ABSTRACT

During drilling and logging, depth alignment of well logs acquired in the same borehole section at different times is a vital preprocessing step before any petrophysical analysis. Depth alignment requires high precision as depth misalignment between different log curve measurements can substantially suppress possible correlations between formation properties, leading to imprecise interpretation or even misinterpretation. Standard depth alignment involves cross-correlation, which typically requires user intervention for reliability. To improve the depth alignment process, we apply deep learning techniques and propose a simple and practical implementation of a 1-dimensional supervised Convolutional Neural Network (1D CNN). We train seven CNN models using different log measurements such as gamma-ray, resistivity, P- and S-wave sonic, density, neutron, and photoelectric factor (PEF), to estimate depth mismatches between corresponding raw logging while drilling (LWD) and electrical wireline logging (EWL) logs of each measurement type. Our deep learning approach avoids manual feature extraction; hence no high level petrophysical knowledge is needed by our algorithms. We use log data from six wells from the Ivar Aasen Field, in the Norwegian North Sea. Four of the six wells constitute the entire dataset for training and model selection, in which we compare three search algorithms during the hyperparameter tuning. Only two wells have both LWD and EWL log suites. These wells are used for depth shift inference. We focus on estimating bulk shifts, and we assume the existence of small pattern differences. We assess our results by visual inspection, and quantitative metrics such as the Pearson correlation and Euclidean distance. We also compare the CNN depth shifts with depth shifts obtained using the classical cross-correlation method. The CNN performs well and is competitive with cross-correlation. CNN performs better for some log types, resistivity for instance, than others. Several factors influence our results including, the quality of the input data, borehole conditions, pattern differences between LWD and EWL, and significant stretch/squeeze effects. Differences between the mean Pearson correlation computed after CNN and cross-correlation depth matching process are of the order of 10^{-1} and 10^{-2}. Our CNN approach is therefore a potential alternative to current depth matching methods, which may reduce the amount of user intervention required from the petrophysicist.

INTRODUCTION

In the oil and gas industry, the value of well log data is well recognized as they yield insights into the subsurface as a function of depth. Real-time well log measurements of the subsurface are typically available to geoscientists, petrophysicists and rock physics practitioners at all stages in exploration and production settings. Well logs provide valuable information about petrophysical properties including porosity, permeability, fluid saturation, and geomechanics related properties, which can be used for reserves estimation, well completion, and reservoir modelling. However, these measurements are imperfect, as they are prone to systematic and random errors, and noise during acquisition. Although most of these problems can be mitigated nowadays using either advanced acquisition techniques or sophisticated processing algorithms, there are other fundamental problems that still pose major challenges in well logging. One clear example is the mismatches in depth between well logs acquired within the same run or different runs (repeated logging passes along the same borehole). The reasons for such mismatches usually depend on the logging system e.g. LWD or EWL. Furthermore, borehole conditions and environment play a role in magnifying misalignment during acquisition. For example, for LWD measurements, noise generated by the drill bit and circulating fluids can pollute the logging data. There are also other factors that increase the uncertainty of LWD measurements, such as the changes undergone by the drilling pipes due to the increase in temperature and pressure with depth, changes in drilling parameters e.g. the weight on the bit, the torque on the pipe, as well as all inherent interactions between the pressure and fluids in the formation and the pressure and drilling fluids along the annulus, inside or outside the tubing strings, etc. The combination of all these factors can lead to variable depth errors ranging from approximately 1 to 10 m (Bolt, 2016, Chia et al., 2006, Theys, 1999, Wilson et al., 2004) in the LWD measurements. On the other hand, the corresponding EWL measurements are very prone to tool sticking and slipping caused by variations in borehole rugosity (mud cake buildup or large changes in the borehole dimensions). This leads to depth errors of up to 12 m (Theys, 1999), and significantly compressed and expanded data sections (squeeze/stretch). Due to the elasticity of the cable, its length is also affected by increasing temperature and pressure, as well as by twisting effects. However, these are considered minor effects which are usually neglected (Sollie and Rodgers, 1994, Theys, 1999).

Depth alignment of log data is an early stage in petrophysical preprocessing workflows. If log measurements are not properly aligned, all possible correlations across different measurements will be partially or completely suppressed (Zangwill, 1982). For instance, density - neutron porosity cross-plots are a well-established tool for lithology identification and gas detection, and can be significantly affected by the log misalignments, thus, leading to sub-optimal or incorrect interpretations (Torres et al., 2020, Zangwill, 1982). To avoid this, repeated measurements are taken with the gamma-ray log, which is selected as a depth reference. This allows, all other log types to be aligned since the gamma-ray captures the general geological patterns very well. This task has been researched for many years, Kerzner (1984) for example, proposed an automatic depth matching across wells via cross-correlation and dynamic programming. However, most current methods which are in theory automatic are not in reality fully automatic, as they require significant user intervention. Too much user intervention in depth matching is often regarded by petrophysicists as a time consuming and cumbersome process, hence they would prefer to focus their time and attention on the analysis of the massive amounts of data becoming available through the digitization of the oil industry. Such analysis may demand a high-level knowledge, and expertise.

Zimmerman et al. (2018) presented a compact artificial neural network (ANN) architecture to perform automated depth matching between gamma ray logs from different runs or logging passes. Their work was inspired by previous application of ANNs for log pattern recognition, which aimed to perform automatic well-to-well correlation within the same field, based on geological datums and relevant markers (Luthi and Bryant, 1997). Two neural networks were proposed. They work in a sequential manner providing geologists with a tool that allows them to speed-up the process interactively, assisted by a neural network that recognizes a pattern in a well and searches for similar patterns in neighboring wells. Zimmermann et al. (2018) and Le et al. (2019), implemented an ANN that mimics the manual procedures carried out by expert petrophysicists for depth matching of well logs. In that sense, they designed a neural network that takes two input depth series: one synchronized (reference) and one desynchronized, with the aim of identifying anchor points in the reference and their corresponding position in the desynchronized signal to be matched. The neural network performs a classification task with 121 classes representing all possible shifts. The class assigned the highest probability by the network corresponds to the shift. The authors emphasized that one major bias in their approach is due to the selection of anchor points. These correspond to peaks, typically

suggested by experts. However, not all peaks necessarily yield the best accuracy for depth matching (Zimmermann et al., 2018).

To reduce the problems related to limited data and to improve the robustness and accuracy of the machine learning model, Le et al. (2019) further developed the approach originally presented by Zimmermann et al. (2018). They used a continuously self-evolving depth-matching framework with a depth matching service in which the user reviews the output matching from the algorithm and performs the necessary adjustments. This feedback is sent back into the algorithm to retrain and improve it over time. The shift suggested by the algorithm passes through an automatic quality control designed by a combination of different metrics before reaching the user, in that way, they assure that the machine outputs high quality results. The inclusion of such anchor filtering into the pipeline circumvents the peak bias effect mentioned previously as now the point selection task is automatically performed by the engine, allowing for generalization when it comes to the log type. However, this approach has only been tested on gamma ray logs. For different log types, a new training step needs to be performed because other log types have different distributions from gamma-ray logs and their patterns might differ. Wang et al. (2020) presented a dynamic depth matching method for geological correlation using a deep neural network with a multitask learning technique. The deep neural network uses a 1D CNN architecture that recognizes patterns from a reference gamma-ray log to find the corresponding depth of this pattern in gamma-ray logs from another well in the field. The multitask problem is implemented by finding the best matching pattern between a query (a pattern in the reference log) and a target gamma-ray log (sliding window along the target gamma-ray log) considering global and local information. They proved the power of CNN to recognize geological patterns across different wells within a field and to accommodate both larger depth shifts than those typically observed between logs from the same well, and the absence and distortion of the patterns due to lateral lithofacies variations between wells.

As an alternative solution to the depth matching problem, we propose a proof of concept machine learning algorithm using a 1D CNN algorithm capable of estimating depth shifts between logs from different runs, specifically between LWD and EWL logs. The main challenge here is the differences between the signals due to significant changes in borehole conditions, formation properties due to invasion effects, depth sampling, log resolution, different tool design and technologies, and different companies' algorithms and procedures. However, the feature extraction power of the CNN allows the extraction of relevant invariant

characteristics of the signals that can retain the important patterns for depth matching regardless of the aforementioned factors. This provides our workflow with more flexibility and generalization capacity. First, we explain the theory behind the CNN machine learning algorithm. Second, we describe the practical implementation of the CNN from data preparation and splitting into train, validation, and test sets. The training process, hyperparameter tuning, model selection, and assessment of results using well data from the Ivar Aasen field in the Norwegian North Sea. Third, we evaluate and discuss the results of the depth shift inference via qualitative visual inspection and comparison of quantitative metrics such as Pearson correlation coefficient and Euclidean distance. We also compare the CNN depth shifts with those obtained using cross-correlation. Finally, we emphasize the main contribution of this work in automating and speeding-up potentially time-consuming interactive processes, as well as the remaining scope for improvements through further work.

TRADITIONAL METHODS

Cross-correlation

The cross-correlation is a well-known and widely used technique in signal processing to detect and match signals. In other words, the cross-correlation is a measure of similarity between two series (e.g. in time or depth) with respect to the displacement of one relative to the other. We use the cross-correlation as a depth matching benchmark between LWD and EWL logs in this work because it is easy to implement and works well when the shape of the patterns is fixed (Wang et al., 2020). The mathematical definition of the cross-correlation is given in Equation (1). In formal terms, given a reference depth series e.g. an EWL log $X=(x_1,\ldots,x_N)$, and a test depth series e.g. an LWD log, $Y=(y_1,\ldots,y_M)$, the cross-correlation (c) between them at depth lag $L= 0,1,\ldots, (||X|| + ||Y|| - 2)$ is:

$$c(L) = \frac{1}{\sqrt{\|X\|\|Y\|}} \sum_{i=1}^{N} x_i y_{i+L} \tag{1}$$

where $||X||$ and $||Y||$ are the lengths of X and Y, respectively, and $N = max\ (||X||, ||Y||)$ (Anderson and Gaby, 1983).

The cross-correlation even though is an easy method to be implemented needs some user intervention to assess the reliability of its solution in presence of excessive noise and distortion of the signal (Zimmerman et al., 2019). Similarly, some drawbacks can arise when the patterns of the signal differ considerably, which can occur when depth matching LWD and EWL logs within the same depth interval due to changes in the borehole environment and conditions through time, leading the cross-correlation to choose a wrong time/depth lag or fail to make a detection (Wang et al., 2020). Another limitation of the cross-correlation is that it works well only if a constant time/depth shift characterizes both signals, in other words the problems is reduced to the correction of the time/depth lag detected by the cross-correlation. In contrast, it will fail to find the best match between signals when the time/depth shift is not stationary or if dynamic shift exists, which cannot be properly addressed by a linear metric (Herrera and van der Baan, 2014). Cycle skipping is another possible problem that cross-correlation might run into whenever borehole depth intervals have low property contrast, suggesting wrong peaks to be matched. In those cases, additional user intervention is needed to ensure optimal signal alignments, for instance manual stretching-squeezing adjustments and definition of relevant patterns to be synchronized. These issues do not allow having a fully automated workflow (without interaction of the petrophysics at any stage of the process) for well log depth matching based on cross-correlation.

MACHINE LEARNING IMPLEMENTATION

1D CNN

Building invariance properties into the structure of neural networks is the basis for convolutional neural networks (LeCun et al., 1989, LeCun et al., 1998). CNN's capability of extracting local features that depend on small subregions of an image are adopted by modern computer vision approaches and other disciplines. These capabilities of CNNs can be obtained using three different mechanisms: (i) local receptive fields, (ii) sharing weights, and (iii) subsampling (Bishop, 2006).

In general, a convolutional neural network can be divided into three different types of layers or blocks: convolutional layers, pooling, and fully connected (FC) layers. The first two perform the feature extraction process, while the latter are like ANNs and do the mapping between the extracted features and

output based, for example, on feedforward propagation. Feature extraction is a sequential process, i.e., the output of one layer is fed into the next such that the extracted features can hierarchically become more complex as we move deeper into the network. The optimization of all learnable parameters is performed during the training process, which can be highly computationally demanding, and is based on minimizing a loss function through an optimization algorithm such as backpropagation and gradient descent (Yamashita et al., 2018).

The mechanism behind 1D CNN is similar to ANN. The main difference is the input preprocessing stage 1D CNN inputs raw data that is processed through convolutional trainable layers to learn an adequate representation of the input. The neurons in a layer are connected to small regions of the previous layer. These small regions are called receptive fields. The input of the 1D CNN is an array representing a segment of the well logs as X. The network is trained to learn a set of parameters θ to map the input to the output prediction S, following a hierarchical feature extraction process given as:

$$S = F\left(X \mid \theta\right) = f_L\left(\ldots f_2\left(f_1\left(X \mid \theta_1\right) \mid \theta_2\right) \mid \theta_L\right) \tag{2}$$

where L is the number of hidden layers in the network. The operation of the l_{th} layer in the convolutional layers is given by:

$$\begin{aligned} S_l &= f_l\left(X_l \mid \theta_l\right) = h\left(W \otimes X_l + b\right), \\ \theta &= \left[W, b\right] \end{aligned} \tag{3}$$

and $\otimes$ represents the convolutional operation, X_l is two-dimensional input of N feature maps, W is a set of N one-dimensional weights or kernels (receptive fields), b is the bias vectors, and $h(.)$ is the activation function, which introduces the no linearity. The output of the final convolutional layer is transformed into a vector and input of several stacked FC layers that are described in Equation (4) (Abdoli et al., 2019).

$$\begin{aligned} S_l &= f_l\left(X_l \mid \theta_l\right) = h\left(WX_l + b\right), \\ \theta &= \left[W, b\right] \end{aligned} \tag{4}$$

CNNs are very popular in two-dimensional and three-dimensional settings; however, it is also possible to use them with one-dimension data types such as time series, e.g. accelerometers, speech, audio, text data, automatic music tagging, speaker identification, and environmental sound classification among others. Their ability to use the raw data instead of using hand-crafted features, which require a high level of expertise in the field makes them very popular (Abdoli et al., 2019). However, the limitations of 2D CNNs Lie in the larger amount of data that is needed for training. In the case of environmental sound recognition, there is a huge scarcity in the data, therefore other techniques such a data augmentation like time stretching, pitch shifting, adding background noise, and dynamic range compression must be carry out as shown by Salamon and Bello (2017). All these additional processes imply also higher levels of expertise and uncertainty. We find in the literature studies implementing 2D CNNs for depth series, for example well logs. Zhu et al. (2018) used the discrete wavelet transform (DWT) as low pass filter based on a cascade frequency bands of the different log types like gamma ray, resistivity, density and acoustic to generate their 2D representation as maps (spectrograms) with different frequency levels as a function of depth. This serves as input maps for 2D CNN supervised lithology classification task. 1D CNN has been successfully implemented for well logs task as lithology classification (Imamverdiyev and Sukhostat, 2019), well to well correlation (Wang et al., 2020, Brazell et al., 2019) and vug facies identification and vug-size classification (Deng et al., 2021). Hoshen et al. (2015) found that time differences between channels in an end-to-end multichannel 1D CNN for speech recognition are indicators of the spatial locations of the inputs. This is an important characteristic that we would like to exploit to synchronize depth series.

1D CNN architecture

Abdoli et al. (2019) showed a compact network representation that extracts both low- and high-level features capturing the most interesting patterns in a signal. In other words, the lower level features are the starting point of a CNN, which is the raw data moving toward a more complex or abstract representation of the data known as a higher-level representation (Lecun et al., 2015). We adopt their architecture as an initial model while enjoying the advantage of an implementation to be used on limited amounts of data. There is no need for exhaustive preprocessing, and we can divide the logs into windows of fixed length. The chosen network topology is based on the following principles: *i)* the first convolutional layer must have a large kernel size or receptive fields in such a way that it can capture the global features of the log signals (long

wavelength content). The successive filters increase in number and decrease in size as we move deeper into the network, extracting the local features (short wavelength content), *ii)* the pooling layers are set in a way that reduce the dimensionality of the feature maps and speed-up the training. Max pooling layers are the most common form of pooling layers, which extract patches of the feature maps and output the maximum value discarding the rest (Yamashita et al., 2018), *iii)* the output of the last max pooling layer is transformed into a single vector or one-dimensional array of numbers (flattened), and it is passed through two fully connected layers or dense layers, in which every input is connected to every output by a learnable weight. The activation function (*h(.)*) for all the convolutional layers and the fully connected layers are rectify linear unit (ReLU) as defined in Equation (5), whereas the output layer has one single unit that estimates the depth shift.

$$\mathrm{Re}\,LU = f\left(x\right) = \max(0, x) \tag{5}$$

To reduce overfitting and speed-up the training process Abdoli et al. (2019) used batch normalization layers after each activation function of the convolutional layers and added a drop-out after each FC layer with a probability of 0.25. The batch normalization layers apply a transformation to keep the mean and standard deviation of the convolutional layers' outputs close to 0 and 1, respectively. By doing so, it accelerates the training process, enabling the use of high learning rates, avoiding the risk of divergence during optimization, and acting as regularization reducing over-fitting of the networks (Ioffe and Szegedy, 2015).

Machine learning algorithm deployment

We proposed a machine learning algorithm based on 1D CNN to depth match LWD and EWL logs, focusing on the correlated depth concept, in which the main goal is to correlate one log measurement LWD with a reference EWL that is assumed to be in the correct position in depth (Theys, 1999). We assume that the depth shift is no larger than 10 ft (≈ 3 m) within a window, therefore the maximum number of data points ranges between -20 and 20. The sampling interval between data points is 0.5 ft (≈ 0.1524 m). This assumption is based on the observed shifts present in the data, which do not reach values larger than 4.5 m. Additionally, this range allows to test our proposed approach faster since it limits the number of simulated shifts without compromising the results acting just as a soft constrain. The main idea is to define segments or windows of

limited length (256 data points ≈ 39 m) for different pairs of log measurements e.g., gamma-ray pairs, resistivity pairs, density pairs, etc. where distinctive log patterns present in both (reference and shifted logs) are identified and matched by the network to improve the alignment between the two signals automatically.

The problem setup can be described as follows. First, we select a pair of logs that measure the same property within a specific depth window, which are used as input to the CNN algorithm. Second, the CNN algorithm finds the distinctive patterns in the reference log (EWL) and identifies the same patterns in the corresponding shifted log (LWD). Finally, we synchronize or align the logs by estimating the number of samples, either positive or negative, that the LWD log needs to be shifted to best match the EWL log. By using the feature extraction capacity of the CNN, the identification of the patterns is performed automatically by the algorithm. We start from an initial model consisting of 3 convolutional layers with ReLU as their activation function, and their associated pooling layers specifically max pooling. In between each convolution-polling layer block, we place a batch normalization layer. These provide the feature extraction components of the network. Then we perform the estimation of the depth shift with two fully connected (FC) layers with ReLU activation functions, with batch normalization and drop-out layers in between each FC layer. The batch normalization and drop-out acts as a regularization method. The output layer has a single unit with the identity function to estimate the depth shift in number of data points, solving a regression task. The best architecture of the network and hyperparameters for each individual pair of logs is determined by testing three different search algorithms, random search, hyperband, and Bayesian Optimization.

This framework is tested on two wells from the Ivar Aasen field in the Norwegian sector of the North Sea, which have available separated LWD and EWL suites of logs. The training of the network is carried out with final petrophysical composite logs (depth matched, spliced, and edited logs) from 4 wells in the same field. Therefore, semi-synthetic data must be generated considering the absence of LWD and EWL logs, and ground truth labels, in the training wells. To generate the training data we take each log measurement from the 4 wells and produce copies that are bulk shifted by a range of numbers of data points up or down (positive or negative), thus the output labels for each semi-synthetic log will correspond to the signed magnitude of the shift in data points. During training, model selection, and preliminary model assessment, we use the mean squared error (MSE) as a metric to evaluate the algorithm on the semi-synthetic data. However, for depth inference with real data, the absence of ground truth labels requires the use of

signal analysis metrics like Pearson correlation and Euclidean distance on the reference and the shifted log before and after depth matching. We also use the values of the same metrics after depth matching using cross-correlation for comparison with the CNN results.

Our initial network structure is shown in Fig. 1. The input of the CNN consists of two channels. The first channel is a window of the reference log (EWL) with a fixed length equal to 256 data points and the second channel is a copy of the reference bulk shifted by a given number of samples. These two channels of the same length represent a single depth window, and a single input sample into the network, which extracts the relevant features and determines the depth shift needed to align the two logs.

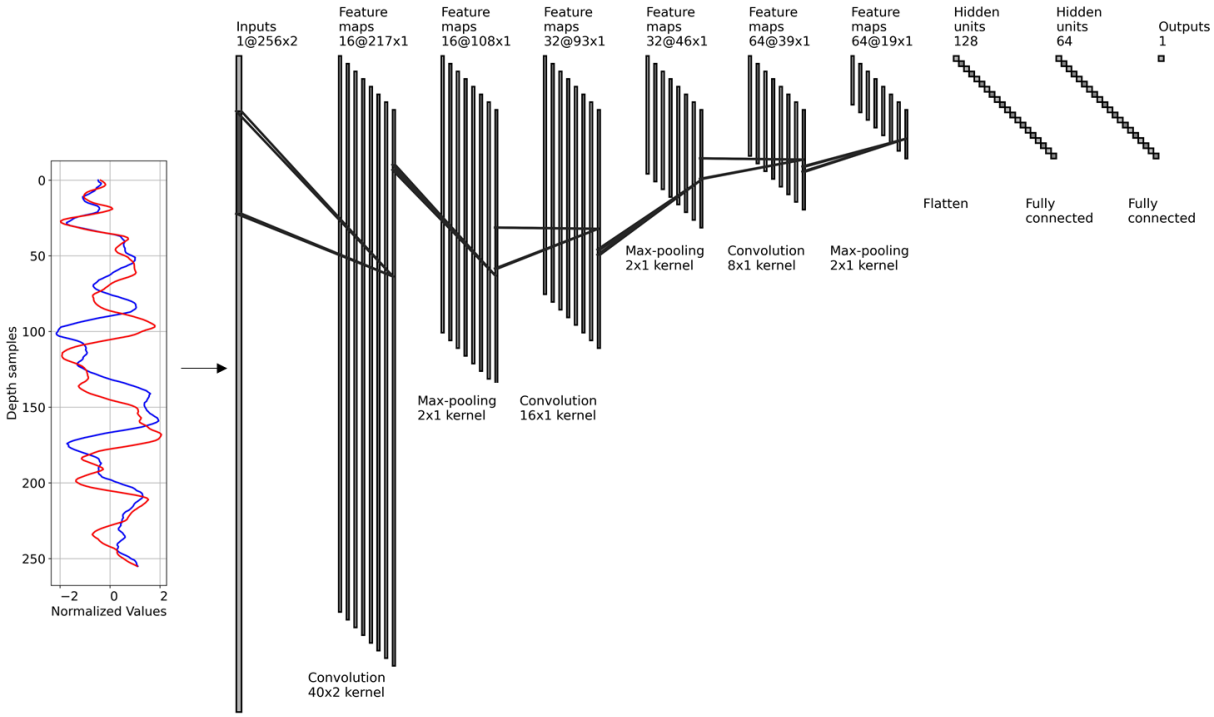

Fig. 1- Sketch of a 1D CNN base model architecture. The input layer is a double channel corresponding to the reference EWL log (solid red) and the misaligned LWD log (solid blue). After each convolution layer we introduce batch-normalization layers (not shown) and after each fully connected layer we introduce batch-normalization and apply a drop-out (not shown). This figure is generated by adapting the code from https://github.com/gwding/draw_convnet.

Dataset and data preparation method

Our dataset consists of 6 well logs acquired in the Ivar Aasen field in the Norwegian sector of the North Sea. The wells are a mixture of exploration and appraisal wells and most of the them are considered vertical with maximum deviations of between 0 and 13 degrees. However, two of them are slightly deviated

with maximum values of 20 – 25 degrees (16/1-22S and 16/1-21S, respectively). Four out of six are introduced as a part of the training dataset while the other two (16/1-9 and 16/1-21 S) are kept as a test set (unseen LWD and EWL logs) for depth shift inference. The training set consists of logs that have already been depth shifted and spliced. They are the final petrophysical composite log suite contained in LAS (Log ASCII Standard) files. This means that no distinction is made between EWL and LWD. For training purposes this is not a problem since we are interested in identifying the shifts needed to depth align the signals. Another important aspect of our approach is removing the need for ground truth labelling by a petrophysicist. We achieve this by training the network using semi-synthetic for the supervised learning process. To generate ground truth labeled data, we artificially simulate an acceptable range of depth shifts expected for all the log types by copying the logs and applying all possible shifts within the range to them. This limits the type of shift applied to be a constant value (bulk shift) for the whole log. An additional augmentation technique is the generation of overlapping windows, which means that each depth window will have an overlap of 50% with the previous one, doubling the number of training samples.

Data preparation consists first of a gentle smoothing with filter parameters chosen after testing. Second, we make multiple copies of each log type equal to the number of bulk depth shifts we want to simulate setting a maximum depth shift of +/- 3 m. This implies the generation of 41 copies for each log type, ranging from -20 to 20 data point shifts including the zero-shift. Thus, we ensure that all depths along the well log can experience all likely shifts. This avoids skewness in the training set. Third, we divide the shifted logs into windows of 256 data points, and we apply a local normalization and standardization of the data in each window. The number of windows depends on the length of each log type (See Table 1). We store the depth shifts applied to each log copy in a depth shift vector as ground truth labels for the training and model selection process. From now on, we will refer to the number of windows into which each log has been divided as the number of sample when we talk about the machine learning implementation. The input samples for the CNN are the log windows consisting of 256 data points.

Training, tuning, and model selection

We use the whole synthetically shifted data derivate from only 4 out the 6 wells available (16/1-2, 16/1-11, 16/1-16, and 16/1-22S) to train and optimize the model and get a first test assessment of the method.

Fig. 2 shows the splitting of the data at each stage and the general workflow. After preparing the dataset for a single log type e.g. gamma-ray (GR), we have a total of 21320 samples. We split these into a training set (70%), a validation set (18%), and a test set (12%). The validation set is used to monitor the training and model selection process. For this set our samples have ground truth labels (signed depth shifts) that allow us to track the training process and evaluate the models. Table 1 summarizes the size of the data sets per log type and their corresponding splits into training, validation, and test sets. Notice that the number of samples per log type varies depending on how the logging plan was designed for each well. For example, usually, gamma ray and resistivity logs are acquired along the full depth range logged while density, neutron and PEF are mostly acquired only over the reservoir/target depth interval. This strategy is very common for appraisal and development wells. It is worth noticing that the models for each log type are trained and tuned individually.

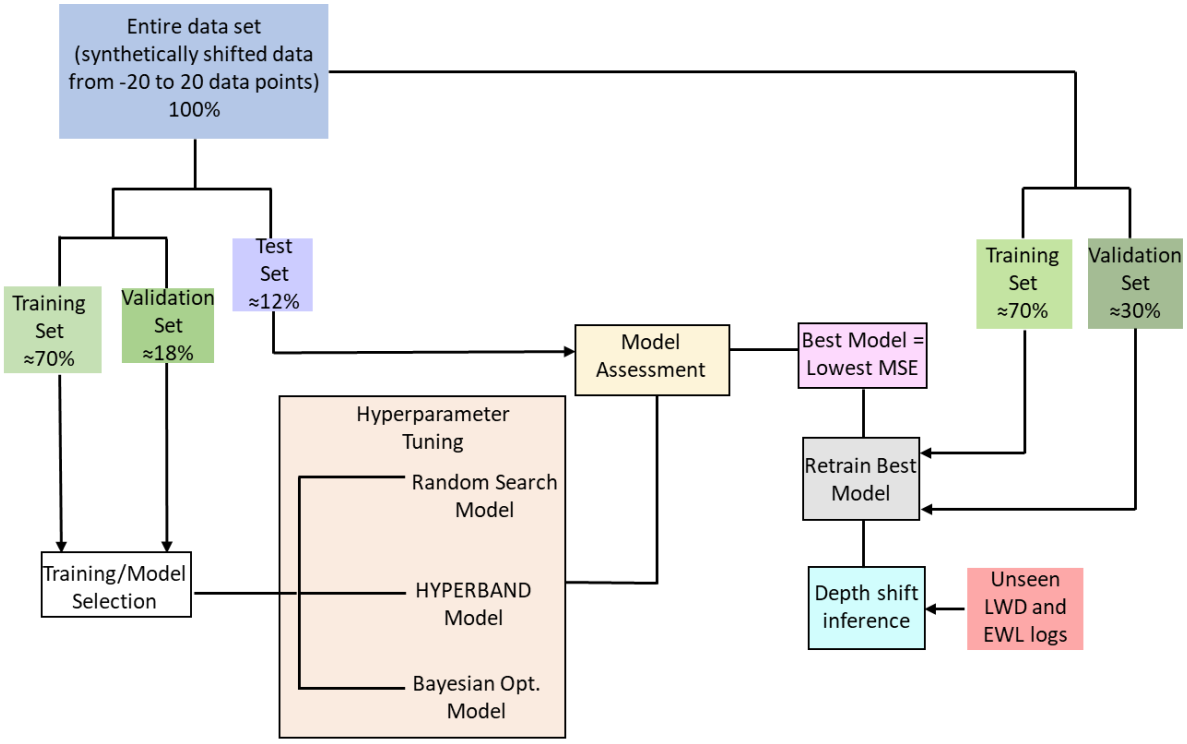

Fig. 2 -Sketch of the data splitting, training -model selection workflow, and final depth shift inference on the completely unseen data from the other 2 wells (16/1-9 and 16/1-21S) kept outside of the training and model selection process.

Table 1 – Data Set Size in Samples for Training and Model Selection from 4 Wells for Each Log Type.

Log Type	Complete data set size (synthetically bulk shifted)	Training set size	Validation set size	Test set size
Gamma-ray	21320	14924	3895	2501
Resistivity	20664	14596	3608	2460
P-Wave sonic	16564	12013	2829	1722
S-wave sonic	7093	5084	1148	861
Density	7995	5617	1394	984
Neutron	8159	5658	1435	1066
PEF	7380	5166	1394	820

We train our initial model using an Adam optimizer, which is a stochastic gradient descent method based on adaptive estimates of first and second order moments (Kingma and Ba, 2015). We use the default parameter settings from *Tensorflow Keras banked* (Chollet, 2015). The learning rate was constant at 0.001. We set the batch size to 128 and the number of epochs to 100. The objective function to be minimized during training is the mean squared error (MSE). We use the MSE to monitor the training and hyperparameter tuning process at each epoch.

From the learning curves (see Fig. 3) we can see that the MSE in the validation set decreases rapidly and reaches stability quite fast, at MSE even lower than those of the training set. This suggests that 50 to 100 epochs are sufficient to train the model without overfitting.

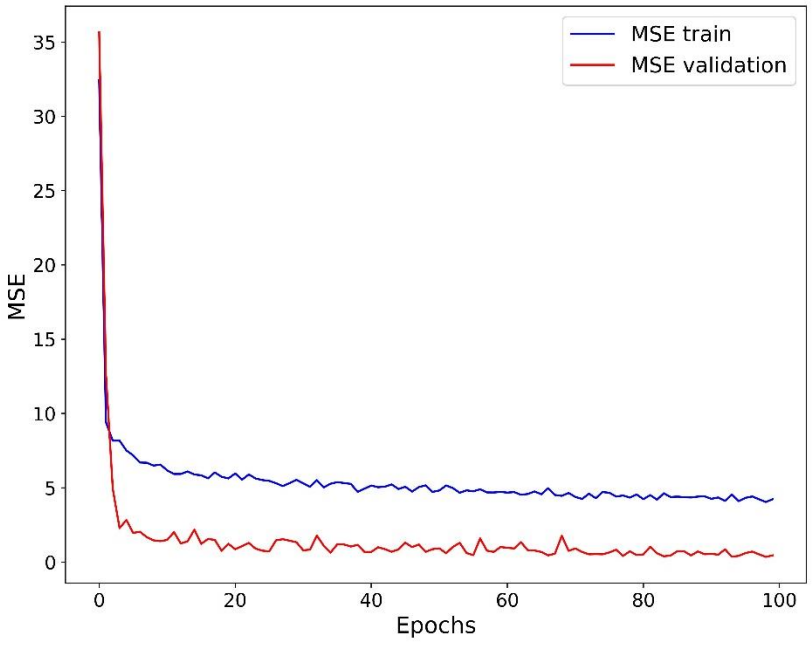

Fig. 3 - Training and validation MSE as a function of the number of epochs for the gamma-ray model.

Before the first assessment of the method, we perform a hyperparameter tuning. This aims to minimize a cost function over a graph-structured configuration space, which involves the identification of optimal network architecture parameters (e.g. the number of convolutional filters/kernels) and algorithm/optimization parameters (e.g. the learning rate), that could be drawn from the space and that can reduce the loss function in further steps (Bergstra and Bengio, 2012. Basically, the main idea is to reduce the MSE, which is evaluated on the validation set by finding the best combination of the model's hyperparameters. This is commonly known as the best trial.

We carry out the tuning process using *Keras tuner* library, which offers three different tuning strategies, random search, hyperband, and Bayesian optimization (O'Malley et al., 2019). Differences between these algorithms are associated with the design of the hyperparameter space, and the hyperparameter search strategy which has an impact on the run time for a given model complexity and dataset size. Random search is one class of non-informed exhaustive parameter search. Hyperband is a modification of random search with an adaptive resource allocation process that speeds up the hyperparameter selection. The concept of adaptive resource allocation means that more resources are allocated for identification and evaluation of hyperparameter configurations that are most likely to be successful while quickly discarding the less successful ones. This allows the exploration of a larger number of hyperparameter configurations than sampling uniformly distributed hyperparameter spaces trained until completion. The Hyperband strategy relies on an early stopping process to allocate resources (Li et al., 2017). Finally, the Bayesian optimization strategy, belongs to the surrogate model category and creates a model of the conditional probability $p(y|\lambda)$ of a configuration's performance on an evaluation metric y e.g. loss, given a set of hyperparameters λ. Bergstra et al. (2011) and Thornton et al. (2013) showed that this technique outperforms random search, but for high-dimensional problems its performance and efficiency degrades and it can then perform similarly to random search (Li et al., 2017). Hyperband can produce good results and has the main advantage of being from 5 to 30 times faster than Bayesian optimization in several deep-learning and kernel-based machine learning problems (Li et al., 2017).

We define a range of values for each hyperparameter that we want to tune, and run the tuning process using the three searching strategies above, selecting the best model based on the final MSE. Each method converges to a different network architecture that is, in most cases, better than the initial model. Table 2

shows the tuning parameters with their corresponding ranges. Note that we keep the number of convolutional layers fixed for all models, as well as their activation functions. For numerical parameters we show the maximum and minimum value that can be selected, and for no numerical values like the type of activation function we refer them as option 1 and option 2.

Table 2- Tuning Parameters Selection Range per Layer.

Layers	Tuning Parameters	Maximum / Option 1	Minimum /Option 2	Default	Step
Convolutional layer 1	# of kernels	128	8	16	8
	kernel size	64	8	32	8
	Activation function	NA	NA	ReLU	NA
Max pooling layer 1	pool size	NA	NA	2	NA
	strides	2	1	1	NA
Convolutional layer 2	# of kernels	128	8	32	8
	kernel size	32	8	16	8
	Activation function	NA	NA	ReLU	NA
Max pooling layer 2	pool size	NA	NA	2	NA
	strides	2	1	1	NA
Convolutional layer 3	# of kernels	128	8	64	8
	kernel size	16	2	16	4
	Activation function	NA	NA	ReLU	NA
Max pooling layer 3	pool size	NA	NA	2	NA
	strides	2	1	1	NA
Dense layer 1	# of units	256	32	128	32
	Activation function	ReLU	Tanh	ReLU	NA
Dropout layer 1	dropout rate	0.5	0	0.25	0.05
Dense layer 2	# of units	256	32	64	32
	Activation function	ReLU	Tanh	ReLU	NA
Dropout layer 2	dropout rate	0.5	0	0.25	0.05
Optimizer	learning rate	1,00E-02	1,00E-04	NA	1,00E-01

Having obtained the improved models after hyper tuning using each of the 3 methods, we assess each of them by estimating the depth shifts on the test set. At this stage we are still using the synthetically shifted samples. We compute the MSE for each model and we choose the one with the lowest MSE. Fig. 4 shows an example of the gamma-ray models, where the initial model regression line a), shows good agreement between the actual depth shift and the predicted depth shift, with a maximum of discrepancy of around ± 3 samples. Similarly, the random search solution b), reduces the general variability of the depth shifts, and has the same R^2 of 0.996 as the initial model. On the other hand, the hyperband c), and Bayesian optimization d), methods produce models with lower performance. There are three samples with the negative depth shifts that seem to be challenging to estimate correctly regardless of the hyperparameter tuning method

used. For this example, based on the gamma-ray measurements, the best model is found using the random search solution with a MSE = 0.47 on the synthetically shifted test set.

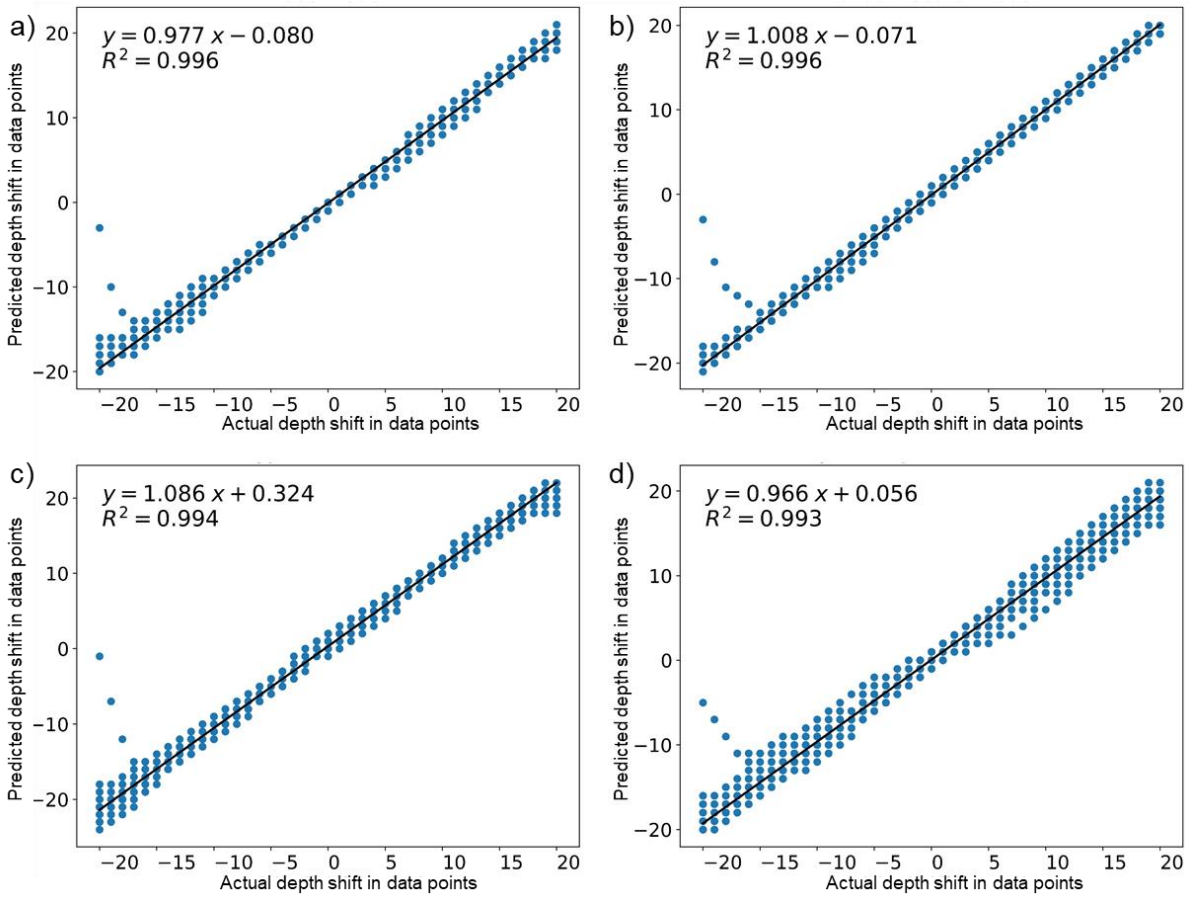

Fig. 4- Regression line on the test set, actual depth shift vs. that predicted with the gamma-ray model; a) initial model with MSE= 0.52; b) random search model with MSE = 0.47: c) hyperband model with MSE = 2.02; d) Bayesian optimization model with MSE= 1.02.

Notice that we do not use the real unseen LWD logs at this stage and use them only when the best model has been selected. Another important aspect is the limited amount of test data that we have, which is around 28 and 55 samples, depending on the well and the log type. For example, well 16/1-9 has the same number of samples for all log types, whereas well 16/1-21S has 55 samples for gamma ray, resistivity, and P-wave sonic, but only 32 samples for density, neutron, and PEF logs.

Depth shift inference

After model selection we take the whole set of synthetic shifted logs and split again into training and validation sets. The percentage of the validation varies between a range from 12 to 30 %, hence the

training also varies between 70 to 88 % depending on the log measurement. We proceed to retrain the best model and follow up the behavior of the MSE at each epoch with the validation set to detect any overfitting problem in addition to the batch normalization and drop-out layers. We perform the depth inference testing of our final models using a completely unseen dataset from well 16/1-9 and well 16/1-21S. These logs are considered raw since they have not been shifted, spliced, or merged, therefore the EWL and LWD/MWD logs are separate, and we use them to define our reference and shifted logs, respectively. The test set is preprocessed in the same way as the training set with the difference that we do not generate overlapping windows. We only extract single windows along the full depth range. Additionally, we perform a cross-correlation between the reference and the shifted log per window to estimate the depth lag, and we save these values in a depth shift vector for further evaluation and comparisons.

Because of the lack of ground truth labels for the two test wells we take a different approach to evaluate the results of the depth shift estimates. We compare some quantitative metrics, such as the Pearson correlation and Euclidean distance, and compute their averages over the total number of samples per log measurement, as well as qualitative visual inspection of log profiles before and after depth matching via CNN and cross-correlation. The Pearson correlation coefficient, *r*, measures how strong the relationship is between two variables or data sets assuming a linear relation between them. It takes values between -1 and 1. A Pearson correlation value of -1 indicates a perfect negative correlation, whereas a value of 1 implies perfect positive correlation. No linear correlation is indicated when the $r = 0$ or $r \approx 0$ (Bulmer, 1979). The Pearson correlation coefficient is calculated using:

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}} \tag{6}$$

where in our case n is window size (256 data points), x_i and y_i are the individual data points of the depth series within the window, and $\bar{x}$ and $\bar{y}$ are their corresponding mean values.

The Euclidean distance (d) is an alternative metric that measures the similarity between series (time or depth). It is calculated using Equation (7) (Herrera and van der Baan, 2014):

$$d_{(x,y)} = \sqrt{\sum_{i=1}^{n} \left(x_i - y_i \right)^2} \tag{7}$$

where in our application $d_{(x,y)}$ is the one-to-one Euclidean distance between the test (LWD/MWD) log x, and the reference (EWL) log y, and the index i represents the individual data points in each windowed depth series.

In addition, we use the Pearson correlation coefficient (r) to compute 5 indicators that help us to judge the quality of the depth matching using the following criteria:

C1. Number of samples in which the CNN matching improves the depth matching compared with the original log position ($r_{CNN} > r_{orig.}$).

C2. Number of samples in which the CNN matching worsens the depth matching compared with the original log position ($r_{CNN} < r_{orig.}$).

C3. Number of samples in which the CNN does not detect any mismatch of the logs ($r_{CNN} = r_{orig.}$).

C4. Number of samples in which the CNN matching performs better than the cross-correlation matching whenever criterion 1 is satisfied, i.e. there is a depth alignment improvement ($r_{CNN} > r_{cross\text{-}correlation}$).

C5. Number of samples in which the CNN and the cross-correlation perform equally well ($r_{CNN} = r_{cross\text{-}correlation}$).

Where $r_{orig.}$, r_{CNN}, and $r_{cross\text{-}correlation}$ are the Pearson correlation coefficients of the log segments before depth matching, after depth matching with CNN, and after depth matching with cross-correlation, respectively. Values of these indicators for well 16/1-9 and 16/1-21 S are shown in Table 3 and Table 4, respectively. The last two columns of the tables show $Ind_{1\%}$ and $Ind_{4\%}$, respectively. These are indicators C1 and C4 expressed as percentages of the total number of samples for each well. They are defined in Equation (8) and Equation (9).

$$\begin{aligned} Ind_{1\%} &= \frac{100}{N} \sum_{i=1}^{N} I\left(r_{CNN} > r_{\,orig.} \right) \\ I &= \begin{cases} if\ \left(r_{CNN} > r_{orig.} \right), I = 1 \\ if\ \left(r_{CNN} \le r_{orig.} \right), I = 0 \end{cases} \end{aligned} \tag{8}$$

$$Ind_{4\%} = \frac{100}{N}\sum_{i=1}^{N} I\left(r_{CNN} > r_{cross-correlation}\right)$$
$$I = \begin{cases} if\left(r_{CNN} > r_{cross-correlation}\right), I = 1 \\ if\left(r_{CNN} \le r_{cross-correlation}\right), I = 0 \end{cases} \quad (9)$$

where N is the total number of samples of each log measurement for each well.

RESULTS

We obtain 7 CNN models corresponding to the 7 log types. The architecture of the models and hyperparameters are the same for most of them except for the gamma-ray and S-wave sonic. However, each model has its own final trained weights. We test each of the 7 models by predicting depth shifts for their corresponding log types in wells 16/1-9 and 16/1-21S. Summaries of the results for these wells are shown in Table 3 and Table 4, respectively.

Table 3 shows that there are five log types for which CNN has achieved depth matching improvements compare to no matching of above 65 %, gamma-ray, resistivity, P and S wave sonic, and neutron. This percentage improvement is computed using Equation (8).Table 3 also shows that for the gamma-ray, resistivity and the P and S wave sonic logs, the CNN correction outperforms the cross-correlation in 46.43, 71.43, 57.14 and 46.43 % of the total number of samples, respectively. For the density and PEF logs the CNN percentage of depth matching improvement over no matching is less than 65 % and For the density, and PEF respectively, the percentages of samples for which CNN depth shifting outperforms the cross-correlation are much lower at 17.86, 35.71 and 21.43 %.

Results for well 16/1-21S (Table 4) are quite promising indicating that CNN can easily recognize the different log type patterns and identify misalignments among them. The CNN percentage of improvement ($Ind_{1\%}$) is above 80 % for all the log types in this well and even higher than 90 % and equal to 100 % for density and resistivity logs, respectively. The 100 % improvement of the resistivity implies that all samples after CNN correction have increased their Pearson correlation in comparison with their original values. The PEF log again shows the lowest improvements due to CNN depth matching as we have seen in the previous well. However, this value is still above 80 %, hence more than 20 % higher than in well 16/1-9.

The average percentage of all samples of all log types for which the CNN correction outperforms the cross-correlation for well 16/1-21S is 53.86 % but only 39.28 % for well 16/1-9. Where although P-wave sonic, neutron and PEF logs have high percentages of samples that improve their alignment after CNN correction, fewer than the half of samples have higher Pearson correlation values than after cross-correlation depth matching. This indicates that CNN can provide a sensible correction for depth misalignments, but we need to evaluate the overall behavior of other metrics to have a more reliable and conclusive assessment.

Table 3- Summary of Depth Matching Correction Using CNN for Well 16/1-9 According to Criteria (C1 -C5) Based on Pearson Correlation Coefficients Before and After Corrections for a Total of 28 Samples

CNN Model	# Improved samples CNN (C1)	# Worsened samples CNN (C2)	# Samples without change (C3)	# Samples CNN better than cross-correlation (C4)	# Samples CNN equal to cross-correlation (C5)	Matching improvements CNN better than no shift (%) ($Ind_{1\%}$)	Matching improvements CNN better than cross-correlation (%) ($Ind_{4\%}$)
GR	23	4	1	13	7	82.14	46.43
Resistivity	21	4	3	20	3	75.00	71.43
P-wave sonic	25	2	1	16	1	89.29	57.14
S-wave sonic	19	9	0	13	3	67.86	46.43
Density	16	10	2	5	6	57.14	17.86
Neutron	19	6	3	10	2	67.86	35.71
PEF	16	10	2	6	3	57.14	21.43

Table 4- Summary of Depth Matching Correction using CNN for Well 16/1-21S According to Criteria (C1 -C5) Based on Pearson Correlation Coefficients Before and After Corrections for a Total of 55 Samples (GR, resistivity and P-wave slowness) and 32 Samples (density, neutron, and PEF)

CNN Model	# Improved samples CNN (C1)	# Worsened samples CNN (C2)	# Samples without change (C3)	# Samples CNN better than cross-correlation (C4)	# Samples CNN equal to cross-correlation (C5)	Matching improvements CNN better than no shift (%) ($Ind_{1\%}$)	Matching improvements CNN better than cross-correlation (%) ($Ind_{4\%}$)
GR	48	6	1	39	4	87.27	70.91
Resistivity	55	0	0	47	6	100.00	85.45
P-wave sonic	45	9	1	23	7	81.82	41.82
Density	30	1	1	20	5	93.75	62.50
Neutron	27	4	1	12	5	84.38	37.50
PEF	26	6	0	8	2	81.25	25.00

For additional assessment of the results, we show example log profiles for some log types. For example, Fig. 5 shows window number 17 for gamma-ray, density, and neutron porosity logs in well 16/1-

9. We can see that the depth shifts suggested by CNN and cross-correlation are quite similar for these example logs and even identical for the neutron porosity log (Fig. 5c). Similarly, we can see the differences in patterns and values in some specific zones of the logs as is the case for the density, Fig. 5b, depth samples between 0 and 125. Despite consistent pre-processing of the data, resolution differences are seen between LWD and EWL logs. From the panels we see that the gamma-ray depth shifts using CNN appears visually better than the cross-correlation depth shifts. This is confirmed by the Pearson correlation values equal to 0.94 and 0.93, for CNN and cross-correlation respectively. In contrast, the CNN depth shifts to the density logs for this specific window is not good for this sample, reducing the Pearson correlation from 0.57 down to 0.47 (Fig. 5b). For the neutron porosity, both CNN and cross-correlation suggest the same correction of -1 data point, hence both increase the Pearson correlation to the same value from 0.69 up to 0.71.

Fig. 6 shows the same example log types for widow number 54 from well 16/1-21S. Again, the differences between the CNN and the cross-correlation depth shifts are small (here only differing by 1 data point), thus the final corrections are quite similar. However, the CNN is slightly visually and quantitatively better than the cross-correlation for gamma-ray and density logs (Fig. 6, and Fig. 6b, respectively). The Pearson correlation is 0.87 and 0.85 for CNN and cross-correlation, respectively for the gamma-ray. Their corresponding values for the density log are 0.99 and 0.98, respectively. For the neutron porosity logs, the panels shown in Fig. 6c are more challenging to judge by visual inspection. However, improvement in the depth alignment after depth matching is clear. In this case, the Pearson correlation value is higher for cross-correlation than for CNN at 0.82 and 0.78, respectively. Comparing the example logs in Fig. 5 and Fig. 6 from these two wells, the patterns in the LWD and EWL logs look more similar to each other in well 16/1-21S than in well 16/1-9.

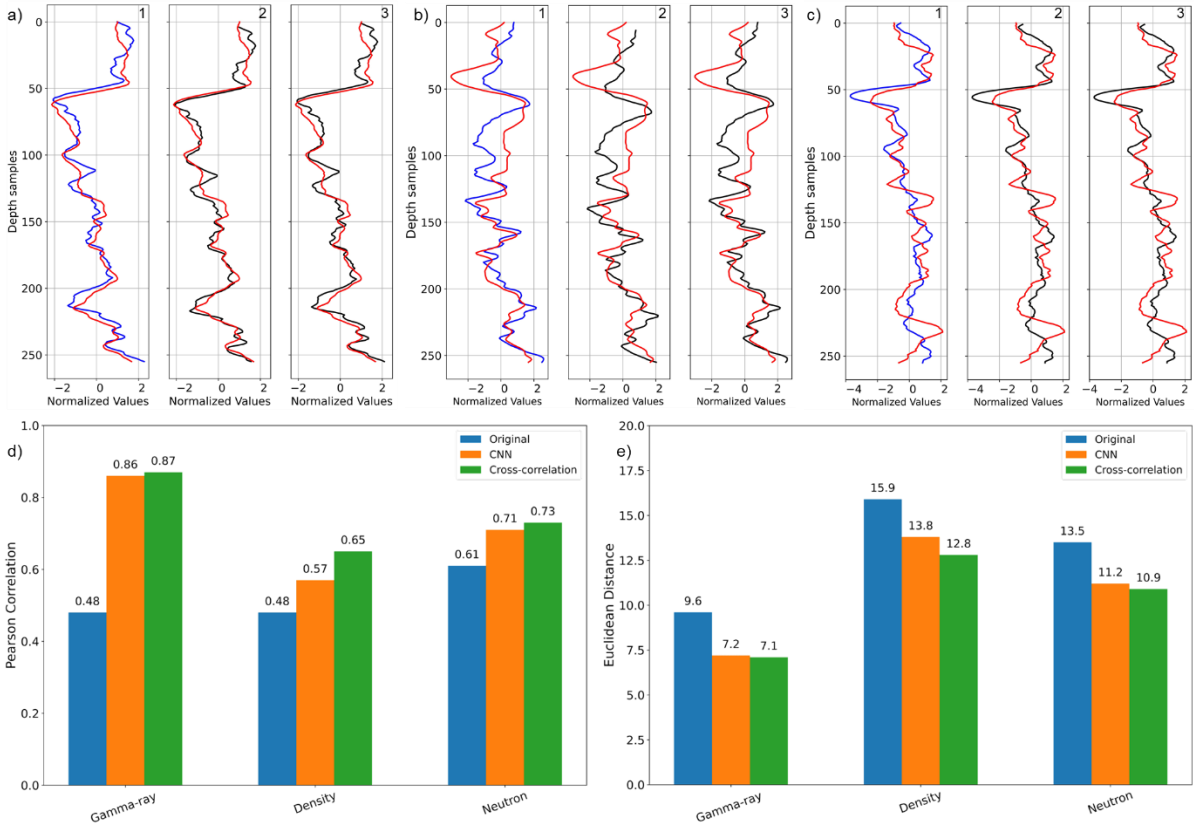

Fig. 5- Log profiles (window number 17) for well 16/1-9 and overall mean Pearson correlation and Euclidean distance metrics for each log type; a-1) Gamma-ray logs without correction, a-2) Gamma-ray after CNN correction of – 4 data points, a-3) Gamma-Ray after cross-correlation correction of -1 data points; b-1) density logs without correction, b-2) density logs after CNN correction of -6 data points, b-3) density logs after cross-correlation correction of 0 data points; c-1) neutron porosity logs without correction, c-2) neutron porosity logs after CNN correction of -1 data points, c-3) neutron porosity logs after cross-correlation correction of -1 data points, d) Pearson correlation overall results, e) Euclidean distance overall results. In panels a) to c) solid blue lines represent the uncorrected LWD curve, red is the reference EWL and black is the corrected LWD.

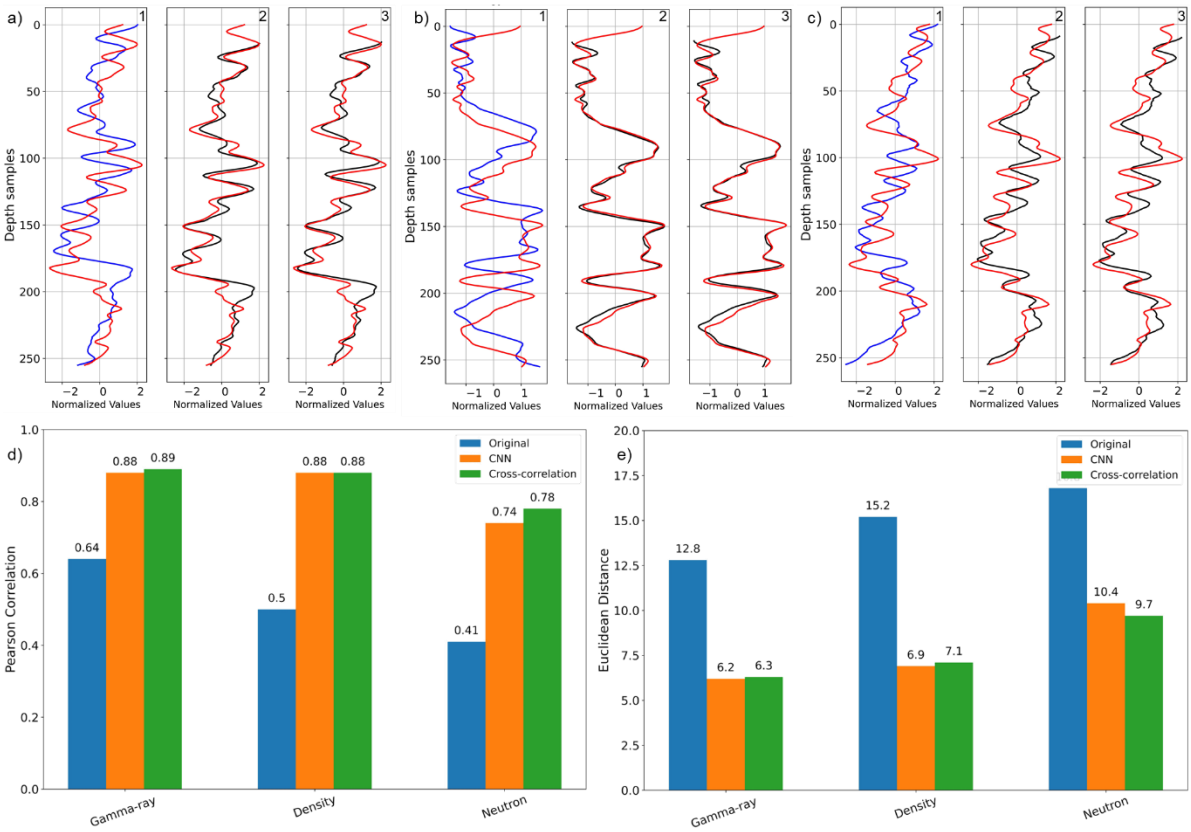

Fig. 6- Log profiles (window number 54) for well 16/1-21S and overall mean Pearson correlation and Euclidean distance metrics for each log type; a-1) Gamma-ray logs without correction, a-2) Gamma-ray after CNN correction of – 14 data points, a-3) Gamma-Ray after cross-correlation correction of -13 data points; b-1) density logs without correction, b-2) density logs after CNN correction of -12 data points, b-3) density logs after cross-correlation correction of -11 data points; c-1) neutron porosity logs without correction, c-2) neutron porosity logs after CNN correction of -9 data points, c-3) neutron porosity logs after cross-correlation correction of -10 data points, d) Pearson correlation overall results, e) Euclidean distance overall results. In panels a) to c) solid blue lines represent the uncorrected LWD curve, red is the reference EWL and black is the corrected LWD.

We want to get a clearer view of how useful the CNN is compared to another common approach to depth matching like the cross-correlation, for instance. We therefore computed the mean of the Pearson correlation, and the Euclidean distance of all the samples in the two test wells before and after depth correction with CNN and cross-correlation and compared them. Fig. 5d, Fig. 5e, Fig. 6d, and Fig. 6e show these comparison for the gamma-ray, density and neutron logs. In well 16/1 21S differences between CNN and cross-correlation results are insignificant. For well 16/1-9 the largest differences in Pearson correlation and Euclidean distance between the two methods are only 0.08 and 1, respectively.

From Fig. 5 and Fig. 6 we see that depth mismatches are successfully reduced by CNN. CNN is performing almost as well as cross-correlation in removing mismatches between LWD and EWL logs in real data. Therefore, we can see CNN as a suitable alternative to cross-correlation for the deployment of automatic workflows capable of handle large amounts of data simultaneously. However, we also see that CNN would benefit from improvement for specific log types where the cross-correlation shows better results

in some individual depth windows. The average value of the metrics over the whole dataset indicates that cross-correlation is superior in most of the cases, but CNN has still the potential to be further improved.

For completeness Fig. 7 shows the mean Pearson correlations and Euclidean distances for all log types in wells 16/1-9 and well 16/1-21S. The Pearson correlation increases after CNN depth matching for all log types. Similarly, the Euclidean distance is reduced for both wells. In well 16/1-9, gamma-ray, resistivity, P-wave sonic, and neutron logs show significant improvements with mean Pearson correlation of 0.86, 0.88, 0.57 and 0.71, respectively. However, for Euclidean distance, only gamma-ray and resistivity achieved values below 10. For well 16/1-9, the mean values of the CNN are only superior to the mean values of those for cross-correlation for the resistivity logs. The other log types such as gamma-ray and neutron yield metrics that are quite close to those from the cross-correlation, but they do not achieve superior values.

For well 16/1-21S we see similar trends to well 16/1-9, where the CNN depth matching improves the alignment significantly, showing values of Pearson correlation across log types above 0.7 for all but one log type. Only the PEF has a low mean Pearson correlation, of 0.36. This indicates poor performance. The mean Euclidean distance shows reductions down to the half of the original distance as is the case for the gamma-ray, resistivity, and density logs. Notice that for well 16/1-21S the CNN depth matching overall performs slightly better than the cross-correlation for the resistivity logs and is quite competitive for the gamma-ray, P-wave sonic, and neutron logs with differences in Pearson correlation of only 0.01, 0.06, and 0.04, respectively. The density log shows the same mean Pearson correlation of 0.88 after CNN depth matching and cross-correlation. The reduction in Euclidean distance in well 16/1-21S observed after CNN depth matching is larger for three log types, gamma-ray, resistivity, and density, than for the others. In general, CNN has a better overall performance in well 16/1-21S than in well 16/1-9. We can also distinguish the log types in which CNN performances better, e.g. resistivity, from those for which it gives poorer results, e.g. PEF.

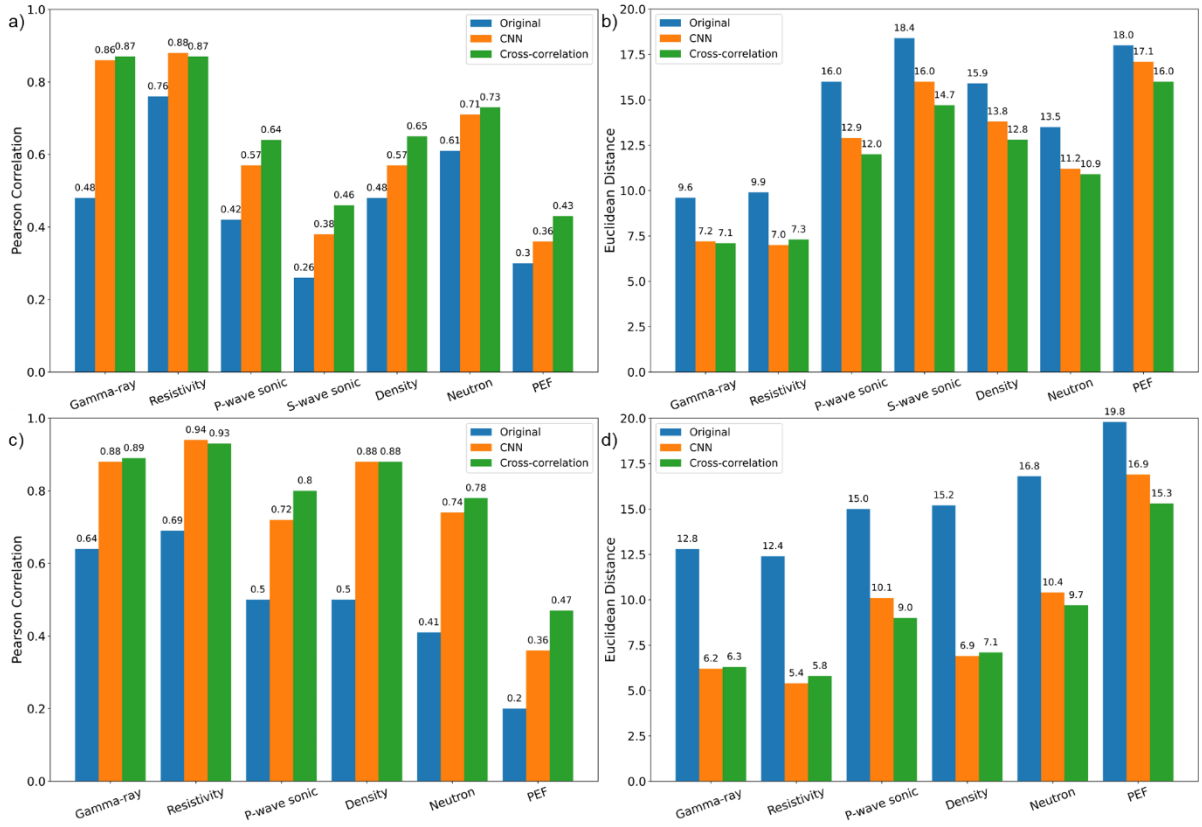

Fig. 7 - Comparison of the overall mean Pearson correlation and Euclidean distance metrics before and after depth correction using CNN and cross-correlation for each log type in well 16/1-9 and well 16/1-21S; a) Pearson correlation overall results for well 16/1-9; b) Euclidean distance overall results for well 16/1-9; c) Pearson correlation overall results for well 16/1-21S; d) Euclidean distance overall results for well 16/1-21S.

DISCUSSION

The hyperparameter tuning is the most expensive stage of any deep learning implementation and the time needed to perform this is highly dependent on the amount of data, the model complexity, the search strategy, and the available computational resources. The results of testing three different search strategies for model selection show their relative efficiency in terms of time and performance on a test set. For most of the log types the best models were found by the Hyperband algorithm. This is also a faster algorithm than random search and Bayesian optimization. Random search is the second-best algorithm for model selection for this specific case. For example, gamma-ray, P-wave, and S-wave sonic logs' best models are obtained using random search. In some cases, we see differences of about 20 - 30 minutes in training time between random search and Hyperband. In general, Bayesian optimization is the costliest process and yields higher MSE errors for all the log type models (See Table 5).

Table 5- Hyperparameter Tunning Execution Times and MSE Values for Each Log Type and Model Selection Algorithm

CNN Model	Model Selection strategy	Execution Time (min)	MSE on the test set
Gamma-ray	Random Search	126.40	0.47
	Hyperband	121.13	2.02
	Bayesian Optimization	237.51	1.01
Resistivity	Random Search	176.29	3.17
	Hyperband	139.74	0.68
	Bayesian Optimization	239.00	4.19
P-wave sonic	Random Search	126.61	0.92
	Hyperband	85.98	0.56
	Bayesian Optimization	157.60	3.89
S-wave sonic	Random Search	184.94	0.73
	Hyperband	49.13	1.00
	Bayesian Optimization	205.29	6.39
Density	Random Search	89.56	1.87
	Hyperband	42.86	1.00
	Bayesian Optimization	101.33	2.38
Neutron	Random Search	69.72	3.75
	Hyperband	44.49	1.13
	Bayesian Optimization	91.26	4.90
PEF	Random Search	90.80	1.24
	Hyperband	45.53	0.57
	Bayesian Optimization	89.29	3.31

For the depth shift results, even though we have relatively few test samples (only 28 and 55/32 for well 16/1-9 and 16/1-21S, respectively) making the establishment of general trends difficult, these results allow us to make some important observations. We see that the differences between EWL and LWD/MWD logs are a key feature that affects the depth matching process based on pattern recognition using our implementation. These differences between the reference EWL and the LWD logs are associated with several factors. For example, their vertical resolution differs due to differences in logging speed since LWD is slower than the EWL, and this for a gamma-ray log means higher accuracy because there is an increase in the number of samples that are averaged and assigned to a given depth. Another example is difference in the tools' characteristics e.g. dual sensor versus single sensor, as is the case for well 16/1-9 where we observe a slightly higher resolution in the LWD compared to the equivalent EWL log. Despite the preprocessing of the logs, which includes filling in of missing data intervals, spike removal, filtering/smoothing and normalization/standardization to compensate for differences in resolution, missing data, spikes, and value range discrepancies, we can still see pattern differences between the LWD and EWL log responses that affect

the CNN performance (See Fig. 5b density panel). In addition, there are environmental factors that change the borehole conditions over time. For this reason, even after standard corrections have been applied to both logs, they might differ due to temporal changes in the borehole conditions between the acquisition of the comparable LWD and EWL logging runs. The impact of this effect depends on how much the relevant borehole conditions have changed during the delay between the comparable LWD and EWL logging runs. This means that depending on the time delay between acquisition of comparable LWD and EWL logging runs the logging measurements might be significantly affected despite proper corrections. This can have additional implications for the way contractors implement environmental corrections to log data.

A clear example of this is shown in Fig. 8. This shows quality control log panels for window 17 in well 16/1-9, which is the same window as is shown in Fig. 5. In Fig. 8, the log values are neither normalized nor standardized, so it is easier to see their values and their pattern differences. In general, the borehole size in this section is larger than the reference bit size (Fig. 8a). However, a larger diameter was measured during the EWL logging in the shaly zones in the shallower part of the section, approximately the first 50 data points. Then there is a cleaner section down to about data point 125 where both diameters are similar and we see overlapping and closer values and similar patterns for density (panel c) and PEF (panel e) logs, respectively. Also, we see smaller differences between the neutron logs in areas with similar borehole diameter. On the other hand, no significant pattern differences are seen between the gamma-ray logs in the whole window (panel b). This is as expected since this log type has a higher depth of investigation, therefore an enlargement of the borehole diameter has a much smaller impact on the gamma-ray log. However, we see a constant shift between LWD and EWL logs that might be associated with different constant corrections for mud type, for instance KCL mud. From sample around 125 we observe gradual increase in shale/clay content, and a steady enlargement of the borehole for EWL, which reduces the density curve and increases the PEF and neutron logs, as well as perhaps causing slight pattern discrepancies. The density correction curve is a good quality control for the density and PEF curves as it is highly sensitive to the borehole diameter and mud type. Here we see that the LWD correction is mostly constant along the section and positive. In contrast, the EWL correction increases and decreases as function of the EWL borehole diameter changes, e.g. in the zone between 50 and 125 samples depth there is almost no correction for the EWL but there is a positive correction for the LWD. Also note that, during LWD, tools are run centered, therefore the stand-off

of the tools must be accounted for. In contrast, the EWL density-PEF, and neutron logs are run eccentered through pads pressed against the borehole wall. Moreover, at this depth they used water-based mud with 12% barite. Barite is an element with a high electron density; hence it has a higher interaction with the gamma-rays emitted from the density source and fewer gamma-rays will be detected by the receptors, thus underestimating the density of the formation. The density correction should be positive to compensate this effect. Similarly, the barite content will have a negative impact on the PEF values since it has such an efficient capacity to absorb gamma-rays in lower energy levels, that it is difficult to measure them accurately. Also, the barite has a PEF of 267 barns/electron compared with most of the lithologies that have PEF values of less than 6 barns/electron. Consequently, the barite content in the mud leads to unreliable PEF values.

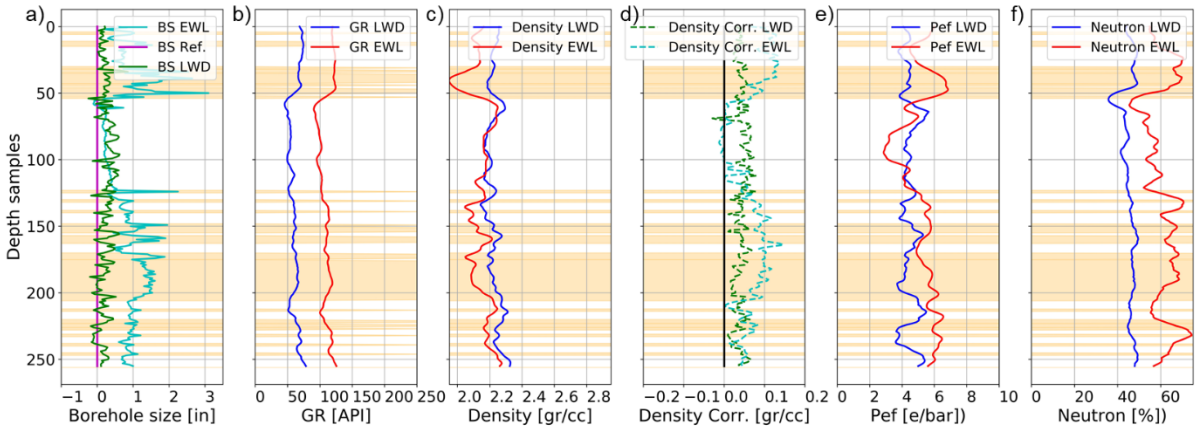

Fig. 8 – Quality control panels showing window 17 from well 16/1-9; a) Borehole size relative to a reference, b) gamma-ray logs, c) density logs, d) density correction logs, e) PEF logs, f) neutron porosity logs. Orange indicates the zones where borehole size during EWL logging is larger than the reference bit size (magenta) by more than 1 inch.

Equivalent quality control panels are presented in Fig. 9 for well 16/1-21S. These show window/sample number 54; hence they are the same logs as are shown in Fig. 6, except that in Fig. 9 the logs are neither normalized nor standardized. This emphasizes any differences in their values. We see from panel a) that during LWD the borehole was in gauge along the whole section and after a couple of hours a mud cake was built-up, which is detected during the EWL acquisition. Small negative corrections are applied to the EWL density log to remove the effects of the mud cake on the density values (panel d). The opposite applies for the LWD where slightly larger positive corrections are needed to compensate the tool stand-off, even though the borehole is in good condition. We can see that only pattern differences between LWD and EWL are very small across the log types, even for those with a shallower depth of investigation, such as

density and PEF. On the PEF logs (panel e), we see a mostly constant shift of the LWD toward higher values than its equivalent EWL. This might be related to the cross-section properties of the mud cake and possible corrections applied to the data before we received it. Note that in this section an oil-based mud was used. It contained some barite traces, which negatively affect the PEF log.

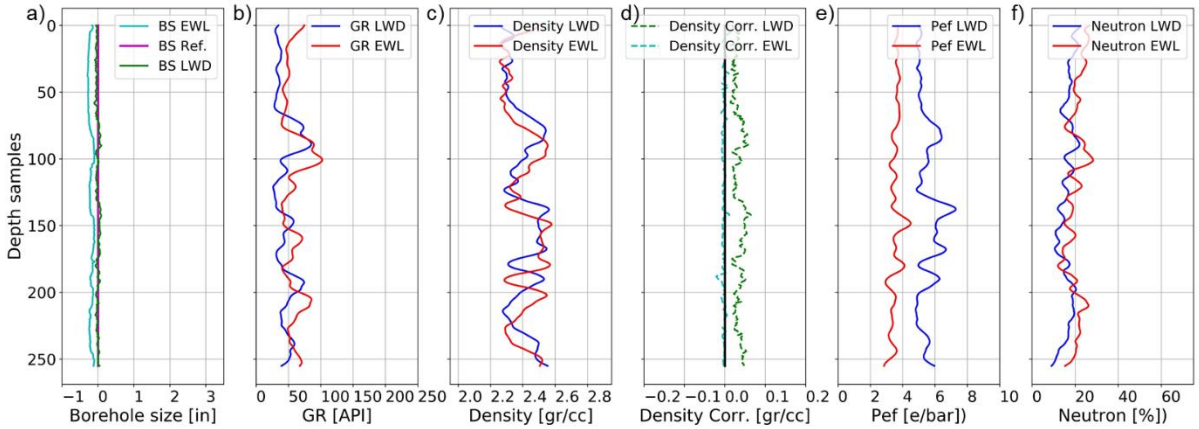

Fig. 9 - Quality control panels showing window 54 from well 16/1-21S; a) Borehole size relative to reference, b) gamma-ray logs, c) density logs, d) density correction logs, e) PEF logs, f) neutron porosity logs.

Large gaps of missing data can also create problems for our approach as we use a quick interpolation technique which might cause pattern differences. Thus, our CNN method can struggle to find a correct depth shift, as we saw for the sonic logs, especially the shear-wave slowness. Another important consideration is the stretch/squeeze effects associated with sticking and slipping of the wireline cable. This effect was not included in the training data but is a common problem during logging process. Examples of that are seen in Fig. 5b, Fig. 5c, Fig. 6b and Fig. 6c. The gamma-ray log seems not to be significantly affected by this, unlike the density, and neutron logs, or even other log types like the sonic logs. This might also be related to the depths at which the tool gets stuck and released, and the relative position of the sensors along the tool string.

In general, our CNN has a better performance on logs from well 16/1-21S than on those from well 16/1-9. This maybe because in the former well, both log acquisition stages were carried out by the same logging company. So even though the tools used are different, due to the nature of the LWD and EWL acquisition processes the general processing workflows and corrections are likely more similar. Thus, it reduces patterns differences considerably. Well 16/1-21S is also newer and more sophisticated tools are likely to have been used, as well as there has been little change in the borehole conditions between LWD and EWL acquisitions. When the differences between the reference log and shifted log are smaller the CNN

and cross-correlation solutions differ slightly, and the CNN outperforms the cross-correlation in several individual samples. However, the overall improvement measured with the metrics used (Pearson correlation and Euclidean distance) seems to favor the cross-correlation in both wells.

From these insights, it is possible to propose changes to our CNN implementation workflow that could potentially improve the performance of the method beyond that of cross-correlation, at the expenses of more complex and realistic training models. For instance, we could try to account for dynamic shifts (stretch/squeeze) and introduce more variability between LWD and EWL signals due to borehole conditions changes into the training data via more sophisticated data augmentation procedures. This might allow the use of CNN to depth match logs within a fully automated workflow.

CONCLUSIONS

We have demonstrated a simple and practical implementation of 1D CNN as a possible tool to perform well log depth matching of LWD and EWL measurements for different log type pairs gamma-ray, resistivity, P-wave and S-wave sonic, density, neutron, and PEF logs. It requires little user intervention when the network is trained. We demonstrated that 1D CNN can detect depth misalignments and suggest sensible depth shift corrections between raw log curves from different runs through the same wellbore interval, based on pattern similarities. The main advantage of our approach is that it does not required extensive and complex pre-processing of the data, nor does it require any feature engineering, which is a time-consuming task and requires a high level of expertise and domain knowledge. The CNN, during the training process automatically performs feature extraction from the raw data and identifies relevant patterns in the logs. We also show that the training and hyperparameter tuning process is the most expensive task in terms of computer resources and time. In general, most of the models used in this work were output by hyperband and random search hyperparameter tuning algorithms, which were much faster training processes than Bayesian optimization. The training and model selection processes for each log type model varies according to the number of samples, model complexity and search strategy, and it takes between 45 minutes and almost 4 hours. However, once the models are in place, the inference process is performed in less than a minute, which is another advantage of our proposed method, potentially saving hours or days of work for a petrophysicist or rock physics practitioner.

Our results were quality controlled through comparing quantitative metrics, qualitative visual inspection of log profiles comparing the original positions of the logs and their updated versions before and after depth matching using CNN or cross-correlation. We demonstrated that most of the windows improved in depth alignment after CNN corrections, however, the overall improvements using CNN are not better than cross-correlation for most of the log pairs. An exception is the resistivity log for both wells. Even though CNN is not superior overall to cross-correlation for well depth shift estimations, several models are quite competitive, which implies that there is room for improvements in the models and in the implementation of the algorithm. We also found that CNN struggles to find good solutions when discrepancies between LWD and EWL logs are large, as well as when substantial stretch/squeeze effects exist. This is related to our training data set having been limited to containing only single bulk depth shifts and the use of shifted copies of the logs for training. This limited the capacity of the CNN to deal with stretch/squeeze and to recognize slightly different patterns between logs. Pattern differences are commonly associated with differences in borehole conditions, tool technology, and processing techniques, between LWD and EWL log runs.

We implemented our CNN models to estimate the depth shifts between logs from two wells in the Norwegian North Sea, as a proof of concept. We plan to further investigate the use of CNN in a more general context by testing whether or not a CNN model can be trained and internally validated on a specific log type, and then used to estimate accurate the depth shifts for other log type pairs. This could save significant amounts of work and time if we can avoid training to build individual models for each log type. We do plan, further investigation with improved training sets that includes more of the variations and effects that are present in real data examples.

ACKNOWLEDGEMENTS

This research is part of the BRU21 – NTNU Research and Innovation Program on Digital and Automation Solutions for the Oil and Gas Industry (www.ntnu.edu/bru21) and is supported by AkerBP. We thank also NTNU-NPD-Schlumberger Petrel ready dataset for borehole data, and Andrew J. Carter for his suggestions for improving this manuscript. We would like to thank the reviewers and corresponding editor for their valuable feedback, as well as their suggestions for improving substantially the content of this paper.

NOMENCLATURE

Abbreviations

ANN = artificial neural network

CNN = convolutional neural network

EWL = electrical wireline logging

GR = gamma ray

KCL = potassium chlorine

LAS = log ASCII standard

LWD = logging while drilling

MSE = mean square error

MWD = measurements while drilling

PEF = photoelectric factor

ReLU = rectify linear unit

REFERENCES

Abdoli, S., Cardinal, P. & Koerich, A. L., 2019, End-to-end environmental sound classification using a 1D convolutional neural network, *Expert Systems with Applications*, **136,** 252-263. DOI: 10.1016/j.eswa.2019.06.040

Anderson, K. R. and J. E. Gaby, 1983, Dynamic waveform matching, *Information Sciences*, **31**(03), 221-242. DOI:10.1016/0020-0255(83)90054-3.

Bergstra, J., Bardenet, R., Bengio, Y. & Kégl, B., 2011, Algorithms for hyper-parameter optimization, Paper presented at NIPS 25th Annual Conference on Neural Information Processing Systems, Granada, Spain, 12-17 December.

Bergstra, J. & Bengio, Y., 2012, Random search for hyper-parameter optimization, *Journal of Machine Learning Research,* **13**(2), 281-305.

Bishop, C. M., 2006, *Pattern recognition and machine learning*, Springer.

Bolt, H., 2016, Wireline Logging Depth Quality Improvement: Methodology Review and Elastic-Stretch Correction, *Petrophysics*, **57**(03), 294-310.

Bulmer, M. G., 1979, *Principles of statistics*, Courier Corporation.

Brazell, S., Bayeh, A., Ashby, M. & Burton, D., 2019, A machine-learning-based approach to assistive well-log correlation, *Petrophysics,* **60**(04)**,** 469-479. DOI: 10.30632/PJV60N4-2019a1

Chia, C. R., H. Laastad, A. V. Kostin, F. Hjortland, and G. A. Bordakov, 2006, A new method for improving LWD logging depth, Paper 102175-MS presented at SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA, 24-27 September. DOI: 10.2118/102175-MS.

Chollet, F., 2015. Keras documentation. *keras. io*, *33*.

Deng, T., Xu, C., Lang, X. & Doveton, J., 2021, Diagenetic Facies Classification in the Arbuckle Formation Using Deep Neural Networks, *Mathematical Geosciences***,** 1-22. DOI: 10.1007/s11004-021-09918-0

Herrera, R. H. and M. van der Baan, 2014, A semiautomatic method to tie well logs to seismic data, *Geophysics*, **79**(03), V47-V54. DOI:10.1190/geo2013-0248.1.

Hoshen, Y., Weiss, R. J. & Wilson, K. W., 2015, Speech acoustic modeling from raw multichannel waveforms, Paper presented at IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, 19-24, April. DOI: 10.1109/ICASSP.2015.7178847

Imamverdiyev, Y. & Sukhostat, L., 2019, Lithological facies classification using deep convolutional neural network, *Journal of Petroleum Science and Engineering,* **174,** 216-228. DOI: 10.1016/j.petrol.2018.11.023.

Ioffe, S. & Szegedy, C., 2015, Batch normalization: Accelerating deep network training by reducing internal covariate shift, Paper PMLR 37:448-456 presented at International conference on machine learning, Lille, France, 7-9, July. 2015.

Kerzner, M. G., 1984, A solution to the problem of automatic depth matching, Paper SPWLA-1984-VV presented at SPWLA 25th Annual Logging Symposium, New Orleans, Louisiana, USA, 10-13 June.

Kingma, D. P. & Ba, J., 2015, Adam: A method for stochastic optimization, Paper presented at ICRL The 3rd International Conference for Learning Representations , San Diego, CA, USA, 7-9, May.

Le, T., L. Liang, T. Zimmermann, S. Zeroug, and D. Heliot, 2019, A Machine-Learning Framework for Automating Well-Log Depth Matching, *Petrophysics*, **60**(05), 585-595. DOI:10.30632/PJV60N5-2019a3.

Lecun, Y., Bengio, Y. & Hinton, G., 2015, Deep learning, *Nature*, **521**(7553)**,** 436-444. DOI: 10.1038/nature14539.

Lecun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. & Jackel, L. D., 1989, Backpropagation applied to handwritten zip code recognition, *Neural computation*, **1**(4)**,** 541-551.

Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P., 1998, Gradient-based learning applied to document recognition, *Proceedings of the IEEE*, **86**(11)**,** 2278-2324.

Li, L., Jamieson, K., Desalvo, G., Rostamizadeh, A. & Talwalkar, A., 2017, Hyperband: A novel bandit based approach to hyperparameter optimization, *The Journal of Machine Learning Research*, **18**(1)**,** 6765-6816.

Luthi, S. M. & Bryant, I. D., 1997, Well-log correlation using a back-propagation neural network, *Mathematical Geology*, **29**(03)**,** 413-425.

O'malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H. & Invernizzi, L., 2019, Keras Tuner, *Github.* Accessed on 28 November 2020.

Salamon, J. & Bello, J. P., 2017, Deep convolutional neural networks and data augmentation for environmental sound classification, *IEEE Signal processing letters*, **24**(03)**,** 279-283. DOI: 10.1109/LSP.2017.2657381

Sollie, F. O. and S. G. Rodgers, 1994, Towards better measurements of logging depth, Paper SPWLA-1994-D presented at the SPWLA 35th Annual Logging Symposium, Tulsa, Oklahoma, USA, 19-22 June.

Theys, P. P., 1999, *Log data acquisition and quality control*, Editions Technip.

Thornton, C., Hutter, F., Hoos, H. H. & Leyton-Brown, K., 2013, Auto-WEKA: Combined selection and hyperparameter optimization of classification algorithms, Paper presented at the 19th ACM SIGKDD international conference on Knowledge discovery and data mining, Chicago, Illinois, USA, 11-14, August. DOI: 10.1145/2487575.2487629.

Torres, V., Duffaut, K., Stovas, A., Westad, F. O. & Johansen, Y. B, 2020, Automation of depth matching using a structured well-log database: Prototype well example in the North Sea, Paper presented at the SEG 90th Annual Meeting, Houston, Texas, USA, 11-16, October. DOI: 10.1190/segam2020-3424928.1.

Wang, S., Shen, Q., Wu, X. & Chen, J., 2020, Automated gamma-ray log pattern alignment and depth matching by machine learning, *Interpretation,* **8**(3)**,** SL25-SL34. DOI: 10.1190/INT-2019-0193.1.

Wilson, H., J. Lofts, G. Page, A. Brooks, and D. Walder, 2004, Depth control: Reconciliation of LWD and wireline depths, standard practice and an alternative simple but effective method, Paper SPE-89899-MS presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas, USA, 26-29 September. DOI: 10.2118/89899-MS.

Yamashita, R., Nishio, M., Do, R. K. G. & Togashi, K., 2018, Convolutional neural networks: an overview and application in radiology, *Insights into imaging,* **9**(4)**,** 611-629. DOI: 10.1007/s13244-018-0639-9.

Zangwill, J., 1982, Depth Matching - A Computerized Approach, Paper SPWLA-1982-EE presented at SPWLA 23rd Annual Logging Symposium, Corpus Christi, Texas, USA, 6-9 July.

Zhu, L., Li, H., Yang, Z., Li, C. & Ao, Y., 2018, Intelligent Logging Lithological Interpretation with Convolution Neural Networks, *Petrophysics,* **59**(06)**,** 799-810. DOI: 10.30632/PJV59N6-2018a5

Zimmermann, T., Liang, L. & Zeroug, S., 2018, Machine-learning-based automatic well-log depth matching. *Petrophysics,* **59**(6)**,** 863-872. DOI:10.30632/PJV59N6-2018a10.

ABOUT THE AUTHORS

Veronica A. Torres C. received a B.Sc. (2010) in geophysical engineering from Central University of Venezuela, Caracas, and a M.Sc. (2017) in petroleum geophysics from Norwegian University of Science and Technology (NTNU) in Norway. She is currently a Ph.D. student at NTNU. Her current research interests include data structuring, petrophysics, rock-physics, seismic inversion, machine learning, statistical learning, seismic attenuation, and data integration. From 2011 to 2012 she worked as a geophysicist at IHS as seismic interpreter and technical support for Kingdom Suit. From 2012 to 2015 she worked as a geophysicist seismic interpreter and velocity model builder for subsalt imaging at PGS, Mexico. From 2017 to 2018 she worked as a research assistant on attenuation estimates and rock physics model simulation. She was the president of the SEG Student Chapter in NTNU from 2019 to 2021, and she is currently student member of SEG, EAGE, and SPWLA.

Kenneth Duffaut received his M.Sc. (1994) in geophysics, at the Norwegian University of Science & Technology (NTNU), Norway, and a Ph.D. (2011) in geophysics, at NTNU. From 1995 to 1997 he worked as a geophysicist in Statoil's petroleum technology department in Norway within field development and geophysical special services. From 1997 to 2015 he worked in Statoil research centre in Trondheim, Norway, as a research geophysicist in various exploration and production programs. Since 2015 he has worked as an associate professor at the Department of Geoscience and Petroleum at NTNU. Current research interests include landscape dynamics through geological time (tectonic uplift), near surface geophysics, 4D seismic monitoring, wellbore geophysics and rock physics analysis combined with machine learning for well and seismic data analysis.

Anis Yazidi received the M.Sc. and Ph.D. degrees from the University of Agder, Grimstad, Norway in 2008 and 2012, respectively. He was a Senior Researcher with NORCE, Norway. From 2014 to 2019, he was an Associate Professor with the Department of Computer Science, Oslo Metropolitan University, Oslo, Norway, where he is currently a Full Professor and leading the Research Group in Applied Artificial

Intelligence. He has participated in different EU and RCN projects, and he is leading the AI work package in the EU AI-Mind project. He is also co-coordinating the Excellence Academic Environment NordSTAR at OsloMet. He holds a Professor II position with the Norwegian University of Science and Technology (NTNU), Trondheim, Norway and he is Senior Researcher at Oslo University Hospital.

Frank O. Westad, born 1959, received his M.Sc. in Chemistry and Data analysis from the University in Trondheim, Norway, in 1988, and completed his Ph.D. thesis "Relevance and parsimony in multivariate modelling" in 2000. His working experience includes Senior Research Scientist positions at GE Healthcare and the Norwegian Food Research Institute (Nofima). He has published 50 papers and has also written several chapters in various textbooks on multivariate data analysis. He is now an adjunct professor at the Norwegian University of Science and Technology in the Department of Engineering Cybernetics, and holds a position as Chief Data Analyst at Idletechs AS.

Yngve Bolstad Johansen received a Cand.Scient. degree in Physics from NTNU, Norway in 2000. He joined Schlumberger in 2001 as Field Engineer and worked with data acquisition in the Middle East and North Sea before starting as Logging Tool Physicist at the Houston Formation Evaluation Center in 2005. In 2009 he started as Principal Petrophysicist for Statoil in field development and later R&D. From 2013 to 2020 Yngve was leading Petrophysics within AkerBP in the position Chief Petrophysicist. Yngve Johansen is currently Principal Advisor and coordinates subsurface R&D across all disciplines in AkerBP.