Reservoir resistivity monitoring through 1D MCSEM data and machine learning

Camila C. S. da Silva, Marcos W. C. Silva, Jéssica L. S. P. Itó - Universidade Federal do Pará (UFPA)

Abstract

The prominence of machine learning to optimize statistical performance using experience has been instrumental in making geophysical predictions in hydrocarbon exploration and monitoring deepwater resistivities using MCSEM data. The training data provide information on the electromagnetic field that is directly related to the resistivity of the medium. Metrics R2, MSE and RMSE, and the analysis of bias and variance of the model about the data show positive results for the ML regression class. Random Forest, K-Nearest Neighbors and Convolutional Neural Network were the algorithms that adopted a more remarkable ability to determine the resistivities of deep targets. The computational cost, predicted values, and error estimates show a better response for the CNN algorithm when it comes to 1D synthetic modeling problems on substrates with six layers.

Introduction

The inverse geophysical problem is ill-posed because often the solution when it exists, is not unique or is not stable. A deterministic approach to solving the inverse problem is based on minimizing an objective function. Regularizers can be used to solve the problem of uniqueness and stability. In the last decade, with the development of intelligent algorithms, machine learning has been increasingly used in the exploration geophysics to obtain geological parameters from geophysical data.

Machine learning (ML) can solve the inverse problem based on statistical analysis using sample data or previous experiments Kim & Nakata (2018). In Tarantola (2005) it is also shown that the machine learning process is inverse, as it can predict an interpreted model from observations. In this sense, among the numerous algorithms associated with ML, we must choose those that best suit the determination of geoelectric properties in the subsurface from electromagnetic geophysical data.

Deep convolutional neural networks (CNN) are gaining popularity in these geophysical studies. Das et al. (2019) propose using CNNs for impedance inversion from seismic data of normal incidence, presenting a strong correlation with the real data. Wu et al. (2019), however, delimited faults from 3D seismic images, also using supervised CNNs. On the other hand, there are several surveys by geoscientists involving the Random Forest (RF) machine learning algorithm. Harris & Grunsky (2015) applied this algorithm to geochemical and magnetic field data to predict lithological maps of a large section of the Hearne Archene tectonic domain in northern Canada. The results indicated great efficiency in producing these maps using RF classifiers. Furthermore, Kuhn et al. (2018) categorized the lithology of an underexplored area adjacent to the Junction gold mine and, through the RF algorithm, identified areas of the map that are more likely to be incorrectly mapped and concluded that the method could be an effective tool for geoscientists in data-limited environments.

To optimize the results for different geophysical data and to determine the resistivities of deep ocean floor drilling targets, we directed this study method to the Marine Electromagnetic Controlled Source (MCSEM), which was developed for study applications of the lithosphere and, with the advancement of electronic equipment since the 2000s, it has shown excellent efficiency in seismic support in hydrocarbon exploration, as it provides information on the resistivity contrast between the host rocks and the reservoir in deep waters (Constable & Srnka, 2007).

MCSEM 1D modeling gives us a quick estimate of the resistivity distribution and can be used in the traditional inverse problem. It can also generate training data for ML algorithms, making it possible to obtain predictions for this distribution. The algorithms used refer to the regression area in the ML, as they are characterized by the search to parameterize relations between dependent and independent variables through statistical methods.

The objective of this work is to obtain the horizontally stratified distribution of subsurface electrical resistivity from MCSEM 1D data applied to reservoir monitoring using K-nearest neighbor (KNN), Random Forest (RF) and Convolutional Neural Network (CNN). The importance of using such algorithms during the work was because we started with the most accessible and most intuitive algorithm (KNN), thereby inducing the use of the most robust algorithms (RF and CNN). To feed them with more information and optimize our responses for any possible resistivity value in the subsurface, especially in the reservoir.

MCSEM 1D modeling

This study used MCSEM 1D synthetic data to determine subsurface electrical resistivity values. The 1D modeling algorithm follows the traditional formulation based on Schelkunov potentials (Ward & Hohmann, 2012). In this sense, we propose the geoelectric model illustrated in figure 1, which consists of a survey line in the x direction with y = 0 m (dashed). The figure shows the horizontal electric dipole (Tx) is 50 m above the seafloor. In this simulation, we use the frequency of 0.25 Hz and the construction of this deep water model will be used to generate the machine learning data. Thus, the range of resistivities of each layer is shown in the figure.

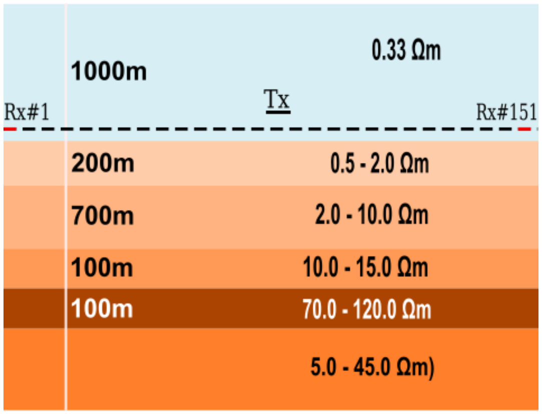

Figure 1 – Representation of horizontally stratified subsurface electrical resistivity distribution from MCSEM 1D data.

To exemplify the MCSEM 1D response, we show in figures 2 and 3 the electric and magnetic fields using the minimum resistivity values of the standard model. In them, we have the amplitude and phase of the fields as a function of the offset.

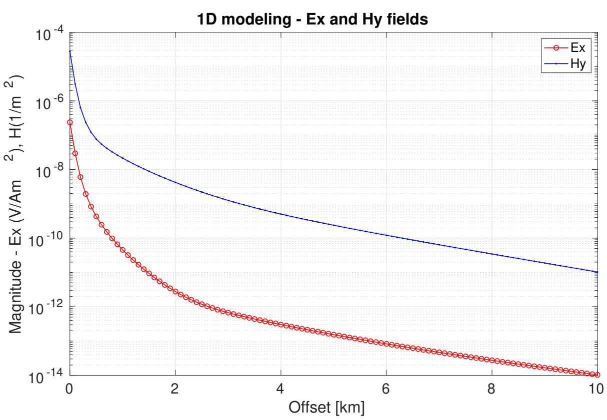

Figure 2 – Amplitude of electric and magnetic fields for horizontally stratified distribution in 1D modeling.

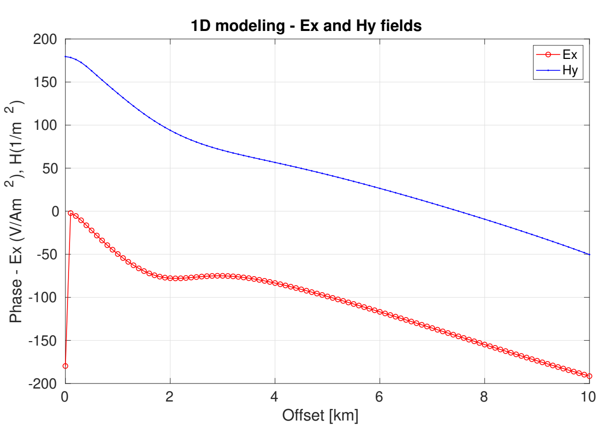

Figure 3 – Phases of electric and magnetic fields for horizontally stratified distribution. 1D modeling.

The next step was to relate the measurement points in x (generated in the modeling) with the corresponding resistivity values in depth z. Tu & Zhdanov (2020) points to the discussion of associating a common midpoint (CMP) family with the transmitter and receiver. We then generate a midpoint as a function of each receiver, resulting in different z-measurement points, as seen in Figure 4.

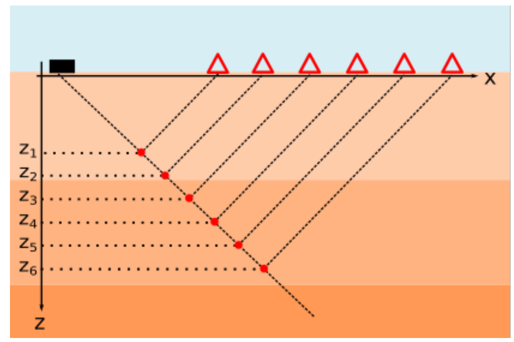

Figure 4 – Relation between the measurement points in x (triangles) regard to the transmitter (rectangle) with the depth z through a family of the common midpoint (CMP).

Machine Learning

The electric and magnetic fields, resulting from the MCSEM field interaction with the marine environment, exemplified in Figures 2 and 3, together with the values in depth for the resistivities presented in figure 4 will be called "training data" or features. We use machine learning (ML) to identify patterns of this information that form a good relationship as we assign a large dataset to algorithms. As a result, we can predict resistivity values for a new input observed (synthetic) data.

The information given to the training set is the amplitudes and phases of E_x and H_y components and the z depth. We split this set into 80% training and 20% validation data to corroborate the model's effectiveness. Initially, 10,000 training data were generated for a random distribution of resistivity in the horizontally stratified medium with three layers and we find the prediction recorded in Figure 5. We infer the need to increase this amount of training when dealing with problems with a more significant number of layers and with more robust algorithms and, consequently, obtain more general answers. Subsequently, 50,000 and 100,000 models were presented to the machine and we verified significant improvements in the Random Forest(RF), K-Nearest Neighbors Regression (KNN-R) and Convolutional Neural Network(CNN) algorithms to perform predictions on the test data.

The choice of algorithms applies to the concept of bias and variance; the first is related to the model's flexibility, while the variance refers to the error of a model about the data. For example, Random Forest combines a series of tree classifiers using bagging and randomization approaches, high classification accuracy, and relatively robust to outliers and noise, Zhang & Lu (2012). Therefore, it can better estimate the essential properties, such as bias and variance of the problem. In the same sense, according to Zhang et al. (2018), the KNN-R algorithm analyzes the nearest neighbors in the training set and works well with non-linear data, attesting to its ability to make good predictions for our problem.

Finally, as mentioned by Goodfellow et al. (2016), CNN has the principle that more similar samples belong to the same layer and have many intermediate layers and operations. Therefore, the information flows within each layer of the network, and an input depends on the previous output, presenting positive responses when related to the error metrics and their bias and variance as executed.

Discussion and Results

The criterion used to corroborate the presence of heterogeneity will follow the principle that sediments saturated by oil or gas have a resistivity of the order of 70-100 ohm.m (Rijo & Almeida, 2005). Therefore, the predictions made with the Random Forest (RF), K-Nearest Neighbors (KNN), and Convolutional Neural Network (CNN) algorithms obtained results very close to the criterion as mentioned above, consequently, to the observed model. Therefore, we affirm the presence of the resistive body and present predictions about the other layers with values close to the observed model, previously known as test data. The peculiarities of each of them were the starting point and justification for the expected solution.

We used a model with only three layers to corroborate the methods applied in this work, commonly reproduced in the numerical modeling of the MCSEM 1D. It served as a stimulus for models with more layers because it indicated a consistent result, as illustrated in Figure 5. Despite being generated with 10,000 models (considered a small number when dealing with robust algorithms associated with the monitoring problem), it ratified the idea of including more training data in the methods that can express better responses. Then, we can include more layers and generalize the study. Covering most of the possible values of random resistivities distributed in the subsurface was our next step, intending that a good prediction is possible for any observed model.

After the discussion, we performed the predictions in models with six layers. We also increased it to 50,000 training data. Figures 6, 7 and 8 show the regression performed by the algorithms Random Forest, K-Nearest Neighbors Regression and Convolutional Neural Network with descriptions of the model in a horizontally stratified marine environment.

The graphs show the predicted resistivities from the water layer, 1,000m thick, to the last layer, which presents the basement's resistivity in three different models. Among them are other resistive layers and the sediment saturated by oil or gas, which has a resistivity of approximately 100 ohm.m in all synthetic data.

Seen as a simple and intuitive algorithm, KNN was the first to induce others to perform, depending on the response generated. It is noteworthy that RF and CNN are more robust and the methods have common characteristics. Furthermore, the prediction duration for the KNN algorithm with 50,0000 models is minimal, about 10 minutes, thus corroborating the importance of starting predictions with KNN. Then, we performed the predictions in the random forests and, finally, in the convolutional neural network.

We can see that despite presenting different synthetic data to the machine, a good response was generated with the convolutional neural network algorithm, since its principle is to have a more significant number of intermediate layers and, consequently, operations, they had a very efficient prediction of resistivity.

We can say that for the monitoring of resistivities, the use of robust algorithms, as is the case of CNN, shows us responses closer to the actual model in all observed models that were tested. However, we have a higher computational cost than the other algorithms. On average, the prediction of this artificial intelligence lasted 10 hours and 33 minutes; it is worth mentioning that the computational cost increases proportionally to the number of training sessions.

The metrics used to validate the model were coefficient of determination (R2), mean squared error (MSE) and mean squared error (RMSE). The R2 determines the percentage of model variance about the predicted data, it is pointed to KNN, CNN and RF, respectively.

Tables 1, 2 and 3 show the model's validation through metrics for the observed data. The MSE is a metric that calculates the average difference between the predicted and actual values, raising it to the square. It performed well in the three algorithms, emphasizing CNN predictions because it adjusted better in all predictions. The RMSE uses the same calculation as the MSE, with the increment of the root to deal with possible problems of the difference between units, and also shows in the best values for CNN.

Table 1 – Metrics for error analysis and validation of the observed model compared to synthetic data 1.

Metrics	KNN	CNN	RF
MSE	0.00225	0.00119	0.00173
RMSE	0.04740	0.03453	0.04154
R2 SCORE	0.99492	0.99731	0.99610

Table 2 – Metrics for error analysis and validation of the observed model compared to synthetic data 2.

Metrics	KNN	CNN	RF
MSE	0.00067	0.00222	0.00044
RMSE	0.02584	0.02117	0.02124
R2 SCORE	0.99686	0.99820	0.99824

Table 3 – Metrics for error analysis and validation of the observed model compared to synthetic data 3.

Metrics	KNN	CNN	RF
MSE	0.00060	0.00044	0.00071
RMSE	0.02459	0.02087	0.02673
R2 SCORE	0.99914	0.99938	0.99898

According to the metrics above and the comparison of the algorithms in Figures 6, 7 and 8, we can identify a better prediction of the Convolutional Neural Network. Associating its result with the execution time, we verified a high computational cost and an excellent prognosis, characteristic of the algorithm known for predicting adequate responses to any observed model according to the amount of information supplied to the machine. For the 50,000 models, we infer that it is more adequate even with noise in the training data. The response found for the other algorithms also satisfied predictions with good variance and bias. In the case of KNN, it presented the lowest computational cost, but it is considered simple and intuitive and unsuitable for robust models. On the other hand, the RF took about 1 hour to predict each observed model, which is a positive point. However, it showed inconsistency when we presented different observed models; we saw that it does not fit so well in all resistivity monitoring despite much training data.

Finally, we analyzed the algorithms' predictions for training with 100,000 models. As we can see in figures 9 and 10, the results confirm the idea that the greater the number of training, the greater the effectiveness of the prediction.

Conclusion

The importance of introducing artificial intelligence in geophysical studies becomes increasingly necessary; we concluded during the processing of the work that machine learning regression algorithms are capable of providing robust program results efficiently. Furthermore, we analyzed that an intelligent choice of the algorithms to be used should be made, mainly for the MCSEM in monitoring the resistivities' variations of a reservoir. The preference for KNN, RF and CNN is related to their characteristics in predicting models with machine training presentation. The interest in selecting robust methods with large training models is to extend the work to other possible observed data, making it more general for different MCSEM models.

Acknowledgment

The authors would like to thank the support of the research grant for project agreement Petrobras/UFPA/Fadesp, ”Estudos de Métodos Eletromagnéticos Aplicados à Exploração de Hidrocarbonetos” and the Graduate Program in Geophysics at UFPA for the computational infrastructure.

References

Constable, S. & Srnka, L. J., 2007. An introduction to marine controlled-source electromagnetic methods for hydrocarbon exploration, Geophysics, vol. 72(2): WA3–WA12.

Das, V., Pollack, A., Wollner, U. & Mukerji, T., 2019. Convolutional neural network for seismic impedance inversioncnn for seismic impedance inversion, Geophysics, vol. 84(6): R869–R880.

Goodfellow, I., Bengio, Y. & Courville, A., 2016. Deep learning, MIT press.

Harris, J. & Grunsky, E. C., 2015. Predictive lithological mapping of canada's north using random forest classification applied to geophysical and geochemical data, Computers & geosciences, vol. 80: 9–25.

Kim, Y. & Nakata, N., 2018. Geophysical inversion versus machine learning in inverse problems, The Leading Edge, vol. 37(12): 894–901.

Kim, Y. & Nakata, N., 2019. A comparison of geophysical inversion and machine learning in inverse problems.

Kuhn, S., Cracknell, M. J. & Reading, A. M., 2018. Lithologic mapping using random forests applied to geophysical and remote-sensing data: A demonstration study from the eastern goldfields of australia, Geophysics, vol. 83(4): B183–B193.

Qadrouh, A., Carcione, J., Alajmi, M. & Alyousif, M., 2019. A tutorial on machine learning with geophysical applications., Bollettino di Geofisica Teorica ed Applicata, vol. 60(3).

Rijo, L. & Almeida, F. L., 2005. Constrained 1-d inversion of mcsem data on resistive oil reservoir, in: 9th International Congress of the Brazilian Geophysical Society, European Association of Geoscientists & Engineers, cp–160.

Tarantola, A., 2005. Inverse problem theory and methods for model parameter estimation, SIAM.

Tu, X. & Zhdanov, M. S., 2020. Robust synthetic aperture imaging of marine controlled-source electromagnetic data, IEEE Transactions on Geoscience and Remote Sensing, vol. 58(8): 5527–5539.

Ward, S. H. & Hohmann, G. W., 2012. 4. Electromagnetic Theory for Geophysical Applications, 130–311, doi: 10.1190/1.9781560802631.ch4.

Wu, X., Liang, L., Shi, Y. & Fomel, S., 2019. Faultseg3d: Using synthetic data sets to train an end-to-end convolutional neural network for 3d seismic fault segmentation, Geophysics, vol. 84(3): IM35–IM45.

Zhang, G. & Lu, Y., 2012. Bias-corrected random forests in regression, Journal of Applied Statistics, vol. 39(1): 151–160.

Zhang, S., Cheng, D., Deng, Z., Zong, M. & Deng, X., 2018. A novel knn algorithm with data-driven k parameter computation, Pattern Recognition Letters, vol. 109: 44–54.

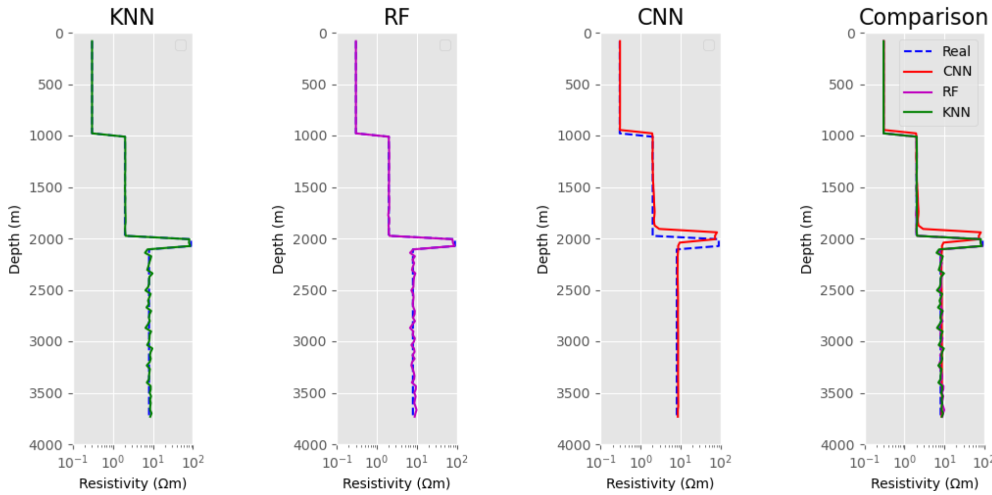

Figure 5 – Prediction with RF, KNN-R and CNN in three layers model, with 10k training models.

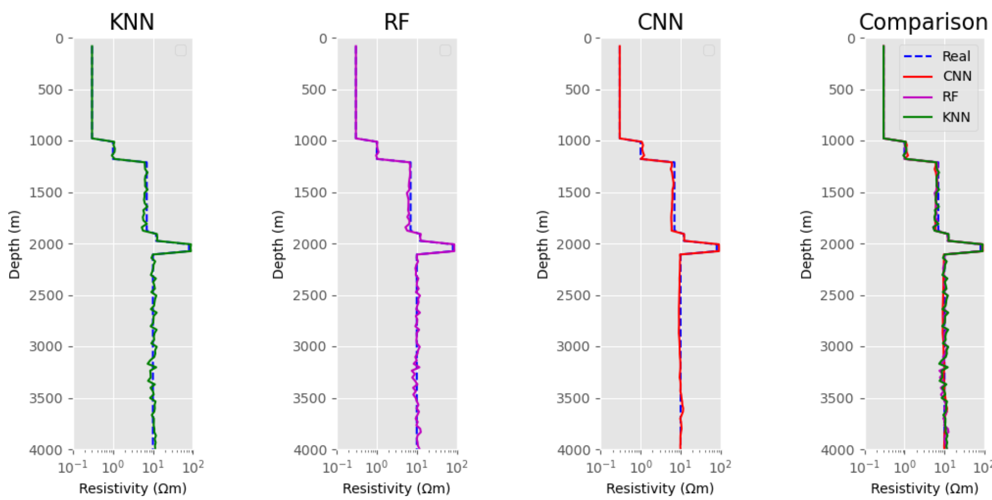

Figure 6 – Prediction with 50k training models in a six layers model, synthetic data 1.

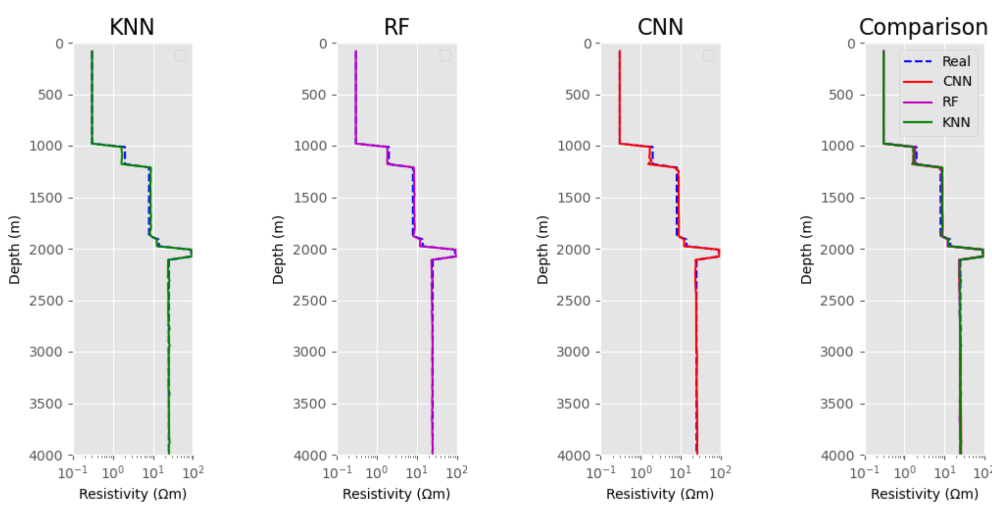

Figure 7 – Prediction with 50k training models in a six layers model, synthetic data 2.

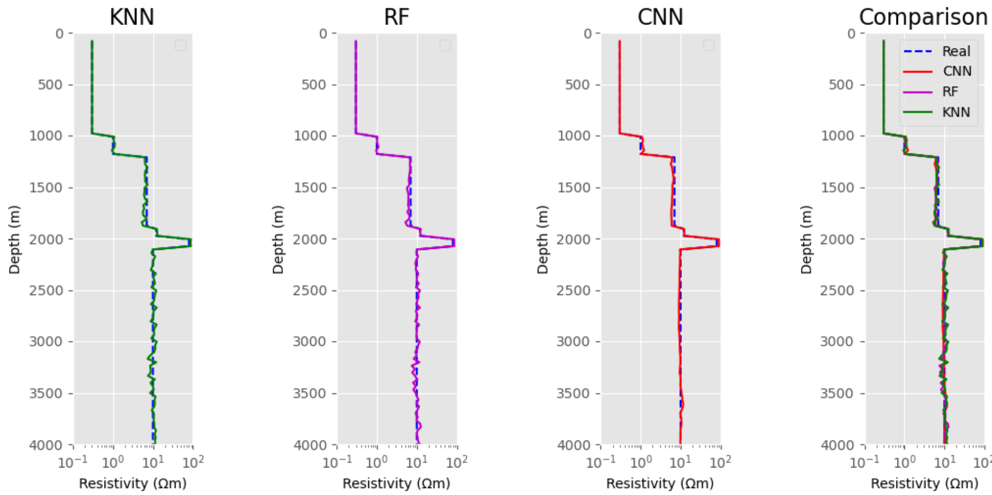

Figure 8 – Prediction with 50k training models in a six layers model, synthetic data 3.

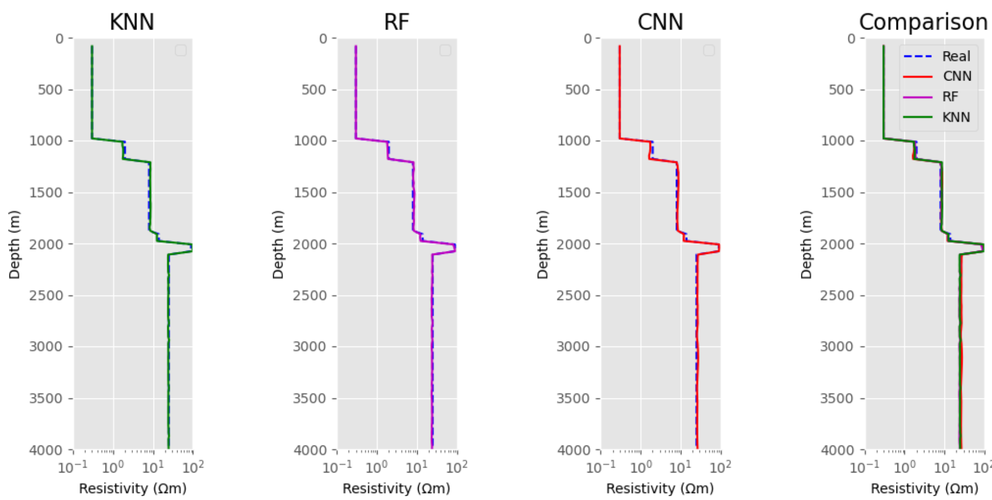

Figure 9 – Prediction with 100k training models in a six layers model, synthetic data 1.

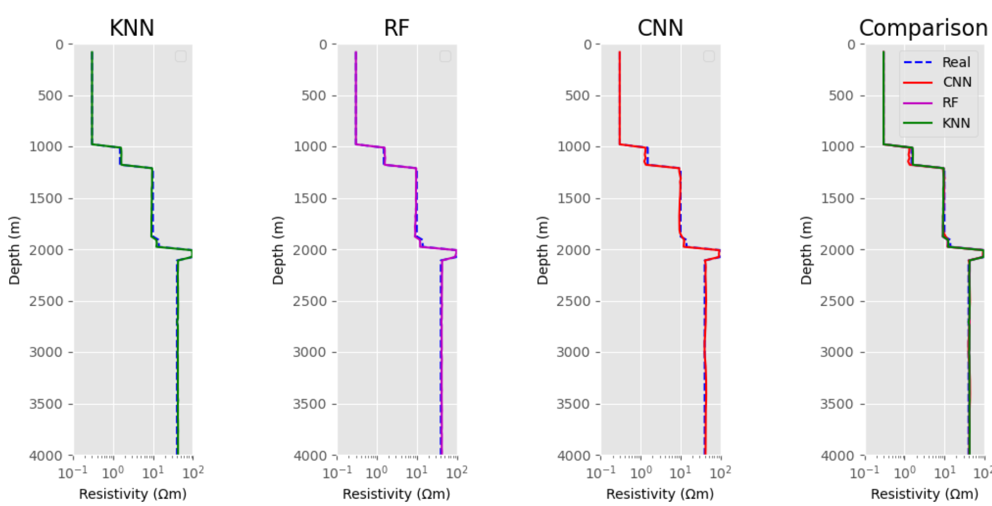

Figure 10 – Prediction with 100k training models in a six layers model, synthetic data 2.