Research article

Geoenergy

https://doi.org/10.1144/geoenergy2025-009 | Vol. 3 | 2025 | geoenergy2025-009

Deep learning-assisted borehole image analysis for enhanced geothermal reservoir evaluation: a case study in the West Netherlands Basin

Attilio Molossi[1]*, Emilio Cecchetti[2,3], Pierre-Olivier Bruna[3] and Michele Pipan[1]

[1] Department of Mathematics, Informatics and Geosciences, University of Trieste, Via Edoardo Weiss 2, 34128 Trieste, Italy
[2] Sproule, Stationsplein 6, 2275 AZ Voorburg, South Holland, The Netherlands
[3] Faculty of Geoscience and Engineering, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands
AM, 0000-0002-0218-8022; EC, 0009-0004-3186-2313; P-OB, 0000-0001-9403-8955; MP, 0000-0003-3095-6305
* Correspondence: attilio.molossi@phd.units.it

Abstract: Evaluating the geothermal reservoir potential often requires fracture analysis, as fractures serve as key pathways for fluid flow in subsurface formations. Borehole images (BHIs) are essential for this analysis, providing 2D representations of boreholes with millimetre-scale resolution. However, their interpretation is highly subjective, leading to uncertainties in the results and the subsequent quantitative assessment of the fracture networks. In the West Netherlands Basin (WNB), accurate fracture characterization is critical for assessing the geothermal viability. However, the traditional manual interpretation of BHIs has shown inconsistencies. This study introduces a supervised deep learning (DL) approach to support fracture analysis using high-resolution formation micro-imager (FMI) data from the Naaldwijk well (NLW-GT-01). The proposed DL-based system integrates a U-Net model (PickNet) for segmentation and a fully connected convolutional network (FitNet) for automated feature extraction. Initially trained on synthetic low-resolution BHIs, the model has been adapted for FMI data using two approaches: (1) transfer learning and (2) a simplified adaptation method that involves resizing the FMI input, leading to some resolution loss. A comparison of these approaches has revealed that the simplified adaptation produces better results, closely aligning with conservative manual interpretations calibrated with core samples while enabling more detailed fracture detection. To enhance reliability, we propose a semi-automated human–machine collaboration framework, where experts validate or refine the automatically detected features. This approach leverages human expertise to improve interpretation accuracy while addressing challenges related to robustness and redundancy in the supervised learning model.

Thematic collection: This article is part of the sustainable future of geoenergy in the hands of early career researchers collection available at: https://www.lyellcollection.org/content/geoenergy-early-career-research

Received 19 March 2025; **revised** 25 June 2025; **accepted** 21 July 2025

Geothermal energy plays a significant role in The Netherlands' sustainable energy portfolio, with recoverable heat resources estimated to be approximately 55 times greater than the country's annual energy demand (Kramers *et al.* 2012; Willems 2017). Notably, the West Netherlands Basin (WNB) stands out as a prime region for geothermal energy development (Wong *et al.* 2007; Kombrink *et al.* 2012). The WNB contains deep aquifers, including Triassic formations such as the Main Buntsandstein Subgroup, where fluid flow is primarily governed by fractures due to limited primary porosity (Pluymaekers *et al.* 2012). Borehole image (BHI) logs and core samples are crucial for geothermal projects as they reveal natural fractures in the subsurface (Kato and Sakagawa 1995; Prensky 1999; Okada and Yamada 2002; Boersma *et al.* 2021; Faiq Adenan *et al.* 2023; Mutonga and Fujimitsu 2024), which are vital for effective fluid movement in complex geological formations and enhancing formation permeability (Vidal and Genter 2018; Boersma *et al.* 2021). However, the interpretation of these data is inherently subjective and can introduce uncertainties that complicate resource evaluation (Bond *et al.* 2008; Bond 2015; Peacock *et al.* 2019; Shipton *et al.* 2020; Alcalde and Bond 2022). Moreover, the interpretation process involves significant technical and financial costs, with fracture detection being particularly challenging (Ozkaya and Mattner 2003; Blanco Valentin 2018). This has led to increasing interest in developing advanced methods to improve the accuracy and consistency of fracture interpretations, which are often hindered by human subjectivity (Akkurt *et al.* 2018; Gupta *et al.* 2019). BHIs, which provide 2D representations of the borehole wall, are used to derive critical petrophysical parameters for reservoir characterization (Basu *et al.* 2019; Maerten *et al.* 2019). Despite their high resolution, BHIs are not exempted from uncertainty, particularly in fracture analysis, as highlighted in several studies (Mondol 2015; Ozol *et al.* 2015; Yang *et al.* 2020; Boersma *et al.* 2021). This uncertainty affects the reliability of fracture characterization, which is essential for subsequent scientific and technical assessments of geothermal reservoirs (Pollack 2007; Maeso *et al.* 2015; Lukawski *et al.* 2016; Witter *et al.* 2019).

Artificial intelligence (AI) has emerged as a valuable tool in geosciences, where it has demonstrated efficacy in addressing complex challenges that human knowledge alone may struggle to solve (Chen *et al.* 2012; Anemangely *et al.* 2019; Roncoroni *et al.* 2022*a*, *b*; Pantaleo *et al.* 2024). The advantages of AI include its ability to process vast amounts of data and generate informative representations from specific inputs. Deep learning (DL), for instance, utilizes input representations to learn how to predict either predefined (supervised learning) or unspecified (unsupervised learning) outputs from raw data across multiple levels of representation (LeCun *et al.* 2015; Agarap 2019). This makes DL particularly suitable for automating or semi-automating BHI analysis, where subjective fracture interpretations benefit from enhanced accuracy and consistency (Gupta *et al.* 2019). The traditional DL applications are normally trained and tested on datasets with the same data distribution, as the performance of a

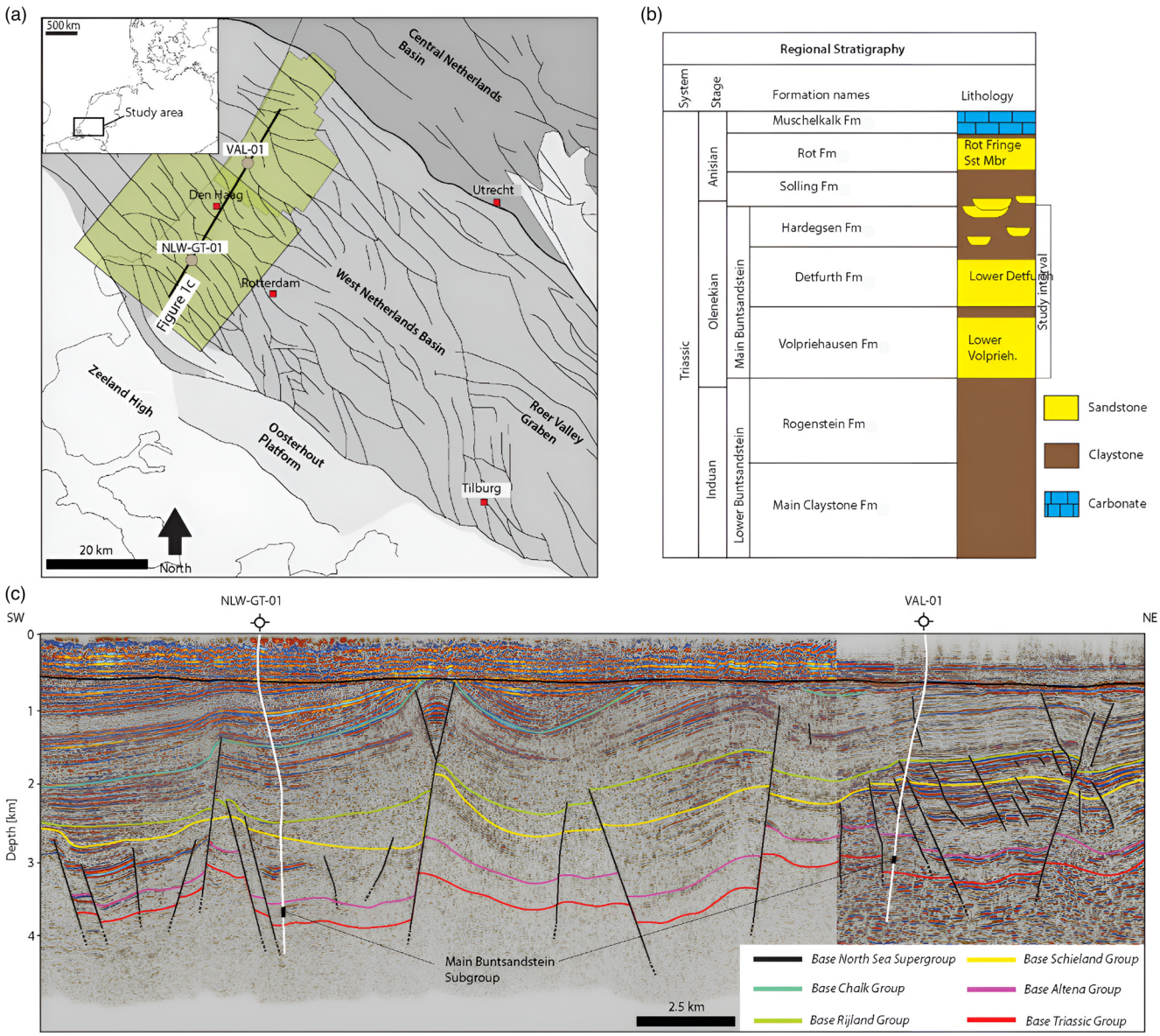

Fig. 1. (**a**) Study area with wells used in the study and seismic surveys (L3NAM1989J and L3NAM1990C) used to interpret the subsurface structures. (**b**) Simplified stratigraphic column for the lower and upper Triassic sediments (after Cecchetti *et al.* 2024). (**c**) Composite seismic section displaying the main structural features and main stratigraphic group horizons.

predictive learner is usually degraded in the testing or prediction phase if the input data distribution differs from the training distribution (Shimodaira 2000; Weiss *et al.* 2016). In many cases, acquiring training data that accurately reflect the feature space and distribution of the test data is challenging and costly. As a result, there is a strong need for methods that enable the development of highly effective models for a target domain by leveraging knowledge from a related source domain. This challenge serves as the foundation for transfer learning, which enhances a model's performance in one domain by utilizing relevant information from another, closely related, domain (Weiss *et al.* 2016).

In recent years, several studies have applied AI to BHIs to support interpreters by providing objective information derived from computational representations. For instance, Blanco Valentin (2018) employed DL for facies recognition and lithology classification at the reservoir scale, while Wedge *et al.* (2015) and Molossi and Pipan (2023) utilized computer vision techniques to detect boundary locations and sinusoidal features in televiewer and logging-while-drilling (LWD) images, respectively. Moreover, Sun *et al.* (2021) combined wavelet transform modulus maxima (WTMM) and recurrent neural networks (NNs) (specifically, long short-term memory networks) to automatically detect sinusoids from raw LWD images, and Molossi *et al.* (2024) proposed a synthetic LWD data approach to train two NNs that replicate manual sinusoid picking in a supervised learning method, with final predictions validated by a human expert. In this study, we modify the approach proposed by Molossi *et al.* (2024) to analyse high-resolution electrical conductivity BHIs (formation micro-imager (FMI)) from well NLW-GT-01 in the WNB (Fig. 1). The results are then compared with a manual interpretation calibrated using core data from Boersma *et al.* (2021). Specifically, we evaluate the effectiveness of a direct application of the NN-based method from Molossi *et al.* (2024) against an approach utilizing transfer learning. We begin with a review of the geological history and setting of the WNB, then describe the methodology outlined in Molossi *et al.* (2024) and explain how we adapted it for the high-resolution FMI from NLW-GT-01, which was originally trained on lower-resolution LWD image logs. Finally, we present and discuss the results of this adapted approach, highlighting its strengths, limitations and potential directions for future research.

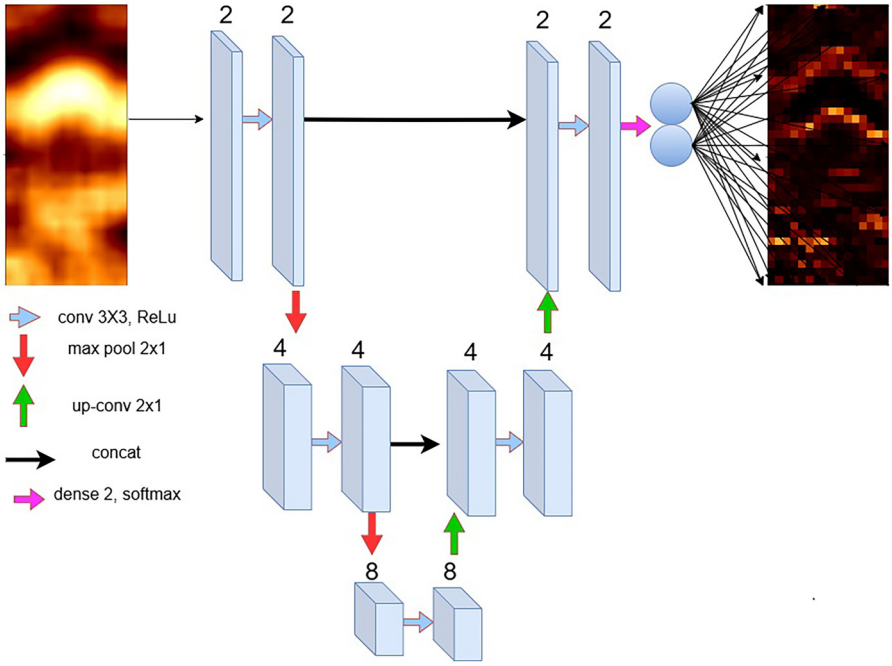

Fig. 2. A U-Net-based architecture for feature extraction, which includes a sequence of convolutions, pooling and upconvolutions, followed by concatenation. At each downsampling step, the number of feature maps (2, 4 and 8) is doubled, and the number of feature maps is halved during the upsampling path.

Geological history

The WNB is a subsurface structural element located in the southwestern Netherlands (Fig. 1). It is bounded by the Central Netherlands Basin to the NE and connected to the Roer Valley Graben in the SE. Along its southern boundary, a series of normal faults separate the WNB from the Oosterhout Platform and the Zeeland High. To the NW, the basin extends into the offshore Broad Fourteen Basin (Geluk *et al.* 1996; Van Balen *et al.* 2000; Kombrink *et al.* 2012; Willems *et al.* 2020).The geological history of the WNB began with the assembly of Pangaea during the Paleozoic (Van Wijhe 1987; Pharaoh *et al.* 2010). During the Late Carboniferous–Early Permian, wrench faulting reactivated Early Paleozoic NW–SE-orientated faults, creating a conjugate fault set with NE–SW to NNE-SSW trends (Van Wijhe 1987; Van Balen *et al.* 2000; Worum *et al.* 2005). Following, in the Triassic, Pangaea began to break-up in response to an east–west extensional regime (Geluk 2005; Wong *et al.* 2007). This rifting phase formed a horst and graben system in the WNB, with dominant north–south to NW–SE fault orientations. By the Early Cretaceous, extensional activity across the North Sea graben system decreased, eventually ceasing during the Aptian–Albian. In the Late Cretaceous, the compressional stress regime related to the Alpine Orogeny caused inversion of the WNB. Pre-existing north–south to NW–SE faults from earlier rifting were reactivated during this phase, followed by regional subsidence.

The Main Buntsandstein Subgroup

The Main Buntsandstein Subgroup in the WNB consists of three formations: Volpriehausen, Detfurth and Hardegsen (Van Adrichem Boogaert and Kouwe 1993). The Volpriehausen and Detfurth formations are formed by a lower sandstone unit of fluvial/aeolian origin and an upper silt/claystone unit deposited in a playa-lake environment. The Hardegsen Formation has a lower net-to-gross (N/G) ratio due to more frequent fluvial sandstone and claystone alternations, as shown by recent sedimentological studies (Cecchetti *et al.* 2024). Deposited from the Induan to the Olenekian stages of the Early–Middle Triassic, the Main Buntsandstein sediments were deposited through fluvial systems in an arid to semi-arid endorheic basin (Ames and Farfan 1993; Geluk 2005). Sediments were largely sourced from a major fluvial system in the southeastern Netherlands, with additional input from the London Brabant Massif (Cecchetti *et al.* 2024). The Main Buntsandstein Subgroup sediments were significantly affected by diagenetic processes such as compaction, cementation and dissolution (Purvis and Okkerman 1996). Compaction and cementation locally deteriorate porosity and permeability, especially where these sediments are deeply buried below 3–4 km. In these settings, where matrix permeability is extremely low (e.g. <0.1 mD), natural fractures can serve as primary conduits for fluid flow. The natural fractures in the Main Buntsandstein Subgroup in the WNB are associated with large-scale tectonic events that shaped the WNB from the Mesozoic onwards. This results in a wide range of natural discontinuities such as veins, joints, shear and stylolites with an overall dominant NW–SE orientation In this context, Tutuarima *et al.* (2023) documented a relationship between fracture density and sedimentary facies, with natural fractures being more prone to develop in intervals characterized by an alternation of sandstone and claystone, rather than in intervals with a dominant sandstone composition (Boersma *et al.* 2021; Tutuarima *et al.* 2023).

Material and methods

Neural networks

The proposed DL method combines a U-Net (Ronneberger *et al.* 2015) and a convolutional network with a final fully connected layer architecture, referred to here as PickNet and FitNet, respectively, to simulate the key stages of manual interpretation. PickNet is designed for automatic segmentation, identifying contrast regions within the images (Fig. 2). This architecture outputs an image of the same dimensions as the input, with each pixel replaced by the probability of it belonging to a contrast region. FitNet then takes the segmentation maps generated by PickNet as input and predicts the sinusoids associated with the segmented edges (Fig. 3). The networks were trained using a synthetic dataset that originally represented 1 m of LWD BHIs, based on the standard specifications of borehole imaging tools. We refer to this pretrained DL system as the 'baseline model'. The latter was validated using field LWD BHI data from different parts of the world, demonstrating the generalization potential of the supervised learning method. This also highlights the importance of combining it with domain-specific

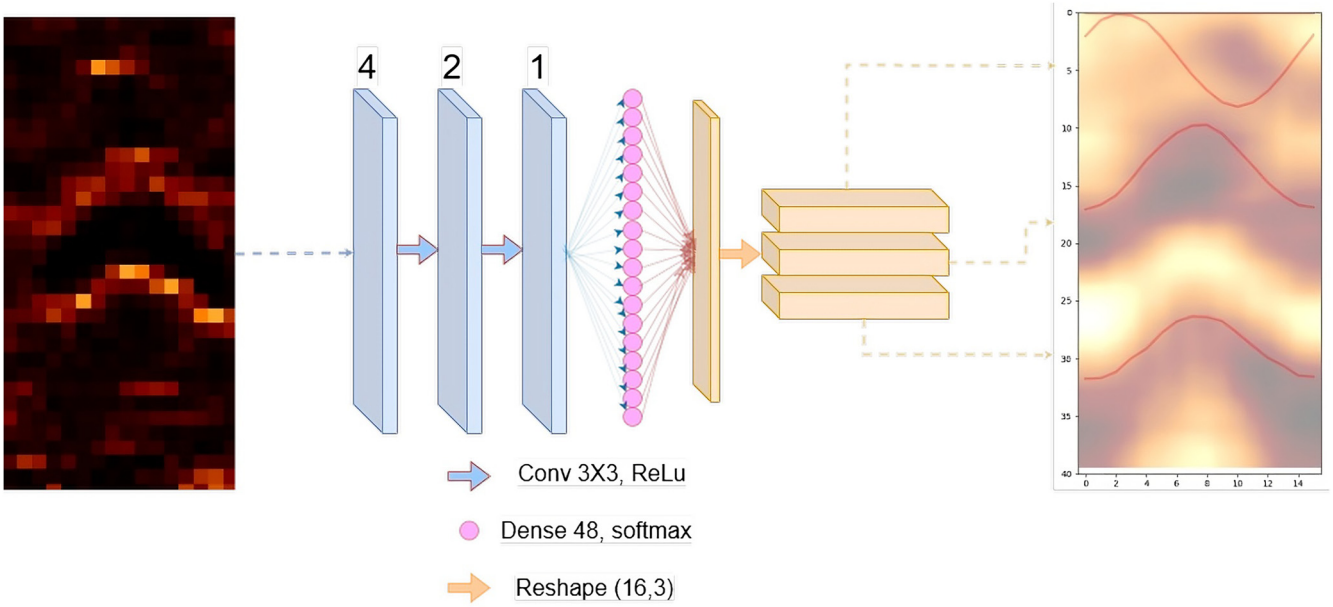

Fig. 3. A convolutional feature extraction module followed by a dense-layer, mapping-input segmentation map of features with sinusoidal symmetry. Note that the final reshape layer in the network implies that the output size is fixed to three features predicted for each input image. However, the network was trained taking into consideration also cases with no edges, to reduce the bias towards feature presence in the input image (refer to Molossi *et al.* 2024 for further descriptions).

human knowledge, following the semi-automation paradigm, to achieve a reliable and quick final data analysis (for further descriptions of the method please refer to Molossi *et al.* 2024).

In this study, we use field FMI images, which, as a wireline technology, offer a consistently higher resolution compared to LWD. We aim to adapt the aforementioned baseline model to test its scalability on these high-resolution BHIs, which are more widely used in geothermal exploration than LWD BHIs. Given a depth sample rate of *c*. 0.05 m for standard LWD tools and 0.003 m for FMI, we observe that 1 m of image corresponds to approximately 20 samples for LWD and 330 samples for FMI. This large increase in resolution brings new challenges, particularly in training data representativeness and model generalization. While LWD images are easier to simulate synthetically – allowing for the creation of effective training datasets that yielded promising results on field data – translating this same approach to the FMI domain is not straightforward. Initially, we assumed that downsampling high-resolution FMI images would preserve sufficient structural information to support realistic predictions from the baseline model.

As the required input image size for the baseline model is 20×16, we had to downsample the 1 m FMI images to enable automated analysis on such high-resolution BHIs. We retained and evaluated the downsampled FMI input strategy as a practical compromise, recognizing that it remains a baseline solution pending the development of more effective synthetic data generation methods or adaptations to the network architecture. Nonetheless, initial efforts to train an enhanced version of the baseline model directly on synthetically generated FMI data are described later in the following 'Transfer learning' subsection.

We tested the baseline model on a FMI dataset consisting of 115 instances of 2 m FMI intervals provided by ConocoPhillips, described in the 'ConocoPhillips dataset' subsection later in this section. Each FMI interval instance is associated with one manually picked fracture, interpreted by experienced petrophysicists, and some fractures display a full sinusoidal curve, while others are partial. The reference depth for these fractures can vary, such as the inflection point, average depth or other criteria. Although manual interpretations that are known to be biased (Shipley *et al.* 2013; Andrews *et al.* 2019; Shipton *et al.* 2020) are used, they remain valuable for deriving quantitative performance metrics that aid in visualizing the scalability test results. In particular, manual interpretations can serve as a useful benchmark for validating the NN outputs, especially in intervals where clearly visible geological boundaries are present. This allows a more objective assessment of misalignments between human and machine performances. In addition, such comparisons can offer insights into potential differences in human perception when supported by mathematically driven analyses from the NN. As detailed in the 'Semi-automation' subsection later in this section, we propose using the NN in a semi-automated workflow; in this context, a direct comparison between manual and machine-assisted interpretations becomes both feasible and meaningful for investigating how AI systems, like the one proposed, influence the subjective bias of analysts.

Applying the same approach to the NLW-GT-01 FMI BHI was more challenging due to the absence of interpreted sinusoids in the open dataset, which prevented direct comparison. Therefore, we referred to the results presented in Boersma *et al.* (2021) for an initial visual comparison, followed by a more in-depth independent analysis of the results. In Boersma *et al.* (2021), fracture planes were identified as features exhibiting a distinct sinusoidal shape and an increase in conductivity in the FMI, compared to the surrounding rock matrix. In addition, authors' interpretation was supported by the resistivity log and the available core data. A notable characteristic of the NLW-GT-01 well data suggested by the authors is that the conductivity image log reveals significantly more interpretable discontinuities than the core data for the same interval. These additional discontinuities observed in the FMI data have been attributed to hydraulically- or drilling-induced fractures. As a result, the image log was interpreted using a highly conservative approach, considering only features that were either visible in the core or which displayed a complete sinusoid.

Transfer learning

As an alternative to the simple approach described above for the scalability assessment, we tried to implement a transfer learning model. We therefore exploited Fitnet as a pretrained model for the automated BHI analysis and added a self-attention block (b in

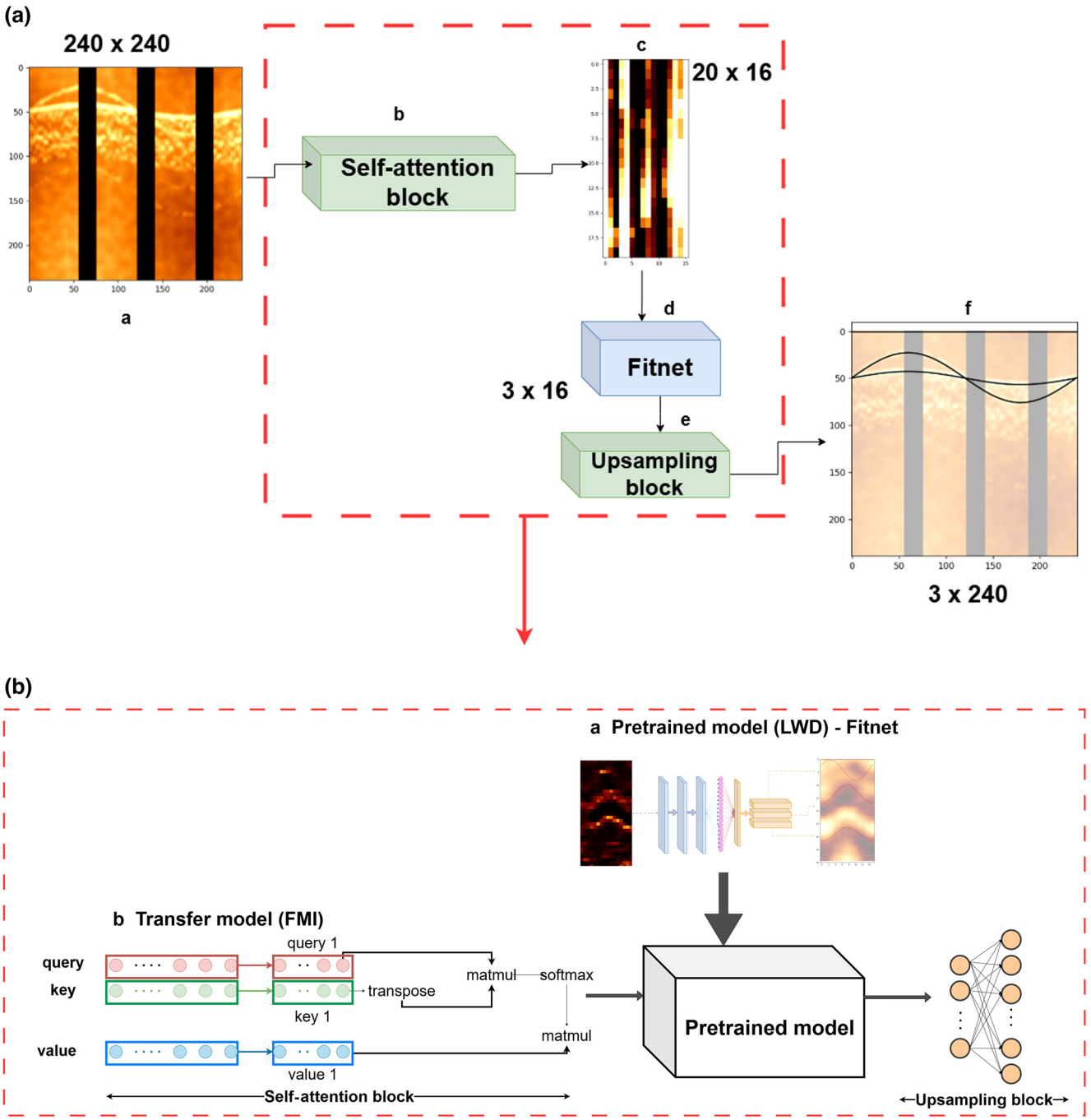

Fig. 4. (**a**) a: Synthetic FMI input image with size 240 × 240. b: self-attention block added before the pretrained model (Fitnet) to select the relevant information to be retained in c, which shows the degraded input with size 20 × 16. d: Fitnet is the pretrained baseline model for feature extraction that corresponds to 3. e: The upsampling block to transform the size of the predicted features from 3 × 16 to 3 × 240, which means that we find a maximum of three features per input image, according to the Fitnet implementation explained in Molossi *et al.* (2024). (**b**) a: Fitnet architecture used as a pretrained model in b, which shows the transfer model.

Fig. 4a, b) before the pretrained model to efficiently select the relevant information in the degraded FMI input. The degradation is caused by the necessary loss of resolution when moving from a 240 × 240 input image to a 20 × 16 image, resulting in a significant reduction in pixel count from 57 600 to 320. This represents a loss of 99.44% of the original resolution, retaining only 0.56% of the initial pixel information. Given this drastic reduction, the self-attention mechanism may help to focus on the most relevant features in the low-resolution FMI before feeding it into the pretrained Fitnet model, improving its ability to extract meaningful patterns despite the loss of detail. Unlike convolutional layers, which operate within fixed local receptive fields, self-attention allows the model to dynamically weigh the importance of features across all positions in the image, regardless of their spatial proximity. This is particularly valuable when much of the fine-grained detail has been lost due to downsampling. In practice, the self-attention block computes attention scores between all pairs of positions in the input. Each position in the input image is projected into three vectors: a query, a key and a value. The attention score between two positions is then determined by the similarity between the query of one and the key of the other, indicating how relevant one feature is to another in the overall context. These scores are then used to weight the corresponding values, effectively aggregating information from across the image. This global aggregation enables the model to

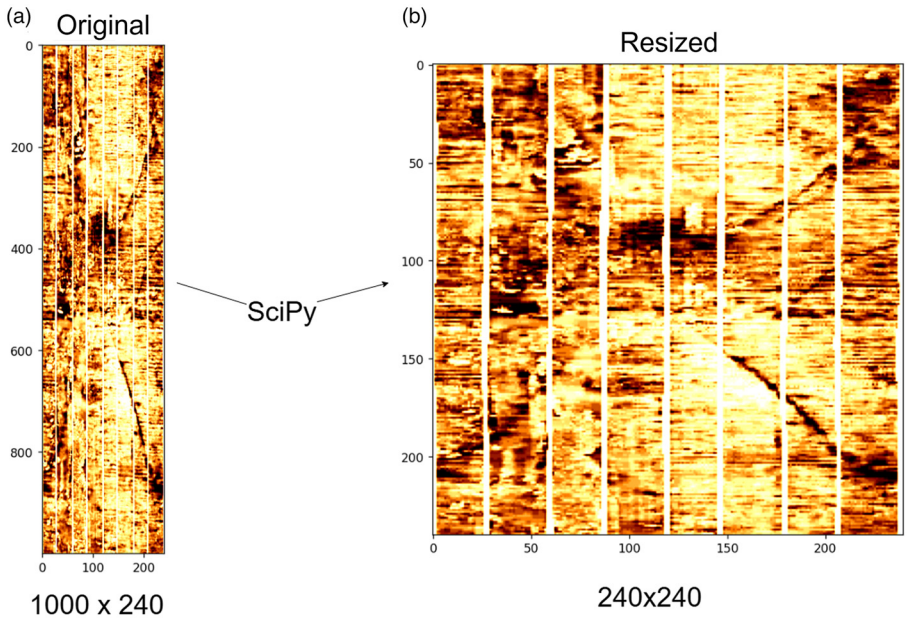

Fig. 5. Resizing of the input PDDA images from (**a**) an original size of 1000 × 240 to (**b**) a resized 240 × 240 with the 76% loss of vertical detail to be fed into the transfer model (Fig. 4).

highlight semantically meaningful features, thus helping to preserve critical structural information despite the resolution loss. After the Fitnet block in the transfer model (b in Fig.4b), two linear fully-connected layers were added to upsample the features that were suggested by Fitnet on the degraded input (i.e. after the self-attention block) associated with the original FMI input (f in Fig. 4a).

To train the additional layers in the transfer model, we used synthetic FMI instances and features generated using an adaptation of the original code introduced in Molossi *et al.* (2024) for synthetic LWD BHI generation. This adaptation aimed to increase the size and resolution of the synthetic BHIs, and to possibly achieve the required training data representativeness of real-world examples. Specifically, we introduced variations such as adding vertical stripes at random positions in the image and with random widths, better simulating the types of artefacts commonly encountered in real FMI data. The training of the transfer model took 500 epochs with a learning rate of 0.0001 on 10 000 synthetic FMI inputs (e.g. a in Fig. 4a). Similarly to the baseline model, the transfer learning model was tested on the PDDA open dataset, which is described later. The method that gave the best results on the latter was then applied on the NLW-GT-01 FMI dataset from the WNB and compared with the interpretation results published in Boersma *et al.* (2021).

ConocoPhillips dataset

The dataset from the SPWLA PDDA SIG 4th Annual Machine Learning Competition was collected; this is available on the GitHub page: https://github.com/pddasig/Machine-Learning-Competition-2024. This dataset, originally provided by ConocoPhillips, contains real FMI images suitable for machine learning (ML) applications. Each 2 m FMI interval is represented as a 1000 × 240 matrix, which must be interpreted as described in the earlier 'Neural networks' subsection. A sliding window of arbitrary depth length (in samples) is then defined and moved sample-wise across the full image. At each depth position, the image region within the window is resized to the NN's accepted input size (20 × 16), and the network predictions for each window are collected. A 2D rectangular assignment algorithm (Crouse 2016) is then applied to identify the best-matching prediction compared to the manual fracture trace, and the mean squared error (MSE) is calculated. As anticipated in the 'Neural networks' subsection, a direct quantitative comparison between the manually interpreted and NN-assisted sinusoid picks for the NLW-GT-01 FMI data – our primary object of study – was not feasible due to the lack of digitized ground-truth sinusoids. To address this, we used the ConocoPhillips dataset as a benchmark with available labelled data, allowing us to quantitatively validate the performance of our DL system on real FMI images before qualitatively applying it to the NLW-GT-01 case. For the transfer learning set-up, to input 240 × 240 images into the NN, the data were resized using the SciPy library in Python, as shown in Figure 5. Although this resizing incurs a 76% loss of vertical resolution, the target sinusoidal patterns remain recognizable in the transformed image, enabling the model to produce valid predictions.

West Netherland Basin FMI data

We used the FMI data from the NLW-GT-01 borehole (see Fig. 1). The data have been extensively studied as part of the DESTRESS program and most of the reporting is freely available at www.nlog.nl (Boersma *et al.* 2021).

Semi-automation

Following the approach used for the ConocoPhillips dataset (see the previous subsection), we analysed the NLW-GT-01 FMI data by segmenting it into windows of varying length (representing different scales) with a sample-wise shift in depth (Fig. 6). However, the absence of digitized interpreted features from Boersma *et al.* (2021) prevented the use of a best-matching algorithm to directly select the most accurate prediction based on a similarity to the manual interpretations.

To address this, we propose a semi-automated approach that combines NN predictions with human validation.

According to Parasuraman *et al.* (2000, p. 287) automation is defined as:

> [T]he full or partial replacement of a function previously carried out by the human operator. This implies that automation is not all or none, but can vary across a continuum of levels, from the lowest level of fully manual performance to the highest level of full automation.

Based on this definition, our approach – combining NN predictions with human validation – positions itself at the midpoint of the automation spectrum. We therefore refer to it as semi-automation. In this workflow, an interpreter reviews the NN-

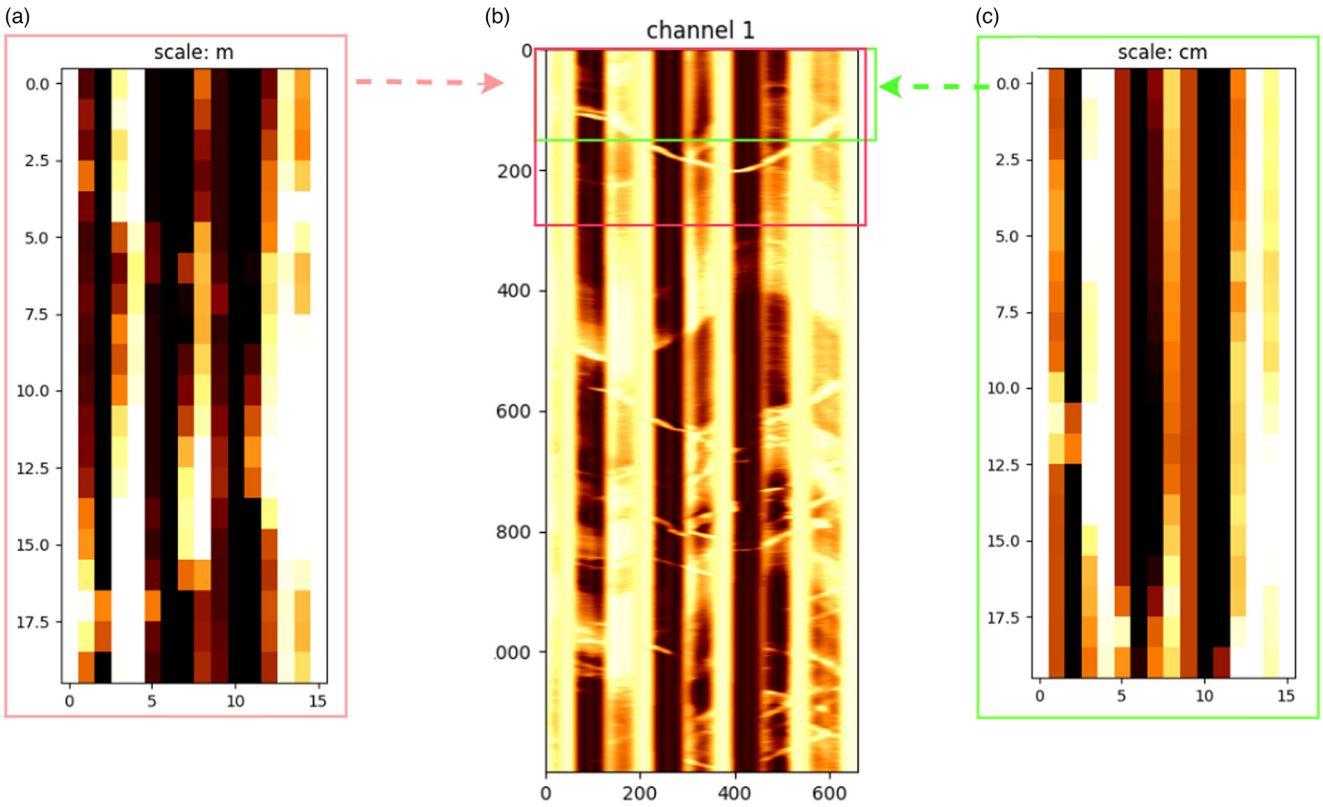

Fig. 6. (**a**) Resized image corresponding to the red box in (b). (**b**) An interval of 1200 samples (3.40 m) showing the three FMI intervals. (**c**) Resized image corresponding to the green box in (b).

suggested features and either confirms or discards them, ensuring a refined final interpretation of the reference intervals.

To facilitate this process, we developed a set of Python functions that allow interpreters to visualize predictions across all considered scales. These functions support the post-prediction validation process by enabling interpreters to validate/discard predictions based on various factors, including prior geological knowledge, expertise and visual features. An example of this validation process can be seen in Figure 7. By leveraging the perceptual strengths of both humans and machines, we aim to empirically demonstrate that a *post-hoc* human–machine collaborative approach can produce realistic and meaningful results. In the long term, the objective is to show that such collaboration can enhance the BHI interpretation experience from the human analyst's perspective – by reducing the time, cognitive fatigue and tedium associated with the task, while introducing mathematically grounded suggestions that help to mitigate subjectivity in the interpretation process.

Results

We first tested both the baseline model ('Neural networks' subsection) and the transfer learning model ('Transfer learning' subsection) on the open PDDA dataset ('ConocoPhillips dataset' subsection), as it allowed a quantitative assessment of the NN's predictions misalignment with available manual interpretations. These approaches were then subsequently applied to the NLW-GT-01 FMI data ('West Netherland Basin FMI data' subsection), but was limited to qualitative evaluation ('Semi-automation' subsection).

Regarding the baseline model, we used windows with depth lengths of 200, 500, 800 and 1000 samples – corresponding to different scales of automated analysis – to compare the MSE of the matched predictions (i.e. the baseline model's predictions that were optimally aligned with the manually picked fracture in the FMI example) (Fig. 8). The starred MSE curves in Figure 8a correspond to low-angle, medium-angle and high-angle fractures in the PDDA dataset (Figs 9–11, respectively). The minimum MSE values for the low-angle and medium-angle fractures (example 80 in Fig. 9 and example 81 in Fig. 10, respectively) were observed at a window length of 200 depth samples. For the high-angle fracture (example

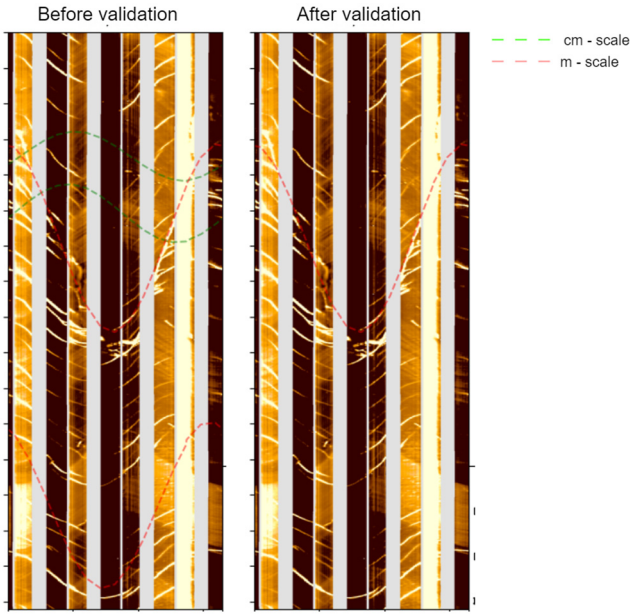

Fig. 7. Example of the semi-automated analysis results overlain on the original FMI. (Left) The DL predictions are shown for a specific window of the full interval. (Right) The results after human intervention are displayed, where the green centimetre-scale predictions are discarded along with one red metre-scale prediction, and only the fracture-associated prediction is validated.

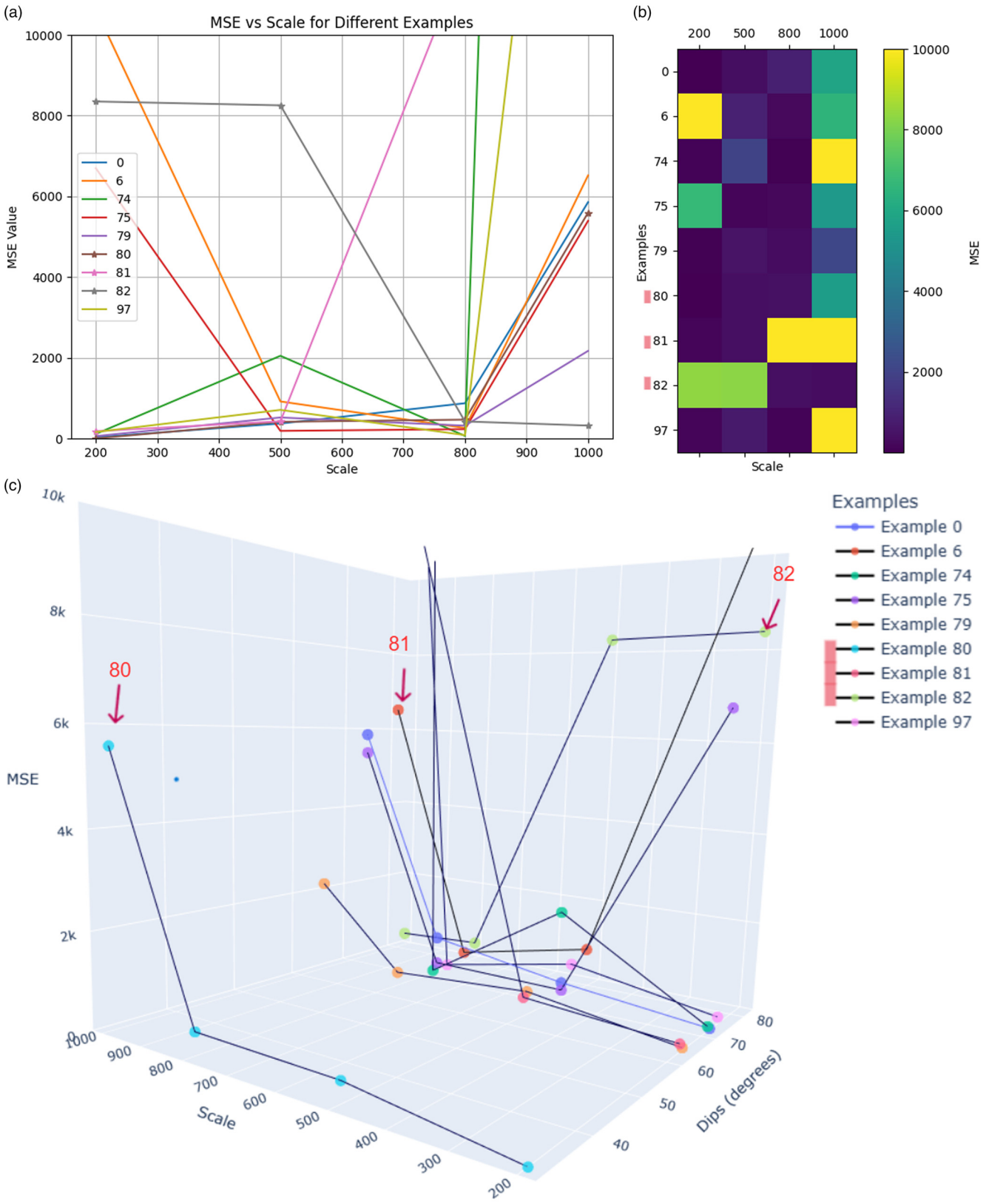

Fig. 8. (**a**) Plot of MSE v. scale and (**b**) heat map of the mean squared error (MSE) across different scales. Examples 80, 81 and 82 represent low-, medium- and high-angle fractures, respectively. Notably, for the high-angle fracture, the MSE decreases as the prediction depth window length (in samples) increases. (**c**) 3D plot showing the relationship between scale, MSE and dip angles for different examples, with points connected by lines to highlight MSE trends within each example. Examples 80, 81 and 82, representing low-, medium- and high-angle fractures, respectively, are marked by red arrows. Notably, for the medium- and low-angle fractures (examples 81 and 80), the minimum MSE corresponds to a window length of 800, whereas for the high-angle fracture (example 82), the minimum MSE occurs at the full length of the interval.

82 in Fig. 11), the minimum error occurred at the maximum window length (1000 depth samples). This highlights the influence of window length – which reflects the scale of automated analysis – on the final results of the baseline model applied on higher-resolution BHIs. These evidences also guided the application choices for the NLW-GT-01 case study.

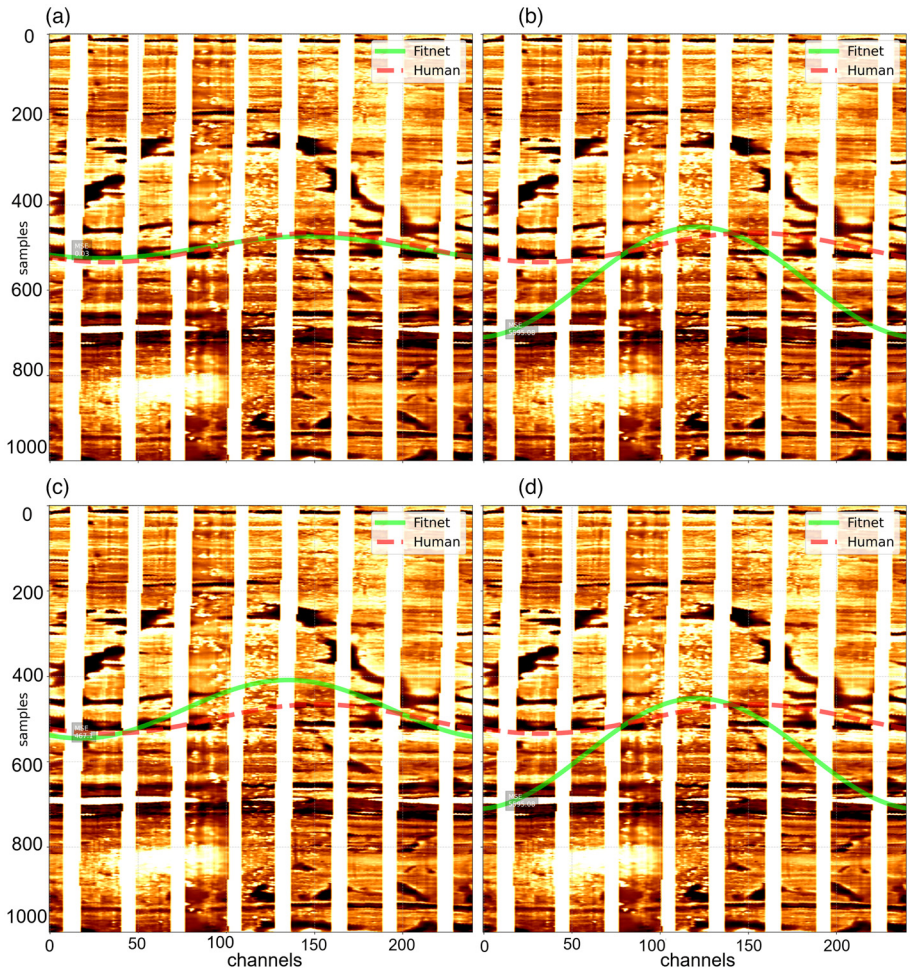

Fig. 9. Example 80. (**a**) Predictions using a window length of 200 samples. (**b**) Predictions using a window length of 500 samples. (**c**) Predictions using a window length of 800 samples. (**d**) Predictions using a window length of 1000 samples.

Regarding the transfer learning model, the training results on the synthetic FMI dataset are presented in Figure 12a–e for batch sizes of 1, 8, 32, 128 and 512, respectively. The loss function's order of magnitude appears to be positively correlated with the batch size, and the trend is confirmed by the visualization of the transfer model's predictions on randomly selected synthetic training examples for different batch sizes (Fig. 13). After training, the optimized transfer model weights for each batch size were saved and then applied to the PDDA dataset to collect predictions (Fig. 14). It is clear from the visualization of the results, that while the transfer model was effectively trained to automatically predict features on synthetic FMI input images, its performance significantly deteriorates when applied to real-case FMI data in accurately positioning the sinusoidal fracture trace.

Therefore, the comparison of the transfer learning and baseline models on the PDDA dataset suggested that the latter is yielding the best results at this point in the development of synthetic FMI generation and transfer model conceptualization. For this reason, we proceeded to apply the baseline model to collect the automatically detected features on the NLW-GT-01 FMI dataset.

The results of a conservative manual interpretation published in Boersma *et al.* (2021) – which only considers fully visible sinusoids – was compared with the validated predictions (see the 'Semi-automation' subsection) for the same intervals in Figure 15.

The baseline model adapted to the FMI data succeeded in extracting the conductive small fractures that were manually picked by Boersma *et al.* (2021). The authors' interpretation within the Volpriehausen Formation focuses on the fractures visible in the core, but other discontinuities in the FMI data have been identified as either hydraulically- or drilling-induced fractures. These additional features may also be detected by the model, as illustrated in Figure 16. Figure 17 presents a detailed visualization of the main fault interpreted in the Volpriehausen Formation using the semi-automated approach, accompanied by core images that provide additional confirmation of its presence.

Discussion

The transfer learning model described in the 'Transfer learning' subsection did not achieve the expected performance on real-world FMI examples from the PDDA dataset. This discrepancy may be due to differences in the data distribution between the synthetic FMI training data and the PDDA test data (Shimodaira 2000). In other words, the PDDA dataset may contain features that are not well represented in the synthetic FMI dataset. For instance, variations in the width and frequency of vertical stripes in the input images might be underrepresented. These are just simple examples of the adjustable parameters in the synthetic FMI generation algorithm. Other, more subtle factors, such as noise levels, sinusoid-related contrast and the complexity of the custom layer added to the pretrained model, may also contribute to the model's reduced accuracy.

Conversely, the trivial adaptation of the baseline model (i.e. the DL system: see the 'Neural networks' subsection), originally trained on synthetic LWD data, to high-resolution FMI data has demonstrated promising results. The semi-automated process (see the 'Semi-automation' subsection) generated results aligned with the conservative interpretations described in Boersma *et al.* (2021) (Figs 15 and 17); therefore, we claim that the semi-automated approach successfully replicated manual interpretations calibrated with core samples. Furthermore, it enabled more detailed analyses by identifying subtle features, such as hydraulically- or drilling-

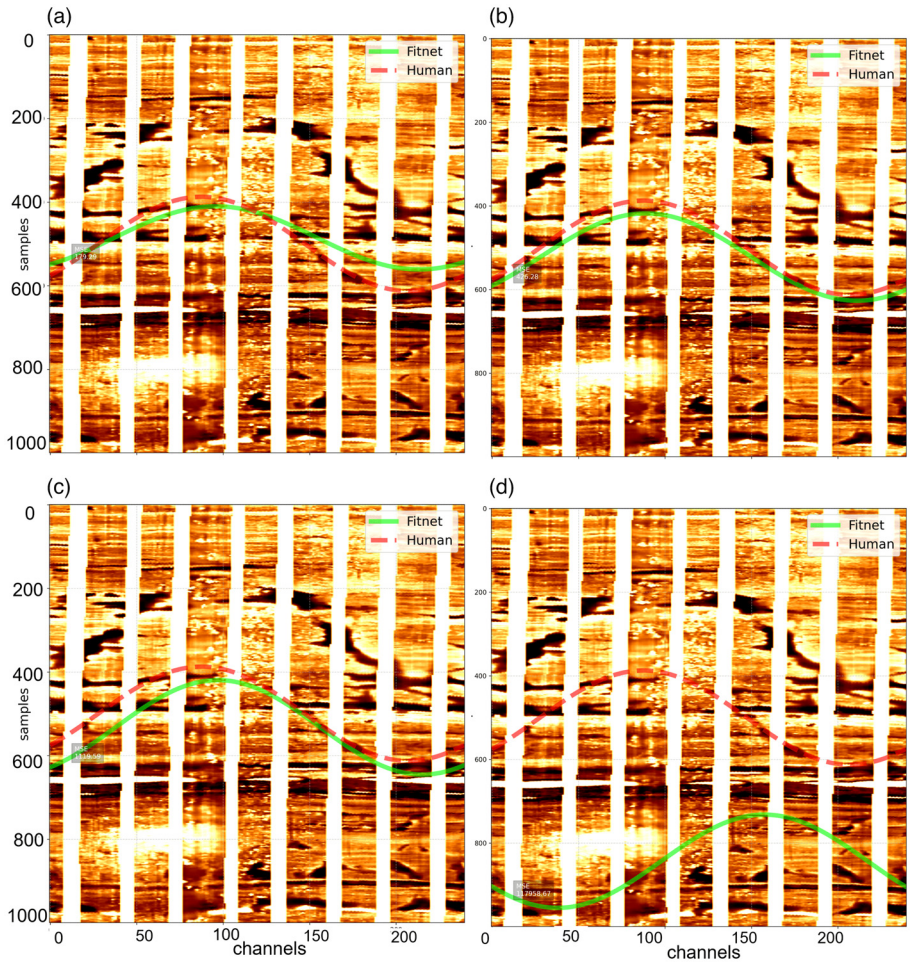

Fig. 10. Example 81. (**a**) Predictions using a window length of 200 samples. (**b**) Predictions using a window length of 500 samples. (**c**) Predictions using a window length of 800 samples. (**d**) Predictions using a window length of 1000 samples.

induced fractures (Fig. 16). As a non-linear regression model, the baseline model maps the 2D BHI input to row vectors ordered by depth, approximating the corresponding geological boundaries as sinusoidal planes. This occurs independently of the type of geological boundary being detected. In this context, the classification of the model's validated predictions is ultimately the responsibility of the human analyst, who interprets them based on geological significance, domain expertise and intuition – factors not available to the model during prediction. This approach exemplifies a positive integration of human and machine perception in the BHI analysis, where machine insight is embedded in the analysis step to address the inherent subjectivity of the process and enhance confidence in the final interpretation. Note that this does not prevent the human analyst from enriching the analysis with additional geological features that they believe the model may have missed – an approach that can also be seen as a form of semi-automation. In such a scenario, a pretrained model like the baseline model could be applied to field data and subsequently fine-tuned based on a hybrid analysis that blends manually picked sinusoids with the model's validated predictions. However, this would represent an additional level of complexity, as implementing such fine-tuning requires extensive evaluation and adjustment of training parameters, as well as consideration of the bias introduced by both manual interpretations and validation processes – an aspect that lies beyond the scope of this work.

Achieving the presented results on field FMI BHIs using a conventional approach required a significant reduction in resolution, as shown in Figure 6, to meet the input size constraints of the DL system. While this loss of resolution presents a limitation, implementing a sliding window technique combined with a multiscale approach helped to mitigate some of its drawbacks. The multiscale strategy, which involved processing data through NNs at varying window depths, effectively addressed the unpredictability and sparsity of geological features in the BHIs. This was particularly valuable given the absence of external constraints, such as outcrop information, that could otherwise provide geological context, something the DL system also inherently lacks. Notably, this approach enabled the accurate detection of fractures dipping a high angle, as had already been observed in the PDDA dataset (Fig. 11), in the WNB FMI data (Fig. 17). However, the sliding window method also introduced challenges, including redundancy in predicted features, which could complicate the semi-automation process and affect the overall quality of results. Despite these challenges, our findings validate the scalability and effectiveness of the DL system represented by the baseline model. They also highlight the potential of training NN architectures on synthetic data, challenging the common perception of DL as a purely data-driven method with limited generalization potential in fracture extraction from BHIs (Wang and Zhou 2023). The model produced insightful and accurate results after validation, even in the presence of a significant loss of resolution. Therefore, the results demonstrate that synthetic-based supervised learning is a viable and effective approach for DL-driven BHI analysis at both LWD and FMI resolutions. This method avoids biases typically associated with using field data and manual interpretations to construct training datasets, which often introduce location-specific subjectivity and epistemic uncertainty. Training a model on data from a single borehole – or even multiple boreholes within the same field – can lead to overfitting to the local geological setting, as the model learns a narrow and potentially unrepresentative data distribution. Unless the goal is explicitly to develop site-specific models that generalize only at the scale of a particular borehole or

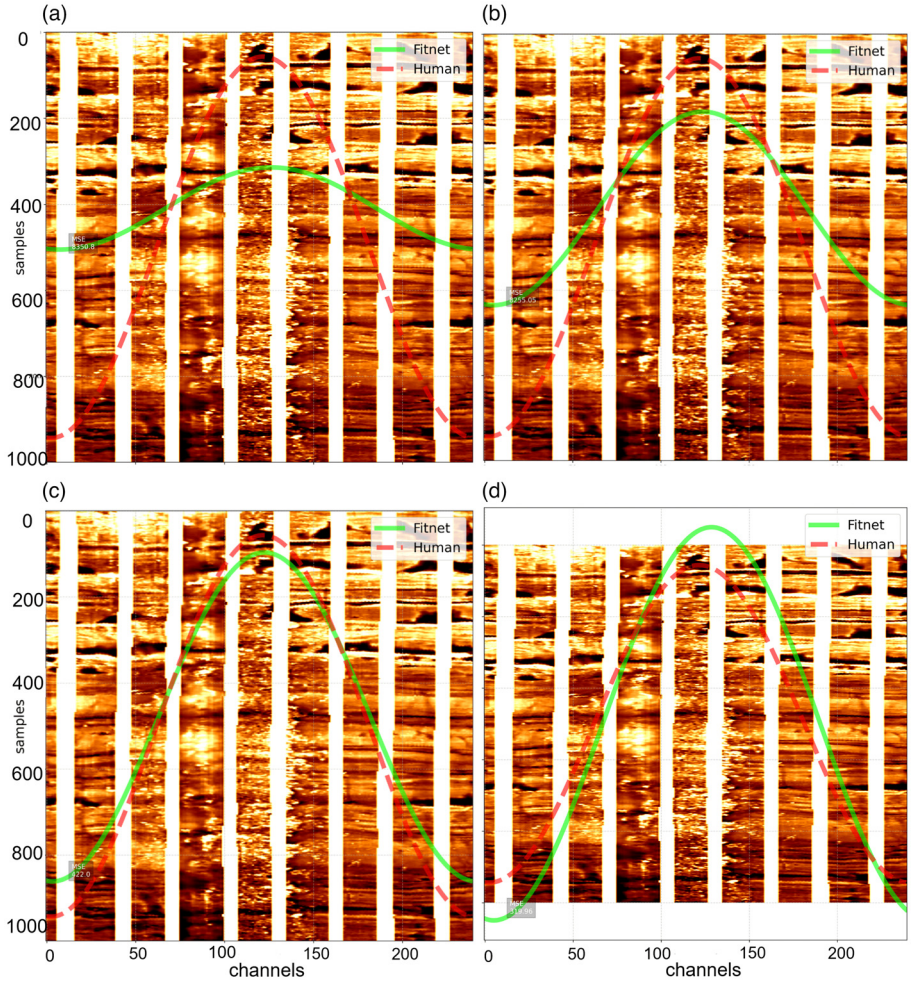

Fig. 11. Example 82. (**a**) Predictions using a window length of 200 samples. (**b**) Predictions using a window length of 500 samples. (**c**) Predictions using a window length of 800 samples. (**d**) Predictions using a window length of 1000 samples.

field, this approach poses a significant limitation for broader applicability and robustness. Moreover, in supervised DL applications, the predictive performance of these algorithms depends heavily on the quality of the labels. This is an especially critical issue in domains such as BHI analysis, where manual annotation is both costly and prone to high inter-observer variability. As Bond

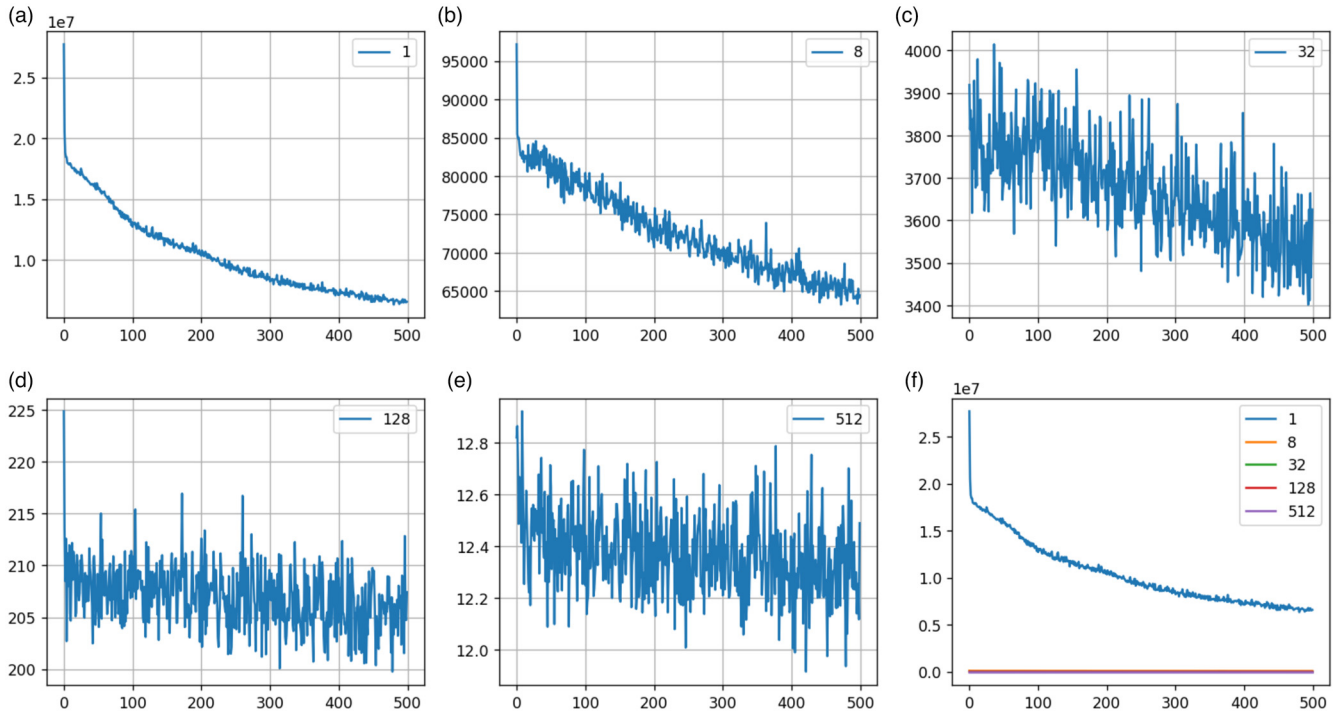

Fig. 12. Loss function over epochs (500) for a batch size equal to (**a**) 1, (**b**) 8, (**c**) 32, (**d**) 128 and (**e**) 512. (**f**) A comprehensive plot of all the loss functions.

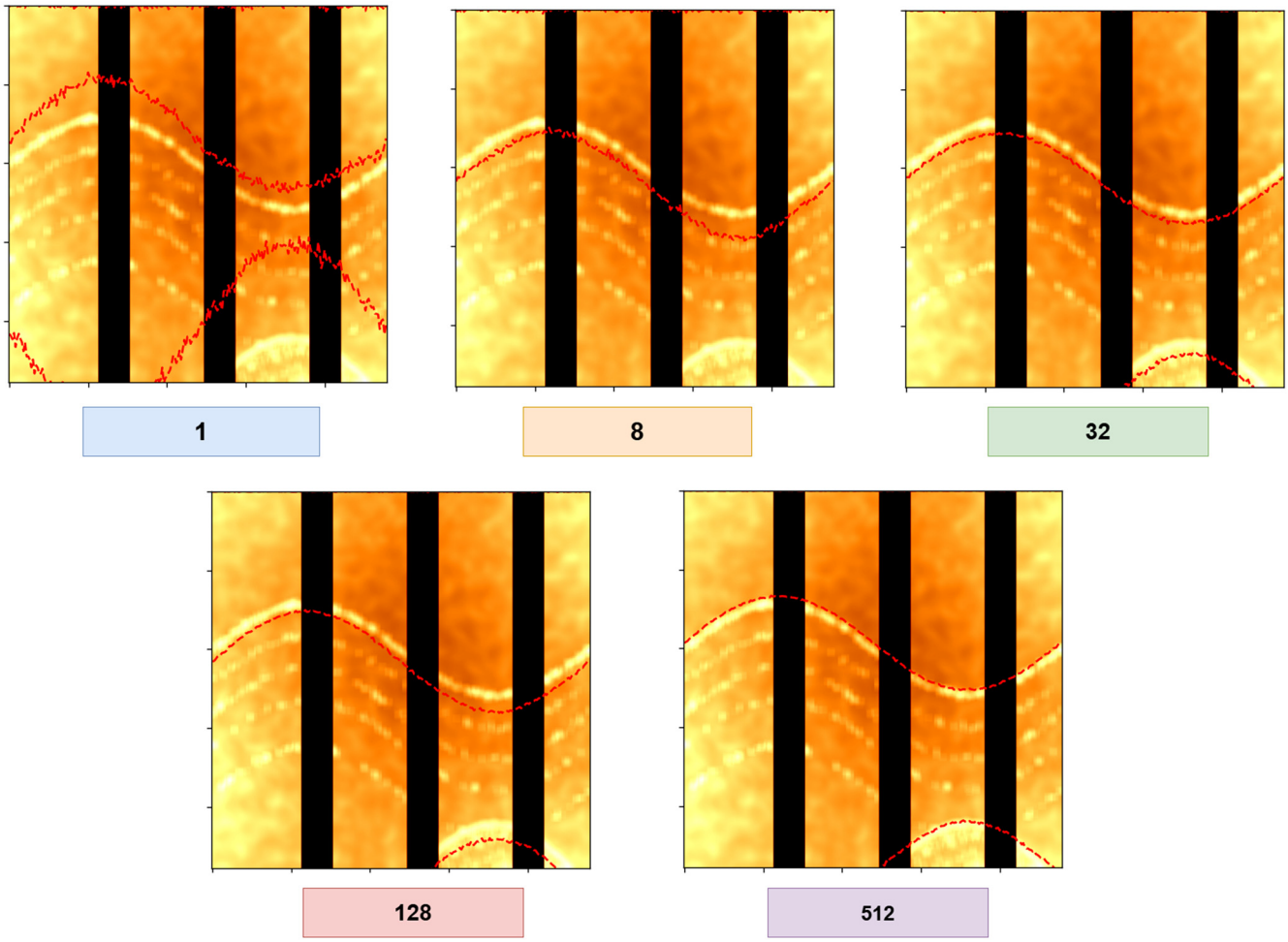

Fig. 13. Transfer model results on training synthetic examples for different batch sizes. The best visual results correspond to the higher batch sizes (from 32 to 512), which is consistent with what the loss functions suggest in Figure 12.

(2015) pointed out, geological interpretations are influenced by cognitive biases such as availability bias, confirmation bias, anchoring bias and optimistic bias. These biases, combined with the non-uniqueness of geological models and limited time, often lead to interpretations that are uncertain and difficult to verify objectively. Moreover, different interpreters can reach different conclusions using the same dataset (Bond *et al.* 2007), which adds further uncertainty.

In addition to uncertainty-related project risks, there is also the critical issue of scientific reproducibility in modelling workflows. If the reservoir's static properties – which serve as the initial conditions for all subsequent modelling stages – are derived from highly uncertain BHI interpretation results, then the entire modelling process becomes dependent on those unreliable foundations. As a result, the outcomes may not be reproducible, which compromises a fundamental principle of scientific research. Moreover, this raises ethical concerns, as decisions based on such models (operational, financial or environmental) may be misleading or unjustified if the underlying assumptions cannot be validated or independently reproduced. The semi-automatic workflow proposed here tries to address some of these challenges by introducing a more objective and repeatable process into the interpretation workflow through the use of NNs. As noted by Tartakovsky *et al.* (2012):

> [T]he accuracy and reliability of [subsurface] predictions are affected by uncertainty from the natural variability of the subsurface, limited and incomplete knowledge, and imperfect mathematical representations of the underlying processes.

NNs, when trained on well-designed datasets and applied consistently, help to reduce subjective uncertainty by standardizing the way in which data are analysed and interpreted. While they cannot remove all uncertainty – especially that linked to the data itself – they offer better reproducibility, greater efficiency and scalability compared to fully manual approaches. Similarly, Zabalza-Mezghani *et al.* (2004) emphasized that uncertainty is present at every stage of subsurface modelling, especially when merging different types of data collected at different times and scales. The semi-automatic approach can help to reduce this uncertainty by integrating these datasets more consistently and minimizing variability introduced by human BHI analysis and interpretation. Importantly, NNs are not intended to replace expert judgement but to complement it – particularly in data-sparse environments – by providing a consistent, repeatable starting point. The success of our workflow, even across domains with significant differences (i.e. LWD to FMI), further supports its use in a broader range of applications. The current baseline model represents a first example of how AI-based methods used in semi-automated workflows can also serve as a foundation for studying human interpretational bias: once multiple ML models are trained and validated for the same task, they can be leveraged to systematically explore the influence on analysts' decisions.

Conclusion

In this study, we explored the scalability of a DL-based method originally tailored for LWD BHI interpretation, extending its

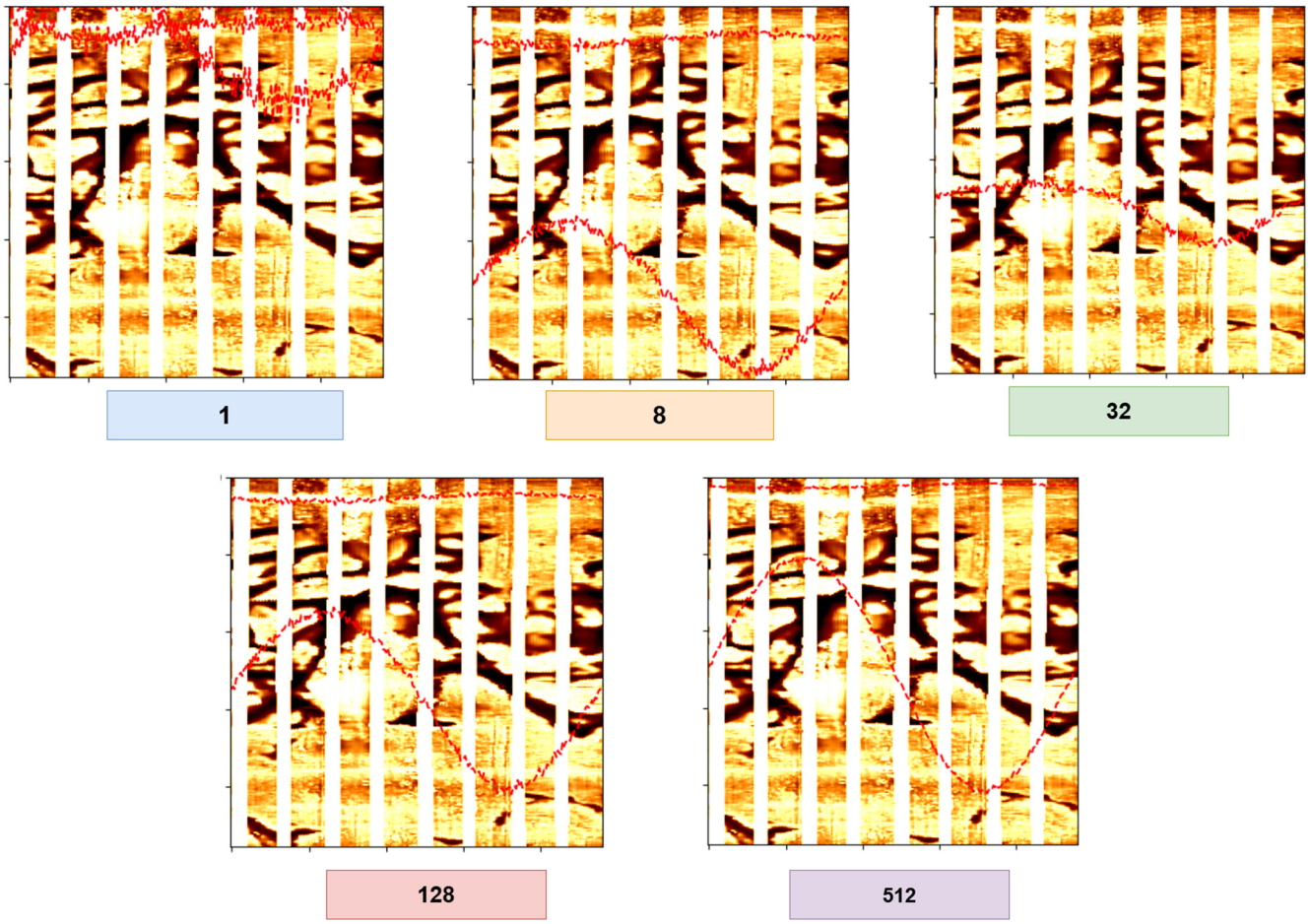

Fig. 14. Transfer learning results on a real PDDA example (example 74: see Fig. 8) highlight the suboptimal performance of the transfer model across all training batch sizes. While the model struggles to accurately position the fracture trace, the sinusoidal prediction at a batch size of 512 appears to be the smoothest compared to the other batch sizes.

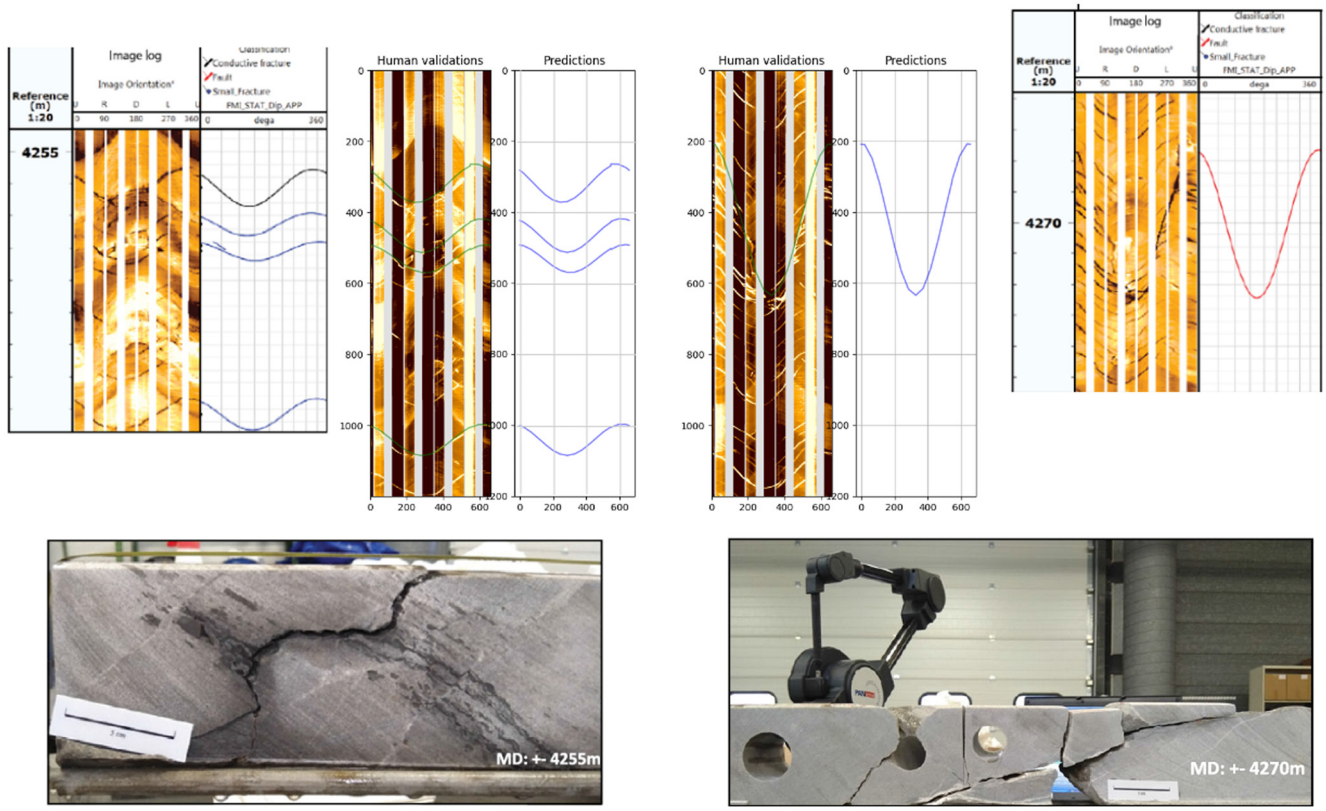

Fig. 15. Comparison of the interpretations published by Boersma *et al.* (2021) with the validated DL predictions. Source: modified from Boersma *et al.* (2021).

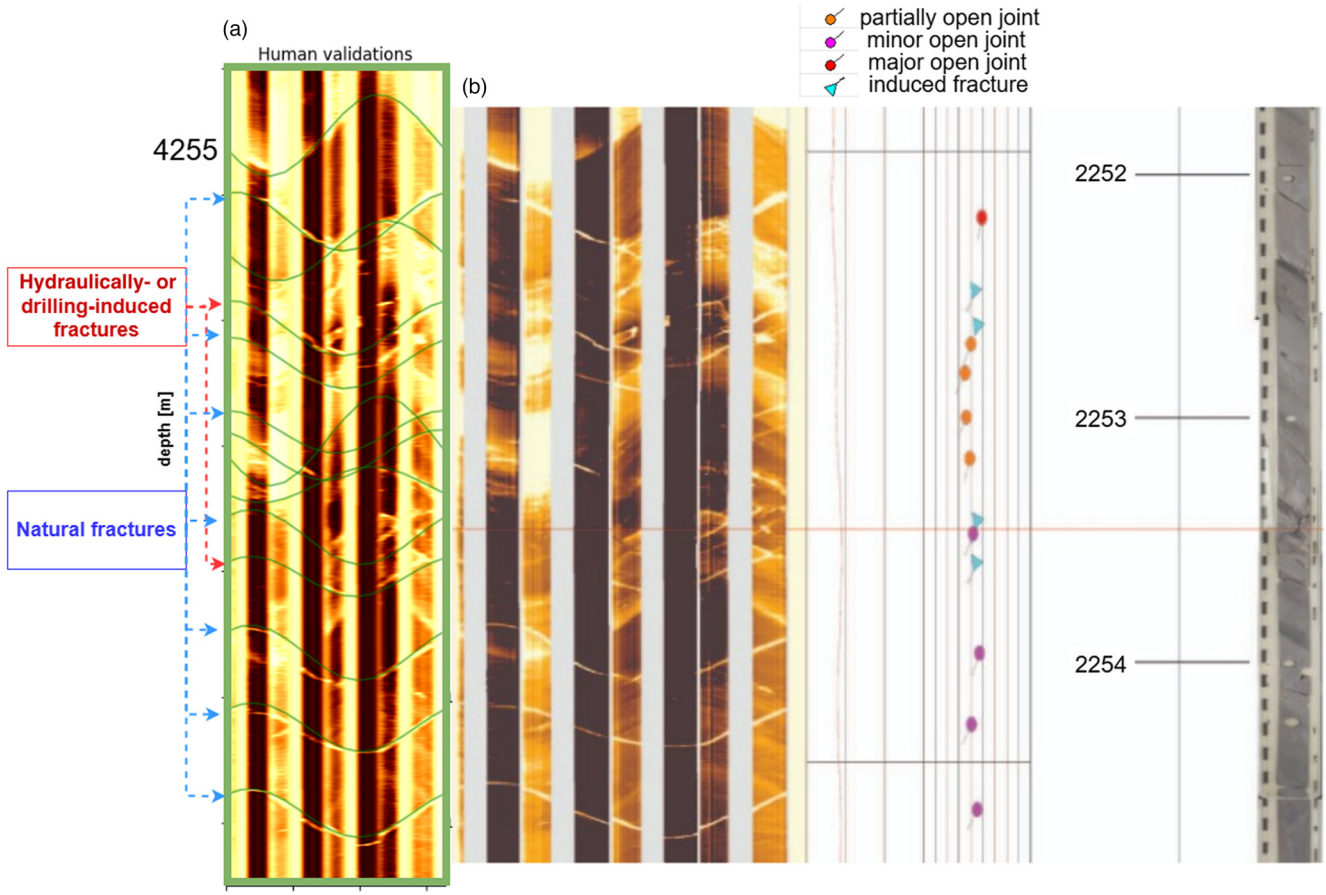

Fig. 16. (**a**) Validated DL predictions using a less conservative approach reveal discontinuities attributed to drilling-induced fractures that are absent in the core data. (**b**) The less conservative interpretation is visualized with tadpoles and core data for the interval starting at a depth of 4255

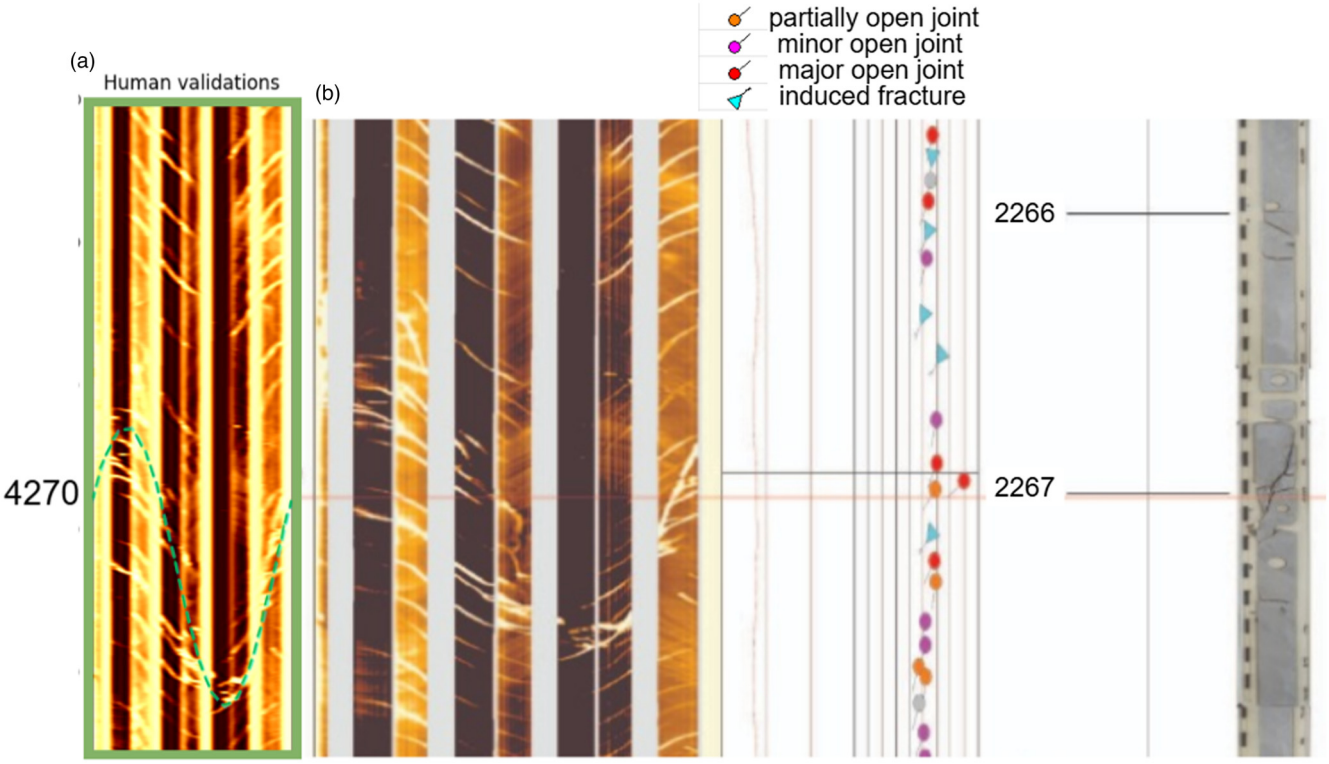

Fig. 17. (**a**) Validated DL predictions replicating the conservative approach in Boersma *et al.* (2021) that revealed a fracture dipping at a high angle visible in core data. (**b**) The less conservative interpretation is visualized with tadpoles and core data for the interval starting at a depth of 4270.

application to the more detailed realm of FMI data used in geothermal exploration. We compared two strategies: (i) a baseline model pretrained on synthetic LWD data, applied to high-resolution FMI inputs through a downsampling preprocessing step; and (ii) a transfer learning model trained on synthetic FMI. The baseline method demonstrated superior performance at this stage, while the transfer model revealed the need for further refinement to both the data and model domains.

We validated our approach using publicly available datasets, including the Naaldwijk geothermal field (well NLW-GT-01), successfully reproducing results from Boersma *et al.* (2021). Remarkably, even with significant resolution loss, the preprocessing preserved enough critical features for the NN to produce geologically credible outputs – reinforcing the promise of synthetic training data to support robust generalization across diverse geological settings and resolutions. The baseline model, initially developed for LWD data, has now been tested across at least four distinct geological environments: Mozambique and the Cascadia margin on the LWD side (Molossi *et al.* 2024), and the PDDA dataset (confidential location), and the Dutch geothermal field in Naaldwijk on the FMI side. This underpins the baseline model's consistent adaptability across different data domains.

More broadly, our findings showcase how DL can serve not as a replacement, but as a less biased analytical partner to human expertise. In this semi-automated framework, NNs generate abstract, mathematically grounded representations of borehole images, assisting experts in the interpretive step while minimizing subjective bias and promoting reproducibility. This interplay of human and machine perception could open new doors, not only for improving efficiency and consistency in BHI analysis but also for future research into interpretational uncertainty. As ML tools are deployed alongside geoscientists with varied expertise and cognitive approaches, they can illuminate how decisions are made, where human biases may subtly shape and how several machine candidates with different characteristics can influence them. Looking ahead, refining synthetic datasets and quantifying model uncertainty will be key to developing more trustworthy, scalable tools for FMI BHI analysis in geothermal reservoir characterization.

Acknowledgements The authors would like to thank the reviewers for their comments, and TU Delft and the University of Trieste for sharing data and supporting this study.

Author contributions **AM**: conceptualization (lead), data curation (lead), methodology (lead), software (lead), writing – original draft (lead); **EC**: resources (equal), writing – original draft (equal), writing – review & editing (supporting); **POB**: methodology (supporting), validation (equal), writing – original draft (supporting), writing – review & editing (equal); **MP**: supervision (lead).

Funding This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability All data generated or analysed during this study are included in this published article (and, if present, its supplementary information files).

References

Agarap, A.F. 2019. Deep learning using rectified linear units (ReLU). *arXiv*, 1803.08375, http://arxiv.org/abs/1803.08375

Akkurt, R., Conroy, T.T., Psaila, D., Paxton, A., Low, J. and Spaans, P. 2018. Accelerating and enhancing petrophysical analysis with machine learning: A case study of an automated system for well log outlier detection and reconstruction. Paper SPWLA-2018-BB presented at the SPWLA 59th Annual Logging Symposium, June 2–6, 2018, London, UK, https://onepetro.org/SPWLAALS/proceedings-abstract/SPWLA18/SPWLA18/D043S008R004/28770

Alcalde, J. and Bond, C.E. 2022. Subjective uncertainty and biases: the impact on seismic data interpretation. *In*: Bell, R., Iacopini, D. and Vardy, M. (eds) *Interpreting Subsurface Seismic Data*. Elsevier, Amsterdam, 103–123, https://doi.org/10.1016/B978-0-12-818562-9.00002-9

Ames, R. and Farfan, P. 1993. The environments of deposition of the Triassic, Main Buntsandstein Formation, in the P and Q blocks, offshore, Netherlands. AAPG Search and Discovery Article #90990, AAPG International Conference and Exhibition, October 17–20, 1993, The Hague, Netherlands, https://www.searchanddiscovery.com/abstracts/html/1993/intl/abstracts/1605a.htm

Andrews, B.J., Roberts, J.J., Shipton, Z.K., Bigi, S., Chiara Tartarello, M. and Johnson, G. 2019. How do we see fractures? Quantifying subjective bias in fracture data collection. *Solid Earth*, **10**, 487–516, https://doi.org/10.5194/se-10-487-2019

Anemangely, M., Ramezanzadeh, A., Amiri, H. and Hoseinpour, S.-A. 2019. Machine learning technique for the prediction of shear wave velocity using petrophysical logs. *Journal of Petroleum Science and Engineering*, **174**, 306–327, https://doi.org/10.1016/j.petrol.2018.11.032

Basu, P., Morris, S., Morris, S., Ritzmann, N. and Zheng, Y. 2019. Categorising risk for drilling and completions using high-resolution acoustic LWD image logs. Paper SPE-195795-MS presented at the SPE Offshore Europe Conference and Exhibition, September 3–6, 2019, Aberdeen, UK, https://doi.org/10.2118/195795-MS

Blanco Valentin, M. 2018. *Deep Learning Methods on Geological Reservoir Borehole Log Images and Applications*. Master's dissertation, Brazilian Centre for Physics Research (CBPF), Rio de Janeiro, Brazil

Boersma, Q.D., Bruna, P.O., de Hoop, S., Vinci, F., Tehrani, A.M. and Bertotti, G. 2021. The impact of natural fractures on heat extraction from tight Triassic sandstones in the West Netherlands Basin: a case study combining well, seismic and numerical data. *Netherlands Journal of Geosciences*, **100**, e6, https://doi.org/10.1017/njg.2020.21

Bond, C.E. 2015. Uncertainty in structural interpretation: Lessons to be learnt. *Journal of Structural Geology*, **74**, 185–200, https://doi.org/10.1016/j.jsg.2015.03.003

Bond, C.E., Gibbs, A.D., Shipton, Z.K. and Jones, S. 2007. What do you think this is? 'Conceptual uncertainty' in geoscience interpretation. *GSA Today*, **17**, 4–10, https://doi.org/10.1130/GSAT01711A.1

Bond, C.E., Shipton, Z.K., Gibbs, A.D. and Jones, S. 2008. Structural models: optimizing risk analysis by understanding concept uncertainty. *First Break*, **26**, 65–71, https://doi.org/10.3997/1365-2397.2008006

Cecchetti, E., Martinius, A.W., Bruna, P.-O., Bender, A. and Abels, H.A. 2024. Structural controls on the Triassic Main Buntsandstein sediment distribution in the Roer Valley Graben, the Netherlands. *Netherlands Journal of Geosciences* **103**, e23, https://doi.org/10.1017/njg.2024.17

Chen, H., Chiang, R.H.L. and Storey, V.C. 2012. Business intelligence and analytics: from big data to big impact. *MIS Quarterly*, **36**, 1165–1188, https://doi.org/10.2307/41703503

Crouse, D.F. 2016. On implementing 2D rectangular assignment algorithms. *IEEE Transactions on Aerospace and Electronic Systems*, **52**, 1679–1696, https://doi.org/10.1109/TAES.2016.140952

Faiq Adenan, M., Fathi, E., Carr, T. and Panetta, B. 2023. Machine learning-based workflow for identifying fractures and baffles from Formation Micro Imager (FMI) log: A practical application in Illinois Basin Decatur Project (IBDP). Paper SEG-2023-3916031 presented at the SEG/AAPG International Meeting for Applied Geoscience & Energy, August 27–September 1, 2023, Houston, Texas, USA, https://doi.org/10.1190/image2023-3916031.1

Geluk, M.C. 2005. *Stratigraphy and Tectonics of Permo-Triassic Basins in the Netherlands and Surrounding Areas*. PhD thesis, Utrecht University, Utrecht, The Netherlands, https://dspace.library.uu.nl/handle/1874/1699

Geluk, M.C., Plomp, A. and van Doorn, Th.H.M. 1996. Development of the Permo-Triassic succession in the basin fringe area, southern Netherlands. *In*: Rondeel, H.E., Batjes, D.A.J. and Nieuwenhuijs, W.H. (eds) *Geology of Gas and Oil under the Netherlands*, Springer, Dordrecht, 57–78, https://doi.org/10.1007/978-94-009-0121-6_7

Gupta, K.D., Vallega, V., Maniar, H., Marza, P., Xie, H., Ito, K. and Abubakar, A. 2019. A deep-learning approach for borehole image interpretation. Paper SPWLA-2019-BB presented at the SPWLA 60th Annual Logging Symposium, June 15–19, 2019, The Woodlands, Texas, USA, https://doi.org/10.30632/T60ALS-2019_BB

Kato, O. and Sakagawa, Y. 1995. *Characteristics of Fractures Based on FMI Logs and Cores in Well WD-1 in the Kakkonda Geothermal Field, Japan*. Technical Report. Geothermal Resources Council, Davis, CA, https://www.osti.gov/biblio/175643

Kombrink, H., Doornenbal, J.C., Duin, E.J.T., Den Dulk, M., Ten Veen, J.H. and Witmans, N. 2012. New insights into the geological structure of the Netherlands; results of a detailed mapping project. *Netherlands Journal of Geosciences*, **91**, 419–446, https://doi.org/10.1017/S0016774600000329

Kramers, L., Van Wees, J.-D., Pluymaekers, M.P.D., Kronimus, A. and Boxem, T. 2012. Direct heat resource assessment and subsurface information systems for geothermal aquifers; the Dutch perspective. *Netherlands Journal of Geosciences*, **91**, 637–649, https://doi.org/10.1017/S0016774600000421

LeCun, Y., Bengio, Y. and Hinton, G. 2015. Deep learning. *Nature*, **521**, 436–444, https://doi.org/10.1038/nature14539

Lukawski, M.Z., Silverman, R.L. and Tester, J.W. 2016. Uncertainty analysis of geothermal well drilling and completion costs. *Geothermics*, **64**, 382–391, https://doi.org/10.1016/j.geothermics.2016.06.017

Maerten, L., Legrand, X., Castagnac, C., Lefranc, M., Joonnekindt, J.-P. and Maerten, F. 2019. Fault-related fracture modeling in the complex tectonic environment of the Malay Basin, offshore Malaysia: An integrated 4D geomechanical approach. *Marine and Petroleum Geology*, **105**, 222–237, https://doi.org/10.1016/j.marpetgeo.2019.04.025

Maeso, C., Dubourg, I., Quesada, D. and ElNour, W.A. 2015. Uncertainties in fracture apertures calculated from electrical borehole images. Paper IPTC-18355-MS presented at the International Petroleum Technology Conference, December 6–9, 2015, Doha, Qatar, https://doi.org/10.2523/IPTC-18355-MS

Molossi, A. and Pipan, M. 2023. Exploiting image logs to reduce drilling hazards: an innovative Artificial Intelligence methodology applied in East Africa. *Geophysical Journal International*, **235**, 942–950, https://doi.org/10.1093/gji/ggad286

Molossi, A., Roncoroni, G. and Pipan, M. 2024. Efficient logging-while-drilling image logs interpretation using deep learning. *Petrophysics*, **65**, 365–387, https://doi.org/10.30632/PJV65N3-2024a5

Mondol, N.H. 2015. Well logging: principles, applications and uncertainties. *In*: Bjørlykke, K. (ed.) *Petroleum Geoscience*. Springer, Berlin, 385–425, https://doi.org/10.1007/978-3-642-34132-8_16

Mutonga, M.W. and Fujimitsu, Y. 2024. Identification of natural subsurface structures using borehole images, temperature logs and pertinent data: a case study of the Menengai geothermal field, Kenya. *Geoenergy Science and Engineering*, **237**, https://doi.org/10.1016/j.geoen.2024.212797

Okada, H. and Yamada, Y. 2002. Fracture distribution in and around intrusive rocks in the Fushime geothermal field, Japan: Evidence from the FMI logging. Paper SPWLA-2002-DD presented at the SPWLA 43rd Annual Logging Symposium, June 2–5, 2002, Oiso, Japan, https://onepetro.org/SPWLAALS/proceedings-abstract/SPWLA-2002/SPWLA-2002/SPWLA-2002-DD/27306?redirectedFrom=PDF

Ozkaya, S.I. and Mattner, J. 2003. Fracture connectivity from fracture intersections in borehole image logs. *Computers & Geosciences*, **29**, 143–153, https://doi.org/10.1016/S0098-3004(02)00113-9

Ozol, C., Climer, E., Aiello, R., Di Tommaso, D., Borghi, M., Cominesi, N. and Medaglia, M. 2015. The use of real-time microimager data to validate structural uncertainties. *In*: *77th EAGE Conference and Exhibition 2015*. European Association of Geoscientists & Engineers (EAGE), Houten, The Netherlands, https://doi.org/10.3997/2214-4609.201412647

Pantaleo, G., Molossi, A. and Pipan, M. 2024. Estimation of CO_2 saturation maps from synthetic seismic data using a deep-learning method with a multi-scale approach. *Geoenergy*, **2**,https://doi.org/10.1144/geoenergy2023-057

Parasuraman, R., Sheridan, T.B. and Wickens, C.D. 2000. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans*, **30**, 286–297, https://doi.org/10.1109/3468.844354

Peacock, D.C.P., Sanderson, D.J., Bastesen, E., Rotevatn, A. and Storstein, T.H. 2019. Causes of bias and uncertainty in fracture network analysis. *Norwegian Journal of Geology*, **99**, 113–128, https://doi.org/10.17850/njg99-1-06

Pharaoh, T.C., Dusar, M. *et al.* 2010. Tectonic evolution. *In*: Doornenbal, J.C. and Stevenson, A.G. (eds) *Petroleum Geological Atlas of the Southern Permian Basin Area*. European Association of Geoscientists & Engineers (EAGE), Houten, The Netherlands, 25–57.

Pluymaekers, M.P.D., Kramers, L., Van Wees, J.-D., Kronimus, A., Nelskamp, S., Boxem, T. and Bonté, D. 2012. Reservoir characterisation of aquifers for direct heat production: methodology and screening of the potential reservoirs for the Netherlands. *Netherlands Journal of Geosciences*, **91**, 621–636, https://doi.org/10.1017/S001677460000041X

Pollack, H.N. 2007. Scientific uncertainty and public policy: Moving on without all the answers. *GSA Today*, **17**, 28–29, https://doi.org/10.1130/GSAT01703GW.1

Prensky, S.E. 1999. Advances in borehole imaging technology and applications. *Geological Society, London, Special Publications*, **159**, 1–43, https://doi.org/10.1144/GSL.SP.1999.159.01.01

Purvis, K. and Okkerman, J.A. 1996. Inversion of reservoir quality by early diagenesis: an example from the Triassic Buntsandstein, offshore the Netherlands. *In*: Rondeel, H.E., Batjes, D.A.J. and Nieuwenhuijs, W.H. (eds) *Geology of Gas and Oil under the Netherlands*, Springer, Dordrecht, 179–189, https://doi.org/10.1007/978-94-009-0121-6_16

Roncoroni, G., Forte, E., Bortolussi, L. and Pipan, M. 2022*a*. Efficient extraction of seismic reflection with Deep Learning. *Computers & Geosciences*, **166**, https://doi.org/10.1016/j.cageo.2022.105190

Roncoroni, G., Forte, E. and Pipan, M. 2022*b*. Multi frequency data merging with bi-directional LSTM. *In*: *19th International Conference on Ground Penetrating Radar, October, Golden, Colorado, USA: Proceedings*. Society of Exploration Geophysicists, Houston, TX, 151–154, https://doi.org/10.1190/gpr2022-069.1

Ronneberger, O., Fischer, P. and Brox, T. 2015. U-Net: Convolutional networks for biomedical image segmentation. *arXiv*, 1505.04597, https://arxiv.org/abs/1505.04597

Shimodaira, H. 2000. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, **90**, 227–244, https://doi.org/10.1016/S0378-3758(00)00115-4

Shipley, T.F., Tikoff, B., Ormand, C. and Manduca, C. 2013. Structural geology practice and learning, from the perspective of cognitive science. *Journal of Structural Geology*, **54**, 72–84, https://doi.org/10.1016/j.jsg.2013.07.005

Shipton, Z.K., Roberts, J.J., Comrie, E.L., Kremer, Y., Lunn, R.J. and Caine, J.S. 2020. Fault fictions: systematic biases in the conceptualization of fault-zone architecture. *Geological Society, London, Special Publications*, **496**, 125–143, https://doi.org/10.1144/SP496-2018-161

Sun, Q., Li, N., Duan, Y., Li, H. and Tang, H. 2021. Logging-while-drilling formation dip interpretation based on long short-term memory. *Petroleum Exploration and Development*, **48**, 978–986, https://doi.org/10.1016/S1876-3804(21)60082-4

Tartakovsky, D.M., Nowak, W. and Bolster, D. 2012. Introduction to the special issue on uncertainty quantification and risk assessment. *Advances in Water Resources*, **36**, 1–2, https://doi.org/10.1016/j.advwatres.2011.12.010

Tutuarima, F., Cecchetti, E., Abels, H.A., Bertotti, G. and Bruna, P.O. 2023. Main controls on natural fracture distribution in the Lower Triassic sandstones of the West Netherlands Basin. *In*: *84th EAGE Annual Conference & Exhibition*. European Association of Geoscientists & Engineers (EAGE), Houten, The Netherlands, https://doi.org/10.3997/2214-4609.2023101051

Van Adrichem Boogaert, H.A. and Kouwe, W.F.P. (eds) 1993. *Stratigraphic Nomenclature of the Netherlands, Revision and Update by RGD and NOGEPA*. Mededelingen Rijks Geologische Dienst, Haarlem, The Netherlands

Van Balen, R.T., Van Bergen, F., De Leeuw, C., Pagnier, H., Simmelink, H., Van Wees, J.D. and Verweij, J.M. 2000. Modelling the hydrocarbon generation and migration in the West Netherlands Basin, the Netherlands. *Netherlands Journal of Geosciences*, **79**, 29–44, https://doi.org/10.1017/S0016774600021557

Van Wijhe, D.H. 1987. Structural evolution of inverted basins in the Dutch offshore. *Tectonophysics*, **137**, 171–219, https://doi.org/10.1016/0040-1951(87)90320-9

Vidal, J. and Genter, A. 2018. Overview of naturally permeable fractured reservoirs in the central and southern Upper Rhine Graben: insights from geothermal wells. *Geothermics*, **74**, 57–73, https://doi.org/10.1016/j.geothermics.2018.02.003

Wang, W. and Zhou, L. 2023. Fracture extraction from logging image using a dual encoder–decoder architecture with Swin Transformer. *Petrophysics*, **64**, 38–49, https://doi.org/10.30632/PJV64N1-2023a3

Wedge, D., Holden, E.-J., Dentith, M. and Spadaccini, N. 2015. Automated structure detection and analysis in televiewer images. *ASEG Extended Abstracts*, **2015**, https://doi.org/10.1071/ASEG2015ab152

Weiss, K., Khoshgoftaar, T.M. and Wang, D.D. 2016. A survey of transfer learning. *Journal of Big Data*, **3**, 9, https://doi.org/10.1186/s40537-016-0043-6

Willems, C.J.L. 2017. *Doublet Deployment Strategies for Geothermal Hot Sedimentary Aquifer Exploitation: Application to the Lower Cretaceous Nieuwerkerk Formation in the West Netherlands Basin*. PhD thesis, Delft University of Technology, Delft, The Netherlands, https://research.tudelft.nl/en/publications/doublet-deployment-strategies-for-geothermal-hot-sedimentary-aqui

Willems, C.J.L., Vondrak, A., Mijnlieff, H.F., Donselaar, M.E. and Van Kempen, B.M.M. 2020. Geology of the Upper Jurassic to Lower Cretaceous geothermal aquifers in the West Netherlands Basin – an overview. *Netherlands Journal of Geosciences*, **99**, e1, https://doi.org/10.1017/njg.2020.1

Witter, J.B., Trainor-Guitton, W.J. and Siler, D.L. 2019. Uncertainty and risk evaluation during the exploration stage of geothermal development: A review. *Geothermics*, **78**, 233–242, https://doi.org/10.1016/j.geothermics.2018.12.011

Wong, Th.E., Batjes, D.A.J. and de Jager, J. (eds) 2007. *Geology of the Netherlands*. Royal Netherlands Academy of Arts and Sciences, Amsterdam.

Worum, G., Michon, L., van Balen, R.T., van Wees, J.-D., Cloetingh, S. and Pagnier, H. 2005. Pre-Neogene controls on present-day fault activity in the West Netherlands Basin and Roer Valley Rift System (southern Netherlands): role of variations in fault orientation in a uniform low-stress regime. *Quaternary Science Reviews*, **24**, 473–488, https://doi.org/10.1016/j.quascirev.2004.02.020

Yang, Y., Zhang, M., Bie, A., Lv, J., Zhang, W., Cui, Z. and Xia, Z. 2020. A case study of subsurface uncertainty analysis in modelling carbonate reservoir. *In*: Lin, J. (ed.) *Proceedings of the International Field Exploration and Development Conference 2018*. Springer, Singapore, 851–864, https://doi.org/10.1007/978-981-13-7127-1_79

Zabalza-Mezghani, I., Manceau, E., Feraille, M. and Jourdan, A. 2004. Uncertainty management: From geological scenarios to production scheme optimization. *Journal of Petroleum Science and Engineering*, **44**, 11–25, https://doi.org/10.1016/j.petrol.2004.02.002