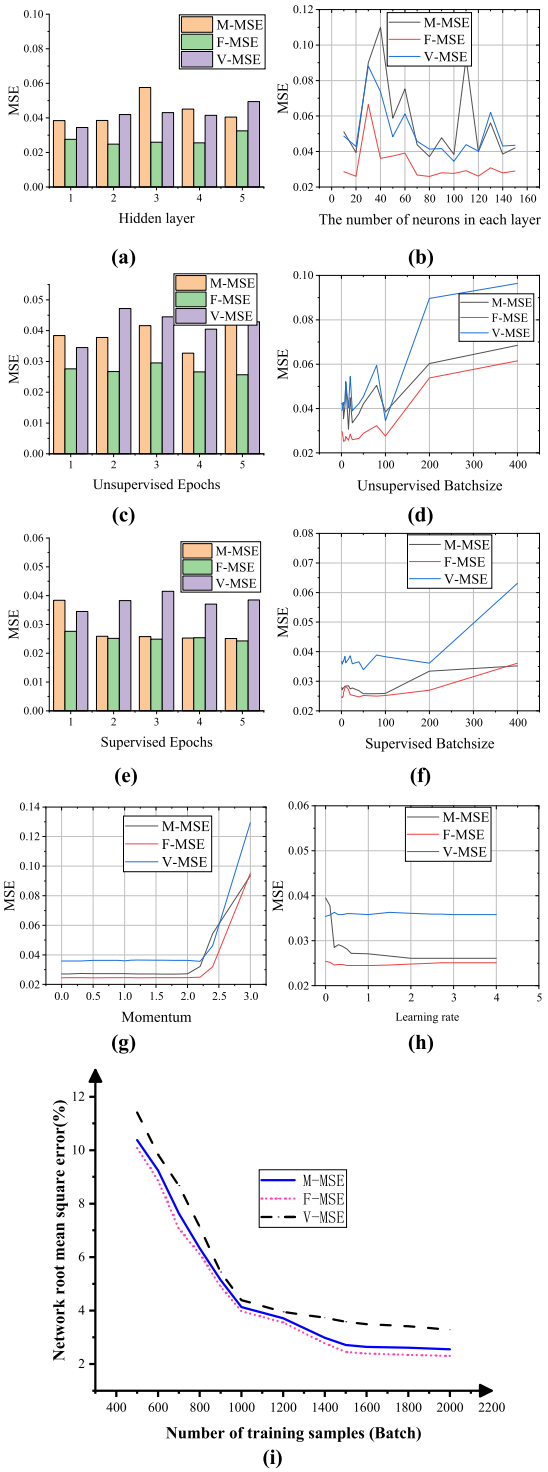


FIGURE 7. The training process of DBN.

V. MODEL INVERSION ACCURACY VERIFICATION

A. ANALYSIS OF SOIL STRUCTURE MODEL OF A 220KV SUBSTATION

Upon obtaining the optimal DBN through training, the soil structure model of a specific area can be inverted using field-measured data. The inversion process is illustrated in

TABLE 6. Adjustment results of network hyper-parameters.

Hyperparameters	Optimal value
hidden layer structure	[100,70]
Unsupervised Epochs	1
Unsupervised Batchsize	100
Supervised Epochs	2
Supervised Batchsize	2
Learning rate	1.3
Momentum	0.2

TABLE 7. The mean square error results of the network.

M-MSE	F-MSE	V-MSE
2.71%	2.45%	3.58%

TABLE 8. Apparent resistivity measured at point C.

l/m	1					
AB/2/m	1.4 10	2.0 14	2.8 20	4.0 28	5.5 40	7.5 55
$\rho_m/\Omega\cdot m$	10258 978	9512 616	8241 221	6798 254	4196 288	2022 303

Fig.8. It is important to note that the apparent resistivity measurement results of the soil in the area must fall within the parameter range of the training samples. Furthermore, the measurement method, number of points, and spacing must be consistent. The inversion model type and model parameter range must also be consistent.

For this study, we selected the apparent resistivity measurement data of an underground substation in Chengdu, as depicted in Fig.9. The yellow hollow area represents the 20 m deep underground substation. At the same time, the blue part denotes the upper layer of the cage grounding system that envelops the substation. Point C, located 2 m from the substation's edge along its middle line, was chosen as the reference point. The apparent resistivity of the soil was measured at this point using the M2124 grounding resistance tester from Metrel company. The measurement involved steel rods with a radius of 1 cm and a length of 80 cm inserted into the soil to a depth of 40 cm. Each of the 12 probe distances was measured thrice, and the results were averaged. The measurement data is summarized in Tab.8. Subsequently, the optimal DBN was employed to invert the aforementioned data, leading to the determination of detailed parameters for the composite double-layer soil model. The soil apparent resistivity of this composite two-layer soil model was then calculated using the boundary element method, and the root-mean-square error was computed by comparing it with the actual measured data. This verification process aims to assess the accuracy of the model.

The data in Tab.8 reveals significant deviations in some of the measured values from the overall dataset. The H value for the measured values is 0.916, and the G values for the first,

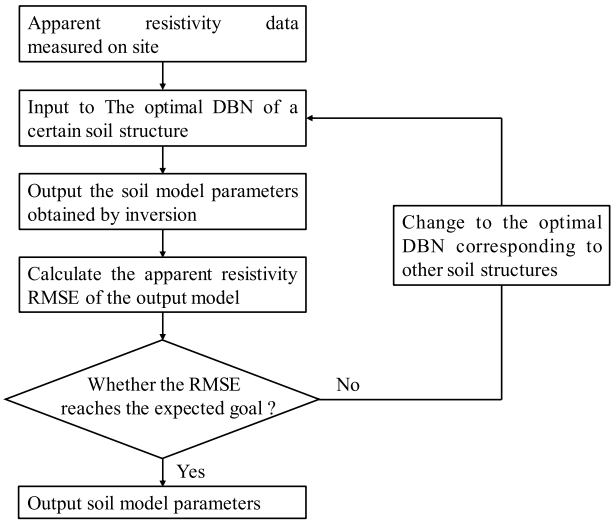


FIGURE 8. Process of predicting soil model's parameters by DBN.

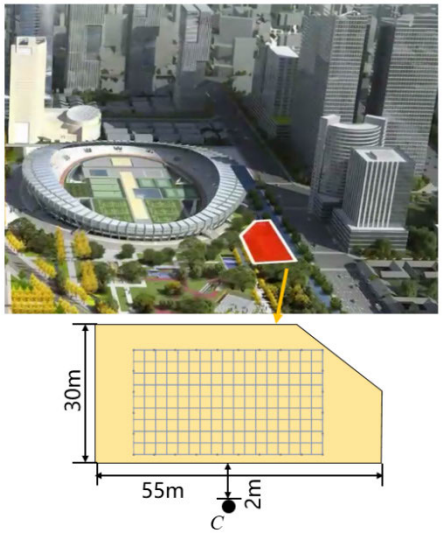
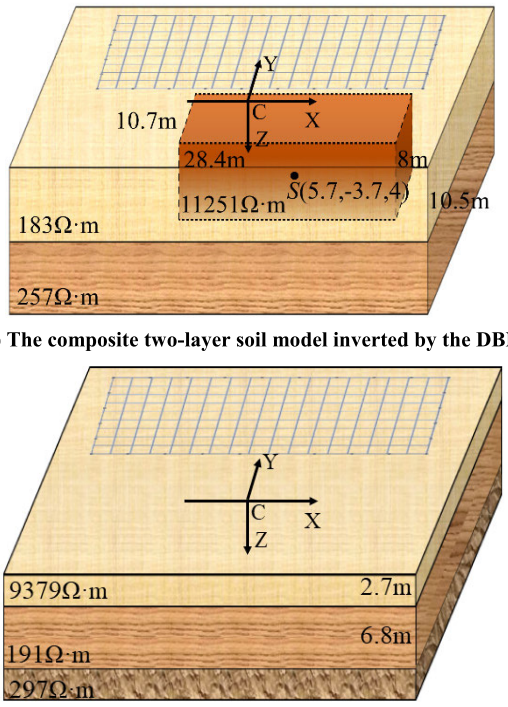


FIGURE 9. Layout of measuring points in the Substation.

second, third, and ninth measured values exceed G0.9 (12) in the Grubbs table, indicating that these measurements exhibit outlier distribution characteristics. Consequently, we input this set of measurement data into the optimal DBN and CDEGS to obtain the composite and horizontal layered soil models, respectively. The specific parameters of these two models are presented in Tab.9, and their corresponding soil structures are depicted in Fig.10. The data in Tab.9 indicates that the RMSE of the apparent resistivity obtained from the CDEGS inversion of the horizontal three-layer layered soil model is 13.6%. Despite increasing the number of horizontal layers, we cannot achieve a reduction in RMSE. In contrast, to enhance the model's accuracy and decrease the RMSE of apparent resistivity to less than 10%, soil resistivity outliers should be accounted for by employing the composite layered soil model proposed in this paper. The RMSE of the

TABLE 9. Parameters' results of two kinds of soil model.

Model parameters	Composite two-layer soil model	Horizontal three-layer soil model
$\rho_1/\Omega\cdot$m	183	9739
h_1/m	10.5	2.7
$\rho_2/\Omega\cdot$m	257	191
h_2/m	/	6.8
$\rho_3/\Omega\cdot$m	/	297
d_1/m	28.4	/
d_2/m	10.7	/
h/m	8	/
x_0	5.7	/
y_0	-3.7	/
z_0	4	/
$\rho_y/\Omega\cdot$m	11251	/
RMSE	8.53%	13.6%



(a) The composite two-layer soil model inverted by the DBN

(b) Horizontal three-layer soil model inverted by CDEGS

FIGURE 10. Two types of inverted soil models.

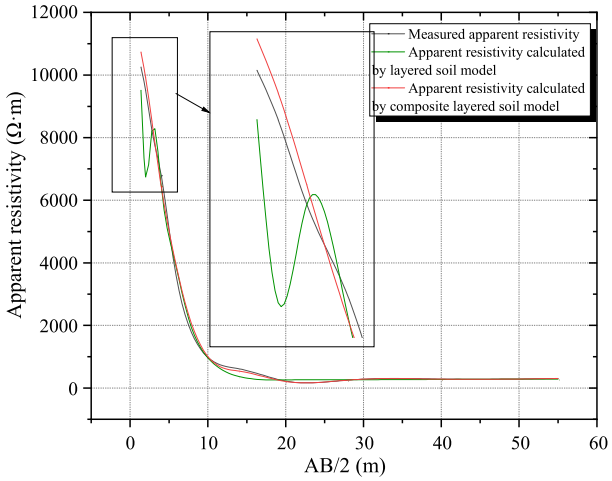


FIGURE 11. Calculation results of apparent resistivity of soil model in this paper and layered soil model.

apparent resistivity in the composite layered soil model is nearly half that in the CDEGS layered soil model, resulting in an RMSE of 8.53%. Fig.11 illustrates the apparent resistivity measured value curve and the calculated value curve. The diagram shows that the apparent resistivity of the first few probe spacings in the composite two-layer soil model is closer to the actual measured value than that of the horizontal three-layer soil model. Thus, the proposed soil model better reflects the influence of the dispersion area with extremely high resistance at the surface on the apparent resistivity. Additionally, the composite layered soil model is more consistent with the actual situation. Moreover, the influence of the high resistivity outlier distribution area can also be considered in the subsequent grounding design of the substation, thereby enhancing the grounding safety performance of this area and ensuring the safety of personnel and equipment in the substation.

Compared to the traditional layered soil model, the composite layered soil model with outlier distribution characteristics proposed in this paper exhibits significantly improved RMSE of apparent resistivity. Furthermore, the accuracy of this method is demonstrated through a case study. The results provide an effective reference for the inversion of soil models with outlier distribution characteristics of apparent resistivity or soil structure models with high resistivity outlier distribution areas near substation sites. This improvement in the calculation accuracy of grounding parameters is crucial.

We utilized a Zc-410A grounding resistance tester for large grounding grids to measure the grounding resistance of the substation. Fig.12 displays the grounding system design of the underground substation, which includes a horizontal grounding grid beneath the substation. Multiple vertical and horizontal grounding electrodes surround the substation, all interconnected with the main grounding grid. The grounding device consists of galvanized round steel with a 20 mm diameter. The horizontal grounding grid reaches a depth of 25 m, while the lowest depth of the substation is 22 m. In the simulation, the outlier distribution area below the station site was excavated, and a soil block replaced the underground substation area with a soil resistivity of 800 Ω·m. The measurement results revealed a grounding resistance of 0.75 Ω. Using the composite layered soil model and horizontally layered soil model, we calculated grounding resistances of 0.82 Ω and 1.03 Ω, respectively. The calculations show that the soil model obtained through this method demonstrates

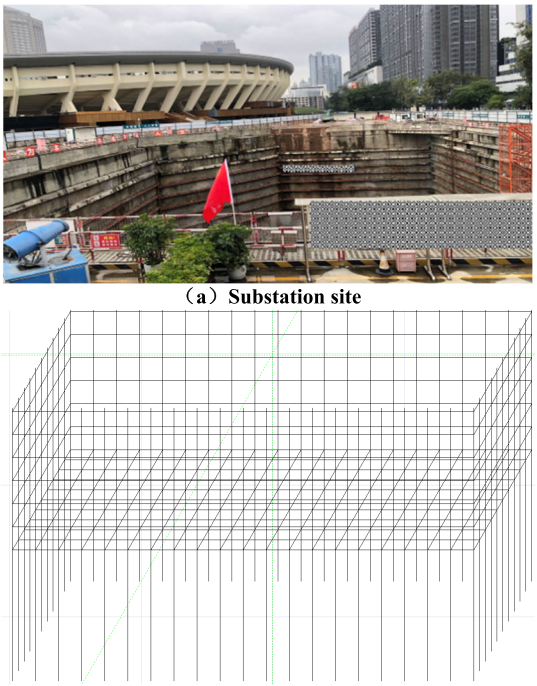


FIGURE 12. Layout of survey points.

higher accuracy in the electromagnetic calculation of the grounding device.

B. SOIL MODEL INVERSION NEAR THE QINGHAI-TIBET RAILWAY

Apart from the 220 kV underground substation, we also conducted apparent resistivity measurements in the permafrost region of the Qinghai-Tibet line at a 220 kV substation area. As depicted in Fig.13, we performed measurements at nine points, and the results are shown in Fig.14.

For the inversion of soil structure model parameters, we constructed 2000 horizontal layered soil models with 3, 4 and 5 layers each. The optimal depth confidence network of the 3, 4 and 5-layer soil models was trained. Additionally, this study employed the GA and the grounding software CDEGS based on the steepest descent method for inversion. The objective function represented by the measured data was optimized using the MATLAB genetic algorithm toolbox (GADs), with parameters manually adjusted to obtain better model parameters. The partial iterative process of the genetic algorithm is presented in Fig.15.

In each iteration, a single parameter is adjusted while keeping the other parameters unchanged until the fitness reaches its optimum. Subsequently, other parameters are adjusted. As seen in Fig.15(a), when the genetic algorithm was employed to solve the horizontally layered soil model at point 1, the average fitness of all individuals in the 50th generation reached 0.981588, with the best individual fitness of this generation being 0.0378755. Despite achieving an error of approximately 3% in the soil model, the iterative

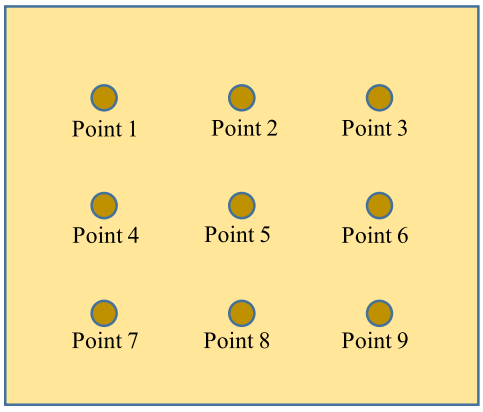


FIGURE 13. Layout of survey points.

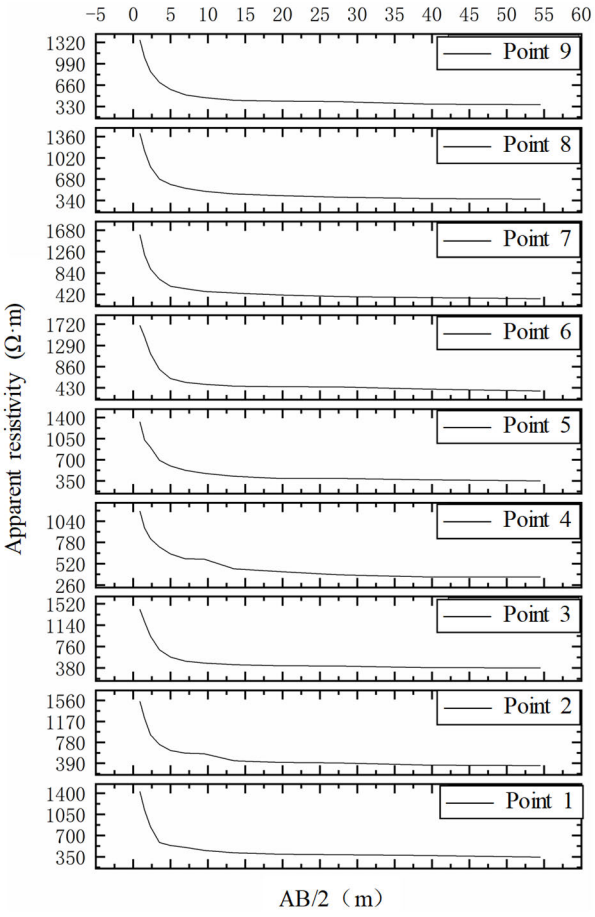
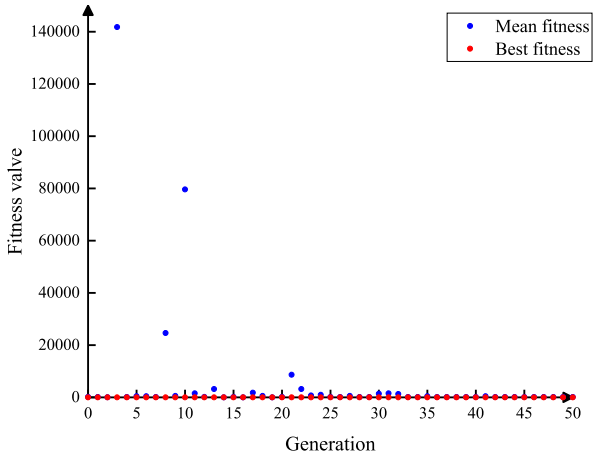


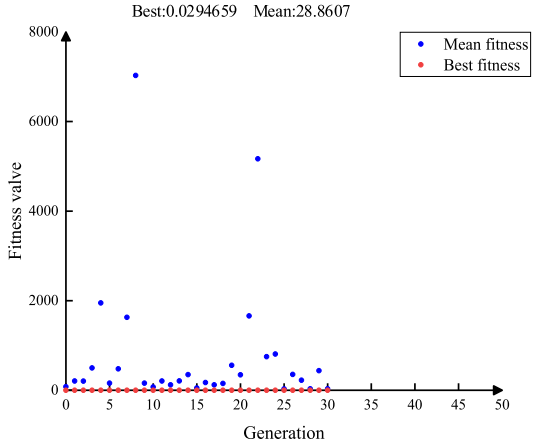
FIGURE 14. The measurement of apparent resistivity at 9 points i.

process was unstable due to significant deviations in the fitness of previous individual groups. Ultimately, a relatively satisfactory result was obtained through manual adjustment.

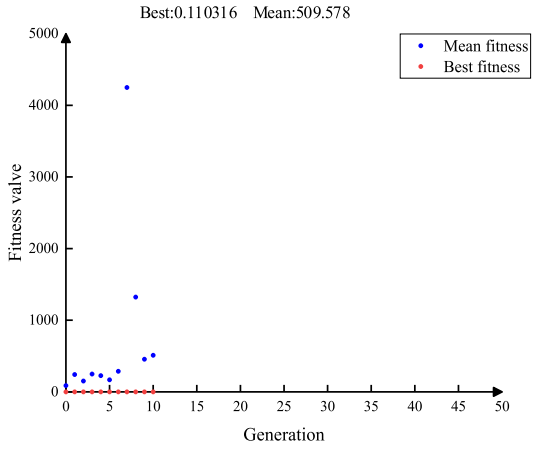
At point 5, the convergence of soil model parameters, as shown in Fig.15(b), was relatively scattered before the 28th generation, with numerous cases of poor individual fitness. Although the best individual error had reached 3%,



(a) Point 1 (The 50th generation)



(b) Point 5 (The 28th generation)



(c) Point 6 (The 11th generation)

FIGURE 15. Iterative process of solving horizontal three-layer soil model by genetic algorithm.

the process remained unstable, necessitating cumbersome manual adjustments.

The seven parameters should be adjusted for each set of measurement data, making the process cumbersome and

TABLE 10. Parameters of genetic algorithm.

Parameters	Point 1	Point 5	Point 6
Population size	500	500	500
Creation function	Constraint dependent	feasible population	uniform
Initial range of individual	[0 0 0 0 0;3000 3000 3000 10 10]	[0 0 0 0 0;2000 2000 2000 10 10]	[0 0 0 0 0;1500 1500 1500 10 10]
Fitness function	Rank	Proportional	Shift linear
Selection function	Remainder	Roulette	stochastic uniform
Mutation function	Constraint dependent	Gaussian	uniform
Crossover function	Constraint dependent	scattered	intermediate

unstable. The resulting soil models obtained through these three methods are depicted in Fig.16. The adaptive complex image method is employed to calculate their apparent resistivity, and the RMSE of apparent resistivity is calculated for all measurement points. The average value of RMSE serves as the final model's RMSE; the results are shown in Fig.17.

As illustrated in Fig.17, the RMSE of the horizontal three-layer, four-layer, and five-layer soil models solved by the GA is greater than the results of the other two methods. The RMSE of the soil model inverted by the DBN is close to that of the extensively used grounding design software CDEGS. Hence, the inversion accuracy of the DBN is comparable to that of CDEGS in the case of horizontal layered soil models, and it outperforms the traditional GA. This method eliminates the need for extensive numerical and iterative calculations, simplifies the solution process for the same type of horizontal multi-layered soil model, and shortens the convergence process, enhancing the inversion efficiency. The comparison of the average calculation time for the nine soil models computed by various methods is presented in Tab.11.

As indicated by the data in Tab.11, the calculation time of the three inversion methods meets engineering needs. Because of its complex iterative process and numerous electromagnetic forward calculation steps, GA takes approximately 5 min for calculation. On the other hand, the main time spent on the DBN is in the preparation work before inversion, specifically, the training of the optimal network. By employing it, soil model parameters can be rapidly inverted based on apparent resistivity measurement values, equivalent to the calculation speed of CDEGS, a software developed over decades. Compared with the traditional method, the inversion efficiency is significantly improved.

C. COMPARISON WITH SEVERAL TRADITIONAL METHODS

To verify the overall prediction performance of our model, we used two sets of measured data, and RMSE alone may not

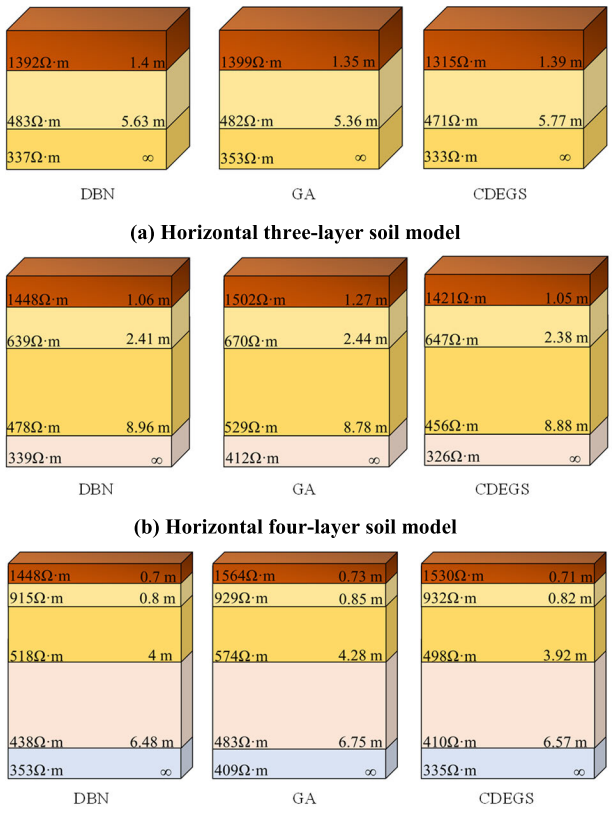


FIGURE 16. Soil model based on three kinds of methods.

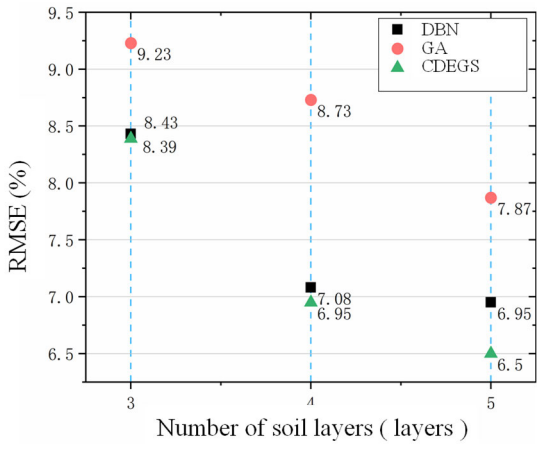


FIGURE 17. The average RMSE of apparent resistivity of soil model inverted by three kinds of methods.

adequately represent the model's performance. Subsequently, we evaluated the prediction power quantitatively using the receiver operating characteristic (ROC) curve method with the Statistical Package for the Social Sciences (SPSS) software. Model evaluation was performed using the testing dataset. The area under the ROC curve (AUC) provides a comprehensive performance measure across all possible classification thresholds. To convert the predictions into a

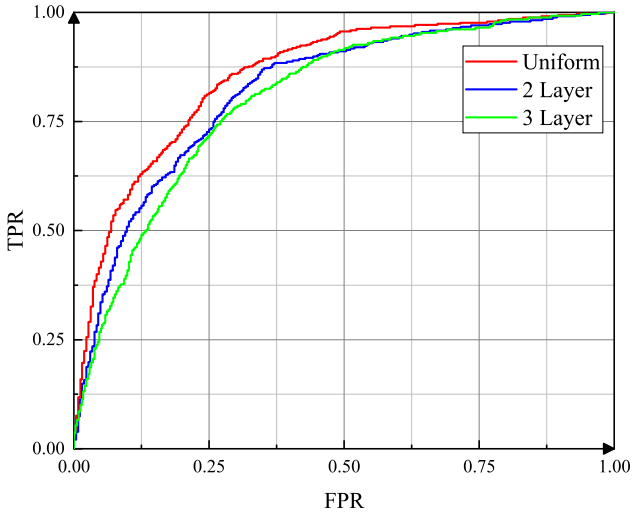


FIGURE 18. The ROC of our model.

TABLE 11. Comparison of calculation time of three kinds of inversion methods.

Soil model	Calculation time (s)		
	DBN	GA	SD
Horizontal three layers of soil	3.2	293	2.8
Horizontal four layers of soil	3.6	318	3.3
Horizontal five layers of soil	4.3	331	3.9

TABLE 12. The three-layer soil model inverted from the data in Table 8.

Model parameters	DBN	SD	BGFS	PSO	GA	SA
$\rho_1/\Omega\cdot$m	183	9739	10545	9803	10722	9779
$h_1/\Omega\cdot$m	10.5	2.75	2.56	2.71	2.38	2.66
$\rho_2/\Omega\cdot$m	257	191	313	215	1021	231
$h_2/\Omega\cdot$m	/	6.76	5.92	6.28	3.61	6.27
$\rho_3/\Omega\cdot$m	/	297	304	274	263	300
d_1/m	28.4	/	/	/	/	/
d_2/m	10.7	/	/	/	/	/
h/m	8	/	/	/	/	/
x_0	5.7	/	/	/	/	/
y_0	-3.7	/	/	/	/	/
z_0	4	/	/	/	/	/
$\rho_y/\Omega\cdot$m	11251	/	/	/	/	/
RMSE/%	8.53	13.6	18.3	14.6	19.8	14.8

classification task, we established the following criteria to determine the success or failure of the predictions:

The RMSE of the apparent resistivity obtained by the optimal DBN must be less than 10%.

The RMSE of the apparent resistivity obtained by the optimal DBN must be less than that from the CDEGS.

The error in each soil parameter, including soil resistivity, the thickness of each soil layer, and parameters of the outlier distribution area obtained using the optimal DBN, must be less than 15%.

TABLE 13. The four-layer soil model inverted from the data in Table 8.

Model parameters	DBN	SD	BGFS	PSO	GA	SA
$\rho_1/\Omega\cdot$m	183	9752	10538	10051	10722	9338
$h_1/\Omega\cdot$m	10.5	2.57	2.31	2.49	2.32	2.52
$\rho_2/\Omega\cdot$m	257	4732	4524	4925	5309	4631
$h_2/\Omega\cdot$m	/	0.36	0.39	0.37	0.69	0.39
$\rho_3/\Omega\cdot$m	/	162	354	237	442	208
$h_3/\Omega\cdot$m	/	3.11	4.15	4.0	2.15	4.41
$\rho_4/\Omega\cdot$m	/	278	288	250	284	296
d_1/m	28.4	/	/	/	/	/
d_2/m	10.7	/	/	/	/	/
h/m	8	/	/	/	/	/
x_0	5.7	/	/	/	/	/
y_0	-3.7	/	/	/	/	/
z_0	4	/	/	/	/	/
$\rho_y/\Omega\cdot$m	11251	/	/	/	/	/
RMSE/%	8.53	13.9	17.9	14.3	18.3	15.2

TABLE 14. The four-layer soil model inverted from the data in Fig.14.

Model parameters	DBN	SD	BGFS	PSO	GA	SA
$\rho_1/\Omega\cdot$m	1448	1421	1144	1591	1502	1425
$h_1/\Omega\cdot$m	1.06	1.05	1.08	1.02	1.27	1.06
$\rho_2/\Omega\cdot$m	639	647	685	638	670	614
$h_2/\Omega\cdot$m	2.41	2.38	2.40	2.37	2.44	2.41
$\rho_3/\Omega\cdot$m	478	456	524	436	529	442
$h_3/\Omega\cdot$m	8.96	8.88	8.75	8.90	8.78	9.09
$\rho_4/\Omega\cdot$m	339	326	327	305	412	355
RMSE/%	7.08	6.66	8.33	5.69	8.73	8.59

TABLE 15. The five-layer soil model inverted from the data in Fig.14.

Model parameters	DBN	SD	BGFS	PSO	GA	SA
$\rho_1/\Omega\cdot$m	1523	1530	1325	1268	1564	1437
$h_1/\Omega\cdot$m	0.70	0.71	0.72	0.71	0.73	0.71
$\rho_2/\Omega\cdot$m	915	932	871	913	929	949
$h_2/\Omega\cdot$m	0.80	0.82	086	0.81	0.85	0.82
$\rho_3/\Omega\cdot$m	518	498	600	579	574	539
$h_3/\Omega\cdot$m	4.00	3.92	4.34	4.01	4.28	3.99
$\rho_4/\Omega\cdot$m	438	411	508	445	483	473
$h_4/\Omega\cdot$m	6.50	6.57	7.22	6.46	6.75	6.41
$\rho_5/\Omega\cdot$m	353	335	414	354	409	359
RMSE/%	6.62	6.50	8.15	6.02	7.87	7.01

Inversion is considered successful if the results satisfy both criteria, and the state variable is assigned a value of 1. Otherwise, the inversion is considered failed, and the state variable is set to 0. In this paper, the diagnostic variable is set as RMSE. Based on these rules, we used 600 test data sets to draw the ROC curve, as shown in Fig.18.

The AUC values for the three optimal DBN models (obtained for uniform, two-layer, and three-layer soils with outlier distribution area) are 0.857, 0.822, and 0.804, respectively. These results indicate that judging the accuracy of the inversion results based on the RMSE of apparent resistivity is feasible. The reason for the decrease in AUC may be due to the increase in the number of variables to be solved, resulting in a specific soil parameter, such as the thickness or resistivity of one layer of soil, easily exceeding 15%.

The apparent resistivity data used for optimal DBN training in this study is obtained from measurements conducted at 12 fixed probe spacings. However, other literature may use different probe arrangements, which makes their apparent resistivity data incompatible for direct input into the optimal DBN. To address this, we compare the performance of several inversion methods, including the steepest descent method (SD), Simulated Annealing (SA), Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, particle swarm optimization (PSO), and GA, using the apparent resistivity measurement data from Sections V-A and V-B as examples.

Tab.12 and Tab.13 present the results for the examples in Section V-A. It is evident from the tables that when the apparent resistivity exhibits outlier distribution characteristics, the DBN exhibits higher inversion accuracy compared to other methods. The two-layer soil model with an outlier distribution area, obtained through DBN, outperforms the three-layer and four-layer horizontally layered soil models obtained through other methods, and increasing the number of soil layers does not significantly reduce the RMSE.

Similarly, Tab.14 and Tab.15 display the results for the examples in Section V-B. It can be observed that when the soil resistivity does not exhibit outlier distribution characteristics, the accuracy of DBN is comparable to other methods, and it still meets the requirements for engineering applications with errors of less than 10%.

VI. CONCLUSION

This study aimed to address the accuracy of soil modeling in cases where an outlier distribution area with extremely high resistivity exists in the local area near the surface. To achieve this, we proposed a statistical criterion to identify outlier distribution characteristics of soil resistivity and developed a composite layered soil model inversion method based on DBN that considers outlier distribution conditions. The results indicated that the three types of mean square errors of the optimal DBN ranged from 2% to 4%. The composite two-layer soil model retrieved by the network for a 220kV substation site was significantly improved compared to the horizontally layered soil model retrieved by CDEGS. Compared with several traditional inversion algorithms, it shows that the inversion accuracy of our method is comparable to that of traditional methods for horizontally layered soil but exhibits a remarkable improvement of approximately 40% when dealing with soil apparent resistivity exhibiting outlier distribution characteristics. The method presented in this paper overcomes the limitation of traditional inversion methods that only allow obtaining horizontal layered soil models using the four-probe method measurement data. Moreover, it makes a crucial contribution by considering the influence of

the outlier distribution area in subsequent grounding design. This novel approach offers a valuable method for soil model inversion with outlier distribution areas. However, there are some limitations in this study. The number of samples used for DBN training was only 2000 groups. We plan to increase the number of samples in future work to enhance model accuracy. Additionally, the training samples used in this study were relatively simple. In future research, we aim to incorporate more complexity into the training samples, such as inclined soil or soil models with non-uniform resistivity distribution along the horizontal direction. Finally, the optimal DBN in this paper is trained under fixed electrode arrangement, so it can only be used in some specific scenarios. In addition to the apparent resistivity, we prepare to add the electrode spacing as input in the future, so that the model can be applied to more cases.

REFERENCES

[1] B. Zhang, J. L. He, and R. Zeng, "State of art and prospect of grounding technology in power system," *High Voltage Eng.*, vol. 41, no. 8, pp. 2569–2582, 2015, doi: 10.13336/j.1003-6520.hve.2015.08.010.

[2] F. Wenner, "A method of measuring earth resistivity," *Nat. Bur. Standards*, vol. 12, no. 3, pp. 469–482, 1915.

[3] F. Dawalibi and C. J. Blattner, "Earth resistivity measurement interpretation techniques," *IEEE Trans. Power App. Syst.*, vol. PAS-103, no. 2, pp. 374–382, Feb. 1984, doi: 10.1109/TPAS.1984.318254.

[4] Y. L. Chow, J. J. Yang, and G. E. Howard, "Complex images for electrostatic field computation in multilayered media," *IEEE Trans. Microw. Theory Techn.*, vol. 39, no. 7, pp. 1120–1125, Jul. 1991, doi: 10.1109/22.85378.

[5] B. Zhang, X. Cui, L. Li, and J. He, "Parameter estimation of horizontal multilayer earth by complex image method," *IEEE Trans. Power Del.*, vol. 20, no. 2, pp. 1394–1401, Apr. 2005, doi: 10.1109/TPWRD.2004.834673.

[6] J. L. del Alamo, "A second order gradient technique for an improved estimation of soil parameters in a two-layer earth," *IEEE Trans. Power Del.*, vol. 6, no. 3, pp. 1166–1170, Jul. 1991, doi: 10.1109/61.85863.

[7] F. P. Dawalibi and F. Donoso, "Integrated analysis software for grounding, EMF, and EMI," *IEEE Comput. Appl. Power*, vol. 6, no. 2, pp. 19–24, Apr. 1993, doi: 10.1109/67.207467.

[8] W. X. Pan and Y. Zou, "Optimization calculation of double-layer soil parameters," *Proc. CSEE*, vol. 16, no. 5, pp. 358–360, May 1996, doi: CNKI:SUN:ZGDC.0.1996-05-015.

[9] H. N. Yang, H. Xie, and Z. X. Li, "Optimization calculation of two-layer soil parameters in grounding design," *Modern Electr. Power*, vol. 16, no. 1, pp. 3–5, Jan. 1999, doi: 10.19725/j.cnki.1007-2322.1999.01.011.

[10] Y. X. Pan, R. Zeng, and J. L. He, "Soil structure analysis for substation sites," *J. Tsinghua Univ. Sci. Technol.*, vol. 42, no. 3, pp. 288–290, 2002, doi: 10.16511/j.cnki.qhdxxb.2002.03.002.

[11] J. Zou, J. L. He, R. Zeng, W. M. Sun, G. Yu, and S. M. Chen, "Two-stage algorithm for inverting structure parameters of the horizontal multilayer soil," *IEEE Trans. Magn.*, vol. 40, no. 2, pp. 1136–1139, Mar. 2004, doi: 10.1109/TMAG.2004.825013.

[12] H. Xu, X. S. Wen, X. Shu, and Q. S. Zou, "New method of computing soil parameters," *High Voltage*, vol. 30, no. 8, pp. 17–19, Aug. 2004, doi: 10.13336/j.1003-6520.hve.2004.08.007.

[13] I. F. Gonos and I. A. Stathopulos, "Estimation of multilayer soil parameters using genetic algorithms," *IEEE Trans. Power Del.*, vol. 20, no. 1, pp. 100–106, Jan. 2005, doi: 10.1109/TPWRD.2004.836833.

[14] J. P. Lee, P. S. Ji, J. Y. Lim, S. S. Kim, A. Ozdemir, and C. Singh, "Earth parameter and equivalent resistivity estimation using ANN," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, vol. 3, Jun. 2005, pp. 2597–2602. [Online]. Available: https://www.webofscience.com/wos/alldb/fullrecord/WOS:000232799402111

[15] L. Zhang, J. Cai, X. Wen, J. Quan, and Z. Pan, "Particle swarm optimization of horizontal multi-layer soil parameters inversion," *High Voltage Eng.*, vol. 40, no. 7, pp. 2018–2023, Jul. 2014, doi: 10.13336/j.1003-6520.hve.2014.07.014.

[16] W. P. Calixto, L. M. Neto, M. Wu, K. Yamanaka, and E. da Paz Moreira, "Parameters estimation of a horizontal multilayer soil using genetic algorithm," *IEEE Trans. Power Del.*, vol. 25, no. 3, pp. 1250–1257, Jul. 2010, doi: 10.1109/TPWRD.2010.2040845.

[17] R. R. A. Coelho, A. E. C. Pereira, and L. M. Neto, "A high-performance multilayer earth parameter estimation rooted in Chebyshev polynomials," *IEEE Trans. Power Del.*, vol. 33, no. 3, pp. 1054–1061, Jun. 2018, doi: 10.1109/TPWRD.2017.2664738.

[18] W. He, R. Q. Zhang, F. Yang, L. W. Zhu, Z. H. Liu, and X. Y. Wang, "Parameter estimation of horizontal multi-layer earth in substation," *Proc. CSEE*, vol. 34, no. 33, pp. 5964–5973, 2014, doi: 10.13334/j.0258-8013.pcsee.2014.33.021.

[19] W. R. Pereira, M. G. Soares, and L. M. Neto, "Horizontal multilayer soil parameter estimation through differential evolution," *IEEE Trans. Power Del.*, vol. 31, no. 2, pp. 622–629, Apr. 2016, doi: 10.1109/TPWRD.2015.2475637.

[20] Z.-X. Li and S.-W. Rao, "The inversion of one-dimensional soil parameters in the frequency domain with considering multilayered earth based on simulated annealing algorithm," *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 2, pp. 425–432, Apr. 2020, doi: 10.1109/TEMC.2019.2906486.

[21] M. Guo, "Study on the measurement of soil structure and inversion of soil model in grounding system," M.S. thesis, Dept. Elect. Eng., Southwest Jiaotong Univ., Chengdu, China, 2018.

[22] O. Ramos-Leaños, F. A. Uribe, L. Valcárcel, A. Hajiaboli, S. Franiatte, and F. P. Dawalibi, "Nonlinear electrode arrangements for multilayer soil resistivity measurements," *IEEE Trans. Electromagn. Compat.*, vol. 62, no. 5, pp. 2148–2155, Oct. 2020, doi: 10.1109/TEMC.2020.2970149.

[23] J. Yang and J. Zou, "Parameter estimation of a horizontally multilayered soil with a fast evaluation of the apparent resistivity and its derivatives," *IEEE Access*, vol. 8, pp. 52652–52662, 2020, doi: 10.1109/ACCESS.2020.2980875.

[24] C. L. Ma, "Research on distribution characteristics of earth potential near UHVDC grounding electrode," Ph.D. dissertation, Dept. Elect. Eng., North China Electr. Power Univ., Beijing, China, 2020.

[25] Q. Xiong, M. X. Wang, H. Huang, Y. Shi, and H. T. Tang, "Establishment of earth model for HVDC earth electrode in complicated terrain," *Proc. CSEE*, vol. 40, no. 7, pp. 2269–2277, 2020, doi: 10.13334/j.0258-8013.pcsee.182115.

[26] J. S. Li, Y. J. Liu, Z. H. Pan, and X. S. Wen, "Site selection of DC grounding electrodes in complex soil and risk assessment of DC bias," *Electr. Meas. Instrum.*, vol. 58, no. 2, pp. 13–18, 2021, doi: 10.19753/j.issn1001-1390.2021.02.003.

[27] G. L. Yue, Y. B. Mu, Z. H. Zhihui, K. K. Yan, T. Xi, and Z. H. Pan, "Exploration and calculation of very fine geodetic model for DC grounding electrodes," *High Voltage Eng.*, vol. 47, no. 11, pp. 3901–3912, 2021, doi: 10.13336/j.1003-6520.hve.20201006.

[28] A. R. Xue, S. G. Ju, W. H. He, and W. H. Chen, "Study on algorithms for local outlier detection," *Chin. J. Comput.*, vol. 30, no. 8, pp. 1455–1463, 2007, doi: 10.3321/j.issn:0254-4164.2007.08.028.

[29] *IEEE Guide for Measuring Earth Resistivity, Ground Impedance, and Earth Surface Potentials of a Grounding System*, IEEE Standard 81, 2012.

[30] J. C. Yang and C. Zhao, "Survey on K-means clustering algorithm," *Comput. Eng. Appl.*, vol. 55, no. 23, pp. 7–14, 2019, doi: 10.3778/j.issn.1002-8331.1908-0347.

[31] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006, doi: 10.1162/neco.2006.18.7.1527.

[32] Y. X. Zhang, "Deep learning models and applications based on the restricted Boltzmann machine," M.S. thesis, Dept. Comput. Math., Univ. Electron. Sci. Technol. China, Chengdu, China, 2016.

[33] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.

[34] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0.

[35] Y. C. Zhang, D. Wen, X. R. Wang, and J. T. Lin, "A method of frequency curve prediction based on deep belief network of post-disturbance power system," *Proc. CSEE*, vol. 39, no. 17, pp. 5095–5104, 2019, doi: 10.13334/j.0258-8013.pcsee.181821.

[36] L. J. Song, "Recognition model of disease image based on discriminative deep belief networks," *Comput. Eng. Appl.*, vol. 53, no. 21, pp. 32–36, 2017, doi: 10.3778/j.issn.1002-8331.1707-0506.

[37] J. S. Lee, J. Park, J. Kim, and H. K. Yoon, "Study of oversampling algorithms for soil classifications by field velocity resistivity probe," *Geomechanics Eng.*, vol. 20, no. 3, pp. 247–258, 2023, doi: 10.12989/gae.2022.30.3.247.
[38] K. Khosravi, F. Rezaie, J. R. Cooper, and Z. Kalantari, "Soil water erosion susceptibility assessment using deep learning algorithms," *J. Hydrol.*, vol. 618, Mar. 2023, Art. no. 129229, doi: 10.1016/j.jhydrol.2023.129229.
[39] A. Niyogi, T. A. Ansari, S. K. Sathapathy, K. Sarkar, and T. N. Singh, "Machine learning algorithm for the shear strength prediction of basalt-driven lateritic soil," *Earth Sci. Informat.*, vol. 16, no. 1, pp. 899–917, Feb. 2023, doi: 10.1007/s12145-023-00950-8.
[40] D. Moghadas and A. Badorreck, "Machine learning to estimate soil moisture from geophysical measurements of electrical conductivity," *Near Surf. Geophys.*, vol. 17, no. 2, pp. 181–195, Apr. 2019, doi: 10.1002/nsg.12036.
[41] Z. Zhang, P. Yu, L. Zhang, C. Zhao, Y. Wang, and Y. Xu, "Application of U-net for the recognition of regional features in geophysical inversion results," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4705207, doi: 10.1109/TGRS.2021.3138790.
[42] Z. Liu, M. C. Liu, W. Wei and R. L. Du, "Gravity anomaly separation based on cellular neural network," *J. China Univ. Petroleum, Natural Sci. Ed.*, vol. 34, no. 4, pp. 57–61, 2010, doi: 10.3969/j.issn.1673-5005.2010.04.010.

XIAOBING XIAO was born in Hubei, China, in 1986. He received the master's degree from the Electrical Engineering Department, Guizhou University, Guizhou, China, in 2011. Currently, he is a Senior Engineer with Guizhou Power Grid Company Ltd.

YONGXIANG CAI was born in Guizhou, China, in 1991. He received the Ph.D. degree from the Electrical Engineering Department, China Agricultural University, Beijing, China, in 2020. Currently, he is an Engineer with Guizhou Power Grid Company Ltd.

XIAOMENG HE was born in Hubei, China, in 1987. She received the master's degree from the Electrical Engineering Department, Wuhan University of Technology, Hubei, in 2013. Currently, she is an Engineer with Guizhou Power Grid Company Ltd.

HUAPENG LI was born in Guizhou, China, in 1973. He received the Graduate degree from the Electrical Engineering Department, Xi'an Jiaotong University, Xi'an, China, in 2014. Currently, he is a Senior Engineer with Guizhou Power Grid Company Ltd.

YUE LI, photograph and biography not available at the time of publication.

XINYI HE was born in Jiangxi, China, in 1998. She received the Graduate degree from the Electrical Engineering Department, Tianshui Normal University, Gansu, China, in 2020. Currently, she is an Engineer with Guizhou Power Grid Company Ltd.

TAO YUAN was born in Sichuan, China, in 1976. He received the Ph.D. degree from the Electrical Engineering Department, Chongqing University, Chongqing, China, in 2010. Currently, he is a Professor with Chongqing University.

QIAN CHEN was born in Chongqing, China, in 1999. He received the Graduate degree from the Electrical Engineering College, Chongqing University, Chongqing, in 2021, where he is currently pursuing the master's degree.