RGNZ/MGPB

Comparison of Porosity Prediction from Seismic Data in the F3 Block, Netherlands using Machine Learning

Rudarsko-geološko-naftni zbornik
(The Mining-Geology-Petroleum Engineering Bulletin)
UDC: 550.3
DOI: 10.17794/rgn.2025.2.2

Preliminary communication

Urip Nurwijayanto Prabowo[1,2]; Sudarmaji[1]; Jarot Setyowiyoto[3], Sismanto[1]*

[1]*Department of Physics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Yogyakarta, Indonesia*
[2]*Department of Physics, Faculty of Mathematics and Natural Science, Universitas Jenderal Soedirman, Purwokerto, Indonesia*
[3]*Department of Geological Engineering, Faculty of Engineering, Universitas Gadjah Mada, Yogyakarta, Indonesia*

Abstract

Porosity is a crucial aspect of reservoir characterization. It can be estimated by measuring rock samples in the laboratory or through indirect methods based on wireline log data. However, both approaches are time-consuming and do not cover large areas. Therefore, this study integrates seismic data with well-logging data for porosity estimation, enabling the coverage of a larger area. Specifically, we estimate porosity in the F3 block of the Netherlands, where there is limited well-logging data available as labelled data. To address this issue, we employ machine learning, consisting of both an inversion generator and a forward generator, to estimate porosity from seismic data. The inversion generator facilitates the process of converting seismic data into porosity, while the forward generator enables forward modelling, transforming porosity data back into seismic data. This forward generator incorporates geophysical knowledge to mitigate the limited amount of labelled data during the training process. Both generators utilize a convolutional neural network-gated recurrent unit network (CNN-GRU). Additionally, various training schemes, such as closed-loop and cycle-GAN (Cycle-Generative Adversarial Network), were employed in this research and compared with an open-loop approach. The data from the F3 block consists of six post-stack seismic lines with an inline spacing of 40 meters and three wells. Data from two wells (F03-4 and F02-1) were used for training, while one well (F06-1) was used for testing. The results indicate that the cycle-GAN produces the most accurate porosity estimates, with a mean square error (MSE) of 0.603 and a Pearson correlation coefficient (PCC) of 0.713. The cycle-GAN utilizes cycle-consistent loss to enhance parameter updates and employs a discriminator to facilitate competitive learning with the generator, thereby improving accuracy. This demonstrates that cycle-GAN can achieve more accurate porosity estimation in locations with insufficient labelled or well-logging data compared to open-loop and closed-loop schemes. However, cycle-GAN requires more computational time than the other methods due to the need to train multiple networks. Nonetheless, this additional training time is justified by the improvement in PCC.

Keywords:
machine learning; porosity; closed-loop; cycle-GAN

1. Introduction

Porosity plays a crucial role in identifying hydrocarbon prospect rocks and conducting reservoir analysis (**Ahmadi and Chen, 2019; Russell et al., 2011; Ali et al., 2021**). A rock's porosity determines its capacity to hold fluids, making high-porosity rocks promising targets for exploration. Porosity estimation involves direct measurements of rock samples in the laboratory and indirect methods based on wireline log data (**Schön, 2011**). While direct measurement is time-consuming and results can be inconsistent due to variations between laboratory and in-situ conditions, the use of wireline logs can offer accurate results, although uncertainties may arise when interpolating porosity between wells (**Yasin et al., 2021**). Hence, integrating seismic data with well-logging data for porosity estimation is often pursued to achieve more robust spatial results.

The porosity and seismic data relationship can be expressed through an empirical equation derived from a cross-plot analysis of porosity from well-logging data and AI (acoustic impedance) from seismic inversion (**Wadas and von Hartmann, 2022; Yasin et al., 2021**). This equation serves to estimate porosity across the entire acoustic impedance cross-section, thereby facilitating the spatial distribution of porosity in both 2D and 3D (**Kushwaha et al., 2020**). **Mojeddifar et al. (2015)** utilized a pseudo-forward equation to establish a nonlinear relationship between AI and porosity, which has been shown to yield more accurate porosity estimations compared to cross-plot methods. However, it's important to note that the equations derived from cross-plots and PFE apply only to specific lithologies, necessitating the

* Corresponding author: Sismanto
e-mail address: sismanto@ugm.ac.id

determination of multiple equations to cover various lithological compositions.

Since a single empirical equation relating porosity to seismic data cannot be applied to all lithologies, several researchers have employed machine learning techniques to estimate porosity using seismic data or seismic attributes. For example, **Kushwaha et al. (2021)** used machine learning as a geostatistical method to extrapolate porosity, density, Vp, and gamma rays from seismic attribute data. **Agbadze et al. (2022)** employed deep neural networks, random forests, and decision tree algorithms to determine porosity using acoustic impedance. **Feng (2020a)** used elastic modulus κ and M to estimate porosity with a convolutional neural network (CNN). **Gholami et al. (2022)** used Vs, Vp, and density to determine porosity through a bat-inspired optimized neural network, bat-inspired optimized support vector regression, bat-inspired optimized fuzzy inference system, and hybrid model based on a committee machine (CM). **Sang et al. (2021)** used a deep extreme learning machine (DELM) to determine porosity from seismic data, using synthetic training data generated through a virtual sample generation method to address the lack of labelled data in the training process. However, since training data can only be generated for specific geological conditions, the trained DELM may not apply to all geological settings. Based on these examples, there are two challenges in porosity prediction using machine learning. First, the accuracy of the seismic attributes used as inputs significantly influences the porosity prediction results (**Feng, 2020a**). Seismic attributes result from the seismic data inversion process, meaning that determining porosity using these attributes involves a two-stages inversion. In a two-stages inversion, the parameter prediction error accumulates from parameter inversion and attribute inversion, leading to larger errors and increased time consumption compared to a one-stage inversion using machine learning (**Das and Mukerji, 2020**). Second, machine learning requires a large number of labelled data points for effective training (**Sang et al., 2021)**.

In this study, we estimate porosity from seismic data of the F3 block of the Netherlands using machine learning. To address the machine learning challenges and improve porosity prediction accuracy, we utilize seismic data as the input for machine learning. We then employ closed-loop and cycle-GAN schemes to train the machine learning models. The closed-loop scheme uses an inversion generator and forward generator simultaneously in the training process and is applied to labelled and unlabelled seismic data (**Alfarraj and AlRegib 2019a; Wang et al., 2021, 2020**). The inversion generator inverts seismic data into porosity and the forward generator reconstructs seismic data from porosity (inversion generator output). The forward model is used as additional information to constrain the training process and overcome the need for large labelled data (**Adler et al., 2021; Wang et al., 2020**). Cycle-GAN is an integration of a forward generator and an inversion generator in a GAN (generative adversarial network) architecture (**Zhu et al., 2017**). In cycle-GAN, the inversion generator and forward generator are referred to as generators and there are 2 additional discriminators (**Wang et al. 2022; Zhang et al., 2022**). The discriminator differentiates between the generator results and the label data. The generator has good capabilities if it can produce output that is very similar to the label data and cannot be differentiated by the discriminator. The generator and discriminator are trained simultaneously so that their ability levels outperform each other during the training process until maximum generator capability is obtained (**Goodfellow et al., 2014**). The inversion and forward generator in this study uses a combination of CNN and GRU to extract high and low-frequency features so that it does not require a low-frequency model as additional input (**Alfarraj and AlRegib, 2019a, 2019b**). High-frequency features denote local patterns of output data within a short data range, while low-frequency features represent global patterns or variations in output data trends across the entire dataset (**Wang and Cao, 2022**). The results of the closed-loop and cycle-GAN schemes will be compared with the open-loop scheme.

2. Geological setting of F3 block

The location and surface geology of the F3 block is shown in **Figure 1**. The F3 offshore block lies to the north of the Netherlands within the central graben, part of the expansive North Sea basin (**Rondeel et al., 1996**). Rifting within the North Sea basin created most of the geological structures in the Netherlands during the late Jurassic to early Cretaceous period (**Rondeel et al., 1996**). However, rifting activities began during the transition between the Permian and Triassic periods (**Ziegler, 1992**). During the earliest Triassic, the rift in the Norwegian-Greenland Sea propagated into the North Sea, leading to the subsidence of the Horda-Egersund half-graben, the Viking and central graben, the Moray Firth-Witch ground graben system. The subsidence of the North Sea rift system continued throughout the early Jurassic period (**Ziegler, 1992**).

During the mid-Jurassic period, a wide arc uplifted in the central North Sea, extending across the central graben and creating the central North Sea dome. This uplift led to a reversal in the subsidence pattern of the central graben. Sediments from the Early Jurassic, Triassic, and even Permian periods that once covered the dome's crest were largely eroded and then redeposited in neighboring basins that were undergoing continuous subsidence, especially within major deltaic complexes. Rifting activity reached its peak at the transition from the Jurassic to the Cretaceous, focusing on the Viking, central graben, and Moray Firth-Witch ground graben systems. This led to rapid subsidence and faulting deformation (**Ziegler, 1992**).

During the Late Cretaceous and Paleocene, the North Sea basin experienced the later stages of rifting, with

rifting activity diminishing further. Regional thermal subsidence, combined with rising sea levels, played a crucial role in eroding the basin margins. During the Cenozoic era, the North Sea basin evolved into a wide, saucer-shaped thermal sag basin (**Ziegler, 1992**). Sediment accumulation was most pronounced in the central graben area during Cenozoic times (**Sorensen et al., 1997**). Hydrocarbon generation within the North Sea rift basin was accelerated by the subsidence that occurred during the late Jurassic to early Cretaceous rifting phases (**Rondeel et al., 1996; Sorensen et al., 1997**). Most of the gas in the basin has been generated from Triassic sediments and Westphalian coals, while the majority of the oil originates from the Lower Posidonia Shale (**De Jager and Geluk, 2007**).

Hydrocarbon generation ceased due to inversion during the Late Cretaceous, which involved uplift and erosion that reduced heat flow. In the Tertiary period, the effects of inversion were limited, and substantial subsidence occurred at the basin margins. In this region, hydrocarbon charge from Westphalian sources resumed and has continued to the present day (**De Jager and Geluk, 2007**).

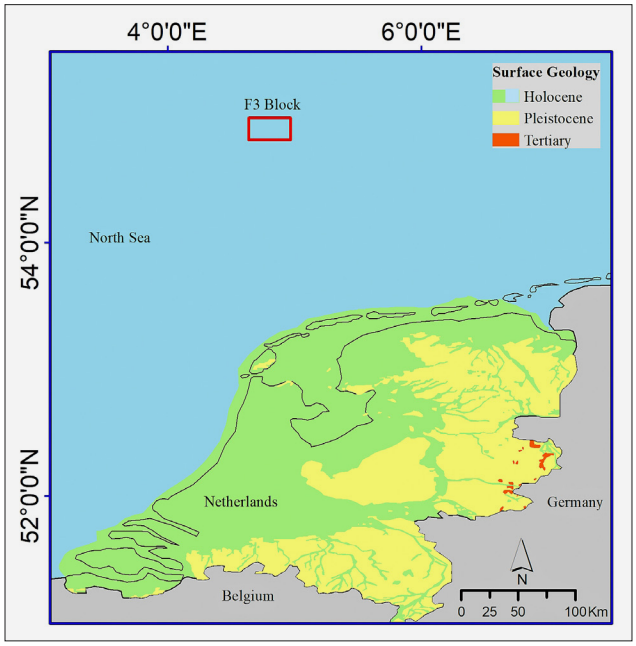

Figure 1: F3 block location map

3. Methods

3.1. Data

We use the F3 data located in the Netherlands, which consists of three wells and six lines of post-stack seismic data (see **Figure 2**). The seismic data ranges from crossline 300 to 1250 without interval steps and inline 244 to 442 with interval 40 steps. The seismic data has a 4 ms sampling rate. In this study, the focus is on the Upper North Sea, where the TWT (two-way travel time) ranges from 600 to 1100 milliseconds (**Alaudah et al., 2019**). We used post-stack seismic data. The seismic line crosses the well location on lines 244, 362, and 442. The F03-4 and F02-1 wells are used as training datasets, and F06-1 is used as testing datasets (see **Figure 2**). The testing dataset is used as a control to evaluate the generalization capability of trained machine learning. Generalization capability means the ability of machine learning to obtain the correct output from unknown data that is not contained in the training data (**Biswas et al., 2019a**). The porosity obtained from well-logging data has a sampling rate six times higher than seismic data.

3.2. Inversion generator

The inversion generator adopts the CNN-GRU network from **Alfarraj et al. (2019a, 2019b)** to estimate porosity from seismic data. The inversion generator input is seismic data while the output is estimated porosity. CNN is used to extract the high frequency features of the input, while GRU extracts low-frequency, eliminating the need for a separate low-frequency model as additional input (**Alfarraj and AlRegib, 2019a, 2019b**). **Figure 3** illustrates that the CNN-GRU networks consist of three CNN layers and three GRU layer. The three CNN layers extract high-frequency information from seismic data using different hyperparameters. The hyperparameters for the first CNN layer are a kernel size of 5, padding of 2, and dilation of 1. The second CNN layer uses a kernel size of 5, padding of 6, and dilation of 3, while the third CNN layer features a kernel size of 5, padding of 12, and dilation of 6. The outputs of the three CNN layers are concatenated and subsequently fed into additional CNN networks. The three GRU layers extract the low-frequency information from seismic data. These simultaneous GRUs are used to achieve desirable low-frequency information because a deeper layer of GRU results in a smoother low-frequency model. The results from the three CNN and three GRU layers are summed and then passed into deconvolution networks to upscale the resolution to match that of the porosity data. This approach allows the network to generate porosity values with a higher resolution than that of the seismic data. The final layer, which regresses the results into the porosity domain, consists of a GRU and a fully connected layer. The hyperparameters of the inversion generator are updated during the training process based on the loss function, which depends on the scheme being used.

3.3. Forward generator

The forward generator adopts the same CNN-GRU network (**Alfarraj and AlRegib, 2019a, 2019b**), but it is used to reconstruct seismic data from porosity, resulting in an inversion generator (see **Figure 4**). Three layers of the CNN network and three layers of GRU are

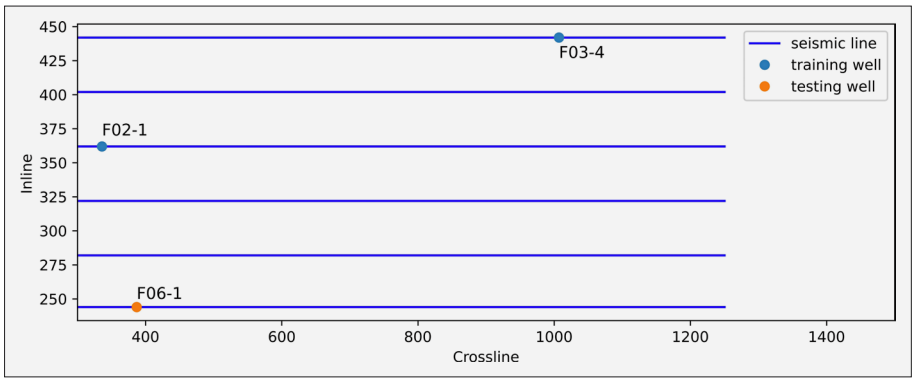

Figure 2: The seismic data and well position

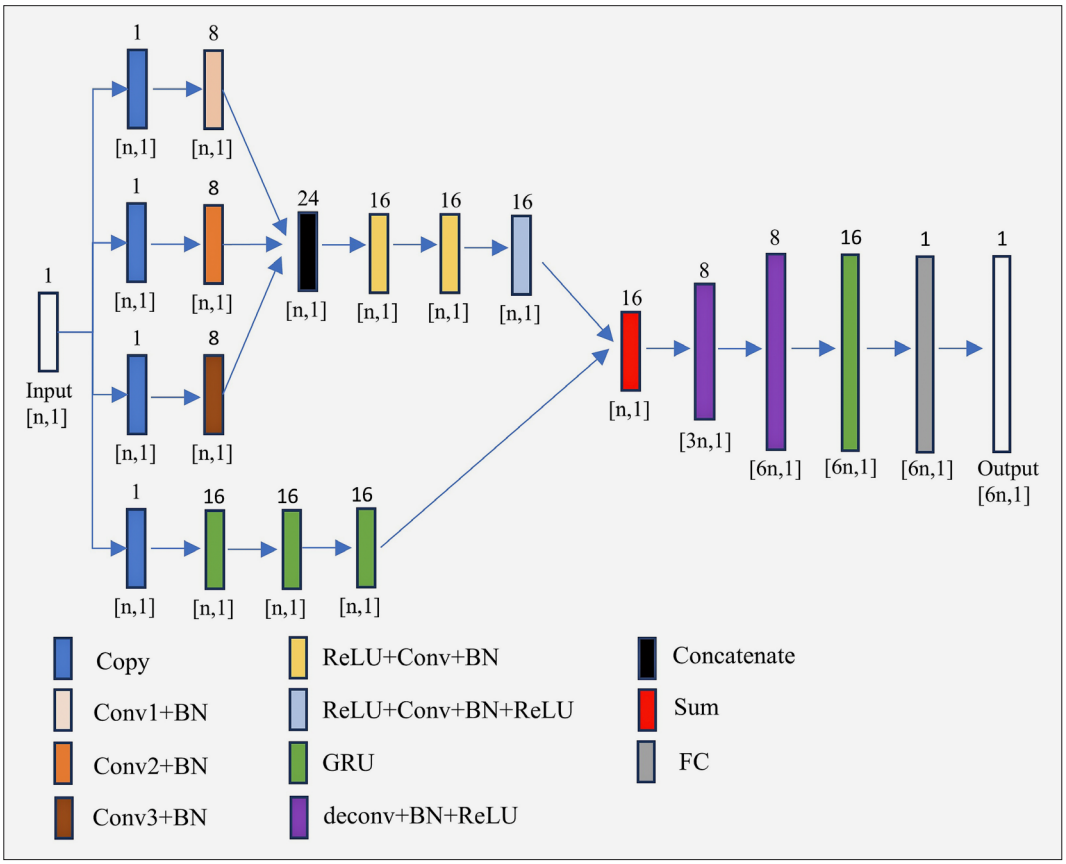

Figure 3: Inversion generator

also used to extract low and high-frequency information from the input. Different from the inversion generator, the deconvolution networks are replaced by downscale calculation in the forward model since the seismic data (output in the forward model) has a lower resolution than porosity (input in the forward model).

3.4. Machine learning architecture

In this study, we compared porosity predictions obtained from a common open-loop scheme with those from a closed-loop scheme and Cycle-GAN. Each scheme utilized the same dataset from the F3 block but differed in architecture and loss functions during model training. The results for each scheme will be evaluated based on mean squared error (MSE), the Pearson correlation coefficient (PCC), and computation time to determine the best machine learning architecture for predicting porosity in the F3 block. The optimal scheme will be the one that achieves a higher PCC and a lower MSE in the testing well.

The Pearson correlation coefficient (PCC) measures the linear correlation between estimated porosity and porosity from well-logging data, based on the following equation:

$$PCC = \frac{1}{N}\frac{1}{\sigma_p \sigma_{p^*}} \sum_{i=1}^{N} \left[P_i - \mu_p\right]\left[P_i^* - \mu_{p^*}\right] \qquad (1)$$

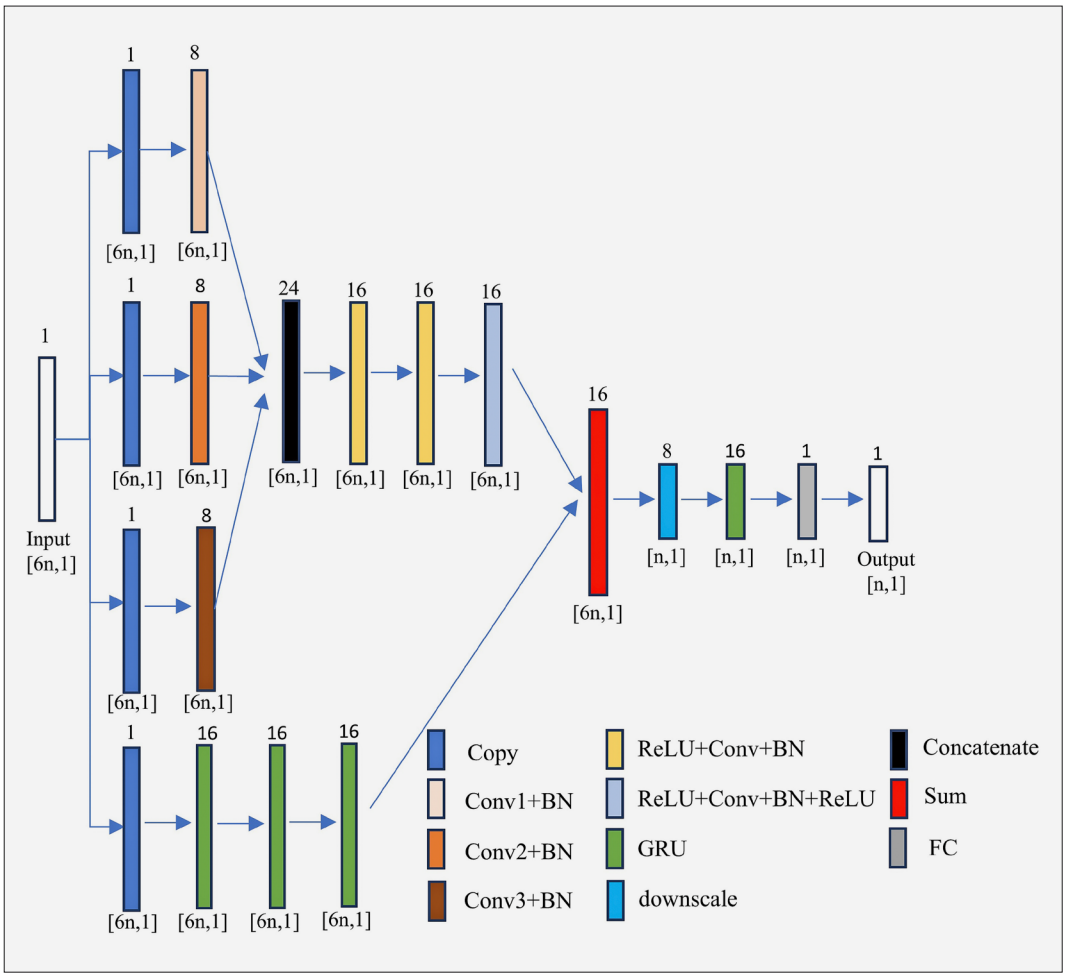

Figure 4: Forward generator

where N is the total number of porosity data from well-logging, P^* is the predicted porosity from the inversion generator, P is the true porosity from well-logging, σ is the standard deviation and μ is the mean value.

MSE measures the average of the squares error between predicted porosity and true porosity based on the equation as follows:

$$\mathrm{MSE} = \left(\frac{1}{N} \sum_{i=1}^{N} \left(\mathrm{P}_i^* - P_i \right)^2 \right) \tag{2}$$

This study uses the same training parameters for all schemes, which are as follows: the learning rate is 0.005, the number of epochs is 1000, the batch size for unlabelled data is 40, and the Adam optimizer is employed (**Kingma and Ba, 2015**) as the optimization method. The parameter was determined through a trial-and-error process to achieve optimal results. All schemes also run on the same laptop (Intel i7 CPU and a 4GB Nvidia GeForce RTX 3050Ti GPU).

3.4.1. Open-loop

The open-loop scheme inverts the seismic data into porosity using an inversion generator (see **Figure 5**). This is supervised learning that depends on the labelled data in the training process. An open-loop scheme needs large, labelled data to obtain good accuracy predictions (**Das et al., 2019**). In the open-loop training process, the loss function is calculated in each iteration to optimize the inversion generator using the backpropagation method (**LeCun et al., 1989**). The open-loop loss is denoted as:

$$L_{open} = \frac{1}{m} \sum_{i=1}^{m} \left[\left(G_I \left(S \right) \right)_i - \mathrm{P}_i \right]^2 \tag{3}$$

where L_{open} is open-loop loss, G_I is the inversion generator that generates porosity from seismic data, m is the number of labelled data, S is seismic used as labelled data, and P is porosity used as labelled data.

3.4.2. Closed-loop

The closed-loop is designed to reduce machine learning's reliance on large volumes of labelled data (**Wang et al., 2020**). The closed-loop scheme consists of an inversion generator and forward generator, which simultaneously models the seismic inversion and forward pro-

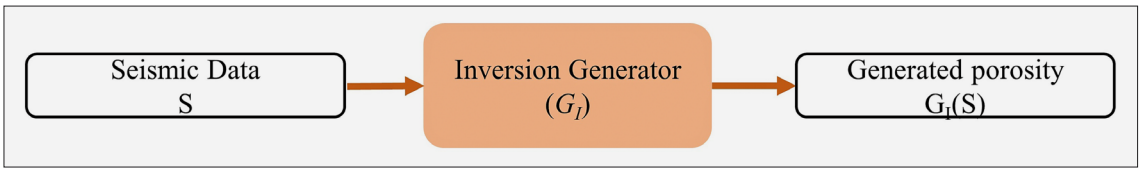

Figure 5: Open-loop scheme

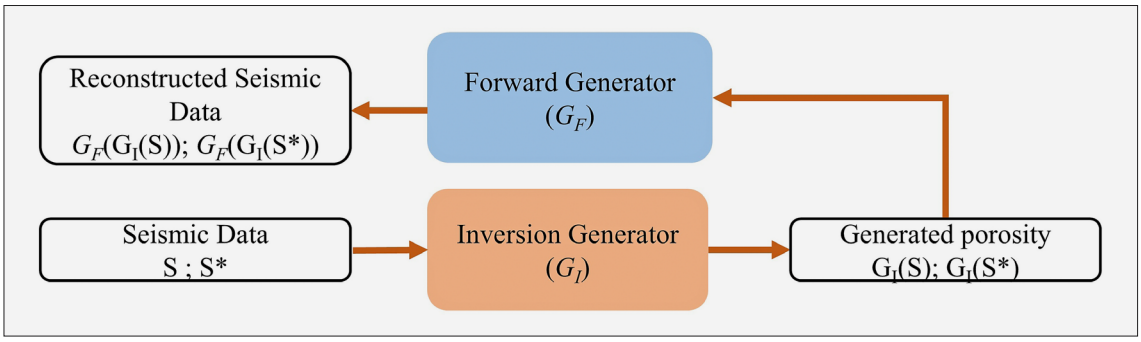

Figure 6: Closed-loop scheme

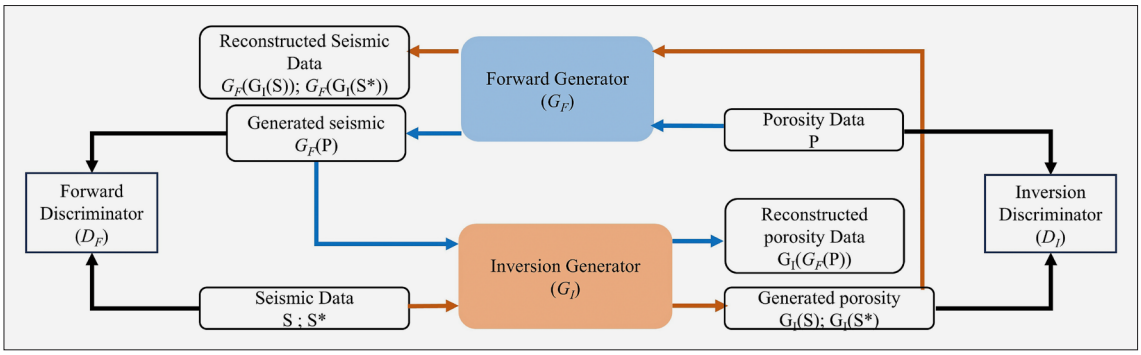

Figure 7: Cycle-GAN

cesses using both labelled and unlabelled data (see **Figure 6**). The closed-loop scheme learns not only from labelled data but also from unlabelled data information. During the closed-loop training process, the loss function is computed in each iteration to optimize both the inversion generator and the forward generator using the backpropagation method (**LeCun et al., 1989**). The closed-loop loss function is denoted as:

$$L_{close} = \alpha L_{label} + \beta L_{unlabel} \tag{4}$$

$$L_{label} = \left(\frac{1}{m} \sum_{i=1}^{m} \left[\left(G_F \left(S \right) \right)_i - \mathrm{P}_i \right]^2 \right) + \\ + \left(\frac{1}{m} \sum_{i=1}^{m} \left[\left(G_F \left(G_I \left(S \right) \right) \right)_i - \mathrm{S}_i \right]^2 \right) \tag{5}$$

$$L_{unlabel} = \left(\frac{1}{m} \sum_{i=1}^{m} \left[\left(G_F \left(G_I \left(S^* \right) \right) \right)_i - S_i^* \right]^2 \right) \tag{6}$$

where L_{close} is closed-loop loss, L_{label} is the loss function of labelled data, $L_{unlabel}$ is the loss function of unlabelled data G_I is the inversion generator, G_F is the forward generator, S is seismic used as labelled data, S^* is seismic used as unlabelled data, P is porosity used as labelled data, m is the number of labelled data, α, and β are the weight coefficient for labelled and unlabelled data. To achieve the best accuracy prediction, we choose α=1 and β=0.2 through a trial-and-error process.

3.4.3. Cycle-GAN

The cycle-GAN was developed by **Zhu et al. (2017)** to mitigate the limitation of paired training data in image-to-image translation problems. **Zhu et al. (2017)** modified the GAN by adding inverse mapping to reconstruct the data. In this study, we implement the cycle-GAN to predict the porosity from seismic data and use a forward generator (see **Figure 4**) as inverse mapping to reconstruct the seismic data. **Figure 7** shows the cycle-GAN scheme consisting of two generators and two discriminators. All generators and discriminators are trained simultaneously in a min-max game so that their ability levels outperform each other until maximum generator capability is obtained. If the generator achieves maximum capabilities, it can produce output that is very

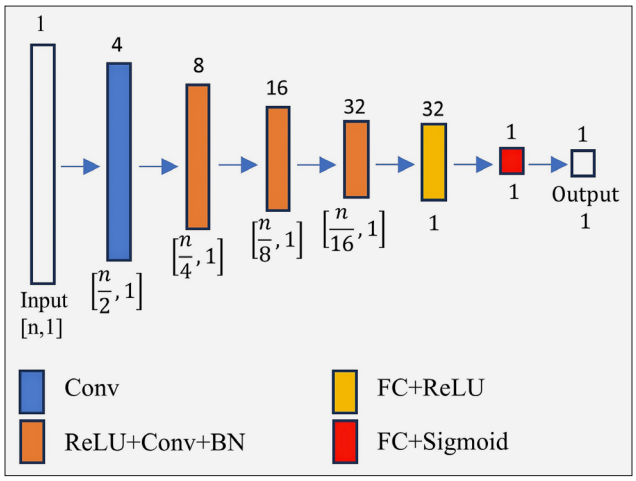

Figure 8: Discriminator

similar to the label data and hard to differentiate by the discriminator. The forward and inversion discriminators use the same CNN networks shown in **Figure 8**.

In the cycle-GAN training process, the loss function is calculated in each iteration to optimize both generators and discriminators using the backpropagation method (**LeCun et al., 1989**). We used a modified generator loss function that is different from (**Zhu et al., 2017**). The modified generator loss function is denoted as:

$$Loss_{GI} = Loss_{GF} = L_{cycle} + \\ + \frac{1}{m}\sum_{i=1}^{m}\left[-\left(D_I\left(G_I(S)\right)\right)_i - \left(D_F\left(G_F(P)\right)\right)_i\right] \quad (7)$$

where $Loss_{GI}$ is the inversion generator loss, L_{cycle} is the cycle-loop loss, $Loss_{GF}$ is the forward generator loss, D_F is the forward discriminator that distinguishes between seismic used as labelled data and seismic generated from forward, D_I is the inversion discriminator that distinguishes between porosity used as labelled data and generated porosity.

The cycle loop loss function is the loss function calculated from a closed-loop framework (semi-supervised) and is denoted as:

$$L_{cycle} = L_1 + L_2 + L_3 + L_4 + L_5 \quad (8)$$

$$L_1 = \frac{1}{m}\sum_{i=1}^{m}\left[\left(G_I(S)\right)_i - P_i\right] \quad (9)$$

$$L_2 = \frac{1}{m}\sum_{i=1}^{m}\left[\left(G_F(P)\right)_i - S_i\right] \quad (10)$$

$$L_3 = \frac{1}{m}\sum_{i=1}^{m}\left[\left(G_F\left(G_I(S)\right)\right)_i - S_i\right] \quad (11)$$

$$L_4 = \frac{1}{M}\sum_{i=1}^{M}\left[\left(G_F\left(G_I(S^*)\right)\right)_i - S_i^*\right] \quad (12)$$

$$L_5 = \frac{1}{m}\sum_{i=1}^{m}\left[\left(G_I\left(G_F(P)\right)\right)_i - P_i\right] \quad (13)$$

where L_{cycle} is the cycle-loop loss, L_1 is the loss of inversion generator result using seismic labelled data, and L_2 is the loss of forward generator result using porosity labelled data. L_3 is the loss of forward generator result using seismic labelled data, L_4 is the loss of forward generator result using seismic unlabelled data, L_5 is the loss of inversion generator result using porosity labelled data, and M is the number of unlabelled data.

The forward discriminator loss and the inversion discriminator loss are denoted as:

$$L_{DF} = \frac{1}{m}\sum_{i=1}^{m}\left[-\left(D_F(S)\right)_i + \left(D_F\left(G_F(P)\right)\right)_i\right] \quad (14)$$

$$L_{DI} = \frac{1}{m}\sum_{i=1}^{m}\left[-\left(D_I(P)\right)_i + \left(D_I\left(G_I(S)\right)\right)_i\right] \quad (15)$$

where L_{DF} is forward discriminator loss, and L_{DI} is inversion discriminator loss.

4. Results and discussion

4.1. Open-loop

Figure 9 shows the inversion results from the open-loop scheme. The porosity estimation in inline 362 shows fairly good spatial continuity around the training well (F02-1), but in inline 442 does not show good spatial continuity between the porosity estimate and the training well (F03-4). The porosity estimation in inline 244 also does not match the testing well. These results show that the estimation results are not good spatially, both around the well that is used as testing and training data. The porosity results show a less smooth or rough pattern with fluctuations in values over a fairly large range. **Table 1** and **Table 2** show that the range of estimated porosity values is wider than the well-logging data for both testing and training data. This indicates that the open loop tends to overestimate porosity values and cannot provide accurate predictions in the F3 block due to the limited labelled data available for training the network. Open-loop systems employ supervised learning, which requires a large amount of labelled data during training to effectively establish the best input-output relationships.

4.2. Closed-loop

Figure 10 shows the inversion results from the closed-loop scheme. The porosity estimation in lines 362 and 442 shows quite good spatial continuity with training wells. Meanwhile, the porosity estimation in line 244 does not match the testing well. These results show good porosity estimation results spatially around the training well but are not accurate on seismic data located far from the training data. Due to the relatively large distance between the well location and the seismic line, the inverted results are suboptimal (**Feng, 2020**). The porosity results show a smoother pattern than the

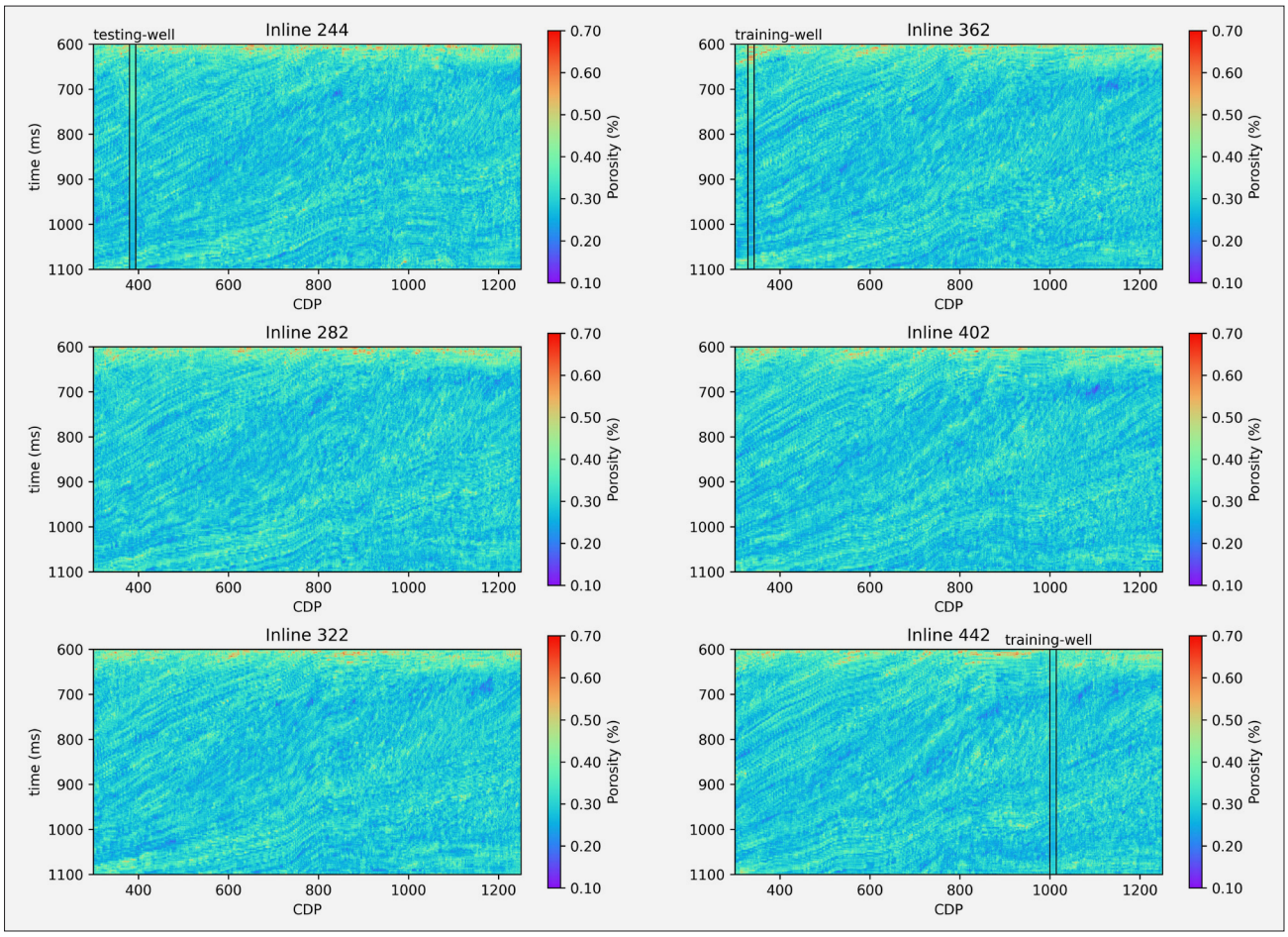

Figure 9: Predicted porosity from open-loop scheme and the spatial correlation with porosity from well-logging data

Table 1: The range value of predicted porosity in Line 244 and true porosity of testing well

	max	Min
Open-loop	0.665	0.149
Closed-loop	0.693	0.209
Cycle-GAN	0.481	0.219
Testing well	0.390	0.240

Table 2: The range value of predicted porosity in Lines 362 and 442 and true porosity of training well

	Line 362		Line 442	
	max	min	Max	Min
Open-loop	0.678	0.164	0.674	0.159
Closed-loop	0.694	0.209	0.691	0.208
Cycle-GAN	0.675	0.218	0.567	0.223
Training well	0.673	0.210	0.371	0.249

Table 3: Evaluation of porosity prediction in testing well

	MSE	PCC	Time (s)
Open-loop	1.020	0.417	17
Closed-loop	0.965	0.421	75
Cycle-GAN	0.603	0.713	164

open-loop results, with slight fluctuations in values at certain points. **Table 1** and **Table 2** indicate a fairly good range of estimated porosity values, with lower values closely aligning with well-logging but showing overestimation at higher values. This suggests that forward modelling can provide additional information during training, resulting in better prediction accuracy compared to open-loop methods. However, the results are still inadequate at locations far from the training data, as evidenced by the low Pearson correlation coefficient (PCC) and high mean squared error (MSE) values in the test data (see **Table 3**). Further improvements are necessary to extract more information from the data, especially since the labelled data in the F3 block is very limited (only two wells).

4.3 Cycle-GAN

Cycle-GAN scheme is the improvement of closed-loop scheme by incorporating discriminators and utilizing cycle consistent loss. **Figure 11** shows the inversion results from the cycle-GAN scheme exhibit sufficient spatial continuity of porosity between the training and testing data. These results show good spatial porosity estimations both around the training wells and in the testing data, which is located far from the training set. The

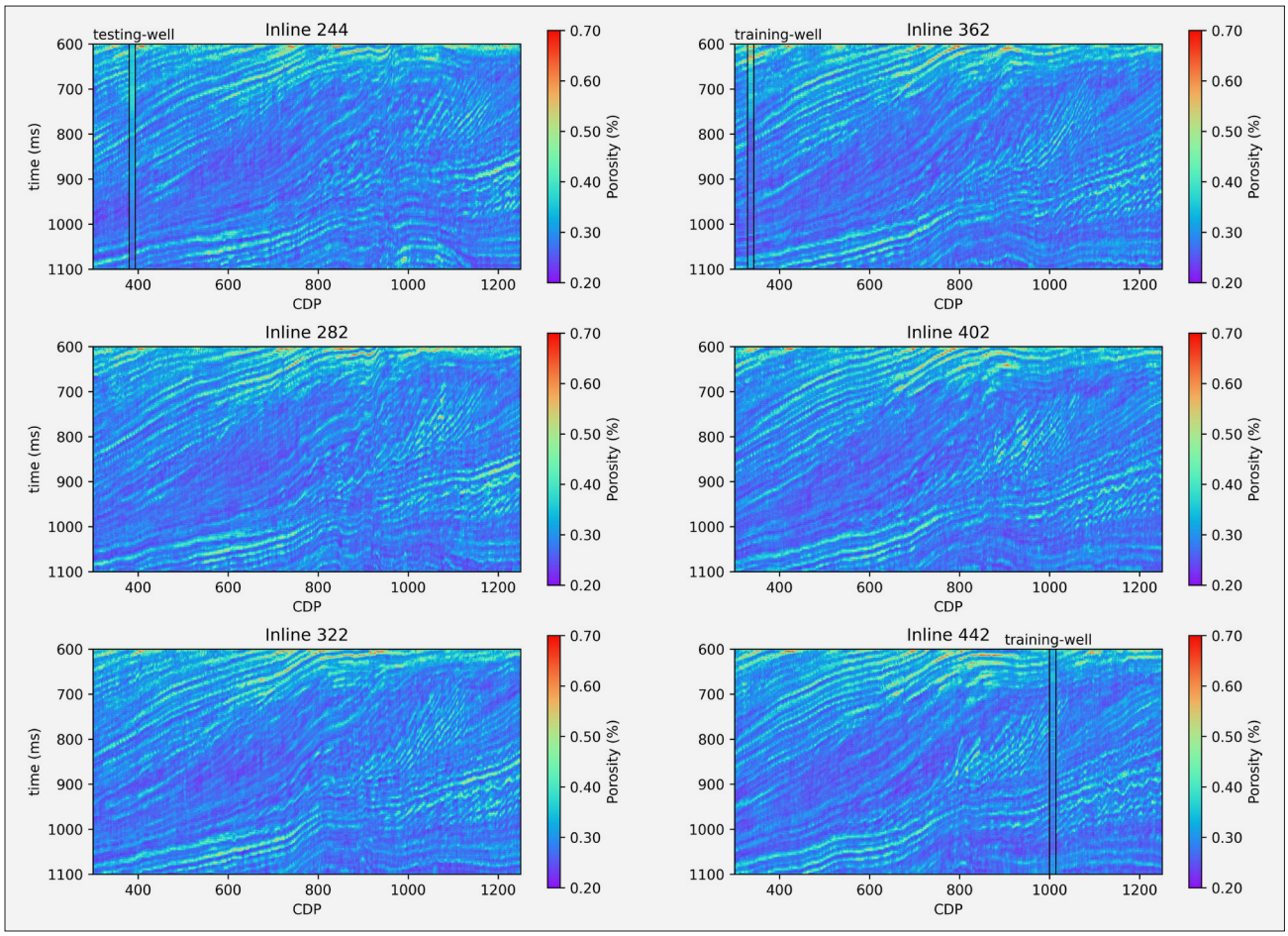

Figure 10: Predicted porosity from closed-loop scheme and the spatial correlation with porosity from well-logging data

porosity results display a smoother pattern compared to those obtained from other schemes. **Figure 12** shows that the estimated porosity pattern at the testing well location from the open-loop and closed-loop results has values that fluctuate suddenly at several points and do not match the true porosity from the well-logging data. This may be due to the amount of label data being too small so that the results follow the pattern of the training data too closely. Meanwhile, the cycle-GAN results have a good pattern, following true porosity without any large fluctuations. Future research can explore varying the number of wells, utilizing either synthetic data or real data with a substantial amount of well-logging data. This approach can help identify the optimal number of labelled data for each training scheme, particularly for the Cycle-GAN scheme.

Table 1 and **Table 2** indicate a good range of estimated porosity values, with both high and low values closely aligning with well-logging data. This demonstrates that the GAN training scheme with cycle-consistent loss and discriminators can extract more information from the data to enhance the training process and enable accurate prediction results even with limited labelled data. In **Table 3**, cycle-GAN has the highest PCC and lowest MSE values, so it is the best scheme compared to the others. This is because cycle-GAN is trained using a more complex loss function, namely the cycle-loss function. In addition, competition between the generator and discriminator in the cycle-GAN training process has also been proven to increase the accuracy of the inversion generator. However, cycle-GAN has a longer processing time than other schemes because it has to train more machine learning (2 discriminators and 2 generators). In an open loop, it only trains 1 generator, while a closed-loop trains 2 generators. However, the cycle processing time is proportional to the prediction accuracy. Since we are currently using a single small GPU, employing multiple larger GPUs for data processing can significantly reduce processing times, particularly with larger datasets.

The data used in this study focuses on the Upper North Sea, which is primarily composed of claystone and sandstone from the Miocene to Quaternary periods. The machine learning models may require adaptation or retraining to be applicable in different lithological contexts. **Table 3** presents the best results from the Cycle-GAN, with a Pearson Correlation Coefficient (PCC) of 0.713, which could be improved for more accurate porosity predictions. Accuracy can be enhanced by adding labelled data or by applying further processing steps to

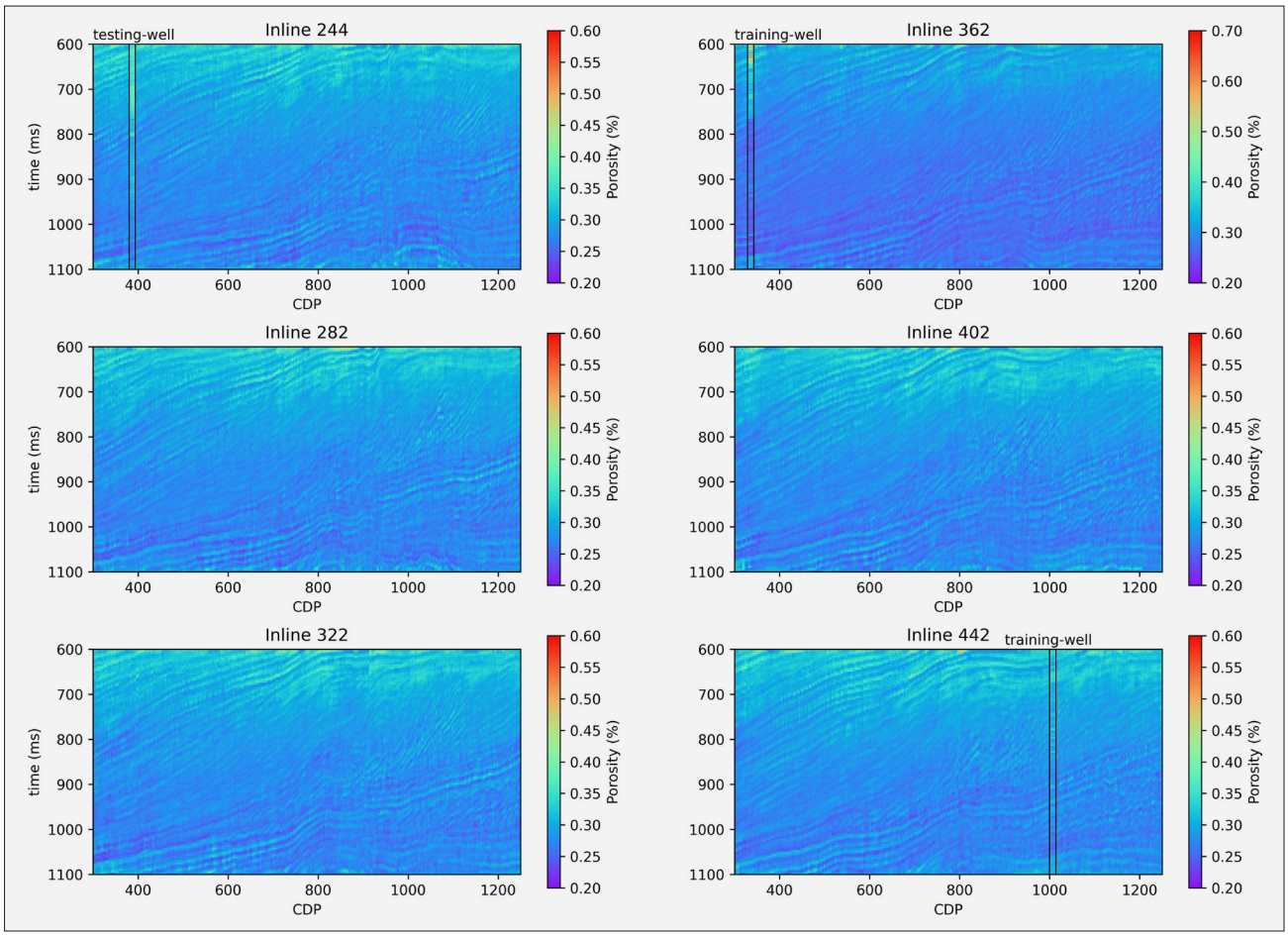

Figure 11: Predicted porosity from cycle-GAN scheme and the spatial correlation with porosity from well-logging data

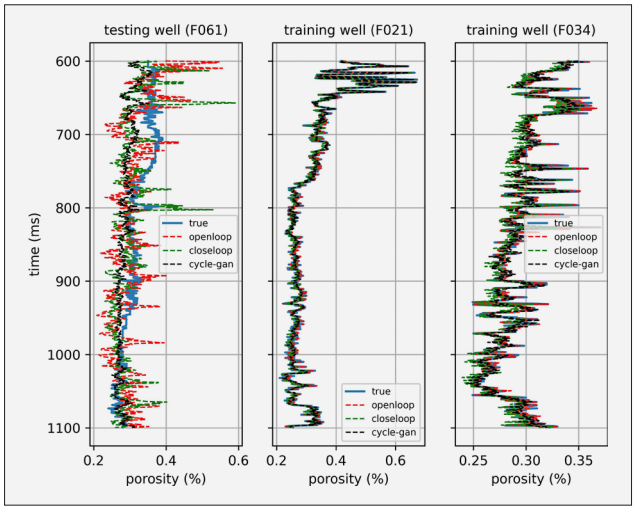

Figure 12: Inversion results in testing well and training wells

the post-stack data to reduce noise in the input data. Since we used unfiltered post-stack data, it may still contain noise that can lead to discrepancies in porosity.

5. Conclusions

Several different schemes of machine learning have been applied in F3 block data with small labelled data to determine the porosity. The results show that the Cycle-GAN scheme provides the best estimation results on test data than closed-loop and open-loop, even when the test data is located at a considerable distance from the training data. This is because cycle-GAN obtains additional information from forward inversion in the form of cycle-consistent loss and there is a competition process between the discriminator and generator.

This demonstrates that cycle-GAN can provide good porosity estimation in locations with insufficient labelled or well-logging data. Accurate porosity prediction is a crucial parameter in reservoir characterization, as it helps evaluate hydrocarbon accumulation and determine flow patterns.

The disadvantage of Cycle-GAN is that it requires longer computational time due to the need to train two discriminators and two generators. However, the accuracy of the obtained results is quite promising, and advancements in GPU technology continue to enhance computational performance. The use of multiple larger GPUs, as well as the development of advanced Cycle-GAN techniques like Cycle-Wasserstein GAN or gradient penalty methods, can be explored in the future to further reduce computational time and enhance the accuracy of the results.

Acknowledgement

We extend our gratitude to dGB Earth Science for publicly providing the F3 Netherlands dataset. This study was supported by the Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Gadjah Mada, under the Collaborative Research Project.

6. References

Adler, A., Araya-Polo, M. and Poggio, T. (2021): Deep Learning for Seismic Inverse Problems: Toward the Acceleration of Geophysical Analysis Workflows. IEEE Signal Process Mag, 38, 89–119. https://doi.org/10.1109/MSP.2020.3037429

Agbadze, O.K., Qiang, C. and Jiaren, Y. (2022): Acoustic impedance and lithology-based reservoir porosity analysis using predictive machine learning algorithms. Journal of Petroleum Science and Engineering, 208, 1–13. https://doi.org/10.1016/j.petrol.2021.109656

Ahmadi, M.A. and Chen, Z. (2019): Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. Petroleum, 5, 271–284. https://doi.org/10.1016/j.petlm.2018.06.002

Alaudah, Y., Michałowicz, P., Alfarraj, M. and Alregib, G. (2019): A machine-learning benchmark for facies classification. Interpretation 7, SE175–SE187. https://doi.org/10.1190/INT-2018-0249.1

Alfarraj, M. and AlRegib, G. (2019a): Semi-supervised Learning for Acoustic Impedance Inversion, in: SEG Technical Program Expanded Abstracts 2019. Society of Exploration Geophysicists.

Alfarraj, M. and AlRegib, G. (2019b): Semi-supervised Sequence Modeling for Elastic Impedance Inversion. Interpretation, 7, 237-249. https://doi.org/10.1190/INT-2018-0250.1

Ali, M., Weihua, B., Kaikai, Y. and Panni, M.S.K. (2021): Pore structures and reservoir characteristics of volcanic rocks in the carboniferous batamayineishan formation in the shuangjingzi area, eastern junggar basin (Western china). Rudarsko Geolosko Naftni Zbornik, 36, 105–119. https://doi.org/10.17794/rgn.2021.5.10

Biswas, R., Sen, M.K., Das, V. and Mukerji, T. (2019a): Prestack and poststack inversion using a physics-guided convolutional neural network. Interpretation, 7, SE161–SE174. https://doi.org/10.1190/INT-2018-0236.1

Das, V. and Mukerji, T. (2020): Petrophysical properties prediction from prestack seismic data using convolutional neural networks. Geophysics, 85, N41–N55. https://doi.org/10.1190/geo2019-0650.1

Das, V., Pollack, A., Wollner, U. and Mukerji, T. (2019): Convolutional Neural Network for Seismic Impedance Inversion. Geophysics. https://doi.org/10.1190/geo2018-0838.1

De Jager, Jan. and Geluk, M.C. (2007): Petroleum geology. In: Wong, Th.E., Batjes, D.A.J., De Jager, J (eds.): Geology of Netherlands. -Royal Netherlands Academy of Arts and Sciences, pp. 241–264.

Feng, R. (2020a): Estimation of reservoir porosity based on seismic inversion results using deep learning methods. J Nat Gas Sci Eng, 77, 1–9. https://doi.org/10.1016/j.jngse.2020.103270

Feng, R. (2020b): Unsupervised learning elastic rock properties from pre-stack seismic data. Journal of Petroleum Science and Engineering, 192, 1–7. https://doi.org/10.1016/j.petrol.2020.107237

Gholami, A., Amirpour, M., Ansari, H.R., Seyedali, S.M., Semnani, A., Golsanami, N., Heidaryan, E. and Ostadhassan, M. (2022): Porosity prediction from pre-stack seismic data via committee machine with optimized parameters. Journal of Petroleum Science and Engineering, 210, 1–14. https://doi.org/10.1016/j.petrol.2021.110067

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014): Generative Adversarial Networks. Advances in Neural Information Processing System, 3, 2672–2680.

Kingma, D.P. and Ba, J.L. (2015): Adam: A Method for Stochastic Optimization, In: 3rd International Conference on Learning Representations, ICLR 2015. pp. 1–15.

Kushwaha, P.K., Maurya, S.P., Rai, P. and Singh, N.P. (2021): Estimation of subsurface rock properties from seismic inversion and geo-statistical methods over F3-block, Netherland. Exploration Geophysics, 52, 258–272. https://doi.org/10.1080/08123985.2020.1815528

Kushwaha, P.K., Maurya, S.P., Rai, P. and Singh, N.P. (2020): Porosity prediction from offshore seismic data of F3 Block, the Netherlands using multi-layer feed-forward neural network. Current Science, 119, 1652–1662. https://doi.org/10.18520/cs/v119/i10/1652-1662

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W. and Jackel, L.D. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation, 1, 541–551. https://doi.org/10.1162/neco.1989.1.4.541

Mojeddifar, S., Kamali, G. and Ranjbar, H. (2015): Porosity prediction from seismic inversion of a similarity attribute based on a pseudo-forward equation (PFE): a case study from the North Sea Basin, Netherlands. Petroleum Science, 12, 428–442. https://doi.org/10.1007/s12182-015-0043-8

Rondeel, H.E., Batjes, D.A.J. and Nieuwenhuijs, W.H. (1996): Geology of Gas and Oil under the Netherlands, Geology of Gas and Oil under the Netherlands. Kluwer Academic, Dordrecht. https://doi.org/10.1007/978-94-009-0121-6

Russell, B.H., Gray, D. and Hampson, D.P. (2011): Linearized AVO and poroelasticity. Geophysics, 73, C19-29. https://doi.org/https://doi.org/10.1190/1.3555082

Sang, K.H., Yin, X.Y. and Zhang, F.C. (2021): Machine learning seismic reservoir prediction method based on virtual sample generation. Petroleum Science, 18, 1662–1674. https://doi.org/10.1016/j.petsci.2021.09.034

Schön, J.H. (2011): Physical Properties of Rocks, Vol. 8: A Workbook (Handbook of Petroleum Exploration and Production), Vasa. Elsevier, The Netherlands.

Sorensen, J.C., Gregersen, U., Breiner, M. and Michelsen, O. (1997): High-frequency sequence stratigraphy of Upper Cenozoic deposits in the central and southeastern North Sea areas, Marine and Petroleum Geology.

Wadas, S.H. and von Hartmann, H. (2022): Porosity estimation of a geothermal carbonate reservoir in the German Molasse Basin based on seismic amplitude inversion. Geothermal Energy, 10, 1–40. https://doi.org/10.1186/s40517-022-00223-5

Wang, J. and Cao, J. (2022): Deep Learning Reservoir Porosity Prediction Using Integrated Neural Network. Arabian Journal for Science and Engineering, 47, 11313–11327. https://doi.org/10.1007/s13369-021-06080-x

Wang, L., Meng, D. and Wu, B. (2021): Seismic inversion via closed-loop fully convolutional residual network and transfer learning. Geophysics, 86, R671–R683. https://doi.org/10.1190/geo2020-0297.1

Wang, Y., Ge, Q., Lu, W. and Yan, X. (2020): Well-Logging Constrained Seismic Inversion Based on Closed-Loop Convolutional Neural Network. IEEE Transactions on Geoscience and Remote Sensing, 58, 5564–5574. https://doi.org/10.1109/TGRS.2020.2967344

Wang, Y.Q., Wang, Q., Lu, W.K., Ge, Q. and Yan, X.F. (2022): Seismic impedance inversion based on cycle-consistent generative adversarial network. Petroleum Science, 19, 147–161. https://doi.org/10.1016/j.petsci.2021.09.038

Yasin, Q., Sohail, G.M., Khalid, P., Baklouti, S. and Du, Q. (2021): Application of machine learning tool to predict the porosity of clastic depositional system, Indus Basin, Pakistan. Journal of Petroleum Science and Engineering, 197, 1–20. https://doi.org/10.1016/j.petrol.2020.107975

Zhang, H., Zhang, G., Gao, J., Li, S., Zhang, J. and Zhu, Z. (2022): Seismic impedance inversion based on geophysical-guided cycle-consistent generative adversarial networks. Journal of Petroleum Science and Engineering, 218. https://doi.org/10.1016/j.petrol.2022.111003

Zhu, J.-Y., Park, T., Isola, P. and Efros, A.A. (2017): Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, in: 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, pp. 2242–2251. https://doi.org/10.1109/ICCV.2017.244

Ziegler, P.A. (1992): North Sea Rift System. Tectonophysics 208, 55–75. https://doi.org/10.1016/0040-1951(92)90336-5

SAŽETAK

Usporedba predviđanja poroznosti iz seizmičkih podataka u bloku F3, Nizozemska, korištenjem strojnoga učenja

Poroznost je ključni aspekt karakterizacije ležišta. Može se procijeniti temeljem ispitivanja uzoraka stijena u laboratoriju ili neizravnim metodama temeljenim na karotažnim podatcima. Međutim, oba navedena pristupa dugotrajna su i ne pokrivaju velike površine. Stoga ovo istraživanje objedinjuje seizmičke i karotažne podatke u svrhu procjene poroznosti omogućujući na taj način obuhvaćanje većega područja. U ovome istraživanju procijenjena je poroznost u bloku F3 u Nizozemskoj za koji su dostupni ograničeni karotažni podatci iz bušotina. Kako bi se riješio taj problem, u ovome su istraživanju korišteni inverzijski generator (engl. *inversion generator*) i prediktivni generator (engl. *forward generator*). Inverzijski generator olakšava proces pretvorbe seizmičkih podataka u poroznost, dok prediktivni generator omogućuje modeliranje pretvarajući podatke o poroznosti natrag u seizmičke podatke. Prediktivni generator primjenjuje znanje iz geofizike kako bi nadomjestio nedostatak potrebnih podataka tijekom procesa obuke. Oba generatorska modela koriste se kombinacijom konvolucionarne neuronske mreže i mreže zatvorenih rekurentnih jedinica (engl. *convolutional neural network-gated recurrent unit network* (CNN-GRU)). Dodatno, u ovome su istraživanju korištene različite sheme obuke, kao što su zatvorena petlja (engl. *closed-loop*) i ciklički GAN (engl. *Cycle-Generative Adversarial Network, Cycle-GAN*), koje su uspoređene s pristupom otvorene petlje. Podatci iz bloka F3 sastoje se od šest naknadno skupljenih (*post-stack*) seizmičkih linija s međulinijskim razmakom od 40 metara i triju bušotina. Podatci iz dviju bušotina (F03-4 i F02-1) korišteni su za treniranje modela, dok su podatci iz bušotine F06-1 korišteni za validaciju. Rezultati pokazuju da ciklički GAN daje najtočnije procjene poroznosti sa srednjom kvadratnom pogreškom (MSE) od 0,603 i Pearsonovim koeficijentom korelacije (PCC) od 0,713. Ciklički GAN primjenjuje gubitak cikličke konzistencije za poboljšanje ažuriranja parametara i diskriminator za olakšavanje konkurentnoga učenja s generatorom, čime se povećava točnost. To pokazuje da ciklički GAN može postići točniju procjenu poroznosti na lokacijama s nedovoljnim brojem označenih podataka ili nedovoljno karotažnih podataka u usporedbi s otvorenim i zatvorenim petljama. Međutim, ciklički GAN zahtijeva više računalnoga vremena od ostalih metoda zbog potrebe za obukom više mreža. S druge strane, navedeno dodatno vrijeme obuke opravdano je poboljšanjem Pearsonova koeficijenta korelacije (PCC).

Ključne riječi:
strojno učenje, poroznost, zatvorena petlja, ciklički GAN

Author's contribution

Urip Nurwijayanto Prabowo (1) (M.Sc, Physics, expertise in microtremor and machine learning) provided the Python code, data analysis, interpretation, and writing the original draft and editing. **Sudarmaji (2)** (PhD, Physics, expertise in seismic exploration and machine learning) performed the data analysis, data collection, editing draft, and supervision. **Jarot Setyowiyoto (3)** (PhD, Geological Engineering, expertise in Petroleum Geology) provided the data interpretation and supervision. **Sismanto (4)** (PhD, Professor, expertise in Seismic Exploration and Rock Physics) performed the data interpretation and supervision.