# Efficient proxy for Time-Lapse Seismic ForwardModeling using U-Net Encoder-Decoder Approach

**Michael Macedo Diniz** ( michael.diniz@ifsp.edu.br )
Federal Institute of São Paulo

**Masoud Maleki**
Universidade Estadual de Campinas

**Marcos Cirne**
Universidade Estadual de Campinas

**Shahram Danaei**
Universidade Estadual de Campinas

**João Oliveira**
Universidade Estadual de Campinas

**Denis José Schiozer**
Universidade Estadual de Campinas

**Alessandra Davolio**
Universidade Estadual de Campinas

**Anderson Rocha**
Universidade Estadual de Campinas

---

**Research Article**

**Keywords:** Time-Lapse Seismic Forward Modeling, U-Net Model, Explained AI.

**Posted Date:** December 8th, 2023

**DOI:** https://doi.org/10.21203/rs.3.rs-3715365/v1



**Additional Declarations:** No competing interests reported.

# Efficient proxy for Time-Lapse Seismic Forward Modeling using U-Net Encoder-Decoder Approach

Michael Diniz[1,4*], Masoud Maleki[2†], Marcos Cirne[1†], Shahram Danaei[2†], João Oliveira[2†], Denis José Schiozer[2,3†], Alessandra Davolio[2†], Anderson Rocha[1†]

[1*]Artificial Intelligence Lab., Recod.ai, Rua Saturnino de Brito, 573, Campinas, 13083-889, SP, Brazil.
[2]Center of Energy and Petroleum Studies, Universidade Estadual de Campinas (UNICAMP), Organization, Rua Cora Coralina, 350, Campinas, 13083-896, SP, Brazil.
[3]School of Mechanical Engineering, Universidade Estadual de Campinas (UNICAMP), Organization, R. Mendeleyev, 200, Campinas, 13083-860, SP, Brazil.
[4]Federal Institute of São Paulo, IFSP, Rod. Pres. Dutra, Km 145 s/n, São José dos Campos, 12223-201, SP, Brazil.

*Corresponding author(s). E-mail(s): michael.diniz@ifsp.edu.br;
Contributing authors: maleki@cepetro.unicamp.br; marcos.cirne@ic.unicamp.br; danaeishahram@gmail.com; j253220@dac.unicamp.br; denis@unicamp.br; davolio@unicamp.br.com; arrocha@unicamp.br;
†These authors contributed equally to this work.

**Abstract**

The time-lapse seismic (4D seismic) forward modeling provides crucial data for calibrating reservoir models through the reservoir data assimilation method. Unfortunately, conventional 4D seismic forward modeling methodology is time-expensive and entails significant computational resource consumption. To address these drawbacks, in this work, our goal is to develop a proxy model for the 4D seismic forward modeling using a class of machine learning algorithm named U-Net encoder-decoder. We applied the developed proxy model to a benchmark carbonate reservoir using an ensemble of reservoir simulation models from UNISIM IV (a synthetic benchmark based on real data of a Brazilian pre-salt field). Moreover,

we aim to introduce seminal strategies for interpreting the proposed proxy model operation, its outputs, and possible correlations between input and output variables. To achieve this, we trained and tested two versions of U-net-based models and applied methods for explainable AI, such as Grad-CAM and Forward Feature Selection. The experiments showed good results when applied to the test dataset. The R-squared values were in the range of 0.7 to 0.9, showing the efficiency of the proxy model to replace the 4D seismic forward modeling. Additionally, the qualitative analysis helped us identify which input properties and regions of the reservoir are more relevant for the model's inference. These results are a valuable preliminary step toward a robust, explainable machine learning-based proxy forward modeling.

**Keywords:** Time-Lapse Seismic Forward Modeling; U-Net Model; Explained AI.

# 1 Introduction

Machine Learning (ML) in geoscience can be reviewed at least for seven decades [1]. Nevertheless, there has been a noteworthy increase in geoscience research that apply these techniques in recent years. Advantages in reducing computational cost and time are feasible when applying ML methods, a positive achievement for time-consuming methodologies related to time-lapse seismic (or 4D seismic) used for reservoir management, monitoring, and development. Time-lapse seismic consists of data acquisition in different periods along a field life while closely monitoring the corresponding production activities [2]. Variations in seismic signals between acquisitions can provide insights related to fluid saturation and pressure changes that might be a result of field development activities, e.g., production and injection of fluids [3–5]. Thus, information obtained from 4D seismic data can be used for qualitative and quantitative interpretations, the first to infer dynamic changes related to 4D seismic anomalies [6, 7] and the later can be used alongside production data to calibrate reservoir models through data assimilation (or history matching) processes [8, 9]. Although time-lapse seismic data is sparsely sampled in time compared to production data, it presents a densely spatial sampling in the field scale [10]. Thus, uncertainties in the reservoir model related to spatial and temporal distributions can be assessed and reduced through 4D seismic data assimilation [11].

In this context, a crucial step in conducting 4D seismic data assimilation is to align the synthetic seismic response of the reservoir model with the observed 4D seismic data, for an ensemble of reservoir models. The synthetic 4D seismic response can be obtained through 4D seismic forward modeling within two steps [12]. The first is the petro-elastic model (PEM) that connects reservoir parameters such as porosity, water saturation, pressure, rock matrix, and reservoir fluid properties to the elastic responses. The second depicts seismic wave propagations using different modeling approaches such as 1D convolutional and finite difference methods [13].

Traditional forward modeling, and consequentially 4D seismic data assimilation, can be an expensive process that demands computational time and effort, especially

when considering the use of an ensemble of models [8, 14–17]. Therefore, alternatives using proxy models were proposed to circumvent this problem and optimize this process. [18] reviewed some of the main applications of proxy models which are mostly used in the inverse problems to estimate saturation-pressure changes from 4D seismic data. However, the application of proxy models in 4D seismic history matching to replace the traditional 4D seismic forward modeling is limited to some studies. [19] proposed a proxy for the replacement of the traditional approach linking 4D seismic signal and pressure and saturation changes through a second-order Taylor series expansion. [16] [17] used a simplified version of [19] proxy formulation to create the PEM-Proxy and DAI-Proxy for 4D seismic data assimilation using ensemble of models.

Lastly, [18] proposed a new methodology for map-based all-in-one proxy (S4D-Proxy) to fully replace the traditional seismic forward modeling using ML algorithms (Extreme Gradient Booster - XGBoost and Deep Neural Network - DNN). The S4D-Proxy model consists of a ML map-based proxy that can predict the root-mean-square amplitude (dRMS) difference using reservoir properties as input features (porosity, net-to-gross ratio - NTG, and saturation and pressure variations). Furthermore, this proxy was applied to an ensemble of models (200 models) from a deepwater post-salt Brazilian offshore field located in the Campos Basin. The S4D-Proxy provided results with high fidelity in comparison to the ground truth (synthetic seismic data generated from the traditional forward modeling), which were evaluated using the coefficient of determination $R^2$-score. In [18], the optimal outcomes were achieved by employing the DNN architecture. This architecture will therefore serve as the reference baseline for our proposed approaches.

In [18], the authors emphasized that although the S4D-Proxy can provide successful results in a sandstone reservoir case, in more complex contexts such as carbonate reservoirs, e.g., UNISIM-IV benchmark [20], the proxy may not provide good predictions and might require changes related to the ML algorithm structure and training set increase. Carbonate reservoir complexity can be appreciated in the PEM formulation for these rocks, as observed in [21]. In addition, [12] highlighted that applying 4D seismic data assimilation in carbonate reservoirs is limited, even though significant oil reservoirs are in carbonates. They observed that this limitation could partially result from the rock matrix's stiffness, making signals related to saturation changes harder to detect compared to the 4D seismic signals of sandstone reservoirs. Thus, dealing with carbonate reservoirs is challenging for 4D seismic studies and, consequently, for the S4D-Proxy application.

Another challenge in the ML proxy use and application is the interpretability/explainability of the results. Although the results can be satisfactory (high fidelity with the traditional approach), understanding input parameters' effectiveness, contribution, and weight in the prediction still needs to be assessed. This could provide meaningful information to improve the ML proxy models that replace traditional seismic forward modeling.

Considering the UNISIM-IV database, this work proposes a deep learning proxy forward model capable of achieving superior performance compared to previously

proposed techniques. Interestingly, we aim to comprehend how the trained models generate predictions and utilize various ML model interpretation techniques to examine the relationships between input and output data.

To achieve this goal and address the listed challenges, deep learning encoder-decoder methods were employed to simultaneously analyze the eight input maps and identify potential features that could serve as good descriptors for reconstructing the dRMS of each simulation model. Furthermore, by utilizing the Pearson Correlation Coefficient (PCC), Forward Feature Selection (FFS) [22], and Gradient-weighted Class Activation Mapping (Grad-CAM) [23], it was possible to establish efficient strategies for a better understanding of the proposed proxy model's operation. The main contribution of this article could be summarized as follows:

- Introduce an adapted version of the encoder-decoder U-Net model aimed at surpassing prior strategies for proxy forward modeling development.
- Propose a fast and efficient proxy forward method for a carbonate reservoir.
- Employ methods to discern the most pertinent variables during the training of each proxy model.
- Propose an adaptation to the Grad-CAM method for encoder-decoder architectures, thereby facilitating the identification of more influential reservoir regions to dRMS prediction.

The primary advantage of our approach compared to previous ones is that the U-Net can process all input data at once and identify any feature/correlation within this data that may contribute to determining dRMS. Furthermore, our proposed model can generate outputs at any resolution, which is highly relevant to our case, given that the resolution of maps generated by 4D seismic data is higher than those generated through reservoir models. Finally, in contrast to prior proposals, we suggest using model interpretation strategies that provide relevance for each property and reservoir region in dRMS determination, thereby enhancing the reservoir and simulation model's understanding.

The article is organized as follows: in Section 2, we present the methodological strategy proposed in this work. A detailed workflow description illustrated in Figure 1 is provided; Section 3 describes the procedure for applying the proposed methodology considering the UNISIM IV benchmark. Specifically, a description of the UNISIM-IV database is provided, covering its main characteristics, data structuring process, and data preprocessing protocol; moreover, the U-Net encoder-decoder architecture and model training/testing procedures is described in detail; in Section 4, we present the quantitative results regarding the performance of the proposed models. Additionally, we detail qualitative interpretations of the models and their results by analyzing the Pearson Correlation Coefficient and applying the Forward Feature Selection and the Grad-Cam method; technique, in Section 5; we conclude and engage in discussions regarding the obtained results.

## 2 Methodology

The research methodology employed in this article can be delineated into three primary phases, as depicted in the workflow of Figure 1.

The initial phase is primarily focused on data preparation (Section 2.1), comprising two crucial steps: the creation of the database and the execution of the data preprocessing protocol. Upon completing this phase, the data is formatted suitably for utilization in the ML model's development.

The second phase encompasses the development of the encoder-decoder ML model based on a supervised learning process (Section 2.2). Within this stage, data is partitioned for use in the model's training, validation, and testing procedures. This phase also entails the expert's determination of the selected model's architectural configuration, hyperparameters, and the parameters governing the training process. Concluding this phase, we obtain a trained ML Proxy model capable of making predictions and being utilized for data interpretation.

In the third phase, we interpret both results and the model's functionality, employing four distinct and independent interpretation tools (Section 2.3). During this stage, we investigate the ML model's development, assessing its predictive capabilities and capacity to offer qualitative insights into the interplay between the input and output variables of the model.

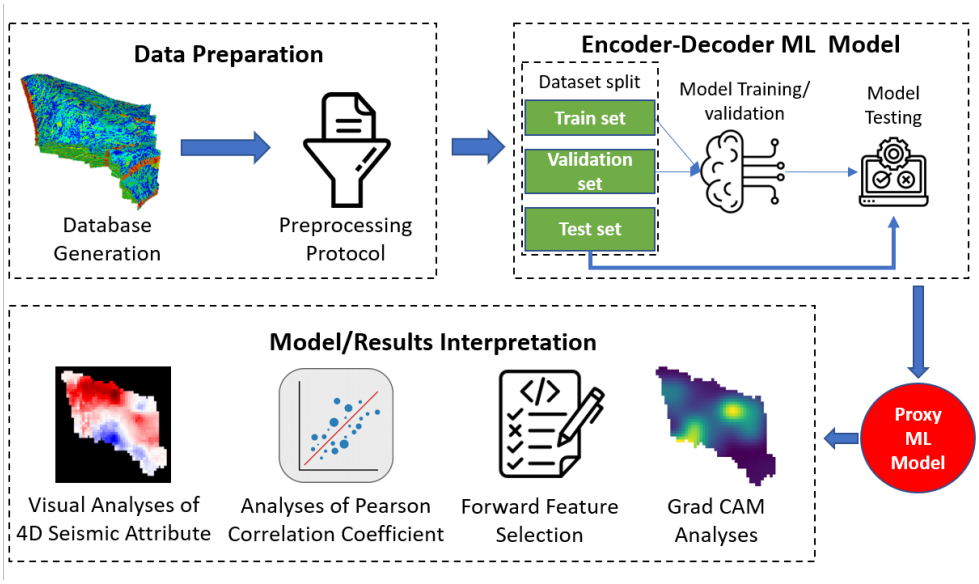



**Fig. 1**: The workflow of the methodology adopted for the development and interpretation of the Proxy Forward Modeling.

## 2.1 Data Preparation

The first stage of the methodology involves organizing the data that will be used to train and evaluate the ML model. For this, averaging maps related to the reservoir properties of each model in the ensemble are generated. The number of averaging maps for each property is defined to appraise the dimensions and heterogeneity of the reservoir. Furthermore, considering the supervised learning process used in the subsequent phase, we must associate a 4D seismic attribute with each generated synthetic model. 4D seismic signals are related to temporal variations of dynamic properties in the reservoir and its usage can enhance the identification of these variations. These 4D seismic attributes can be obtained through real or synthetic data.

In the latter case, it is essential to conduct standard forward seismic modeling using a PEM plus seismic model to generate the dataset associated with the ground truth 4D seismic attribute. Initially, a calibrated PEM is employed to estimate impedance data from simulation models. Subsequently, reflectivities are calculated and convolved with a field-extracted wavelet to generate seismic amplitude volumes. Finally, maps of 4D seismic attributes, such as dRMS maps, are extracted. To streamline the process and avoid the time-consuming preparation of the training and testing dataset, it is recommended to utilize automatic and prompt procedures for computing reservoir properties and synthetic 4D seismic attributes of the simulation models [24].

Typically, the data obtained in generating the database exhibit some inconsistencies (especially in the reservoir edge region) due to the computational/numerical methods involved. Hence, it is imperative to undertake data preprocessing to eliminate any disparities and ensure the consistent and accurate training of the machine learning model. Finally, it is necessary to transform, organize, and store the data in files whose format is compatible with the training framework of the ML model.

## 2.2 Encoder-Decoder Machine Learning Model

The second stage of the methodology involves training, validation, and testing an Encoder-Decoder (ED) ML model. The EDs are deep learning architectures structured from two main parts, as shown in Figure 2: the encoder, responsible for transforming the original input into a latent representation (feature vector), and the decoder, which uses this representation to reconstruct or generate the desired output.

Within the framework of proxy forward modeling, the input comprises a compiled set of property maps originating from simulation models. These maps undergo processing and encoding through the encoder to yield a latent feature vector. Subsequently, utilizing the latent feature vector, the decoder reconstructs the 4D seismic attribute corresponding to the initially processed inputs.

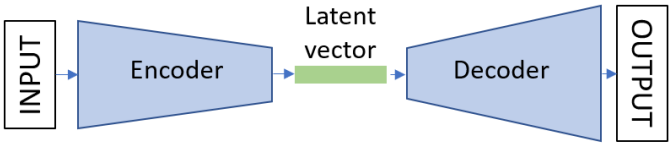



**Fig. 2**: Encoder-Decoder diagram.

One of the advantages of using ED architectures is that the dimensions of the input variable do not need to be the same as the dimensions of the output variables. In the context of forward modeling, this is highly relevant as the resolution of input property maps is generally different (lower) than the resolution of 4D seismic attribute maps. Therefore, in the model development stage, we consider working with two versions of the ED architecture. In the first version, denoted simply by "ED", the output seismic attribute maps' resolution will equal the input maps' resolution. In this case, resizing the 4D seismic attribute maps is usually necessary during the data preprocessing stage. This version is necessary to enable the performance comparison with other machine learning methods that require equality of resolution between input and output. In the second version, the 4D seismic attribute maps will be reconstructed by the decoder at their original resolution. This version of the architecture is denoted as "ED-HighRes".

The first step in the model development process is to randomly divide the preprocessed dataset into three disjoint subsets: training, validation, and test sets. Despite the randomness of the data split, it is performed only once, ensuring comparability across subsequent experiments

The ML model's structure and hyperparameters are defined through experiments conducted using the training and validation sets. Once the optimal setup has been defined, the ML model is trained based on the training and validation set. Finally, the model is evaluated on the test set to determine the performance of the trained model on data not used in the training process. The model's performance is quantitatively defined using metrics such as $R^2$ score and SMAPE, as shown in the next section.

### 2.2.1 Performance Metrics

To assess the performance of the models, two metrics are used, the well-known $R^2$ score and the Symmetric Mean Absolute Percentage Error - SMAPE. The $R^2$ score and SMAPE are given by Equation 1

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \overline{y})^2} \qquad SMAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{(|\hat{y}_i| + |y_i|)/2} \tag{1}$$

where $y_i$ is a observed data, $\hat{y}_i$ is a predicted data, and $\overline{y}$ is the observed mean value. Considering the scatter plot between observed and predicted data, $R^2$ measures how

closely the points on the graph align with the bisector of the first quadrant. This metric provides an absolute approximation between the observed and predicted data.

The sum argument in the SMAPE equation represents the absolute error percentage relative to the mean between the predicted and observed values. SMAPE is a measure of relative error, where the closer the observed value is to zero, the more sensitive the absolute error becomes in this measure. The lower the SMAPE value, the better the model's performance evaluation.

## 2.3 Model/Results Interpretation

The most natural interpretation of the model results is achieved through the visual inspection of the outputs generated by the model and its respective ground truth. This step provides the spatial distribution of the model's performance, highlighting the aspects of the 4D seismic attribute in which the model faces more difficulty/ease in making predictions.

In recent years, machine learning models have been used for inference and interpreting the relationship between variables [25]. Furthermore, with the new Explainable AI tools [26], it is possible to enhance the understanding of how the model operates and elucidate how it relates to input and output.

This article explores the relationship between the input and output variables in two ways. In the first approach, we seek to quantify the importance of each input properties in predicting 4D seismic attribute. To do this, the Pearson Correlation Coefficient (PCC) [27] and the Forward Feature Selection (FFS) [22] technique are used. In the second approach, we apply a method called Gradient-weighted Class Activation Mapping (Grad-CAM) [23] to identify which reservoir regions were most heavily relied upon in the model's inference process.

### 2.3.1 Pearson Correlation Coefficient

The most common way to measure the degree of correlation between input and output variables is by calculating the Pearson Correlation Coefficient (PCC) ($\rho$) between them [27]. The $\rho$ formula is given by Equation 2, as follows

$$\rho = \frac{n \sum_{k \in \mathbb{I}} x_k y_k - \left( \sum_{k \in \mathbb{I}} x_k \right) \left( \sum_{k \in \mathbb{I}} y_k \right)}{\sqrt{\left( n \sum_{k \in \mathbb{I}} x_k^2 - \left( \sum_{k \in \mathbb{I}} x_k \right)^2 \right) \left( n \sum_{k \in \mathbb{I}} y_k^2 - \left( \sum_{k \in \mathbb{I}} y_k \right)^2 \right)}}, \tag{2}$$

where $\mathbb{I}$ is the set of grid cell map coordinates, $n$ is the number of grid cells, $x_k$ and $y_k$ are the values of the properties and 4d seismic attribute at coordinate $k \in \mathbb{I}$, respectively.

It is important to highlight that, in our context, a high correlation value ($|\rho| \approx 1$) indicates a local dependence between input and output, meaning that a variation in a specific coordinate in the input map implies a proportional variation in the same coordinate in the output map. This interpretation is thoroughly discussed in Section 4.2.1.

### 2.3.2 Forward Feature Selection

U-Net-based models are capable of identifying complex dependencies between input and output. A widely used technique for identifying the most important variables of a model is Forward Feature Selection (FFS) [22]. This technique involves training the model using one feature at a time and identifying the one that leads the experiment to better performance.

Through this technique, we can identify which property inherently holds more relevant information for predicting the 4D seismic attribute map using the encoder-decoder-based models. Considering the encoder-decoder's operational nature, these pieces of information stem from the spatial distribution of property values throughout the reservoir.

### 2.3.3 Gradient-weighted Class Activation Mapping (Grad-CAM)

Deep learning models are commonly characterized as black box systems due to their inherent lack of interpretability in elucidating the transformation of input data into output inferences. In contrast to models grounded in physical principles, the parameters within deep learning architectures lack discernible qualitative significance, thereby confining their utility primarily to inference tasks.

In recent years, there has been a notable surge in endeavors aimed at augmenting the interpretive capacity of models and elucidating their mechanisms for processing input data. For instance, when considering models that take images as input, a relevant qualitative aspect is to identify which region of the image is most crucial for determining the model's response. In this regard, various methods such as Local Binary Pattern (LBP) [28], Shapley Additive Explanations (SHAP) [29], Saliency Maps [30], and Gradient-weighted Class Activation Mapping (Grad-CAM) [23] were proposed. Based on the U-Net architecture, the Grad-CAM method was chosen for our model interpretation.

Typically, Grad-CAM is employed in the context of neural architectures designed for image input and output probabilities or scalar values. The architectural representation corresponding to this category is succinctly depicted in the left diagram of Figure 3.

The Grad-CAM heatmap is derived as a weighted summation of feature maps generated by the final convolutional layer within the feature extraction process. Determining these weights involves computing gradients for each feature map about a specific target output neuron, yielding a set of gradient maps that correspond to the final layer's feature maps. The computation of weight factors for the aggregation process employs the Global Average Pooling (GAP) technique applied to each gradient map. The resulting weighted average generates a singular two-dimensional map that should be resized to match the input resolution.

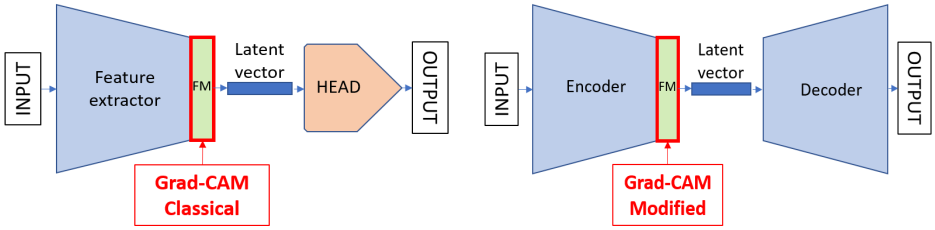



**Fig. 3**: In the conventional approach (left diagram), Grad-CAM is typically employed on the final Feature Maps (FM) of the feature extractor, commonly used for regression or classification tasks. In our study, we customized Grad-CAM to operate within an encoder-decoder framework (right diagram), where it is applied to the ultimate FM of the encoder.

The Grad-CAM method was adapted in this work to be applied to Encoder-decoder-based models. The weighted average of the feature maps was not performed in the final layer of the architecture but rather in the last layer of the encoder (see Figure 3). Another adaptation is that our problem is a multitask regression case, so the Grad-CAM representing all the outputs was calculated through the average of the Grad-CAMs computed for each grid cell. To use Grad-CAM in a multitask problem like ours, the average of the Grad-CAMs obtained for each output grid cell was calculated.

To the best of our knowledge, this type of adaptation of Grad-CAM to explain an encoder-decoder is unprecedented and stands as one of the contributions of this work.

# 3 Application on UNISIM IV database

In this section, we elucidate the procedural application of the previously introduced methodology considering UNISIM-IV database.

## 3.1 UNISIM-IV and Data Preparation

This section includes a detailed description of the UNISIM-IV development process, highlighting its key features and the preprocessing protocol implemented to train ML models.

### 3.1.1 UNISIM-IV Description

UNISIM-IV is an open-source benchmark case based on a light-oil carbonate Brazilian pre-salt field with a high oil-gas ratio and CO2 content [20]. The benchmark comprises a reference model representing the true earth model and an ensemble of models. Mimicking what happens in a real field, the reference model is used only to extract field-measured data, such as production history, well logs, and 4D seismic data. The "measured" data are used to build the ensemble of uncertain models.

The reference model has cell dimensions of 200 x 200 x 5 m and 47 x 39 x 291 blocks (i, j, and k directions). The model has 533403 blocks in total, whereas 77004 are

active. In the first phase of the production strategy, the model considered six vertical producers and seven vertical WAG-CO2 injectors (Figure 4b). The base seismic data was generated at 608 days of production and the monitor seismic data at 1583 days of production. The history period starts at 10/2018 [20].

The ensemble of models consists of 100 uncertain dynamic models. As also highlighted by [18], it is noteworthy that the ensemble of models used here were not submitted to any data assimilation process. To prepare the data used to train the proxy model, three average maps of each property (porosity, NTG, and saturation/-pressure changes) were generated to more accurately appraise the dimension of the reservoir. Each map presented a resolution of 46x46 cells. The averaging was also performed between four horizons that considered geological heterogeneities of the reservoir (Figure 4a). Horizon-1 to Horizon-70 (Map1) is considered a first set of microbialites at the top of the reservoir. Horizon-70 to Horizon-200 (Map2) is considered a second set of microbialites. Horizon-200 to Horizon-291 (Map3) is considered a set of coquinas at the base of the reservoir.

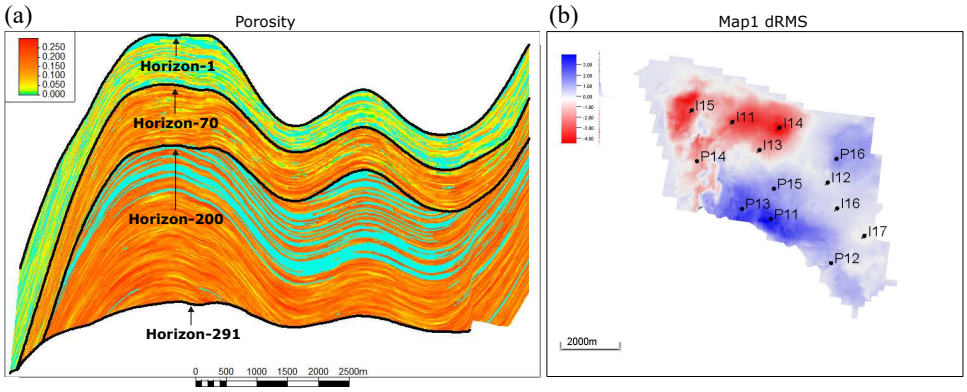



**Fig. 4**: (a) Porosity of an inline section in the reservoir. The four horizons used to build the average maps are displayed. (b) Map1 dRMS for Model 96 of the ensemble. dRMS prior to preprocessing of data preparation. Well locations for the first phase of production are displayed on the map.

| Property | Nº of models per map | Number of Maps | Resolution |
|---|---|---|---|
| dRMS (output) | 100 | 3 | 94 x 78 |
| Variation of Pressure ($\Delta$P) | 100 | 3 | 46 x46 |
| Variation of gas saturation ($\Delta$Sg) | 100 | 3 | 46 x46 |
| Variation of water saturation ($\Delta$Sw) | 100 | 3 | 46x 46 |
| Net-to-Gross (NTG) | 100 | 3 | 46 x46 |
| Porosity (POR) | 100 | 3 | 46 x46 |
| Pressure base (Pr_Base) | 100 | 3 | 46 x46 |
| Gas saturation base (Sg_Base) | 100 | 3 | 46 x46 |
| Water saturation base (Sw_Base) | 100 | 3 | 46 x46 |

**Table 1**: Summary of the UNISIM-IV database.

Once the reservoir property maps were generated, generating the synthetic dRMS of the models was also necessary using the traditional forward modeling process. The dRMS is necessary to train the models and adjust the parameters and hyper-parameters during the generation of the proxy model. The proxy evaluation is also done using the dRMS from the evaluation set of models (ground truth) generated by the traditional process.

### 3.1.2 Data preprocessing

The input data is rearranged for each simulated model in a tensor (Input Block) with an 8x46x46 shape, where the first dimension is related to each input property. Moreover, the latter two regard the resolution of each input map. The dRMS resolution is 94x78, which differs from the resolution of the input maps (46x46). For the application of certain ML methods, including the DNN proposed by [18], resizing the dRMS to a resolution of 46x46 is necessary. As a consequence, for Map1, Map2, and Map3, the number of points in each dRMS map reduced from 3858 to 881 (22.83%), 3624 to 804 (22,18%), and 2671 to 676 (25,31%) respectively. In a qualitative sense, this process involves the loss of some significant information. These issues are discussed in detail in Section 4.1.

With this in mind, encoder-decoder-based models can produce output in any desired dimension. Consequently, as established in the methodological process (Section 2.2), we work with two encoder-decoder model versions, ED and ED-HighRes, that produce output dRMS maps of 46x46 and 94x76 resolution, respectively. Therefore, the first preprocessing step is to resize the dRMS maps to be used in the training process of the ED model.

The second step is to exclude data inconsistencies according to the following conditions. For each Map described in Section 3.1 and for a specific grid cell coordinate:

- If one of the models is missing data, we disregard the data from that coordinate from all model properties, including dRMS for the resized case.
- NTG, water saturation, and gas saturation must be between 0 and 1. Otherwise, we disregard the data from that coordinate from all model properties including dRMS for the resized case.

After executing the preprocessing steps, the data is ready for the ML model training process. Figure 5 provides an example of a graphical representation, via heatmap, of property maps derived from simulation models and their respective dRMS in both their original and resized dimensions.

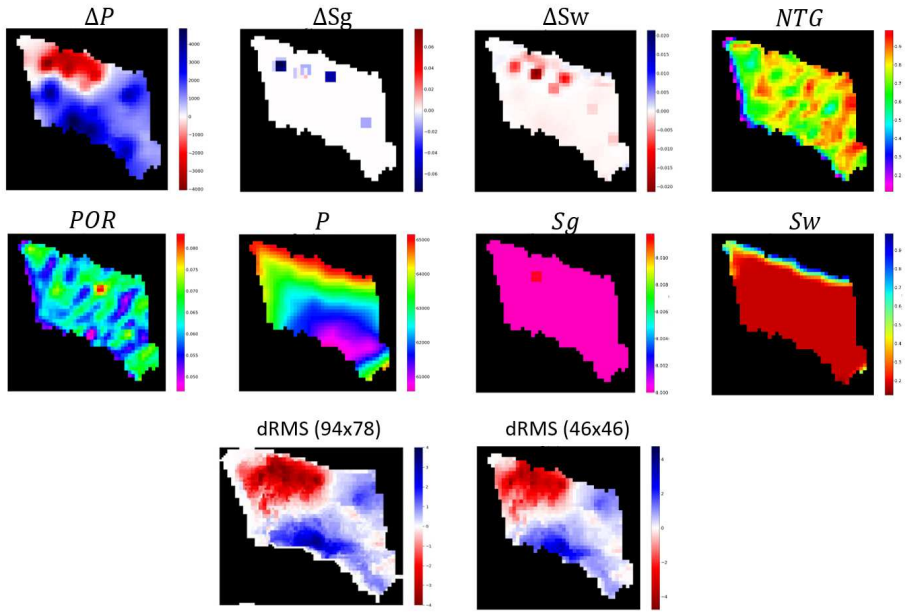



**Fig. 5**: Heatmap of preprocessed input properties and output signals from simulation model 79 of Map1.

## 3.2 U-Net model

In the literature, there are numerous variations of architectures defined as encoder-decoder. For our application context, the architecture used as the encoder-decoder is the U-Net. It is a U-shaped encoder-decoder neural network architecture proposed in [31]. This architecture is widely used for tasks such as semantic segmentation, image resolution enhancement, synthetic image generation, among others [32]. Typically, the U-Net is trained to model functions that take images as input and generate another image as output; however, for this study, we adapted the architecture to take the Input Blocks (Section 3.1.2) as input and generate the dRMS maps as output. Therefore, our architecture addresses the problem through a multitask regression approach.

### 3.2.1 The architecture description

We denote as "U-Net46" the architecture that plays the role of the ED and provides output in resized dimension (46x46). Along the same line, we denote as "U-Net94" the architecture that acts as the ED-HighRes and generates output in the original resolution (94x78), as outlined in the methodological process. The architectures of U-Net46 and U-Net94 are summarized in Figures 6 and 7, respectively.

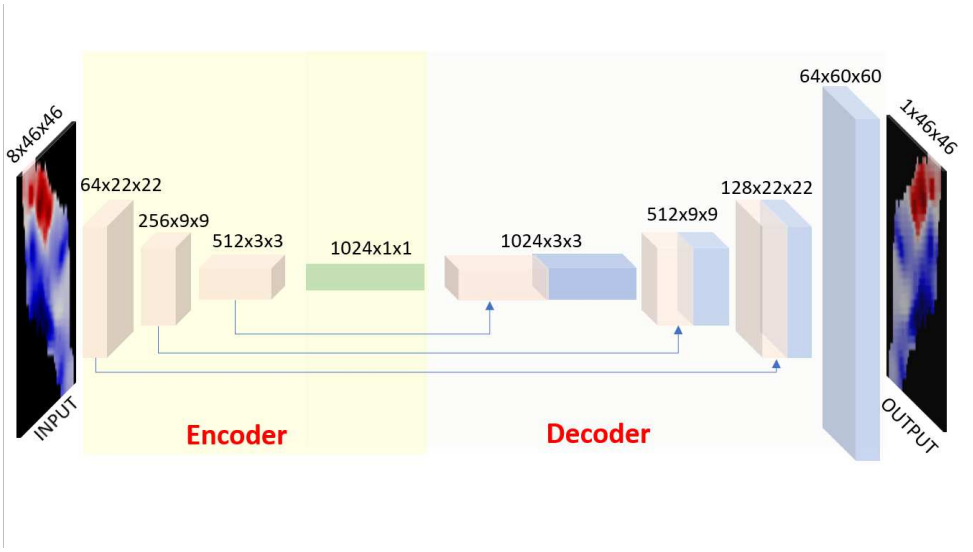



**Fig. 6**: U-Net46 architecture: On the last feature map volume, a convolutional operation with stride 1 is applied, and the resulting map is cropped at the center in 46x46 shape.

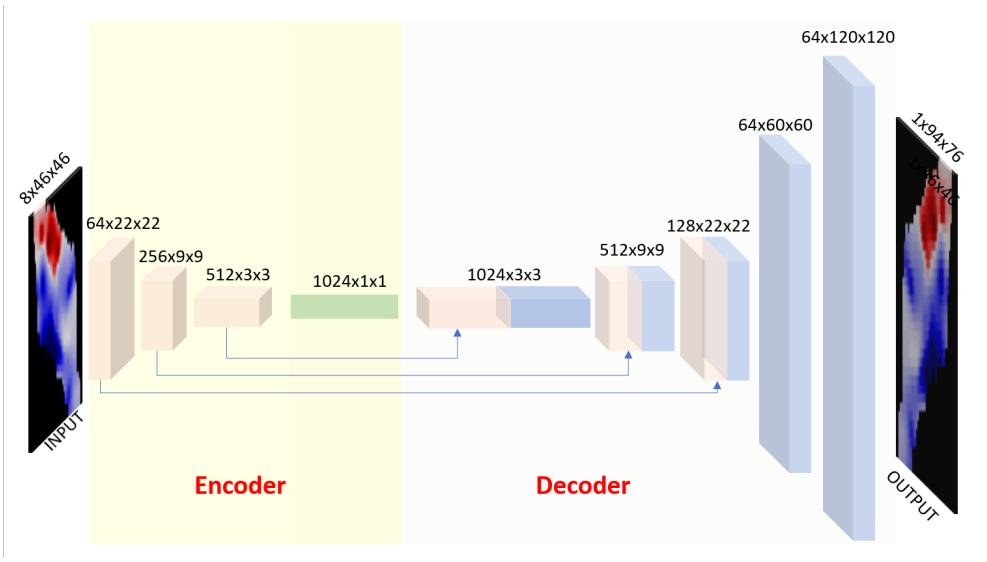



**Fig. 7**: U-Net94 Architecture: On the last feature map volume, a convolutional operation with stride 1 is applied and the resulting map is cropped at the center in 94x78 shape.

The encoder part is the same for both architectures and is responsible for extracting features from the input maps and summarizing them into a reduced-dimension vector, called Latent Vector. Between the feature maps volumes, convolutional and max pooling operations are applied.

The decoder part is responsible for reconstructing the desired output map from the Latent Vector generated by the encoder. The Latent Vector is upsampled through transposed convolution operation. Each decoder feature map is concatenated for both architectures with the same dimensional feature map from the encoder. The objective is to leverage features from all complexity levels of the encoder to reconstruct the decoder output.

### 3.2.2 U-Net Training

The U-Net-based models were trained and tested individually for each of the three maps. For each map, out of the 100 available simulation models, 70% were allocated for training, 10% for validation, and 20% for testing. The Adam's optimizer was used with $10^{-3}$ of learning rate. As a minimization criterion, we employed the Huberloss function.

The models were trained for 500 epochs in all training sessions, achieving the lowest validation loss used for testing. In the case of U-Net46, the lowest validation loss was observed at epochs 131, 40, and 102 for each map, respectively. For U-Net94, the lowest validation loss occurred at epochs 189, 201, and 220 for each map. All training sessions were conducted on a GPU and lasted approximately one hour each.

# 4 Experiments and Results

In this section, we present the quantitative and qualitative outcomes pertaining to the proposed models' training, validation, and testing processes. We also showcase the analyses resulting from deploying machine learning model interpretation tools. Given that the Deep Neural Network (DNN) achieved the most favorable performance in [18], we employ it as the benchmark and compare it with our novel propositions.

## 4.1 Quantitative and Visual Analyses

The U-Net46, U-Net94 and DNN models were trained and tested for each map. Figure 8 presents the boxplots of the evaluated models.

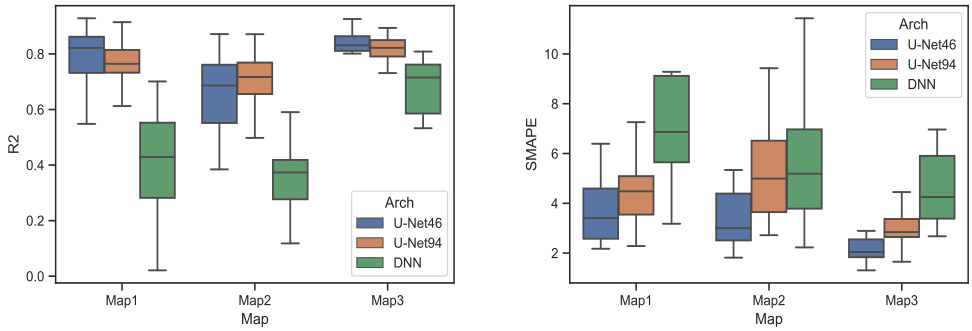



**Fig. 8**: Boxplots of performance metrics ($R^2$-score on the left and SMAPE on the right) for the test set's U-Net46, U-Net94 and DNN models.

To complement the information provided in the boxplots of Figure 8, the Wilcoxson hypothesis test [33] was applied. This is a non-parametric statistical hypothesis test to evaluate whether two groups have different medians. In our context, the Wilcoxon test is used to assess evidence to conclude the superiority of the methods based on U-net and the performance indifference between U-Net46 and U-net94. Table 2 shows the p-values of the conducted tests.

| Models | MAP 1 | | MAP 2 | | MAP 3 | |
|---|---|---|---|---|---|---|
| | $R^2$ | SMAPE | $R^2$ | SMAPE | $R^2$ | SMAPE |
| DNN × U-Net46 | $< 10^{-4}$ | $< 10^{-4}$ | $< 10^{-4}$ | $< 10^{-4}$ | $< 10^{-4}$ | $< 10^{-4}$ |
| DNN × U-Net94 | $< 10^{-4}$ | $< 10^{-4}$ | $< 10^{-4}$ | 0.475 | $< 10^{-4}$ | $< 10^{-4}$ |
| U-Net46 × U-Net94 | 0.294 | 0.00271 | 0.083 | $< 10^{-4}$ | 0.522 | $< 10^{-4}$ |

**Table 2**: The p-values of the Wilcoxon test. The three evaluated models were compared pairwise regarding $R^2$ and SMAPE. The p-value indicates the probability of an error when rejecting the null hypothesis.

Considering the $R^2$-score, the boxplots in the left graph on Figure 8 and the p-values of the Wilcoxson test on Table 2 indicate that, for the three maps, U-net-based methods outperform the DNN method. Furthermore, no statistical evidence indicates a difference between the performance of the U-Net46 and U-Net94 methods. In all cases, 0.01 significance level is considered. In other words, concerning absolute error (refer to Section 2.2.1), U-Net-based networks outperformed the baseline in all the examined scenarios.

Considering the SMAPE metric, the boxplots in the right graph in Figure 8 and the p-values of the Wilcoxson test in Table 2 indicate that U-Net46 outperformed DNN in all three maps. Moreover, U-Net94 outperformed DNN in Map1 and Map3 but, in Map2, despite the slight difference between the medians and variability, no evidence was found at the adopted significance level to confirm the superiority of U-Net94 compared to the DNN. Finally, in all the maps, considering the level of

significance of the adopted test, statistical evidence demonstrated the superiority of U-Net46 compared to U-Net94 in terms of SMAPE metric.

It is important to remark that the SMAPE metric is sensitive to values close to zero; therefore, since the output of U-net94 has more values close to zero than the output of U-net46 and DNN, it was expected that U-net94 would have more difficulty in performance considering the SMAPE metric.

A visual comparison between the predicted and observed maps is provided in Figure 9 for a more robust and qualitative evaluation of the model performance. The 4D signals displayed in the dRMS maps of Figure 9 are related to hardening effects in blue and softening effects in red. The dRMS considered here is $dRMS = RMS(base) – RMS(monitor)$, to appraise a soft reservoir in comparison to their surroundings [3].

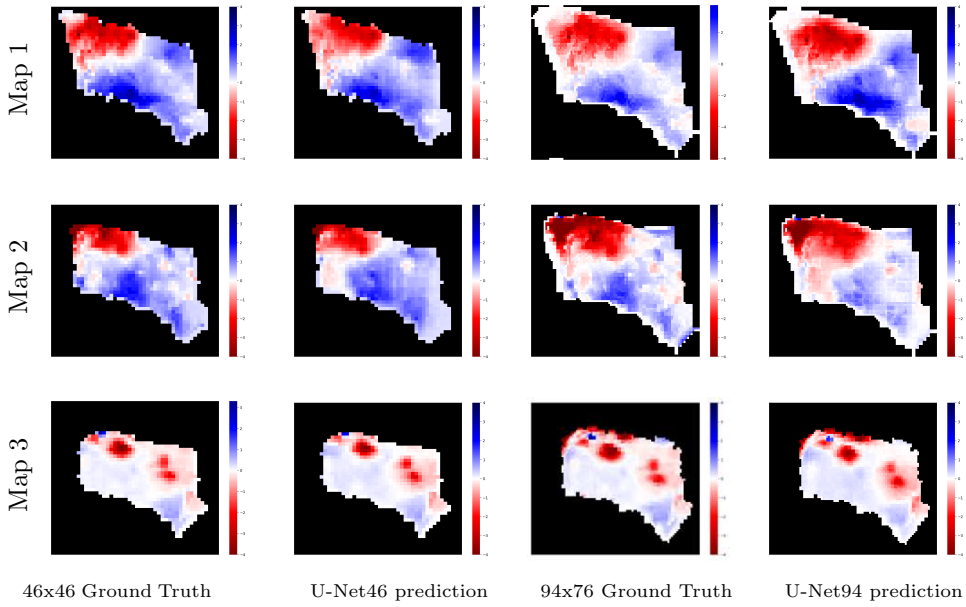



**Fig. 9**: Plots of the observed and predicted dRMS. The dRMS for Map1 and Map2 are related to the Test Model 79, and the dRMS for Map3 are related to the Test Model 3. 4D seismic signal hardening effects are displayed in blue and softening effects in red.

Considering the ground truth maps on Figure 9 (columns 1 and 3), it is visually apparent that certain information/data is lost during the resizing and preprocessing procedures outlined in Section 3.1.2. When comparing the observed dRMS maps with the predicted ones, it becomes evident that, on the whole, the models achieved a commendable level of prediction accuracy. Notably, the greatest challenges are encountered in the dRMS values for Map2, which is consistent with the quantitative analysis (see Figure 8) of model performance.

## 4.2 Model Interpretation

In this section, we conduct a qualitative analysis of the results and the correlation between input and output variables. This analysis aims to deepen our understanding of the proposed models by pinpointing the most pertinent input properties and frequently accessed reservoir areas during the inference process. This represents the first application of such analysis in Proxy Forward Modeling. Leveraging explanatory tools, our goal is to demonstrate that, in addition to their operational efficiency and speed, U-Net-based proxy models can enrich our understanding of spatial relationships and inter-property dynamics within each reservoir.

### 4.2.1 PCC Analysis

To assess the significance of each input property in predicting dRMS, for all test models, we computed the PCC (see Section 2.3.1) between each input property and its corresponding dRMS. The distribution of these values is depicted in the boxplots in Figure 10.

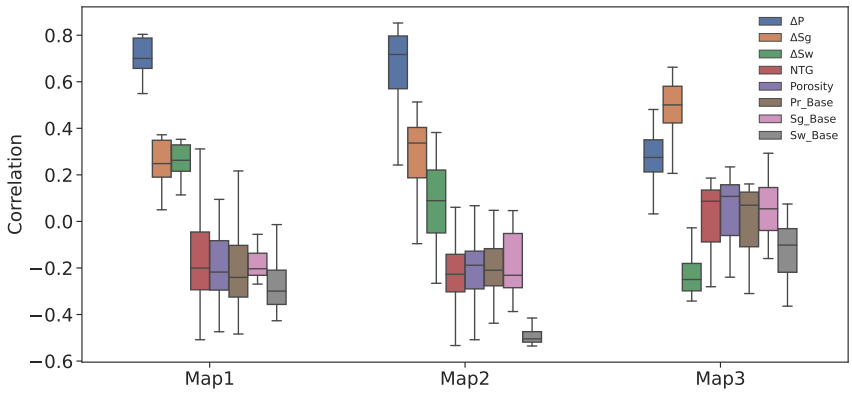
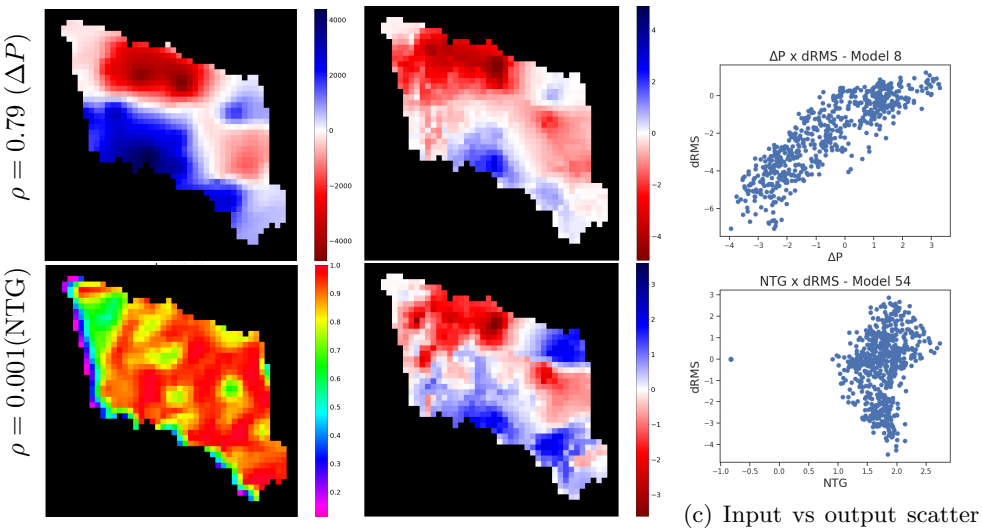



**Fig. 10**: Boxplot of PCC values calculated for each input property and dRMS.

By the graph of Figure 10, the $\Delta P$ is the property most correlated with dRMS in Map1 and Map2. The 4D signals observed in the ground truth maps (Map1 and Map2) of Figure 9 predominantly consist of softening effects (indicated by red signals) in the location of injectors. These signals are likely attributed to increased pore pressure from injection, gas saturation increase, or a combination of both factors. The hardening signals in these maps (depicted by blue) are mainly caused by pore pressure decrease around producers. According to our interpretation, pressure variations - whether a drop or an increase - are important for shaping the 4D signals in Map1 and Map2. This evidence might support our finding in Figure 10 in which $\Delta P$ was the most influential attribute in predicting dRMS.

Conversely, in Map3 of Figure 10, the property most correlated is ΔSg. For our context, the Pearson coefficient measures a local correlation between input properties and dRMS. To illustrate this, we can observe the graphs in Figure 11.




(a) Input. (b) Output. (c) Input vs output scatter plot.

**Fig. 11**: On the top row, we have the ΔP and dRMS maps for Model 8 and their scatter plot. On the bottom row, we have the NTG and dRMS maps for Model 54 and their scatter plot.

The top row of Figure 11 shows the graphs of ΔP and dRMS of Map1 are shown. The scatter plot in the same row represents the values of ΔP and dRMS for each reservoir coordinate. For this case, the PCC is 0.79, indicating a strong correlation between the variables. Notably, there is a visual correspondence between the ΔP and dRMS graphs. Furthermore, the scatter plot reveals a positive linear correspondence between the quantities.

On the other hand, the bottom row shows the graphs of Net-to-Gross and dRMS of Map1, along with the scatter plot of these two variables. In this case, $\rho = 0.001$, indicates the absence of correlation between NTG and dRMS. The scatter plot reinforces this conclusion.

The Pearson coefficient is limited to measuring local correspondences between corresponding regions of variables; however, models like the U-net can extract more complex correspondences between inputs and outputs that cannot be perceived visually or measured by the Pearson coefficient. For this context, we apply the method of the next subsection.

### 4.2.2 Forward Feature Selection

The Forward Feature Selection (FFS) (see Section 2.3.2) method involves training the model with each input variable individually and determining the one with the best performance as the most important for prediction. Figure 12 shows the distribution of the resulting $R^2$ - score from applying the technique.

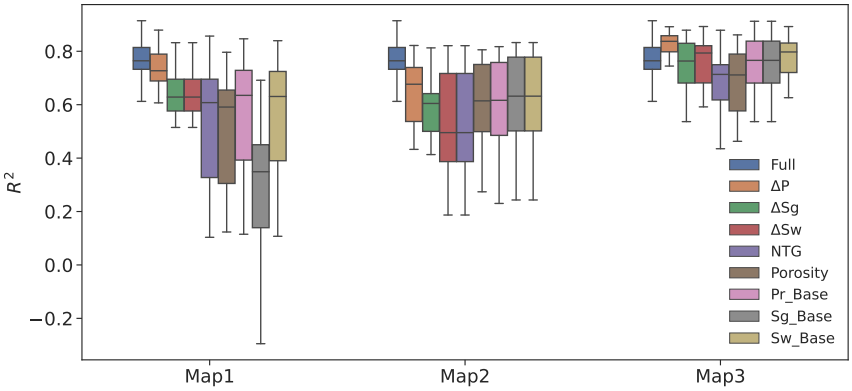
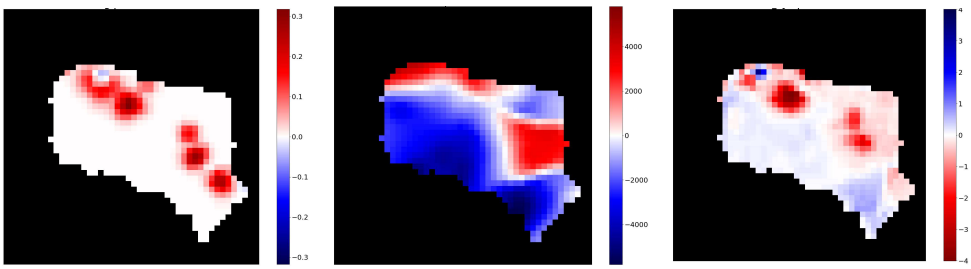



**Fig. 12**: Boxplot of $R^2$ - scores from FFS. For each map, nine models were trained, one called "Full" with all eight input properties and eight others with only one of the available input properties.

For Map1 and Map2, the model trained with all properties outperformed all models trained with a single property input. The property with the best performance was ΔP, but its inferior performance compared to the 'Full' model indicates that combining the other seven properties enhances the model's performance.

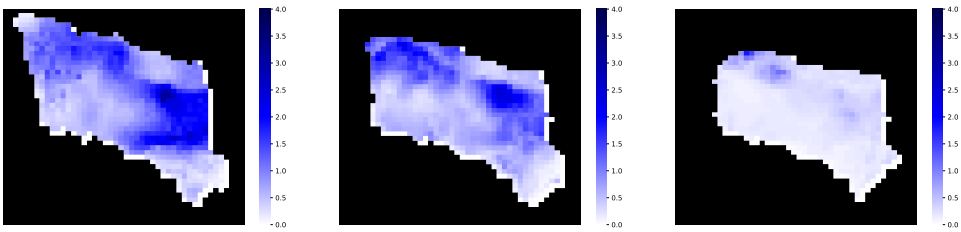
For Map1 and Map2, the property ΔP proved the most relevant in the PCC-based analysis and FFS. This demonstrates that ΔP exhibits local and global correlation with dRMS in these maps. However, in Map3, the property ΔP emerged as the most significant predictor of dRMS through U-net, while ΔSg displayed the highest PCC. We can observe from Figure 13 that there is a visual (local) correlation between ΔSg and dRMS, which is not the case between ΔP and dRMS. This means that, despite ΔSg having a higher local correlation with dRMS, the U-Net can extract more complex correlations from ΔP that result in better dRMS prediction. Furthermore, it can be observed from Figure 13 that ΔSg exhibits a strong local correlation with dRMS at specific locations within the reservoir, albeit not predominantly.



(a) ΔSg (b) ΔP dRMS

**Fig. 13**: Maps of $\Delta Sg$, $\Delta P$ and dRMS of Model 3.

The boxplots in Figure 12 related to Map3 suggest that the model has an easier time predicting the dRMS in this map regardless of the input. Additionally, by analyzing the standard deviation maps of Figure 14, we observe that the dRMS of Map3 varies less among the test samples when compared to Map1 and Map2. Therefore, the dRMS of Map3 is easier to predict compared to the other maps, which justifies the superior performance of the U-net model on this map, considering the quantitative results from the boxplots of Figure 8 (Section 4.1).

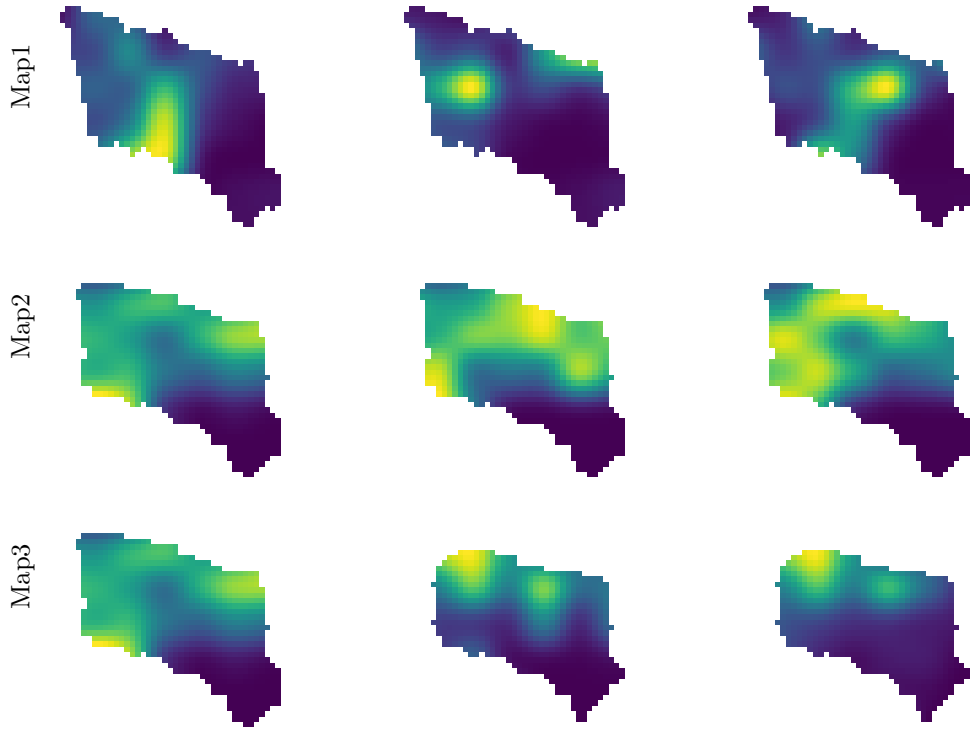


Standard Deviation of Map1. Standard Deviation of Map2. Standard Deviation of Map3.

**Fig. 14**: Standard deviation of dRMS maps.

### 4.2.3 Grad-CAM Analysis

To enhance our comprehension of the U-net's functionality within our specific framework, it is imperative to discern the areas of the reservoir that played a pivotal role in the inference of the dRMS for the test dataset. This inquiry employs the modified Grad-CAM methodology (see Section 2.3.3), and the outcomes of its application are depicted in Figure 15.




(a) Model 87 (b) Model 54 (c) Model 23

**Fig. 15**: Grad-CAM visualizations for Models 87, 54, and 23. The regions with the highest intensity on the visualizations correspond to the activated reservoir areas with the most significant contributions to dRMS prediction.

In Figure 15, it is evident that the most activated regions in Map1 exhibit a more localized pattern, while in Map2 and Map3 the activations are more diffuse, indicating that these maps involve larger combinations of information for inference. It is also worth noting that, in all three maps, the southern region of the reservoir remains inactive during the inference process.

It is also noteworthy that, for each test model, the activated regions in Grad-CAM maps are different, which indicates that the U-net has learned to delineate the most relevant regions according to the specific features of each input information.

It is important to emphasize that the Grad-CAM is an attention heatmap applied to combining the eight input properties. Consequently, specific conclusions about individual properties cannot be drawn from the calculated Grad-CAM. Rather, it provides an overview of regions with heightened activity in the inference of dRMS for each tested model.

In forthcoming studies, we anticipate that Grad-CAM can be examined from various perspectives. For instance, in our upcoming research endeavors, we intend to

explore potential correlations between Grad-CAM maps and the geological characteristics specific to each reservoir region. This investigation promises to enhance our comprehension of the impact of geological factors on determining the dRMS map. We will also investigate techniques for predicting the dRMS based solely on the most highly activated regions within Grad-CAM. In essence, we will evaluate the viability of predicting the complete dRMS map using only a subregion of the input maps, thus reducing computational overhead.

# 5 Conclusion

In this work, an ML proxy model was developed to replace 4D seismic forward modeling. U-Net architecture was trained and tested on the UNISIM-IV database in our approach. The quantitative results indicate that our proposal outperformed the baseline approach. This paper also introduced innovative strategies for interpreting the results and the proxy's operation.

Analyzing the models' performance on the test dataset, the non-parametric Wilcoxon test (Table 2) indicated that, in general, both U-Net46 and U-Net94 outperformed the baseline. The performance of both networks also exhibited similarities in almost all tests. It must be highlighted that this proposed innovative process, the U-Net94, can be a very good solution for real-case applications where maps extracted from 4D seismic data usually have higher spatial resolution than the maps built from reservoir models. This means a great contribution in that we could match reservoir models and 4D seismic data at the resolution of seismic provided maps.

The results based on SMAPE highlight that U-Net46 achieved better results than U-Net94. This is likely due to the sensitivity of the SMAPE metric to values close to zero; the dRMS maps with original resolution contain significantly more zero values than the rescaled ones.

Considering the PCC analysis, we identified the most locally correlated property with dRMS in Map1 and Map2 as $\Delta$P, and in Map3, as $\Delta$Sg. Based on the results of PCC and FFS methods, it becomes evident that the performance of the U-Net model is not solely reliant on the local correlations. U-Net models leverage spatial and inter-property information for their predictions, surpassing the baseline model's performance.

It is also worth noting that, in Map1 and Map2, combining all eight properties yields better results than individual property usage. However, in Map3, the model achieved similar outcomes when utilizing the eight input properties and just $\Delta$P. According to standard deviation analysis, Map3 exhibits minimal output variability, which indicates that this map is less challenging than the other two and justifies that the ML models achieve favorable results by using all properties individually.

By applying Grad-CAM, we identified the key reservoir regions that are crucial for predicting each model. As anticipated, individual samples triggered distinct reservoir activations, yet our observations revealed a concentration of these activations within the central and upper regions of the reservoir. Thus, Grad-CAM analysis suggests that the input property values south of the reservoir do not significantly influence the overall reservoir dRMS.

Based on the promising results of this article, we believe that we can further enhance the performance of the proxy model by leveraging other encoder-decoder architectures already published in the literature.

The interpretability of deep learning models remains a pervasive issue across all applications. In this article, we believe we have taken a significant seminal step towards a better understanding of the operation and outcomes of the 4D seismic forward modeling proxy model. Nevertheless, numerous challenges in terms of interpretability still need to be explored.

Finally, all the analyses conducted herein are confined to synthetic data and still require real-world data assimilation process validation. Furthermore, the proposed methodology can be applied in more general cases and is not restricted to the UNISIM IV context. Using our methodology, other 4D seismic attributes, different databases, and alternative encoder-decoder model architectures can be considered.

## Acknowledgments

This work was carried out in association with the ongoing project registered under ANP number 23202-5 as the "Frequent 4D seismic data acquisition for enhanced reservoir management in Brazilian pre-salt fields – Phase 3" (Universidade Estadual de Campinas [UNICAMP]/Shell Brazil/ANP) funded by Shell Brasil Petróleo Ltda. under the ANP R&D levy for "Investment Commitment to Research and Development". The authors are grateful for the support of the Center for Energy and Petroleum Studies (CEPETRO-UNICAMP/Brazil), the Reasoning for Complex Data Research Group (RECOD.ai/UNICAMP/Brazil), the Department of Energy (DE-FEM-UNICAMP/Brazil), the Research Group in Reservoir Simulation and Management (UNISIM/UNICAMP/Brazil), Energi Simulation, and CNPq. In addition, a special thanks to SLB, GeoSoftware, and CMG for software licenses.

## References

[1] Dramsch, J.S.: 70 years of machine learning in geoscience in review. Advances in geophysics **61**, 1–55 (2020)

[2] Maleki, M., Schiozer, D.J., Davolio, A., Lopez, J.: A workflow for high-resolution reservoir characterization using multiple 4d seismic datasets. In: ECMOR 2022, vol. 2022, pp. 1–13 (2022). European Association of Geoscientists & Engineers

[3] Stammeijer, J., Hatchell, P.: Standards in 4d feasibility and interpretation. The Leading Edge **33**(2), 134–140 (2014)

[4] Maleki, M., Davolio, A., Schiozer, D.J.: Quantitative integration of 3d and 4d seismic impedance into reservoir simulation model updating in the norne field. Geophysical Prospecting **67**(1), 167–187 (2019)

[5] Maleki, M., Davolio, A., Schiozer, D.J.: Using simulation and production data to resolve ambiguity in interpreting 4d seismic inverted impedance in the norne

field. Petroleum Geoscience **24**(3), 335–347 (2018)

[6] Johann, P., Sansonowski, R., Oliveira, R., Bampi, D.: 4d seismic in a heavy-oil, turbidite reservoir offshore brazil. The Leading Edge **28**(6), 718–729 (2009)

[7] Maleki, M., Davolio, A., Schiozer, D.J.: Qualitative time-lapse seismic interpretation of norne field to assess challenges of 4d seismic attributes. The Leading Edge **37**(10), 754–762 (2018)

[8] Emerick, A.A.: Analysis of the performance of ensemble-based assimilation of production and seismic data. Journal of Petroleum Science and Engineering **139**, 219–239 (2016)

[9] Davolio, A., Schiozer, D.J.: Probabilistic seismic history matching using binary images. Journal of geophysics and engineering **15**(1), 261–274 (2018)

[10] Rosa, D.R., Schiozer, D.J., Davolio, A.: Evaluating the impact of 4d seismic data artefacts in data assimilation. Petroleum Geoscience **29**(3), 2022–069 (2023)

[11] Correia, M.G., Maleki, M., Silva, F.B., Gomes, A.D., Schiozer, D.J.: Integrated approach to improve simulation models in a deep-water heavy oil field with 4d seismic monitoring. Petroleum Geoscience **29**(1), 2022–048 (2023)

[12] Oliver, D.S., Fossum, K., Bhakta, T., Sandø, I., Nævdal, G., Lorentzen, R.J.: 4d seismic history matching. Journal of Petroleum Science and Engineering **207**, 109119 (2021)

[13] Amini, H., MacBeth, C., Shams, A.: Seismic modelling for reservoir studies: a comparison between convolutional and full-waveform methods for a deep-water turbidite sandstone reservoir. Geophysical Prospecting **68**(5), 1540–1553 (2020)

[14] Emerick, A.A., Reynolds, A.C.: Ensemble smoother with multiple data assimilation. Computers & Geosciences **55**, 3–15 (2013)

[15] Maschio, C., Schiozer, D.J.: Probabilistic history matching using discrete latin hypercube sampling and nonparametric density estimation. Journal of Petroleum Science and Engineering **147**, 98–115 (2016)

[16] Danaei, S., Neto, G.M.S., Schiozer, D.J., Davolio, A.: Using petro-elastic proxy model to integrate 4d seismic in ensemble based data assimilation. Journal of petroleum science and engineering **194**, 107457 (2020)

[17] Danaei, S., Neto, G.M.S., Schiozer, D.J., Davolio, A.: Substituting petro-elastic model with a new proxy to assimilate time-lapse seismic data considering model errors. Journal of Petroleum Science and Engineering **210**, 109970 (2022)

[18] Danaei, S., Cirne, M., Maleki, M., Schiozer, D.J., Rocha, A., Davolio, A.: All-in-one proxy to replace 4d seismic forward modeling with machine learning

algorithms. Geoenergy Science and Engineering **222**, 211460 (2023)

[19] MacBeth, C., Geng, C., Chassagne, R.: A fast-track simulator to seismic proxy for quantitative 4d seismic analysis. In: SEG Technical Program Expanded Abstracts 2016, pp. 5537–5541. Society of Exploration Geophysicists, ??? (2016)

[20] Botechia, V., Correia, M., Rios, V., Davolio, A., Santos, S., Hohendorff Filho, J., Schiozer, D.: Unisim-iv: Benchmark proposal for light oil carbonate reservoir with high co2 content. Brazilian Journal of Petroleum and Gas **16**(4) (2023)

[21] Xu, S., Payne, M.A.: Modeling elastic properties in carbonate rocks. The Leading Edge **28**(1), 66–74 (2009)

[22] Cai, J., Luo, J., Wang, S., Yang, S.: Feature selection in machine learning: A new perspective. Neurocomputing **300**, 70–79 (2018)

[23] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 618–626 (2017)

[24] Maleki, M., Cirne, M., Schiozer, D.J., Davolio, A., Rocha, A.: A machine-learning framework to estimate saturation changes from 4d seismic data using reservoir models. Geophysical Prospecting **70**(8), 1388–1409 (2022)

[25] Rudin, C.: Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature machine intelligence **1**(5), 206–215 (2019)

[26] Velden, B.H., Kuijf, H.J., Gilhuijs, K.G., Viergever, M.A.: Explainable artificial intelligence (xai) in deep learning-based medical image analysis. Medical Image Analysis **79**, 102470 (2022)

[27] Jebli, I., Belouadha, F.-Z., Kabbaj, M.I., Tilioua, A.: Prediction of solar energy guided by pearson correlation using machine learning. Energy **224**, 120109 (2021)

[28] Hadid, A.: The local binary pattern approach and its applications to face analysis. In: 2008 First Workshops on Image Processing Theory, Tools and Applications, pp. 1–9 (2008). IEEE

[29] Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. Advances in neural information processing systems **30** (2017)

[30] Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034 (2013)

[31] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241 (2015). Springer

[32] Siddique, N., Paheding, S., Elkin, C.P., Devabhaktuni, V.: U-net and its variants for medical image segmentation: A review of theory and applications. Ieee Access **9**, 82031–82057 (2021)

[33] Cuzick, J.: A wilcoxon-type test for trend. Statistics in medicine **4**(1), 87–90 (1985)