Assessing the Accuracy of Fault Interpretation using Machine Learning Techniques when Risking Faults for CO_2 Storage Site Assessment

E. A. H. Michie[1]*, B. Alaei[2, 1], A. Braathen[1]

[1] Department of Geosciences, University of Oslo, Sem Sælands Vei 1, Oslo 0371, Norway

[2] Earth Science Analytics AS, Strandveien 35, Oslo 1366, Norway

***Corresponding author:** e.m.haines@geo.uio.no

ABSTRACT

Generating an accurate model of the subsurface for the purpose of assessing the feasibility of a CO_2 storage site is crucial. In particular, how faults are interpreted is likely to influence the predicted capacity and integrity of the reservoir; whether this is through identifying high risk areas along the fault, where fluid is likely to flow across the fault, or by assessing the reactivation potential of the fault with increased pressure, causing fluid to flow up the fault. New technologies allow users to interpret faults effortlessly, and in much quicker time, utilizing methods such as Deep Learning. These Deep Learning techniques use knowledge from Neural Networks to allow end-users to compute areas where faults are likely to occur. Although these new technologies may be attractive due to reduced interpretation time, it is important to understand the inherent uncertainties in their ability to predict accurate fault geometries. Here, we compare Deep Learning fault interpretation versus manual fault interpretation, and can see distinct differences to those faults where significant ambiguity exists due to poor seismic resolution at the fault; we observe an increased irregularity when Deep Learning methods are used over conventional manual interpretation. This can result in significant differences between the resulting analyses, such as fault reactivation potential. Conversely, we observe that well-imaged faults show a close similarity between the resulting fault surfaces when both Deep Learning and manual fault interpretation methods are employed, and hence we also observe a close similarity between any attributes and fault analyses made.

INTRODUCTION

Deep Learning

The process by which seismic is interpreted has developed significantly over the years. Initially, seismic interpretation involved manual picking using printed seismic sections (Sheriff, 1981), which has since developed to provide users with the ability to interpret using a suite of digital environments (e.g. Al-Shuhail et al., 2017). The ease and accuracy of seismic interpretation is continually increasing, associated with advancing geophysical and rock physics tools (Chopra and Marfurt, 2007; Avseth et al., 2010), as well as the increased use of automated and semi-automated technologies (e.g. Araya-Polo et al., 2017; Bugge et al., 2018). While technology has progressed to allow users to quickly interpret horizons using facilities such as auto-tracking, the ability for machine learned algorithms for automated fault extraction has, until recently, been lacking. New technology has emerged that is based on multi-attribute support vector machine (SVM) classification, or that uses multi-layer perceptron (MLP) neural network, or more recently, Deep Learning (DL), i.e. Deep Neural Network inspired machine learning, to automatically extract faults from seismic, with minimal manual seismic fault interpretation (e.g. Zheng et al., 2014; Di et al., 2017; Huang et al., 2017; Zhao and Mukhopadhyay, 2018; Wu et al., 2020). DL used for automated fault extraction is based on machine learning (ML) techniques developed over decades, building on the well-known methods such as neural networks (Dramsch 2020). The use of neural networks in geophysics was first reviewed in McCormack (1991), as a use for predicting geological information such as lithologies. The use of neural networks in geology and geophysics mimics biological neurons, breaking down the process into several layers. In the simplest terms, neural networks contain an input layer that receives information from input data, which then

provides the data to hidden layer(s) that identifies certain parameters, before passing the information to the output layer. The output value can then act as another input value to retrain the data, or it can act as the final output answer. This iteration process is known as deep feedforward networks (Goodfellow et al., 2016).

The use of neural networks has recently been built on, improving the ability to identify locations of features that artificial neural networks don't have the ability to do. Further, multiple hidden layers between the input and output layer may be required for such processes, known as Deep Neural Networks (DNNs). Convolutional Neural Networks (CNN) are a type of DNN network that have certain sampling and convolutional layers that enable the ability to learn hierarchical feature representations from data sources, such as seismic images (Saha, 2018). A CNN extracts successively larger features into hierarchical layers, such that CNNs have fewer connections and parameters than feedforward networks, and hence they are easier to train. CNNs are used within this study, and the main type of CNN used is a modified version of the U-NET Architecture (Ronneberger, 2015). Initially developed for biomedical image analysis, the process has been utilized for geological analysis, as it is crucial for automatically identifying the locations of faults (Huang et al., 2017; Wu et al., 2020; Moseer et al., 2020). The U-NET Architecture relies on down-sampling, which is similar to that in a neural network, but only passes on the maximum value to the next layer, which allows the image size to be compressed such that data processing becomes highly efficient. Down-sampling is followed by up-scaling which restores the image to the original size and identifies the feature in the correct location (Ronneberger, 2015).

Importance of Deep Learning in CO2 storage site assessment

Carbon capture and storage (e.g. Birol, 2008; Rogelj et al., 2016) has been identified as one of the methods to reach the 2°C goal of the Paris Agreement. Irrespective of fault interpretation method on seismic data, an accurate structural framework model for fault analysis is crucial in terms of CO_2 storage site assessment, where the storage site is bound by faults.

One identified candidate for CO_2 storage is the Smeaheia site within the Northern Horda Platform, in the Norwegian North Sea (Halland et al., 2011; Statoil, 2016; Lothe et al., 2019). The prospect for this site is bound by a deep-seated basement fault, known as the Vette Fault Zone (VFZ) (Mulrooney et al., 2020; Wu et al., 2021). Fault analysis, such as assessing the likelihood for fault reactivation, is critical for all fault-bound CO_2 storage sites. When CO_2 is injected into the subsurface, fluid pressures are increased, which may cause the fault to become unstable and reactivate, allowing the CO_2 to migrate up the fault to shallower levels (e.g. Barton et al., 1995; Ferrill et al., 1999a; Streit and Hillis, 2004; Rutqvist et al., 2007; Chiaramonte et al., 2008). The ability of faults to reactivate upon an increased pore pressure is heavily dependent on the orientation and dip of the fault relative to the in situ stress conditions (e.g. Ferrill et al., 1999a; Mildren et al., 2005). Hence, an accurate subsurface model of the faults is critical for this analysis. Since methods are progressing towards automated technologies, we present results comparing structural interpretation of faults within the Smeaheia site using a traditional, manual interpretation approach with an automated approach using: i) a modified U-NET (a supervised 2D CNN algorithm) and ii) a pretrained synthetic 3D CNN utilized to identify faults within seismic images. This will provide an assessment of any potential discrepancies between methods, and an improved understanding of the influence of methodology on further geomechanical, i.e. fault

reactivation, analysis. The accuracy of defining fault structures in the subsurface is crucial for any fault-controlled prospect, and may be detrimental when considering a potential site for CO_2 storage.

STUDY AREA

Smeaheia is located approximately 40 km northwest of the Kollsnes processing plant, and around 20 km east of Troll East, in the Northern Horda Platform (Halland et al., 2011; Statoil, 2016; Lauritsen et al., 2018; Lothe et al., 2019; Mulrooney et al., 2020; Figure 1A). The Northern Horda Platform is a structural high along the eastern margin of the northern North Sea (Færseth, 1996; Whipp et al., 2014; Duffy et al., 2015; Mulrooney et al., 2020; Wu et al., 2021; Figure 1A), containing several deep-seated, west-dipping, basement faults. These basement faults, with km-scale throws, generate several half grabens across the Horda Platform (Badley et al., 1988; Yielding et al., 1991; Færseth, 1996; Bell et al., 2014; Whipp et al., 2014). These faults were first active during the Permo-Triassic, associated with an east-west phase of extension, and were subsequently reactivated during the Jurassic (Badley et al., 1984; Roberts et al., 1995; Færseth et al., 1995; Færseth 1996; Odinsen et al., 2000; Ter Voorde et al., 2000; Deng et al., 2017). Two first-order thick-skinned faults bound the half graben Smeaheia site: the VFZ and the Øygarden Fault Complex (ØFC) (Mulrooney et al., 2020; Figure 1A, B). The ØFC bounds the Beta prospect in its hanging wall, and the VFZ bounds the Alpha prospect in its footwall. Following a drilling campaign that recorded no oil shows in well 32/4-1 (Goldsmith, 2000; PL 205 Licence Group Well

32/4-1 T2 Final Well report, 1997), the site was proposed as a potential storage site for CO_2 within the saline aquifer.

The Jurassic to Cretaceous rifting event within the northern North Sea also saw the formation of thin-skinned northwest-southeast striking faults within the footwall of the VFZ displacing post-Upper Triassic stratigraphy (Mulrooney et al., 2020; Wu et al., 2021). These faults generally have low displacements (<100 m), and hence are not considered important for sealing CO_2. They may, however, be crucial for any potential fault reactivation. This study is focused on the Alpha prospect, hence the primary aim of this study is to accurately capture the properties of the bounding VFZ. While these thin-skinned faults are not the primary focus of the study, they have been used for supervised DL purposes, for labelling as well as for quality control (QC) of the final fault extraction model. Further, one footwall fault has been used as an example of fault extraction of a well-imaged fault.

The Sognefjord Formation of the Viking Group is the main reservoir for targeted storage. This formation is a shallow marine sandstone with very high permeability and porosity, of 440 - 4000 mD and 31% - 39%, respectively (Dreyer et al., 2005; Holgate et al., 2013; Patruno et al., 2015; Statoil 2016; Ringrose, 2017; Mondol et al., 2018). The primary caprock for the Alpha prospect within the Smeaheia site is the Upper Jurassic Draupne Formation. This is an organic-rich mudstone with a low permeability and high capillary threshold pressure (Skurtveit et al., 2015). However, in the immediate Footwall of the VFZ, the Draupne Formation thins significantly to less than 100 m and hence the overlying deep water marls, carbonates and shaley units of the Cromer Knoll and Shetland Groups may also become key caprocks (Nybakken and Bäckstrøm, 1989; Isaksen and Ledjie, 2001; Justwan and Dahl, 2005; Gradstein and Waters, 2016; Figure 1B).

Further, the Cromer Knoll Group is displaced next to the Sognefjord Formation along the VFZ, and hence also acts as an important side seal (Mulrooney et al., 2020).

DATA AND METHODOLOGY

The VFZ has been interpreted using the GN1101 3D survey. The GN1101 3D survey is a prestack time-migrated dataset covering the Smeaheia potential CO_2 storage site, in the footwall of the VFZ (Figure 1A). However, it is important to note that this survey does not extend far enough to the north and south to interpret the entire fault structure. Hence, only the part of the fault that is observed on GN1101 survey is analysed. The GN1101 survey was shot in 2011 by Gassnova SF, with an inline spacing of 25 m and a crossline spacing of 12.5 m, covering an area of 442.25 km^2. Crosslines are oriented 065°, and inlines oriented 155°. Shooting direction is almost normal to the trend of the VFZ. An increase in acoustic impedance corresponds to positive amplitude and data quality is overall good, with a resolution of roughly 16 m at the Sognefjord level. In places, poor seismic imaging of the VFZ occurs, whereas minor faulting within the footwall of the VFZ shows good seismic imaging. Data has subsequently been depth converted using a velocity model that has been created using QC'd Time-Depth curves from 15 wells from the Troll and Smeaheia area: 31/2-1, 31/2-2R, 31/2-4R, 31/2-5, 31/2-8, 31/3-1, 31/3-3, 31/5-2, 31/6-1, 31/6-2R, 31/6-3, 31/6-6, 32/2-1, 32/4-1 T2 and 32/4-3 S (Figure 1A). Other wells in the area have no velocity data.

Our approach has two main steps: structural interpretation and geomechanical analysis. We have used DNN and manual fault picking for the structural interpretation stage to provide the 3D

geometry of the VFZ. Fault surfaces have been created from each method, with geomechanical analysis performed on both fault surfaces, as a way to compare the two different fault interpretation approaches. The workflows that have been used in the two approaches are described in the following.

Fault interpretation using DNN Method

Supervised learning technique is implemented in this study to train multiple machine learning models to predict faults. The software EarthNet, an independent software provided from Earth Science Analytics, has been used for this analysis. The main steps of the applied workflow is shown in Figure 2, with example results shown in Figure 3. Supervised learning is where a known truth has been labelled (i.e. the interpreted faults), which is used along with a feature set (in this case the full stack seismic data). The training is carried out on a subset of data (85%) and test is performed on the remaining 15%. Once the training is complete, the algorithm predicts on the other portion of data that it has not trained on. The error between the predicted fault and the labelled faults is known as the loss function, and the new neuron weights are calculated according to the gradient descent method (Goodfellow et al., 2016) to optimize the model. Through the optimization process the loss is taken and is minimized. The optimizer used within this study is the Nesterov-accelerated Adaptive Moment Estimation (NADAM) (Dozat, 2016). Two loss functions have been used to detect errors within this study: Categorical Cross-Entropy and Jaccard. More than one loss functions have been used to minimize the combined loss in the optimization process, where the average of the multiple losses has been taken. Categorical Cross-Entropy is based on the accuracy, which is the ratio of the sum of successful predictions

over the sum of all predictions (Mannor et al., 2005). This Categorical Cross-Entropy function has subsequently been modified, producing a balanced loss function to tackle the imbalance between fault (one in a binary mask) and background (zero in a binary mask) (e.g. Wu et al., 2019). Jaccard is based on the Intersection over Union (IoU), which is the ratio of true positives over the population of the sample (Bertels et al., 2019). Different metrics for both training and test data are calculated and used for model verifications. The simplest method to examine error is using a confusion matrix. The confusion matrix is the errors created within different classifications. Fault imaging is a binary classification task that segments the input data into fault and non-fault (background) classes. Errors that occur when background is predicted as fault and fault is predicted as background are presented within a confusion matrix as percentage when these scenarios occur.

In this study we have used efficient and light U-NET algorithms which are modified versions of U-NET (e.g. Mosser et al., 2020). The U-NET Architecture (Ronneberger, 2015) has been originally developed for medical image segmentations and later on used in many other fields. However, as it currently exists, the architecture of U-NET is not suitable for fault interpretation on seismic data due to limited number of parameters in seismic data compared to medical imaging. Hence, a modified U-NET is used that contains less convolutional operations compared to the original one. This modified U-NET is a supervised 2D CNN. The light U-NET uses the standard skip connection of U-NET algorithm. Features are passed from the encoder, to the decoder, where they are concatenated with decoder feature maps. Different hyperparamters for the network architecture have been tried for optimising the process. The network architecture has 3-4 layers, up to 24 initial filters, 0.2 to 0.3 drop out, 2 filter multiplier and 0.3 overlap. When using U-NET

Architecture, the output of the neural network is a binary image containing either zero or one, depending on the prediction of non-fault or fault, respectively.

Like all other supervised ML approaches, the quality of the predicted faults using the DL depends on the quality of ground truth labels provided for the training. The labelling or fault interpretation is the most time consuming part of ML based fault interpretation. One of the possible alternatives is to apply pretrained synthetic fault models that train on either synthetic data (e.g. Wu et al., 2020) or faults from neighbouring areas. We have predicted faults using real seismic data pretrained fault model which trained on 20 surveys (excluding the GN1101 survey). This approach is so-called model assisted labeling. This enabled us to quickly label and QC a few inlines and crosslines, and trained multiple ML models using different model hyperparameters to predict faults. The only feature used in this study has been full stack seismic data. However, the same approach can be used with several features such as partial stack volumes. We have then applied the models that trained on few lines to generate fault volumes. In the next step we combined the fault volumes (taking the average of all volumes) to generate fault probability volumes. Such ensemble fault volumes have been used to increase and refine fault labels, by turning predictions from the model(s) that have passed a human QC step into more labels, and unrealistic predicted faults have been filtered. These labels are then used to retrain to further enrich the models built in the first step. This so-called ML assisted labelling enable us to label few lines and use the trained models to increase the training data. In this example we have labelled 23 inlines (at an equal spacing of 625 m, i.e. every 25 lines), and 23 crosslines. An example of a labelled inline is shown in Figure 3A. The chosen 23 crosslines that have been labelled are focused immediately around the VFZ, i.e. where the density of faulting is at the highest. Here,

the faults have been picked at a spacing of 625 m, i.e. every 50 lines. Labels have been added for *all* interpreted faults on each line for those faults penetrating the Shetland Group or lower, i.e. labels have not been added for the polygonal faults within the overburden (Figure 3A). This is to aid the training of the deeper structures, of interest in this study. Labels and hyperparameters used in the ML algorithms have been tweaked and altered at different stages in order to optimize modelling returns (Table 1). QC has been performed using the confusion matrix accuracy score, IoU score (example shown in Figure 3C) as well as visual examination (Figure 3B, D). Comparing the labelled faults with the predicted faults (Figure 3A, B), as well as showing predicted faults on sections with no labels (blind test) (Figure 3D), allows us to visually assess the quality of the model.

Once an acceptable model has been created, by assessing the accuracy using the confusion matrix and by visual examination of the predicted faults, the model is applied on the full 3D to generate a fault volume. Here, we have created two different models using the same input labels but altering one hyperparameter that alters the output results, i.e. the patch size. One model had a patch size of 320x320 and the other with 256x256. Differences in patch sizes changes the accuracy of the model, but also the segmentation of predicted fault segments. The patch size is the smallest size of data that is filtered through the CNN process, and has been shown to influence final results of the model (e.g. Hamwood et al., 2018). Hence, larger patch sizes are expected to produce more continuity in the imaged larger faults. Since a 2D CNN approach is adopted for this study, this is illustrated by the 2D patch sizes presented within this contribution. The two fault volumes created from these models have been used to create a probability volume by arithmetically averaging the two seismic volumes (Figure 3E, F). This probability volume

provides a confidence limit to the predicted faults, highlighting areas that are considered low/high uncertainty, where low/high probability is observed, respectively. This probability volume, which covers every inline and cross line, is then used to extract fault segments using the predicted fault locations (Figure 3G). These labels are transferred into the software T7 as fault segments to assess any differences in further fault analyses, such as fault reactivation risk, between manual and automatic fault extraction methods.

The process of extracting fault segments from the fault volumes of DNN models uses several parameters: probability threshold value (between 0 and 1), minimum length (where any segments with a length in time less than this are removed. This number is related to pixels and seismic sampling vertically and laterally), prune length (where any segments with leafs less than this are removed) and poly approximate (where the number of nodes along the segment are reduced). Here, we used a threshold of 0.1, minimum length of 1 (i.e. 1 pixel, where the model searches 1 pixel in the vertical direction for faults), prune length of 1 and poly approximate of 1. These parameters were considered to provide the best approximation of the extracted fault volume within the DNN models for this case study. Each fault segment is flagged as being separate faults. Hence, these fault segments were subsequently grouped by correlation into one fault, which is required to be done manually. Note, the fault segments were initially correlated with little to no QC, to assess purely the ML technique with minimal manual interpretation.

Machine learned fault extraction was used to create a fault surface comprising of fault segments that have been extracted on every line, every 4th line and every 32nd line, in order to assess the best line spacing extraction method, as well as to compare with manual interpretation using different line spacing (e.g. see Michie et al., 2021).

Fault interpretation using manual picking

Since fault interpretation of the VFZ using ML generates fault segments on every inline, corresponding to 25 m spacing, the VFZ has also been manually picked on every line. Rigorous QC-ing of the manual interpretation of the VFZ has been done in order to maintain continuity between each inline. Fault surfaces have also been created using fault segments that have been picked on every 4^{th} line and every 32^{nd} line.

Fault Framework generation

Fault segments generated from both manual picking and fault interpretation using ML have been used to generate a fault surface of the VFZ using unconstrained triangulation. Unconstrained triangulation honours all data points and generates a fault surface that allows the triangulation of fault segments without forcing the surface to conform to the lines between adjacent points on the same segment.

Geomechanical Analysis

Fault strike, dip and stability (specifically slip tendency and dilation tendency) attributes are calculated and mapped onto the fault surfaces at a resolution of 8 m lateral by 4 m vertical, providing an optimum resolution at the seismic scale without the need to extend processing time for a finer resolution.

Assessing the differences in predicted fault stability between manual versus automated fault picking strategy has been performed. Understanding differences between practical methodologies and how this may influence predicted reactivation potential of any bounding or intra-basin fault is crucial when considering any pressure increase from CO_2 injection.

An unpublished Equinor data package using four wells within the Horda Platform (31/6-3, 31/6-6, 32/4-1 and 32/2-1) has been used to define the *in situ* stress regime (Statoil, 2016). Vertical stress (Sv) has been determined from the overburden gradient; minimum horizontal stress (SHmin) has been determined from extended leak-off tests; pore pressure (Pp) has been measured as hydrostatic. The maximum horizontal stress (SHmax) is less well-constrained, but is assumed to be the same/similar as SHmin, using the documented stress orientation and faulting regime data based on nearby exploration and production wells. Further, shallower levels, <5 km, in this area of the northern North Sea are recorded as being within a normal faulting regime with almost isotropic horizontal stresses (Hillis and Nelson, 2005; Andrews et al., 2016; Skurtveit et al., 2018). Borehole breakout data provides information concerning the orientation for SHmax, and is recorded as trending E-W (Brudy and Kjørholt, 2001; Skurtveit et al., 2018). The in situ regime is shown in Table 2 and Figure 1C.

Using information regarding the predicted SGR, as documented by Michie et al. (2021), we can estimate the cohesion and frictional coefficient of the fault as 0.5 MPa and 0.45, respectively, using previously published values relating the proportion of clay material and the mechanical properties (Meng et al., 2016, and references therein). The SGR has been calculated as approximately 40% where the Sognefjord Formation is observed in the footwall.

Results of slip tendency and dilation tendency are shown within this paper. Slip tendency is the ratio of resolved shear stress (τ) to normal stress (σ_n) on a plane, where an increase in the value will predict an increase in the likelihood of the fault to reactivate by shear failure (Morris et al., 1996). Shear failure is generally taken as occurring at approximately 0.6, corresponding to the coefficient of static friction. In this example, the likelihood of the fault to slip depends solely on the stress field and orientation/dip of the fault surface. Dilation tendency is the relative probability of a plane to dilate within the current stress field (Ferrill et al., 1999b). This is a ratio between 0 and 1, where the higher the value, the more likely a fault will go into tensile failure. Dilation tendency takes into consideration the cohesion and tensile strength of the fault rock, as well as the stress field and orientation/dip of the fault surface, whereas slip tendency analysis assumes a cohesionless fault.

RESULTS

Comparisons have been made between fault surfaces that have been picked using traditional manual picking methods and ML techniques, at different picking intervals. Here, we show how fault surfaces and the subsequent attributes and fault analyses vary when picked on every line, every 4th line and every 32nd line. We subdivide results according to faults identified by different seismic quality, which influence the results from ML automated fault extraction. The VFZ has relatively poor seismic resolution, with a wide fault zone shown by decreased seismic quality. Conversely, minor faults surrounding the VFZ have significantly improved seismic quality compared to the VFZ. Differences in seismic quality show variations in the results of DNN models.

Vette Fault Zone: Poorly imaged fault

Starting with the coarse picking strategy of every 32nd line, we can see significant disparities in the modelled fault surfaces between manual verses ML fault picking (Figure 4). Despite the overall smoothing that tends to exist when a coarse line spacing is chosen, there remains a high propensity of the fault to appear as highly irregular when ML techniques are used, as compared to manual picking. The irregularity of the fault surface increases when the spacing for the picking strategy is decreased, as seen when examining the results for picking on every 4th line and on every line. When fault segments are picked manually using every crossing line, this can lead to a highly irregular fault surface, despite rigorous fault QC that is performed in order to improve correlation of fault segments between each line. However, a fault surface with increased irregularity is formed when ML techniques are used in all picking strategy scenarios, associated with increased segmentation. This is observed by the increased number of triangles, wider spread in triangle size as well as clustering of different sized triangles that is formed through ML techniques, related to irregular fault segments (Figure 4). The increased irregularity observed from ML techniques and also when using a narrower line spacing may lead to potential inaccuracies during any further fault analyses performed, if not viable.

The extents of the fault surface are also observed to differ with picking strategy, and between ML and manual methods (Figure 4). Specifically, there are places where the height of the fault surface is observed to be increased or decreased when ML methods are used. For example, the top of the fault is observed to extend to shallower levels in only a portion of the southern section of the fault when ML methods are used, which is not observed through manual interpretation.

This could be associated with over-correlation with polygonal faulting within the overburden. Since identifying the location of the base of the fault is highly ambiguous due to poor seismic resolution at depth, ML techniques create a fault surface with an increase irregular fault base, where the depth of the fault base varies considerably across the entire fault (Figure 4). However, through model-driven perception when manually picking the fault, the base of the fault is much smoother.

Dip

Fault surface irregularity is significantly higher when ML techniques are employed over manual interpretation, as described above. This is reflected in the increased number of irregular 'bulls-eye' patches of varying dip values displayed on the fault, observed on all three line spacing scenarios: picking on every line, every 4th line and every 32nd line (Figure 5). Moreover, these patches of irregular dips are predominantly steeper than the surrounding. Conversely, the patches of irregular dip when manual interpretation is performed is a combination of both steeper and gentler dips than the surrounding. The fault is smoothed such that no patches of irregular dip is observed when the fault is picked manually using a picking strategy of 32nd line spacing. The irregularities remain at the 32nd line spacing when ML techniques are used. However, they show both lower and higher dip compared to the background and are larger in size compared to irregularities observed in dip attributes at narrower line spacing.

Strike

Strike is shown to vary both with picking strategy and between manual interpretation and ML techniques (Figure 6). An increased variation to the strike observed along the fault occurs when every line is used to pick the fault. However, this increased irregularity is further enhanced when ML techniques are used. Specifically, more alternations to the fault orientation is observed along-strike, particularly in the southern section of the fault (from 10000 to 18000 m from the north), and the increase in orientation variation is observed in all ML line spacing instances. For example, similarities in the observed strike variation occur when results from the DNN model picked on every 4th line are comparable with manual interpretation using every line spacing (Figure 6).

Geomechanical analysis

Upon examining how the predicted fault stability (dilation tendency and slip tendency) of the VFZ changes with manual verses ML picking techniques, we can see distinct differences. Regardless of picking strategy spacing, we can observe an increased predicted dilation tendency when ML techniques are employed over manual interpretation (Figures 7, 9). Manual interpretation using different line spacing shows a gradual increase in predicted dilation tendency with a decrease in line spacing, where a fault is predicted to be closer to the failure envelope at every line spacing (i.e. higher predicted dilation tendency values), and further away from the failure envelope at every 32nd line spacing. This trend of a decrease in fault stability with a decrease in line spacing used for interpretation is also observed with increased line spacing when ML techniques are used. However, in all scenarios the fault is predicted to be at the failure envelope, regardless of

line spacing used for fault surface generation. Specifically, all three line spacing scenarios (every line, every 4th line and every 32nd line) show areas on the fault where the dilation tendency is 1 or over. This means that any increase in pore fluid pressure, e.g. through CO_2 injection, is likely to cause the fault to fail (under these specific input parameters).

We can observe that the bulls-eye patches of higher dips correspond to those areas of high dilation tendency (Figure 5 verses Figure 7), and hence it is these areas of steep dip that have an increased likelihood of failure upon injection of CO_2. Since these irregular high dip patches occur to a lesser degree when manual interpretation is used, the likelihood of the fault to reactivate by tensile failure is interpreted to be lower, particularly for those scenarios where the fault has been picked using a coarser line spacing. However, a question is posed as to whether these patches of irregularity are geologically accurate or simply a product of poor picking, human error and/or triangulation method.

A different observation is recorded for the resulting slip tendency (Figures 8, 9). Similar values of slip tendency are observed between manual and ML techniques. However, the highly irregular nature of the fault surfaces produced using every line spacing, regardless of technique used, creates a resulting slip tendency attribute which in places is near to, or at, the critical failure envelope of 0.6. The spread of slip tendency is also increased when ML techniques are used, associated with the increased surface irregularity. Hence, there are also areas along these faults where the interpreted stability is shown to be increased. Increasing the line spacing creates a narrower range to the slip tendency, and decreases the values away from the critical failure envelope of 0.6 (Figure 9).

Well imaged minor faulting

While machine learning techniques have shown to be challenging for areas of poor seismic quality creating significantly disparities between the two methodologies (manual vs. ML), and hence where manual picking based model-driven perceptions may produce a more viable result, other smaller faults that are better imaged show improved identification. Specifically, minor faults (up to 100 m displacement) within the footwall of the VFZ show accurate identification and have a significantly reduced segmentation through ML techniques (Figure 10), despite in several places not showing any sharp cutoffs, but rather identified by subtle folding. To qualitatively and quantitatively assess this improved automated fault extraction of the minor faults, we compare calculated dilation tendency using ML techniques with manual interpretation for one fault within the footwall of the VFZ: fault 'FW 01' (see Mulrooney et al., 2020 for location details of this fault). We can observed that FW 01 has significantly less segmentation than those picked for the VFZ, and in fact, only one segment is observed on the majority of lines for this fault. Moreover, the predicted dilation tendency is very similar between the ML and manually interpreted fault surfaces, which would lead to the same overall interpretation of fault stability (Figure 10). The only slight difference between ML and manual interpretation is the extent of the fault: ML techniques do not extrapolate deeper than manual interpretation, which is simply a product of the poor seismic resolution at greater depths.

DISCUSSION

Significant uncertainty occurs when interpreting structural data in the subsurface (Bond, 2015), and can be influenced by issues such as seismic quality (Alcalde et al., 2017), cognitive bias (Bond et al., 2007), picking strategy used (Tao and Alves 2019; Michie et al., 2021), or simply by flattening of sections leading to the creation of artifacts (Novoa et al., 2000). Specifically, cognitive bias manifests itself as conceptual models used to interpret the subsurface, which is guided by the interpreters' individual training (Bond et al., 2007; Alcalde et al., 2019; Shipton et al., 2020). This can lead to differences in the chosen interpreted location of the fault surface (Faleide et al., 2021), which can create significant differences in the resulting fault analysis, such as reactivation potential (Michie et al., 2021). Chosen line spacing for manual interpretation has also shown to influence any resulting fault analyses. Picking on every line tends to create a high irregular fault surface that is often a product of human error and/or triangulation method extrapolating the fault segments into a fault surface due to the higher number of data points (Michie et al., 2021). Further to these well-known concerns when interpreting subsurface data, interpreters need to be aware of how the speed of modern technology, such as auto-tracking and fault extraction methods, may impact the resulting framework model. Although these technologies significantly reduce the time needed to spend on certain tasks (Silva et al., 2005), they may introduce further uncertainty, which requires attention and acknowledgement when performing any further analyses. Here, we have presented the results of manual interpretation of faults picked using different line spacing compared with fault interpretation using ML. Ensuring the correct picking strategy has been chosen when manually interpreting faults, and understanding the uncertainties involved in both manual and automated fault extraction

methods, is fundamental in providing the most likely estimate of any subsequent fault stability analysis, which in turn is crucial for any assessment of the viability of a fault-bound CO_2 storage site.

Utilising modern advances in fault picking, i.e. automated fault extraction using supervised DNN, is a fast approach to the normally time-consuming manually interpretation strategy. However, as with any new technology, it is crucial to understand the influence of how, and to what extent, ML techniques may influence subsequent fault analyses such as fault stability. Differences in fault attributes between manual interpretation and ML are recorded, and hence understanding these disparities is fundamental when assessing a fault-bound CO_2 storage site, along with understanding their potential cause. In our case study, ML results showed more irregularity to the fault surfaces, especially for the poorly imaged VFZ. It could be argued that fault surfaces are highly irregular in nature and hence the overly irregular faults produced by DNN method could be due to automated extraction picking every kink or bend in a fault, that may be missed by manual interpretation. Hence, the increased number of data points will result in an increased fault surface irregularity. Large amounts of data exist in geophysical studies, thereby a neural network might find hidden irregularities in the data that manual interpreters may have overlooked. On the other hand, the manual fault pick is model-driven, reflecting the interpreter's perception of what a fault should look like. **Error! Reference source not found.**Further, discrepancies may appear due to areas of false positive interpretation by the DNN models that we applied, particularly in areas of poor seismic resolution. As with manual fault picks, errors associated with ML techniques will be integrated in the fault model. Hence, it is crucial to identify areas of high uncertainty that are not best capture through ML techniques, and would require

further attention through manual QC. A similar result has previously been presented by Qi et al. (2019) who stated that, similar to autopicked horizons, it is recommended that the interpreter spend time to manually QC the results to ensure no incorrectly joined fault segments or picked artifacts have occurred. For the poorly imaged VFZ, we can say that confidence is low when performing ML fault extraction due to the high disparity between manual versus ML interpretation, creating high uncertainty. Specifically, identifying short-comings through manual QC may indicate the need for continued ML retraining to improve the model, by either further hyperparameter tuning and/or improving the input labels, and more ensemble modelling. On the contrary, well-imaged minor faulting show excellent correlation between manual and machine learned fault picking, with the resulting predicted dilation tendency being very similar between these two methods. Hence, for our case study, improved seismic quality reduces the uncertainty and increases our confidence when utilizing advancing ML technologies. This would lead to the assumption that much less later-stage manual QC would be required.

Areas of high fault complexity along fault strike also lead to a decrease in seismic quality, and hence create a high degree of ambiguity and uncertainty, and often leading to poor picking through ML techniques. For example, areas surrounding the relay to the southern end of the VFZ (known as Vette_2, see Mulrooney et al., 2020**Error! Reference source not found.**) have been poorly identified (Figure 11B). The high complexity assumed for the basin-scale VFZ results in a poorer seismic resolution at the fault. This leads to poor predictions of the location of the fault, and even some areas of the fault being missed. One assumption for this case study is that the larger faults generally contain higher complexity than the minor faults. For example, areas of breached or un-breached relays, asperities, large fault zones etc., and all these complexities

contribute to the deterioration of seismic quality. Hence, for larger-scale faulting, a higher uncertainty remains when performing ML fault extraction, in comparison to the model-driven manual interpretation approach. Conversely, the lower complexities often observed within minor fault zones manifests itself as a higher seismic quality. Hence, users will have an increased confidence and lower uncertainty when utilizing ML technologies.

Seismic data quality is an issue, which influences both manual and DNN results. The survey used in this study is prestack time migrated which is not the best imaging approach in cases with structural complexity, as well as large lateral variations in velocity. Fault shadow effect is observed in the data which deteriorate the quality of seismic data underneath large fault planes and causes challenges for both manual and DNN based fault picking. Our applied DNN models in case of large faults provided vertically segmented images. The networks use features on seismic data that can represent discontinuity. This can happen for example in areas with weak discontinuities on seismic data (e.g. packages with very low reflectivity). Hence, this likely to be an explanation for the increased irregularity of the VFZ. In manual interpretation, we usually ignore such weak reflectivity areas and extend picks over them. One approach to overcome this issue was to take ensemble of multiple DNN results. Figure 11B and D show the improvement of vertical and lateral continuity by using ensemble results, however some irregularity, and hence high uncertainty, remain.

If the highly irregular nature of the fault surface produced through the use of DNN models are assumed to be correct, then this would lead to the interpretation of a fault that is highly unstable, potentially rendering the CO_2 storage site unfeasible. However, if it is acknowledged that the irregularity is associated with potential short-comings of the ML technique and further QC was

performed, then this may lead to the interpretation of a fault with an increased stability, increasing the viability of the CO_2 storage site. Since interpretation technique and picking method has a crucial influence on the resulting predicted fault stability, it is detrimental to understand any and all uncertainties with each technique to improve interpreted fault stability for CO_2 storage feasibility assessment.

Further to the inherent uncertainties to the automated fault extraction method, the process of transforming the DNN models into fault segments may increase the fault surface irregularity further. This process uses the parameters: probability threshold value, minimum length, prune length and poly approximate (see section 3.1. for more detail). Certain parameters, such as poly approximate (where the number of nodes along the segment are reduced), may lead to an increased segmentation to a single fault where a low value is used. However, if a larger poly approximate value was used, the precise location of the fault would no longer be represented as the number of nodes on the fault segment is reduced and the fault location is approximated. It is important to note that since the process of fault segment creation from DNN models will be non-unique due to the varying parameters, we have only showed one example of the final fault surface that may be created from a single DNN model.

While uncertainties to the faults extracted from DNN models can lead to irregular surfaces, ML may also provide the opportunity to better understand and interpret the inner complexities of fault zones without the creation of fault framework models. Through manual interpretation, the interpreter will generally utilize a model driven approach, essentially ignoring any observed complexities of faults as these are often very difficult to visualize and to use as input into a framework model. Using DNN fault volumes, we can identify areas where fault complexities, i.e.

where multiple slip surfaces have been defined, which have not been picked through manual interpretation (Figure 12). Here, we have utilized a synthetic pretrained 3D CNN model to identify areas as faults. This enhanced fault extraction provides the opportunity to better understand the fault growth areas where possible 'risk' for across- or along-fault fluid flow may occur. While these complexities may not be possible to input into a framework model, they can be used to improve our ability to interpret where to precisely pick the principal slip surface (PSS) of a fault zone. Note that increased detail is observed in the results when a pretrained model is used over the 3D synthetic trained CNN. Not all fault zone complexity is identified when a 3D synthetic trained CNN is used, with more noise identified. As Michie et al. (2021) has showed, two different interpreters picking on either the hanging wall or footwall side of a fault may have significant effect on the resultant fault stability results, where one fault may be interpreted as stable and the other as unstable. Hence, using DNN models to aid interpretation of the correct location of a PSS may be crucial for future fault analyses such as fault stability analysis.

CONCLUSION

While machine learning techniques provide an optimum methodology to interpret sizeable areas of seismic data rapidly, there are some crucial pitfalls that require attention to reduce uncertainties when performing any subsequent fault analyses. Specifically, it has been noted that automated methods of fault extraction are sensitive to the quality of seismic data. Poorer seismic imaging of faults, such as the Vette Fault Zone, create fault surfaces with increased irregularity when compared to manual interpretation, leading to high uncertainty. For this fault

example, this increased irregularity lead to predictions of higher dilation tendency values in all line spacing scenarios. As such, an unstable fault would be interpreted, which may not be correct but may be detrimental when assessing a potential CO_2 storage site. The poor seismic imaging of this fault may simply be a product of the large nature of this fault: showing km-scale offset and 10's km in length, which will create a wide fault zone with severe heterogeneity and irregularity. Using ensemble models, larger coverage of faults were imaged using Deep Neural Networks with additional information of the confidence of predicted faults. Further fine-tuning of hyperparameters and fault label picks can also potentially improve the results for these poorly imaged faults. On the other hand, minor faults that are well imaged show excellent correlation between machine learned and manual interpretation techniques, improving our confidence and lowering any uncertainty in using Deep Neural Networks as a rapid fault extraction method. Understanding the uncertainty created by seismic quality on machine learned results will aid the assessment of any fault-bound CO_2 storage site.

Through continued technological advances, it is likely that accurate geological models can be produced in much shorter timeframes than through manual interpretation. However, quality control through manual interpretation will remain a necessity to ensure accurate fault framework generation, reducing our uncertainty in any further fault analyses such as fault stability analysis. This is crucial for evaluating any storage or production site.

ACKNOWLEDGEMENTS

This is a contribution of the FRISK project, supported by the Research Council of Norway (RCN# 295061). Support from the NCCS Centre is acknowledged, performed under the Norwegian research program Centres for Environment-friendly Energy Research (FME). The authors acknowledge the following partners for their contributions: Aker Solutions, Ansaldo Energia, CoorsTek Membrane Sciences, EMGS, Equinor, Gassco, Krohne, Larvik Shipping, Lundin, Norcem, Norwegian Oil and Gas, Quad Geometrics, Total, Vår Energi, and the Research Council of Norway (RCN# 257579/E20). Earth Science Analytics AS and Badley Geoscience Ltd. are thanked for their academic license of EarthNet and T7, respectively. Great thanks is given to four anonymous reviewers that helped to improve the quality of this work.

REFERENCES

Alcalde, J., Bond, C.E., Johnson, G., Ellis, J.F. and Butler, R.W, 2017, Impact of seismic image quality on fault interpretation uncertainty: GSA Today, **27**, no. 2, 4-10.

Al-Shuhail, A.A., Al-Dossary, S.A. and Mousa, W.A, 2017, Seismic data interpretation using digital image processing: John Wiley & Sons, 1-184.

Andrews, J.S., Fintland, T.G., Helstrup, O.A., Horsrud, P. and Raaen, A.M, 2016, June, Use of unique database of good quality stress data to investigate theories of fracture initiation, fracture propagation and the stress state in the subsurface: In 50th US Rock Mechanics/Geomechanics Symposium, American Rock Mechanics Association, ARMA-2016-887.

Araya-Polo, M., Dahlke, T., Frogner, C., Zhang, C., Poggio, T. and Hohl, D, 2017, Automated fault detection without seismic processing; The Leading Edge, **36**, no.3, 208-214.

Avseth, P., Mukerji, T. and Mavko, G, 2010, Quantitative seismic interpretation: Applying rock physics tools to reduce interpretation risk: Cambridge university press, 1-359.

Badley, M.E., Egeberg, T. and Nipen, O, 1984, Development of rift basins illustrated by the structural evolution of the Oseberg feature, Block 30/6, offshore Norway: Journal of the Geological Society, **141**, no.4, 639-649.

Badley, M.E., Price, J.D., Dahl, C.R. and Agdestein, T, 1988, The structural evolution of the northern Viking Graben and its bearing upon extensional modes of basin formation: Journal of the Geological Society, **145**, no.3, 455-472.

Barton, C.A., Zoback, M.D. and Moos, D, 1995, Fluid flow along potentially active faults in crystalline rock: Geology, **23**, no.8, 683-686.

Bell, R.E., Jackson, C.L., Elliott, G.M., Gawthorpe, R.L., Sharp, I.R. and Michelsen, L, 2014, Insights into the development of major rift-related unconformities from geologically constrained subsidence modelling: Halten Terrace, offshore mid Norway: Basin Research, **26**, no.1, 203-224.

Bertels, J., Eelbode, T., Berman, M., Vandermeulen, D., Maes, F., Bisschops, R. and Blaschko, M.B, 2019, Optimizing the Dice score and Jaccard index for medical image segmentation: Theory and practice: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, Cham, 92-100.

Birol, F, 2008, World energy outlook. Paris: International Energy Agency, **23**, no.4, 329.

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**, no.11, 4-10.

Bond, C.E, 2015, Uncertainty in structural interpretation: Lessons to be learnt: Journal of Structural Geology, **74**, 185-200.

Brudy, M. and Kjørholt, H, 2001, Stress orientation on the Norwegian continental shelf derived from borehole failures observed in high-resolution borehole imaging logs: Tectonophysics, **337**, no.1-2, 65-84.

Bugge, A.J., Clark, S.R., Lie, J.E. and Faleide, J.I, 2018, A case study on semiautomatic seismic interpretation of unconformities and faults in the southwestern Barents Sea: Interpretation, 6, no.2, SD29-SD40.

Chiaramonte, L., Zoback, M.D., Friedmann, J. and Stamp, V, 2008, Seal integrity and feasibility of CO2 sequestration in the Teapot Dome EOR pilot: geomechanical site characterization: Environmental Geology, **54**, no.8, 1667-1675.

Chopra, S. and Marfurt, K.J, 2007, Seismic attributes for prospect identification and reservoir characterization: Society of Exploration Geophysicists and European Association of Geoscientists and Engineers, 1-464.

Deng, C., Fossen, H., Gawthorpe, R.L., Rotevatn, A., Jackson, C.A. and FazliKhani, H, 2017, Influence of fault reactivation during multiphase rifting: The Oseberg area, northern North Sea rift: Marine and Petroleum Geology, **86**, 1252-1272.

Di, H., Shafiq, M.A. and AlRegib, G, 2017, Seismic-fault detection based on multiattribute support vector machine analysis: SEG International Exposition and Annual Meeting, 2017, OnePetro.

Dramsch, J.S, 2020, 70 years of machine learning in geoscience in review: Advances in Geophysics, **61**, 1-55.

Dreyer, T., Whitaker, M., Dexter, J., Flesche, H. and Larsen, E, 2005, From spit system to tide-dominated delta: integrated reservoir model of the Upper Jurassic Sognefjord Formation on the Troll West Field: Geological Society, London, Petroleum Geology Conference series, Geological Society of London, **6**, no.1, 423-448.

Duffy, O.B., Bell, R.E., Jackson, C.A.L., Gawthorpe, R.L. and Whipp, P.S, 2015, Fault growth and interactions in a multiphase rift fault network: Horda Platform, Norwegian North Sea: Journal of Structural Geology, **80**, 99-119.

Færseth, R.B., Gabrielsen, R.H. and Hurich, C.A, 1995, Influence of basement in structuring of the North Sea basin, offshore southwest Norway: Norsk Geologisk Tidsskrift, **75**, no.2-3, 105-119.

Færseth, R.B, 1996, Interaction of Permo-Triassic and Jurassic extensional fault-blocks during the development of the northern North Sea: Journal of the Geological Society, **153**, no.6, 931-944.

Faleide, T. S., Braathen, A., Lecomte, I., Mulrooney, M. J., Midtkandal, I., Bugge, A. J. and Planke, S, 2021, Impacts of seismic resolution on fault interpretations: Insights from seismic modelling: Tectonophysics, **816**, 229008.

Ferrill, D.A., Winterle, J., Wittmeyer, G., Sims, D., Colton, S., Armstrong, A. and Morris, A.P, 1999a, Stressed rock strains groundwater at Yucca Mountain, Nevada: GSA Today, **9**, no.5, 1-8.

Ferrill, D.A., Stamatakos, J.A. and Sims, D, 1999b, Normal fault corrugation: Implications for growth and seismicity of active normal faults: Journal of Structural Geology, **21**, no.8-9, 1027-1038.

Goldsmith, P.J, 2000, Exploration potential east of the Troll Field, offshore Norway, after dry well 32/4-1: Norwegian Petroleum Society Special Publications, Elsevier, **9**, 65-97.

Goodfellow, I., Bengio, Y., Courville, A. and Bengio, Y, 2016, Deep learning: Cambridge: MIT press, **1**, no.2.

Gradstein, F.M. and Waters, C.N, 2016, Stratigraphic guide to the Cromer Knoll, Shetland and Chalk Groups, North Sea and Norwegian Sea: Newsletters on Stratigraphy, **49**, no.1, 71-280.

Halland, E.K., Johansen, W.T. and Riis, F, 2011, CO2 storage atlas: Norwegian North Sea: Norwegian Petroleum Directorate, PO Box 600, NO–4003 Stavanger, Norway. URL http://www.npd.no/no/Publikasjoner/Rapporter/CO2-lagringsatlas

Hamwood, J., Alonso-Caneiro, D., Read, S.A., Vincent, S.J. and Collins, M.J, 2018, Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of OCT retinal layers: Biomedical optics express, **9**, no.7, 3049-3066.

Hillis, R.R. and Nelson, E.J, 2005, In situ stresses in the North Sea and their applications: Petroleum geomechanics from exploration to development: Geological Society, London, Petroleum Geology Conference series, Geological Society of London, **6**, no.1, 551-564.

Holgate, N.E., Jackson, C.A.L., Hampson, G.J. and Dreyer, T, 2013, Sedimentology and sequence stratigraphy of the middle–upper Jurassic Krossfjord and Fensfjord formations, Troll Field, northern North Sea: Petroleum Geoscience, **19**, no.3, 237-258.

Huang, L., Dong, X. and Clee, T.E, 2017, A scalable deep learning platform for identifying geologic features from seismic attributes: The Leading Edge, **36**, no.3, 249-256.

Isaksen, G.H. and Ledje, K.H.I, 2001, Source rock quality and hydrocarbon migration pathways within the greater Utsira High area, Viking Graben, Norwegian North Sea: AAPG bulletin, **85**, no.5, 861-883.

Justwan, H. and Dahl, B, 2005, Quantitative hydrocarbon potential mapping and organofacies study in the Greater Balder Area, Norwegian North Sea: Geological Society, London, Petroleum Geology Conference series, Geological Society of London, **6**, no.1, 1317-1329.

Lauritsen, H., Kassold, S., Meneguolo, R. and Furre, A, 2018, Assessing potential influence of nearby hydrocarbon production on CO2 storage at Smeaheia: Fifth CO2 Geological Storage Workshop, European Association of Geoscientists & Engineers, **2018**, no.1, 1-5.

Lothe, A.E., Bergmo, P.E. and Grimstad, A.A, 2019, Storage resources for future European CCS deployment; A roadmap for a Horda CO2 storage hub, offshore Norway: TCCS–10. CO2 Capture, Transport

and Storage. Trondheim 17th–19th June 2019. Selected papers from the 10th International Trondheim CCS Conference, SINTEF akademisk forlag, **4**, 39-48.

Mannor, S., Peleg, D. and Rubinstein, R, 2005, The cross entropy method for classification: Proceedings of the 22nd international conference on Machine learning, 561-568.

McCormack, M.D, 1991, Neural computing in geophysics: The Leading Edge, **10**, no.1, 11-15.

Meng, L., Fu, X., Lv, Y., Li, X., Cheng, Y., Li, T. and Jin, Y, 2017, Risking fault reactivation induced by gas injection into depleted reservoirs based on the heterogeneity of geomechanical properties of fault zones: Petroleum Geoscience, **23**, no.1, 29-38.

Michie, E.A., Mulrooney, M.J. and Braathen, A, 2021, Fault interpretation uncertainties using seismic data, and the effects on fault seal analysis: a case study from the Horda Platform, with implications for CO2 storage: Solid Earth, **12**, no.6, 1259-1286.

Mildren, S.D., Hillis, R.R., Lyon, P.J., Meyer, J.J., Dewhurst, D.N. and Boult, P.J, 2005, FAST: a new technique for geomechanical assessment of the risk of reactivation-related breach of fault seals: AAPG Hedberg Series, Evaluating Fault and Cap Rock Seals, **2**, 73-85.

Mondol, N.H., Fawad, M. and Park, J, 2018, Petrophysical analysis and rock physics diagnostics of Sognefjord Formation in the Smeaheia area, Northern North Sea: Fifth CO2 Geological Storage Workshop, European Association of Geoscientists & Engineers, **2018**, no.1, 1-5.

Morris, A., Ferrill, D.A. and Henderson, D.B, 1996, Slip-tendency analysis and fault reactivation: Geology, **24**, no.3, 275-278.

Mosser, L., Purves, S. and Naeini, E.Z, 2020, Deep Bayesian neural networks for fault identification and uncertainty quantification: First EAGE Digitalization Conference and Exhibition, European Association of Geoscientists & Engineers, **2020**, no.1, 1-5.

Mulrooney, M.J., Osmond, J.L., Skurtveit, E., Faleide, J.I. and Braathen, A, 2020, Structural analysis of the Smeaheia fault block, a potential CO2 storage site, northern Horda Platform, North Sea: Marine and Petroleum Geology, **121**, 104598.

Novoa, E., Suppe, J. and Shaw, J.H, 2000, Inclined-shear restoration of growth folds: AAPG bulletin, **84**, no.6, 787-804.

Nybakken, S. and Bäckstrøm, S.A, 1989, Shetland Group: stratigraphic subdivision and regional correlation in the Norwegian North Sea: Correlation in Hydrocarbon Exploration, Springer, Dordrecht, 253-269.

Odinsen, T., Reemst, P., Van Der Beek, P., Faleide, J.I. and Gabrielsen, R.H, 2000, Permo-Triassic and Jurassic extension in the northern North Sea: results from tectonostratigraphic forward modelling: Geological Society, London, Special Publications, **167**, no.1, 83-103.

Patruno, S., Hampson, G.J., Jackson, C.A.L. and Whipp, P.S, 2015, Quantitative progradation dynamics and stratigraphic architecture of ancient shallow-marine clinoform sets: a new method and its application to the Upper Jurassic Sognefjord Formation, Troll Field, offshore Norway: Basin Research, **27**, no.4, 412-452.

PL 205 Licence Group Well 32/4-1 T2 Final Well report, 1997, https://factpages.npd.no/pbl/wellbore_documents/2918_32_4_1_COMPLETION_REPORT_AND_COMPLETION_LOG.pdf

Qi, J., Lyu, B., AlAli, A., Machado, G., Hu, Y. and Marfurt, K, 2019, Image processing of seismic attributes for automatic fault extraction: Geophysics, **84**, no.1, O25-O37.

Ringrose, P.S., Thorsen, R., Zweigel, P., Nazarian, B., Furre, A.K., Paasch, B., Thompson, N. and Karstad, P.I, 2017, Ranking and risking alternative CO2 storage sites offshore Norway: Fourth Sustainable Earth Sciences Conference, European Association of Geoscientists & Engineers, **2017**, no.1, 1-5.

Roberts, A.M., Yielding, G., Kusznir, N.J., Walker, I.M. and Dorn-Lopez, D, 1995, Quantitative analysis of Triassic extension in the northern Viking Graben: Journal of the Geological Society, **152**, no.1, 15-26.

Rogelj, J., Den Elzen, M., Höhne, N., Fransen, T., Fekete, H., Winkler, H., Schaeffer, R., Sha, F., Riahi, K. and Meinshausen, M, 2016, Paris Agreement climate proposals need a boost to keep warming well below 2 C: Nature, **534**, no.7609, 631-639.

Ronneberger, O., Fischer, P. and Brox, T, 2015, U-net: Convolutional networks for biomedical image segmentation: International Conference on Medical image computing and computer-assisted intervention, Springer, Cham, 234-241.

Rutqvist, J., Birkholzer, J., Cappa, F. and Tsang, C.F, 2007, Estimating maximum sustainable injection pressure during geological sequestration of CO2 using coupled fluid flow and geomechanical fault-slip analysis: Energy Conversion and Management, **48**, no.6, 1798-1807.

Saha, S, 2018, A comprehensive guide to convolutional neural networks—the ELI5 way: Towards Data Science, 1-15.

Sheriff, R.E, 1981, Structural interpretation of seismic data: American Association of Petroleum Geologists, **23**.

Silva, C.C., Marcolino, C.S. and Lima, F.D, 2005, Automatic fault extraction using ant tracking algorithm in the Marlim South Field, Campos Basin: 2005 SEG Annual Meeting. OnePetro.

Skurtveit, E., Grande, L., Ogebule, O.Y., Gabrielsen, R.H., Faleide, J.I., Mondol, N.H., Maurer, R. and Horsrud, P, 2015, Mechanical testing and sealing capacity of the Upper Jurassic Draupne Formation, North Sea: 49th US Rock Mechanics/Geomechanics Symposium, American Rock Mechanics Association, ARMA-2015-331.

Skurtveit, E., Choi, J.C., Osmond, J., Mulrooney, M. and Braathen, A, 2018, 3D fault integrity screening for smeaheia CO2 injection site: 14th Greenhouse Gas Control Technologies Conference Melbourne, 21-26.

Statoil, 2016. Subsurface Evaluation of Smeaheia as part of 2016 Feasibility study on CO_2 storage in the Norwegian Continental Shelf. OED 15/1785. Document A – Underground report Smeaheia (Internal Report – Available on Request Only).

Streit, J.E. and Hillis, R.R, 2004, Estimating fault stability and sustainable fluid pressures for underground storage of CO2 in porous rock. Energy, 29(9-10), pp.1445-1456.

Tao, Z. and Alves, T.M., 2019. Impacts of data sampling on the interpretation of normal fault propagation and segment linkage: Tectonophysics, **762**, pp.79-96.

Ter Voorde, M., Færseth, R.B., Gabrielsen, R.H. and Cloetingh, S.A.P.L, 2000, Repeated lithosphere extension in the northern Viking Graben: a coupled or a decoupled rheology?: Geological Society, London, Special Publications, **167**, no.1, 59-81.

Whipp, P.S., Jackson, C.L., Gawthorpe, R.L., Dreyer, T. and Quinn, D, 2014, Normal fault array evolution above a reactivated rift fabric; a subsurface example from the northern Horda Platform, Norwegian North Sea: Basin Research, **26**, no.4, 523-549.

Wu, X., Liang, L., Shi, Y. and Fomel, S, 2019, FaultSeg3D: Using synthetic data sets to train an end-to-end convolutional neural network for 3D seismic fault segmentation: Geophysics, **84**, no.3, IM35-IM45.

Wu, X., Geng, Z., Shi, Y., Pham, N., Fomel, S. and Caumon, G, 2020, Building realistic structure models to train convolutional neural networks for seismic structural interpretation: Geophysics, **85**, no.4, WA27-WA39.

Wu, L., Thorsen, R., Ottesen, S., Meneguolo, R., Hartvedt, K., Ringrose, P. and Nazarian, B, 2021, Significance of fault seal in assessing CO2 storage capacity and containment risks—an example from the Horda Platform, northern North Sea: Petroleum Geoscience, **27**, petgeo2020-102.

Yielding, G., Badley, M.E. and Freeman, B, 1991, Seismic reflections from normal faults in the northern North Sea: Geological Society, London, Special Publications, **56**, no.1, 79-89.

Zheng, Z.H., Kavousi, P. and Di, H.B, 2014, Multi-attributes and neural network-based fault detection in 3D seismic interpretation: Advanced Materials Research, Trans Tech Publications Ltd, **838**, 1497-1502.

Zhao, T. and Mukhopadhyay, P, 2018, A fault detection workflow using deep learning and image processing: SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists, 1966-1970.

FIGURES

Figure 1. (a) Location of the Smeaheia site within the Northern Horda platform, indicated by the Alpha prospect, partially covering the GN1101 survey. Graben-bounding faulting shown, along with the hydrocarbon contact depth of the Troll field (grey outline). Seismic extents of the 3D survey used in the analysis is outlined in a black dashed outline: GN1101. Wells used in the analysis shown. Blocks shown. Norwegian coastline outlined in green with the Kollsnes processing plant highlighted for reference. From Norwegian Petroleum Directorate Fact Maps (http://factmaps.npd.no/factmaps/3_0/). (b) Example seismic section of row 1200 from GN1101 showing the main lithostratigraphic units and faults. Location of row 1200 highlighted in figure (a). Seismic section shows a colourmap with amplitudes from -127 to 127. (c) In situ stress field shown using the combined stresses (in MPa). Pp: pore pressure. SHmin: Minimum horizontal stress. SHmax: Maximum horizontal stress. Sv: Vertical stress.

Figure 2. Fault imaging workflow using supervised CNN, showing the main iterative processes used.

Figure 3. Seismic sections showing key steps of the fault imaging workflow, along with the results shown as ML fault segments. Seismic sections and timeslices show a colourmap with amplitudes from -127 to 127. (a) Example of picked labels on inline 1150, with corresponding predicted faults on the same inline (b) allowing for a first pass visual QC. (c) Predicted faults on an inline (1072) with no labels, used as visual QC. Confusion matrix and IoU scores of the model shown (d), used for additional QC. (e) and (f) SEGY probability cube combining two models with differing patch sizes shown on inline 1072 and timeslice 1652, respectively. (g) Extracted fault segments from the SEGY cube. (h) SEGY probability cube showing predicted faults using 3D synthetic trained CNN rather than a supervised 2D CNN. Note the different probability cutoff values for the scale bar between figures (e) and (f), and (h); synthetic trained CNN does not predict fault locations as

well as supervised 2D CNN for this example, and picks up a lot of 'noise' from the basement as faults.

Figure 4. Triangulated fault surfaces of the VFZ, using unconstrained triangulation, showing the triangle elements when manual interpretation vs. machine learned techniques are used, picked at differing line spacing (every line, every 4th line and every 32nd line). No fault segments shown.

Figure 5. Fault plane diagrams showing fault dip attribute displayed on the fault surfaces for different picking strategies: manual interpretation verses machine learned techniques, at different line spacing: 1, 4, and 32 lines. Fault dip is observed to vary with line spacing used for fault picking, as well as picking technique used. A highly irregular fault surface is observed with highly variable dip attributes when every line is used for picking, when compared to the overly smooth surface when every 32nd line is used for picking. The irregularity of the fault surface and hence also dip attribute is increased when machine learned techniques are employed over manual interpretation, in all spacing instances. Fault surface irregularity often manifests itself as bulls-eye patches of varying fault dip. Note that unconstrained triangulation is used for fault surface generation.

Figure 6. Fault plane diagrams showing fault strike attribute displayed on the fault surfaces for different picking strategies: manual interpretation verses machine learned techniques, at different line spacing: 1, 4, and 32 lines. Fault strike is observed to vary with line spacing used for fault picking, as well as picking technique used. The highly irregular fault surface observed with increased variation in strike attribute occurs when every line is used for picking, compared to the overly smooth surface when every 32nd line is used for picking. The irregularity of the fault surface, and hence also strike attribute, is increased when machine learned techniques are employed over manual interpretation, such that the varying strike attribute observed when every 4th line spacing is used for machine learned techniques is similar to that observed when every line spacing is used for manual interpretation. Note that unconstrained triangulation is used for fault surface generation.

Figure 7. Fault plane diagrams showing dilation tendency attribute displayed on the fault surfaces for different picking strategies: manual interpretation verses machine learned techniques, at different line spacing: 1, 4, and 32 lines. Dilation tendency is observed to vary with line spacing used for fault picking, as well as picking technique used. The irregular fault surfaces created with manual interpretation using narrower line spacing, as well as using machine learned techniques with any line spacing, produce fault surfaces that are interpreted to be unstable, i.e. the dilation tendency in places exceeds 1 (white areas). Conversely, the overly smoothed fault

surface when every 32nd line spacing is manually picked creates an interpreted relatively stable fault, with no areas where the dilation tendency exceeds 1. Note that unconstrained triangulation is used for fault surface generation.

Figure 8. Fault plane diagrams showing slip tendency attribute displayed on the fault surfaces for different picking strategies: manual interpretation verses machine learned techniques, at different line spacing: 1, 4, and 32 lines. Slip tendency is observed to vary with line spacing used for fault picking, as well as picking technique used. Although patches occur when the slip tendency is low, and hence more stable, the irregular fault surfaces created with using narrower line spacing (i.e. picked on every line) produce areas where the slip tendency is close to, or exceeds 0.6, and hence an unstable fault is interpreted. Increasing the line spacing creates a fault that is interpreted to be less likely to slip by shear failure. The increased irregularity when machine learned techniques are employed creates a wider spread to the predicted slip tendency of the fault, showing both high and low slip tendency values in all line spacing scenarios. Note that unconstrained triangulation is used for fault surface generation.

Figure 9. Histograms showing frequency of dilation tendency (top) and slip tendency (bottom) for scenarios picked on every line (green), every 4th lines (blue) and every 32nd line (red), for both manual interpretation (darker colour) and machine learned techniques (lighter colour). In all line spacing instances, machine learned techniques creates a fault that has higher predicted dilation tendency and a wider range of slip tendency. The fault is only predicted to fail by shear failure when machine learned techniques are used at a line spacing of 1. However, the fault is predicted to fail by tensile failure in all line spacing instances when machine learned techniques are used, as the values reach or exceed 1.

Figure 10. Comparison of manual verses machine learned techniques for a minor fault in the footwall of the Vette Fault (fault FW_01, see Mulrooney et al., 2020 for location details). Upper left: Row 1403 showing an example of over-segmentation of the Vette Fault Zone but no vertical segmentation of the minor faults within the footwall. Seismic section shows a colourmap with amplitudes from -127 to 127. Upper right: Histogram showing dilation tendency for the FW_01 fault for both manual interpretation (blue) and machine learned (red) picking technique. Similar values are observed between both techniques. Lower left: Fault plane diagram showing dilation tendency for the manual interpretation technique picked on every 4th line. Lower right: Fault plane diagram showing dilation tendency for the machine learned technique picked on every 4th line. Note the similarities between these two results. Note that unconstrained triangulation is used for fault surface generation.

Figure 11. Examples of fault predictions using the supervised DNN models. The predicted faults are ensemble of different models and the colour represents the confidence of predicted faults. The yellow colour means that all models predicted faults. (a) Inline without predicted results (b) Inline with predicted faults. The area inside the ellipse shows poor quality seismic zone. (c) Time slice without predicted faults (d) Time slice with predicted faults. Note the areas of the VFZ with less confidence in the prediction. Seismic sections and timeslices show a colourmap with amplitudes from -127 to 127.

Figure 12. Timeslice 1652 showing results from a pre-trained (on real data) DNN model and results from a pretrained synthetic 3D CNN model. Results highlight complexities within the fault zone, showing multiple fault strands that are unlikely to be picked through manual interpretation alone. Two examples of fault complexities are shown: Green (left) shows multiple slip surfaces around inline 1140; Red (right) shows multiple slip surfaces around inline 1250. Location of inlines 1140 and 1250 shown as vertical white lines on the timeslice images. (a) Overview timeslice showing location of the two studied areas, highlighted by the green and red boxes, showing results from the pre-trained DNN on real data. (b) Detailed timeslice from green area highlighted on Figure (a), showing results from a pre-trained model on real data. (c) Detailed timeslice from red area highlighted on Figure (a), showing results from a pre-trained model on real data. (d) Detailed timeslice from green area highlighted on Figure (a), showing results from a 3D pretrained synthetic CNN model. (e) Detailed timeslice from red area highlighted on Figure (a), showing results from a pretrained 3D synthetic CNN model. Note the decrease in output data quality when using 3D synthetic CNN rather than a pre-trained model on real data. (f) Inline 1140 showing fault zone complexity around the highlighted green area with no interpretation. (g) Inline 1250 showing fault zone complexity around the highlighted red area with no interpretation. Seismic sections and timeslices show a colourmap with amplitudes from -127 to 127.

TABLES

Table 1. Parameters used for each supervised iteration, along with test results to show accuracy.

Table 2. *In situ* stress data used for geomechanical analysis.

Model ID	Orientation	Patch Size	Noise	Epochs	Train Score %	Test Score %	Confusion Matrix Epoch 79 Fault-Fault Prediction	Confusion Matrix Validation Epoch 79 Fault-Fault Prediction	Notes
3498	Inline	256,256	0.15	250	99.6	99.6	82.8	41.7	All labels picked
3499	Inline	320,320	0.15	250	99.7	99.5	83	47.2	All labels picked
3500	Xline	256,256	0.15	250	99.5	99.3	86.1	15.2	All labels picked
3501	Xline	320,320	0.15	250	99.6	99.2	80.2	15.6	All labels picked
3518	Xline	320,320	0.15	250	99.6	99.1	90.4	16.4	Re-pick labels due to erroneous modelling predictions
3715	Inlines	320,320	0.1	250	99.7	99.5	87.6	44.7	Labels remained the same
3716	Inlines	256,256	0.1	250	99.7	99.5	86.8	46.3	Labels remained the same
3720	Xlines	320,320	0.1	250	99.7	99.3	91.4	17.5	Labels remained the same
4027	Inlines	256,256	0.1	250	99.7	99.5	88.5	41.8	Labelled tweaked and improved
4028	Inlines	320,320	0.1	250	99.7	99.6	90.1	42.5	Labelling tweaked and improved

Table 1.

	Gradient (MPa/m)	Stress (MPa)	Depth (m)	Direction (degrees)
SHmin	0.0146	23.07	1699.5	090
SHmax	0.0146	23.07	1699.5	180
Sv	0.0215	32.37	1699.5	
PP	0.01	16.94	1699.5	

Table 2

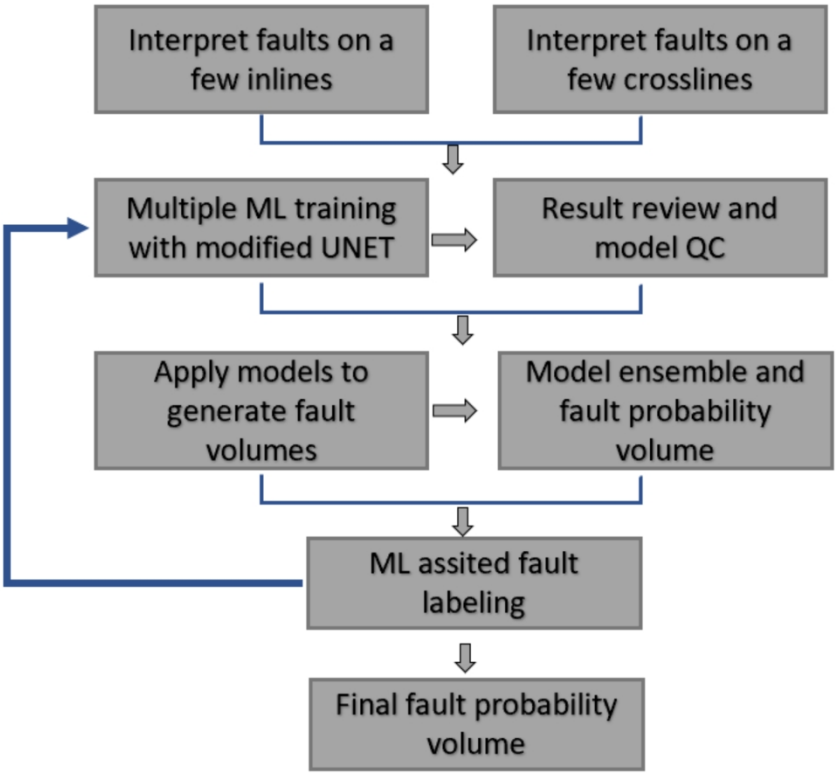

Figure 2. Fault imaging workflow using supervised CNN, showing the main iterative processes used.

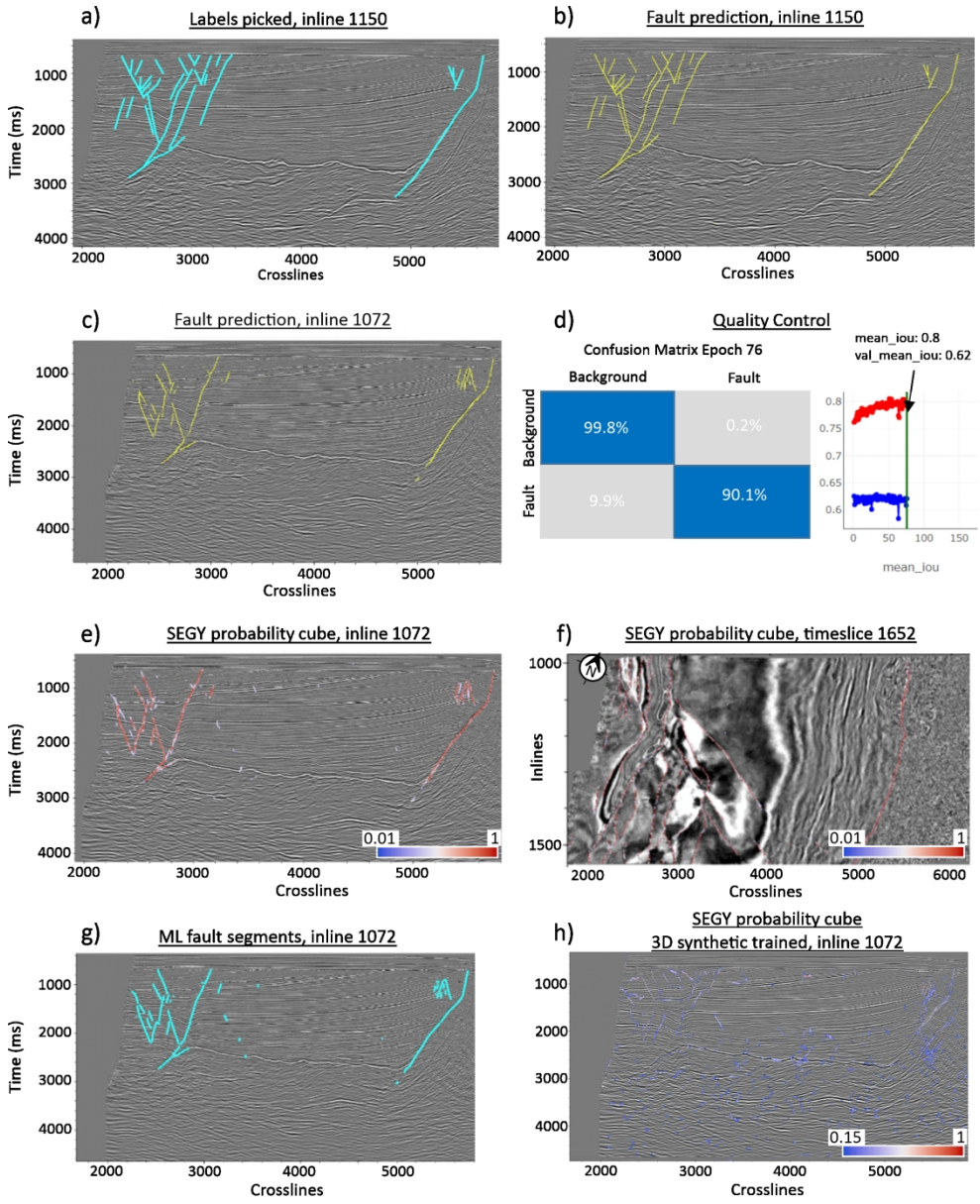

Figure 3. Seismic sections showing key steps of the fault imaging workflow, along with the results shown as ML fault segments. Seismic sections and timeslices show a colourmap with amplitudes from -127 to 127. (a) Example of picked labels on inline 1150, with corresponding predicted faults on the same inline (b) allowing for a first pass visual QC. (c) Predicted faults on an inline (1072) with no labels, used as visual QC. Confusion matrix and IoU scores of the model shown (d), used for additional QC. (e) and (f) SEGY probability cube combining two models with differing patch sizes shown on inline 1072 and timeslice 1652, respectively. (g) Extracted fault segments from the SEGY cube. (h) SEGY probability cube showing predicted faults using 3D synthetic trained CNN rather than a supervised 2D CNN. Note the different probability cutoff values for the scale bar between figures (e) and (f), and (h); synthetic trained CNN does not predict fault locations as well as supervised 2D CNN for this example, and picks up a lot of 'noise' from the basement as faults.

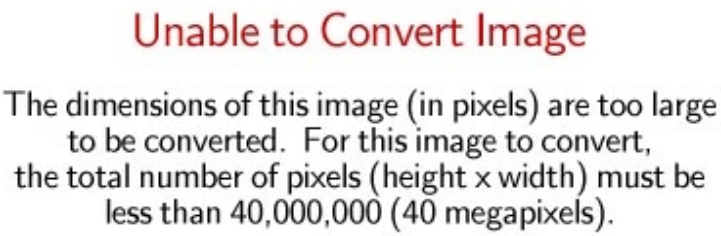

Figure 4. Triangulated fault surfaces of the VFZ, using unconstrained triangulation, showing the triangle elements when manual interpretation vs. machine learned techniques are used, picked at differing line spacing (every line, every 4th line and every 32nd line). No fault segments shown.

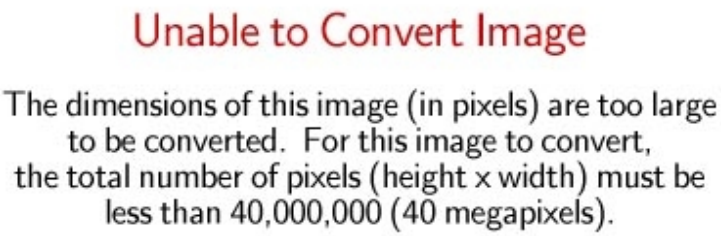

Figure 5. Fault plane diagrams showing fault dip attribute displayed on the fault surfaces for different picking strategies: manual interpretation verses machine learned techniques, at different line spacing: 1, 4, and 32 lines. Fault dip is observed to vary with line spacing used for fault picking, as well as picking technique used. A highly irregular fault surface is observed with highly variable dip attributes when every line is used for picking, when compared to the overly smooth surface when every 32nd line is used for picking. The irregularity of the fault surface and hence also dip attribute is increased when machine learned techniques are employed over manual interpretation, in all spacing instances. Fault surface irregularity often manifests itself as bulls-eye patches of varying fault dip. Note that unconstrained triangulation is used for fault surface generation.

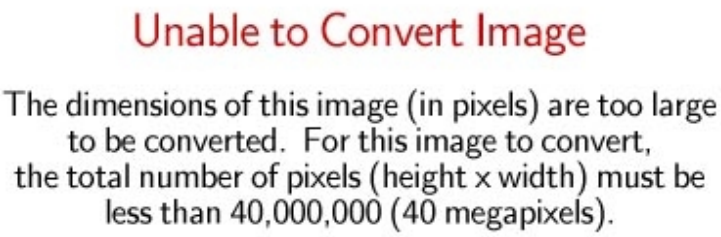

Figure 6. Fault plane diagrams showing fault strike attribute displayed on the fault surfaces for different picking strategies: manual interpretation verses machine learned techniques, at different line spacing: 1, 4, and 32 lines. Fault strike is observed to vary with line spacing used for fault picking, as well as picking technique used. The highly irregular fault surface observed with increased variation in strike attribute occurs when every line is used for picking, compared to the overly smooth surface when every 32nd line is used for picking. The irregularity of the fault surface, and hence also strike attribute, is increased when machine learned techniques are employed over manual interpretation, such that the varying strike attribute observed when every 4th line spacing is used for machine learned techniques is similar to that observed when every line spacing is used for manual interpretation. Note that unconstrained triangulation is used for fault surface generation.

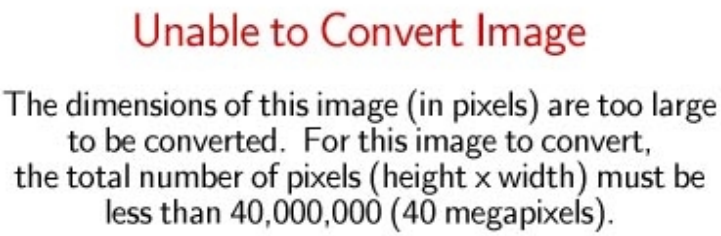

Figure 7. Fault plane diagrams showing dilation tendency attribute displayed on the fault surfaces for different picking strategies: manual interpretation verses machine learned techniques, at different line spacing: 1, 4, and 32 lines. Dilation tendency is observed to vary with line spacing used for fault picking, as well as picking technique used. The irregular fault surfaces created with manual interpretation using narrower line spacing, as well as using machine learned techniques with any line spacing, produce fault surfaces that are interpreted to be unstable, i.e. the dilation tendency in places exceeds 1 (white areas). Conversely, the overly smoothed fault surface when every 32nd line spacing is manually picked creates an interpreted relatively stable fault, with no areas where the dilation tendency exceeds 1. Note that unconstrained triangulation is used for fault surface generation.

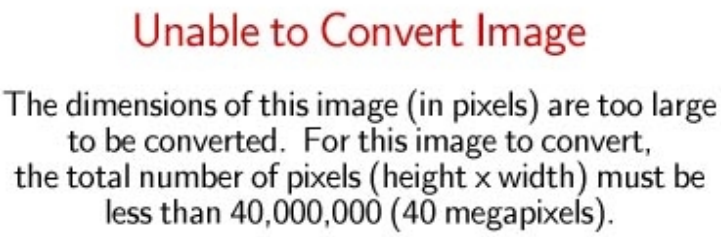

Figure 8. Fault plane diagrams showing slip tendency attribute displayed on the fault surfaces for different picking strategies: manual interpretation verses machine learned techniques, at different line spacing: 1, 4, and 32 lines. Slip tendency is observed to vary with line spacing used for fault picking, as well as picking technique used. Although patches occur when the slip tendency is low, and hence more stable, the irregular fault surfaces created with using narrower line spacing (i.e. picked on every line) produce areas where the slip tendency is close to, or exceeds 0.6, and hence an unstable fault is interpreted. Increasing the line spacing creates a fault that is interpreted to be less likely to slip by shear failure. The increased irregularity when machine learned techniques are employed creates a wider spread to the predicted slip tendency of the fault, showing both high and low slip tendency values in all line spacing scenarios. Note that unconstrained triangulation is used for fault surface generation.

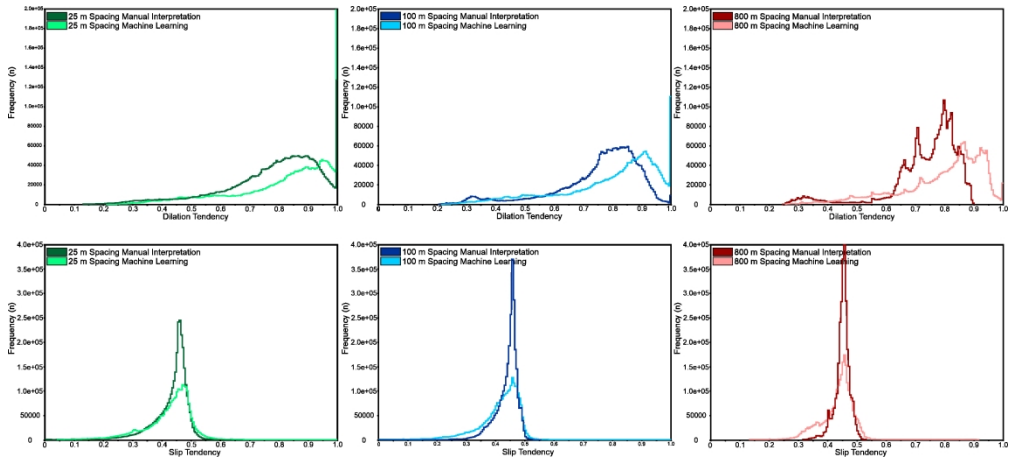

Figure 9. Histograms showing frequency of dilation tendency (top) and slip tendency (bottom) for scenarios picked on every line (green), every 4th lines (blue) and every 32nd line (red), for both manual interpretation (darker colour) and machine learned techniques (lighter colour). In all line spacing instances, machine learned techniques creates a fault that has higher predicted dilation tendency and a wider range of slip tendency. The fault is only predicted to fail by shear failure when machine learned techniques are used at a line spacing of 1. However, the fault is predicted to fail by tensile failure in all line spacing instances when machine learned techniques are used, as the values reach or exceed 1.

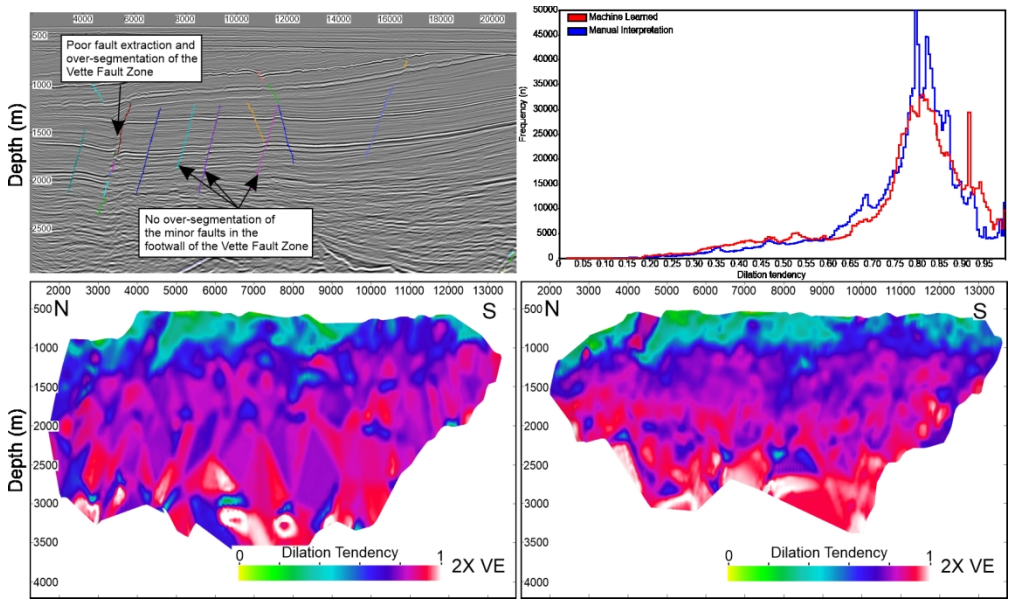

Figure 10. Comparison of manual verses machine learned techniques for a minor fault in the footwall of the Vette Fault (fault FW_01, see Mulrooney et al., 2020 for location details). Upper left: Row 1403 showing an example of over-segmentation of the Vette Fault Zone but no vertical segmentation of the minor faults within the footwall. Seismic section shows a colourmap with amplitudes from -127 to 127. Upper right: Histogram showing dilation tendency for the FW_01 fault for both manual interpretation (blue) and machine learned (red) picking technique. Similar values are observed between both techniques. Lower left: Fault plane diagram showing dilation tendency for the manual interpretation technique picked on every 4th line. Lower right: Fault plane diagram showing dilation tendency for the machine learned technique picked on every 4th line. Note the similarities between these two results. Note that unconstrained triangulation is used for fault surface generation.

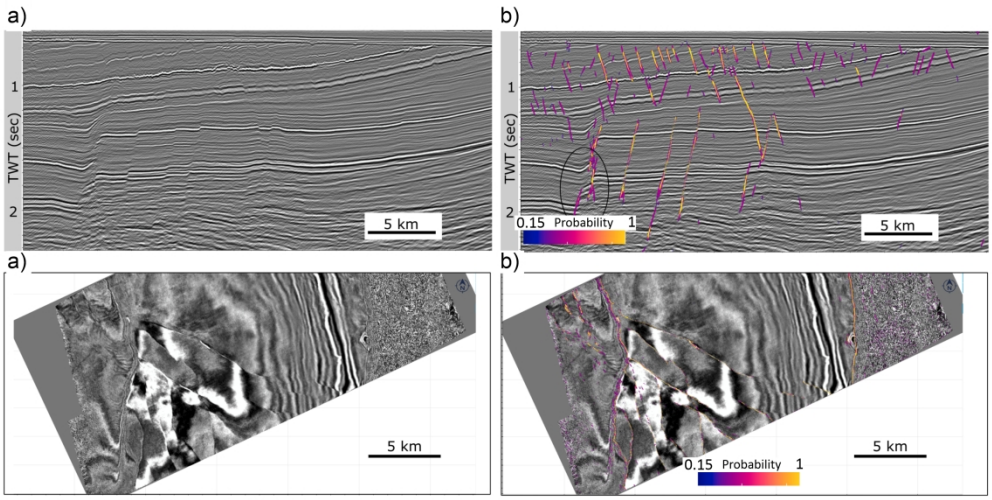

Figure 11. Examples of fault predictions using the supervised DNN models. The predicted faults are ensemble of different models and the colour represents the confidence of predicted faults. The yellow colour means that all models predicted faults. (a) Inline without predicted results (b) Inline with predicted faults. The area inside the ellipse shows poor quality seismic zone. (c) Time slice without predicted faults (d) Time slice with predicted faults. Note the areas of the VFZ with less confidence in the prediction. Seismic sections and timeslices show a colourmap with amplitudes from -127 to 127.

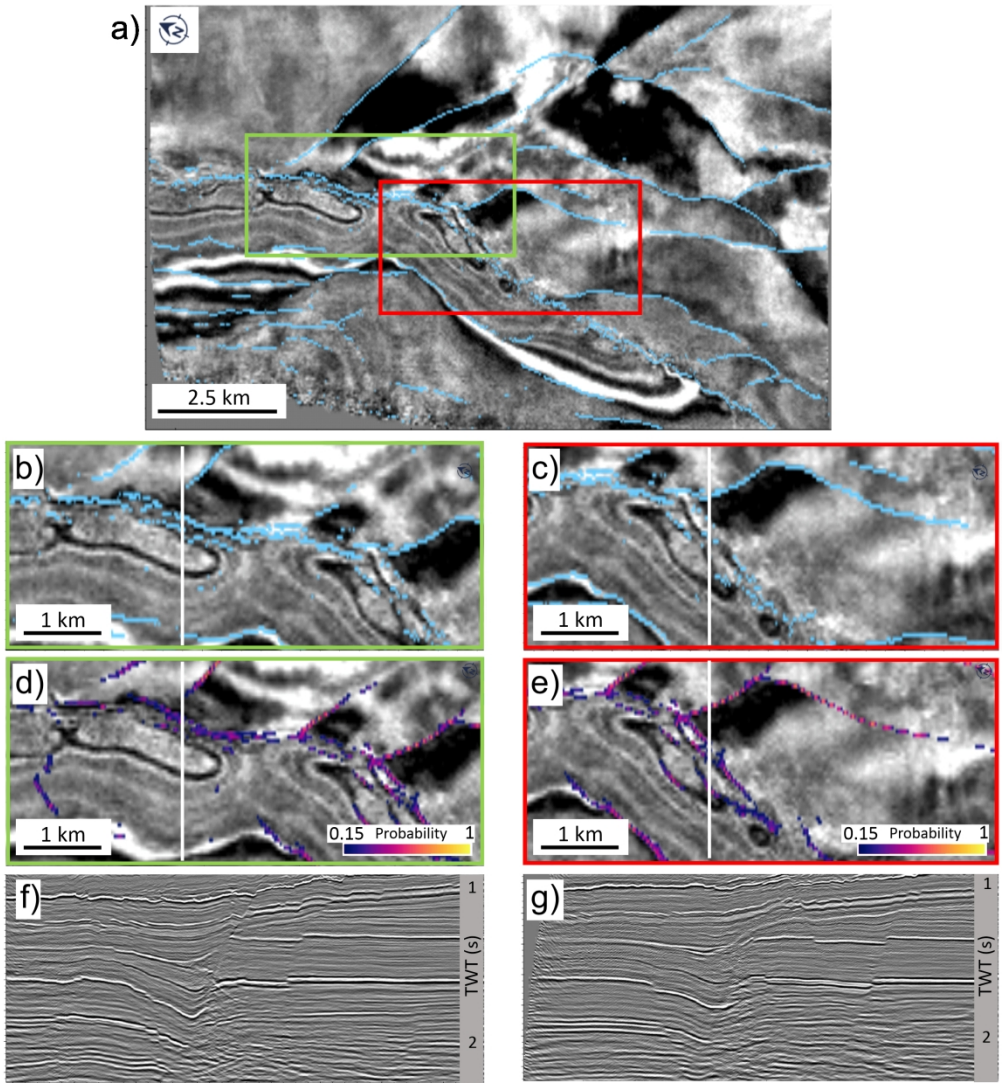

Figure 12. Timeslice 1652 showing results from a pre-trained (on real data) DNN model and results from a pretrained synthetic 3D CNN model. Results highlight complexities within the fault zone, showing multiple fault strands that are unlikely to be picked through manual interpretation alone. Two examples of fault complexities are shown: Green (left) shows multiple slip surfaces around inline 1140; Red (right) shows multiple slip surfaces around inline 1250. Location of inlines 1140 and 1250 shown as vertical white lines on the timeslice images. (a) Overview timeslice showing location of the two studied areas, highlighted by the green and red boxes, showing results from the pre-trained DNN on real data. (b) Detailed timeslice from green area highlighted on Figure (a), showing results from a pre-trained model on real data. (c) Detailed timeslice from red area highlighted on Figure (a), showing results from a pre-trained model on real data. (d) Detailed timeslice from green area highlighted on Figure (a), showing results from a 3D pretrained synthetic CNN model. (e) Detailed timeslice from red area highlighted on Figure (a), showing results from a pretrained 3D synthetic CNN model. Note the decrease in output data quality when using 3D synthetic CNN rather than a pre-trained model on real data. (f) Inline 1140 showing fault zone complexity around the highlighted green area with no interpretation. (g) Inline 1250 showing fault zone complexity around the highlighted red area with no interpretation. Seismic sections and timeslices show a colourmap with amplitudes from -127 to 127.