Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Geoscience and Petroleum

Charles Rutherford Ildstad

Exploration of Sediment Hosted Massive Sulfide (SHMS) deposits in the Barents Sea

Using machine learning to automate the exploration process

Master's thesis in MTTEKGEO
Supervisor: Steinar Løve Ellefmo
May 2018

NTNU
Norwegian University of
Science and Technology

Charles Rutherford Ildstad

Exploration of Sediment Hosted Massive Sulfide (SHMS) deposits in the Barents Sea

Using machine learning to automate the exploration process

Master's thesis in MTTEKGEO
Supervisor: Steinar Løve Ellefmo
May 2018

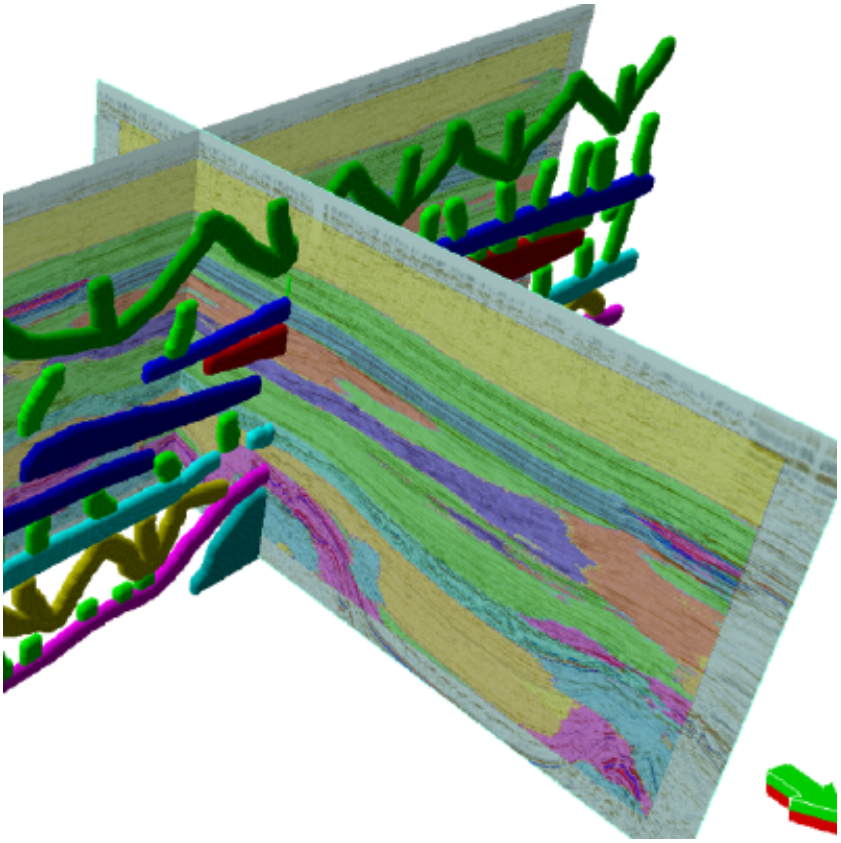
Norwegian University of Science and Technology
Faculty of Engineering
Department of Geoscience and Petroleum

NTNU
Norwegian University of
Science and Technology

TGB4905 Resource Geology, Master's Thesis
Spring 2018

Exploration of Sediment Hosted Massive Sulfide (SHMS) deposits in the Barents Sea

Using machine learning to automate the exploration process



Charles Rutherford Ildstad (748729)

Supervisor: Steinar Løve Ellefmo (NTNU)
Co-Supervisor: Michael Nickel (Schlumberger)

January 7, 2019

Abstract

The geology of the Barents Sea has previously been extensively studied, and shown to hosts several hydrocarbon reservoirs, making it an economically important part of the Norwegian continental shelf. In this thesis, the region is explored in terms of a new economical resource, namely Sediment Hosted Massive Sulfide(SHMS) deposits, with a concrete exploration strategy presented. In addition, novel ways to automate the steps in the exploration process are investigated and a new machine learning algorithm for automation is suggested.

The results show that the Barents Sea contains the necessary genetic components for a whole range of SHMS deposits, with the Sedimentary Exhallative - Mississippi Valley Types favoured, due to the prevalence of carbonates, and scarcity of volcanics. The most likely type, based on exploratory findings, are Late-Permian to Early-Triassic Irish Valley Type deposits in the vicinity of the Leirdjupet fault complex in the Bjørnøya basin.

An ideal exploration strategy for such deposits includes magnetic, gravimetric, electromagnetic, and seismic data, and targets the mineral deposit both directly and indirectly. Due to data availability constraints, the exploration strategy utilized was exclusively based on seismic data and consisted of direct detection by seismic amplitude analysis, as well as indirect detection by comprehensive data visualization and interpretation.

Finally, a machine learning method called a 3D-Convolutional Neural Network was used for automating both the mapping of host rock and fault detection from seismic data. This algorithm showed promising results and additional ways to improve on it, as well as other potential applications, are suggested.

Coverphoto: Auto-classified cube with 9 facies classes. (by Peter Bormann and Charles Rutherford Ildstad, 2017[1]).

Sammendrag

Barentshavets geologi hjemmer flere hydrokarbonreservoarer, som gjør den til en økonomisk viktig del av norsk sokkel, og er tidligere blitt grundig undersøkt på grunn av dette. I denne oppgaven blir regionen utforsket med bakgrunn i en ny økonomisk ressurs, nemlig sedimentbaserte, massive sulfidforekomster(SHMS), og en konkret letestrategi er presentert. I tillegg undersøkes nye måter å automatisere trinnene i leteprosessen på, og en ny maskinlæringsalgoritme for automatisering blir foreslått.

Resultatene viser at Barentshavet inneholder de nødvendige genetiske komponentene for en rekke SHMS-forekomster, med Sedimentære Exhalative(SEDEX) - Mississippi Valley Typer(MVT) foretrukket på grunn av utbredelsen av karbonater og mangel på vulkanisme. Den mest sannsynlige typen, basert på funn, er sen-permisk til tidlig-triassisk Irish Valley Type(IVT)-forekomster i nærheten av Leirdjupet-forkastningskomplekset i Bjørnøya-bassenget.

En ideell letestrategi for slike forekomster inkluderer magnetiske, gravimetriske, elektromagnetiske og seismiske data, og detekterer mineralforekomsten både direkte og indirekte. På grunn av datatilgjengelighetsbegrensninger ble utforskningsstrategien utelukkende basert på seismiske data og besto av direkte deteksjon ved seismisk amplitudeanalyse, så vel som indirekte deteksjon ved omfattende datavisualisering og tolkning.

Til slutt ble en maskininnlæringsmetode kalt et 3D-konvolusjonelt neuralt nettverk(CNN) brukt for å automatisere både kartlegging av vertsbergart og forkastningsdeteksjon i seismiske data. Resultatene fra denne algoritmen viste gode resultater, og ytterligere forbedringer, samt andre potensielle applikasjoner, ble foreslått.

Pretext

We are drowning in information and starving for knowledge.

–Rutherford D. Rogers

Project development

This Master's thesis is written at The Department of Geoscience and Petroleum (IGP) at The Norwegian University of Science and Technology (NTNU) during the Spring of 2018. The work is the final result of a cooperation between NTNU and Schlumberger, and is the continuation of a project paper written by the author during the Fall of 2017[2]. All geophysical data has been provided by Schlumberger and is either the property of Schlumberger itself, one of its partners (primarily OMV (former Österreichische Mineralölverwaltung) or TGS (TGS-NOPEC Geophysical Company ASA)), or The Norwegian Petroleum Directorate (NPD). In addition, the seismic data used has been previously interpreted by a research team at Schlumberger.

As epitomized in the quote by the great American librarian Rutherford D. Rogers, one of the biggest challenges in georesource exploration is extracting the underlying meaning in extremely vast amounts of data. This thesis addresses exactly this, with the goal of investigating novel ways to automate the interpretation of data. The main method of automation utilized in this project is a Machine Learning (ML) technique called Convolutional Neural Network (CNN), as it has shown to be useful in other fields and has shown a recent emergence in georesource exploration[3].

The outline for the problem statement was first developed through the author's machine learning research in the Earth Resources Lab at The Massachusetts Institute of Technology (MIT), given a proof of concept through a summer internship at ConocoPhillips Norway, and further refined from the initial project paper with the help of Steinar Ellefmo, Associate Professor at IGP and Michael Nickel, Geophysics Discipline Manager at Schlumberger.

Problem statement

The candidate is expected to make a regional map describing the favorability for the existence of a Sediment Hosted Massive Sulfide (SHMS) deposit on a section of the Norwegian continental shelf in the Barents Sea. The method should be data driven and focus should be put on possible ways to automate the process.

There are many types of SHMS deposits, and they are normally tightly linked to areas of tectonic rifting, heightened geotherms and high fluid flow rates. There are many locations on the continental shelf in the Barents Sea capable of accommodating such deposits, and the candidate is expected to narrow the study down to a specific type of SHMS deposit, and choose a well defined Region of Interest (ROI) within the shelf more likely to contain this deposit type.

Once the ROI is better understood and the characteristics of a deposit are defined the candidate should apply this knowledge to develop an exploration strategy and map the favorability of these deposits within the ROI, and finally create the section map desired.

The ROI should be defined such that the cooperating business, Schlumberger, has available geophysical data for the majority of the area.

Acknowledgements

I am very aware that I would not be in the position I am in today without the help of others and I would like to extend my deep appreciation to a number of people.

Firstly, I would like to thank NTNU and Schlumberger for the opportunity to pursue a Master's degree, and my supervisors, Steinar Løve Ellefmo and Michael Nickel, in particular for their continued guidance throughout the process.

Secondly, I am extremely grateful for the opportunity to take a year abroad at MIT. The sponsorship from the Fulbright Foundation, the opportunity to do research and receive funding from the Earth Resources Lab (MIT), and the advice from professor Laurent Demanet has been invaluable.

Thirdly, I would like to thank Statoil, Lundin Norway, ConocoPhillips Norway, OMV and Schlumberger for giving me internships, making available real data sets, and partnering with me in my education over the past 3 years.

Finally, I am grateful to my parents for their constant support, for their many sacrifices, and for raising me to believe in myself.

This project has been a great learning experience of which I hope to benefit throughout my career.

Contents

1 Introduction

1.1 Rationale

The world population is increasing at an ever growing rate, and with the increasing standards of living the mineral consumption of the world is skyrocketing like never before[4]. As more and more of the on-shore mineral deposits are being mapped and developed there has been a recent surge in interest around marine minerals to fill the need for resources, as exemplified by the emergence of projects like NTNU Ocean[5] and the Blue Mining initiative[6]. These projects focus mainly on Mid-Ocean Ridge (MOR) settings, but there are also many good arguments for moving the areas of focus onto deposits hosted in the continental shelves as well. These include greater abundance of geophysical data on the shelves, tighter connections to familiar on-shore deposits, and likely shallower deposits[7]. It is in particular the abundance of geophysical data that makes this setting interesting, especially when considering ML applications that require large amounts of data, and motivates the focus on the shelf in this thesis.

ML is the ability of a machine to acquire knowledge by extracting patterns from raw data. As computing has become increasingly fast and more readily available, there have been great advancements in the field of ML, and one of its biggest advantages is the ability to handle very big data sets and produce abstract results[8]. This makes the utilization of ML in georesource exploration particularly interesting, where one of the challenges is exactly this; interpreting large 3D data sets. More specifically, the applicability of ML for *3D object recognition* has been shown extensively in medical imaging with processing of 3D data sets like Magnetic Resonance Imaging (MRI) and Computed Axial Tomography (CAT) scans[9]. Recent studies have also had success using machine learning to solve simple object recognition problems in 3D geophysical data as well[10], giving a proof of concept and motivating its use in this thesis.

Finally, the Norwegian continental shelf is an ideal candidate to perform this study as the region has historically had to overcome large technological challenges in its off-shore petroleum operations, and has a strong drive for automation due to its high labour costs and recent drop in oil price. The Barents Sea in particular is suited for this study as petroleum Exploration and Production (E&P) has resulted in large amounts of both seismic and Electromagnetic (EM) surveys, especially in the southern part. In addition, the Barents Sea has a history of tectonic and rifting activity that can have caused [11].

1.2 Research questions

The problem statement, given in the pretext, is comprehensive and has several stages that need to be divided into research questions.

This is summarized below, and will guide the rest of the thesis.

- Choose a specific SHMS-deposit type anticipated to be prevalent on the Barents Sea.
- Define an appropriate ROI with the necessary mechanisms for these SHMS-deposits.
- Design an exploration strategy for this SHMS-deposit type in the ROI.
- Find a way to automate the steps in the exploration strategy by using machine learning.
- Apply the exploration strategy, with the automated steps where possible, in the ROI to create the final favourability map.

1.3 Structure of the thesis

In order to answer the research questions, tackle the problem statement, and create a favourability map the thesis will be structured in the following way;

1. **Introduction** - Motivation and scope for the project as well as objectives going forward.
2. **Background** - Definition of fundamental terminology, introduction to and statistics of SHMS-deposits, description of the location and setting of the study area, and essentials of ML.
3. **Geological setting** - Description of the specific ROI, its main structural elements, and relevant geological formations.
4. **Theory** - Literature review on genetic modeling of SHMS-deposits, exploration models for SHMS-deposits and other resources in the ROI, in addition to automation by machine learning in geophysics and medical imaging as a proxy.
5. **Data** - Overview of the available data for the ROI along with visualizations if available.
6. **Method** - Description and explanation of selected exploration strategy and machine learning algorithm used for automation.
7. **Results** - Presentation of received results from the exploration strategy and automation algorithm.
8. **Discussion** - Analysis of the results, possible sources of error, and reflections around the chosen methodology.
9. **Conclusion** - Statement of the conclusions drawn from the results.
10. **Further work** - Suggestions for further work to improve the exploration strategy presented, improve the machine algorithm utilized, and novel applications of the methodology presented.

2 Background

In the following chapter the necessary background knowledge to embark on the research questions will be provided. This includes fundamental terminology, introduction to and basics statistics of SHMS-deposits, description and setting of the study area, and essentials of ML. The work in sections 2.2.1, 2.4 and 2.5 is rephrased from work done in the project paper from Fall 2017[2].

2.1 Exploration Strategy

An exploration strategy is simply put an answer to 3 basic questions[12]:

- **What?** - What commodity and deposit type is being explored?
- **Where?** - What is the specific area that is being explored?
- **How?** - What would a genetic model for the deposit be in this area, and how will this model manifest itself in geophysical data (also known as an exploration model)?

None of these questions are trivial and all need to be answered to make a complete exploration strategy. SHMS-deposits are a large family of deposits and a single type, or suite of deposits, need to be defined, preferably along with commodity, that is most favoured in the Barents Sea.

Where to look for these deposits obviously depends on several factors; what deposit is being investigated, where is there available data, what area contains the necessary mechanisms for accomodating a deposit, and more. A combination of these should result in a clearly defined ROI.

Finally, to determine the best way to look for the deposit there first needs to be a comprehensive genetic model for a given type of deposit in the ROI. This is be done by investigating lithostratigraphy, regression-transgression (T-R) cycles, geological structures and geodynamic events in the area, along with any other data deemed relevant. A hypothesis is then formulated for how a combination of these factors can support the forming, accumulation and preservation of a resource. If the hypothesis withholds to scrutiny from the data available, it is a viable genetic model for a resource deposit in the area. Once this is defined a new model, called the exploration model, is created for how the deposit and geologic framework outlined in the genetic model will present itself in geophysical data.

2.2 Sediment Hosted Massive Sulfide (SHMS) deposits

2.2.1 The Sediment Hosted Massive Sulfide (SHMS) continuum

The term SHMS-deposit is often ambiguously used and should be properly defined before moving forward in this thesis. An SHMS-deposit is a sulfide mineral deposit caused by the circulation of a hydrothermal fluid and deposited in a sedimentary host rock. This includes both exhalative deposits, like Sedimentary Exhalative (SEDEX) deposits where a highly concentrated metalliferous brine is expelled into a large sulfide loaded body of water[13], and deposits where the mineral deposition happens exclusively in the subsurface, like the Mechernich-Laisvall type deposits in the Pan-African Orogenic Belt (PAOB), where diagenetic replacement or impregnation of clastic metasediments causes ore formation[14].

As discussed by Reynolds and Muhling, SHMS deposits can now be further categorized into a continuum of two suites[15]. One is the **SEDEX-Volcanic Hosted Massive Sulfide (VHMS) suite** with increasing degrees of volcanic components, that occur in a setting where fluid flow is driven by magmatic intrusion. The magmatic extreme of the suite is a pure VHMS deposit (also known as a Volcanogenic Massive Sulfide (VMS)-deposit) where the mineralization is hosted by volcanic and/or volcano-sedimentary rock. A modern day analogue for this suite is where the East Pacific Rise extends onto the North-American plate to form the San Andreas fault system. It here goes from ocean floor black smokers (VHMS-deposits) on the Rise, to Besshi-type ores in the back arc setting of the Gulf of California (mix of SEDEX- and VHMS-deposit), to the Salton Sea deposits in the Colorado Desert (SEDEX-deposit).

The other is the **SEDEX-Mississippi Valley Type (MVT) suite** with increasing degrees of carbonatic components, that occur in a basinal setting where fluid flow usually is driven by larger regional tectonic forces. The carbonatic extreme is called a MVT-deposit where the mineralization is hosted by a carbonate wall rock and precipitation occurs epigenetically. There are many deposits that today are interpreted to be a part of this continuum, like the lead-zinc deposits of the São Francisco Craton in Brazil where deposits occur in dolomites[16], the Irish Valley Type (IVT) (also known as Irish Type) deposits where deposition occurs both like the MVT deposits and by exhalation, and Lead-zinc deposits in the PAOB described above[14].

The continuum of SHMS-deposits, along with a few famous type deposits, is shown in figure 1, and is discussed further in the presentation by Neal Reynolds and Peter Muhling[15]. Genetic models for the continuum is described in more detail in section 4.1.

2.2.2 Statistics

SHMS depsits account for some of the most economically important sources of zinc, lead and silver deposits in the world. The specific deposit characteristics varies depending on the source rock, the metals available to the circulating fluid, and the setting in which it occurs, but in general, most deposits show major Iron(Pyrite/Pyrhotite), Lead(Galena) and Zinc(Sphalerite), minor Barium(Barite) and Copper(Chalcopyrite), and trace amounts of Platinum Group Elements (PGE), Gold and Silver[13].

The characteristics, like common mineralogy and relative abundance, of 14 of the largest, most studied type deposits for the SHMS continuum are given in the table below, and is discussed in greater detail in the paper by MacIntyre, 1991, with additional statistics on other SHMS deposits[13]).

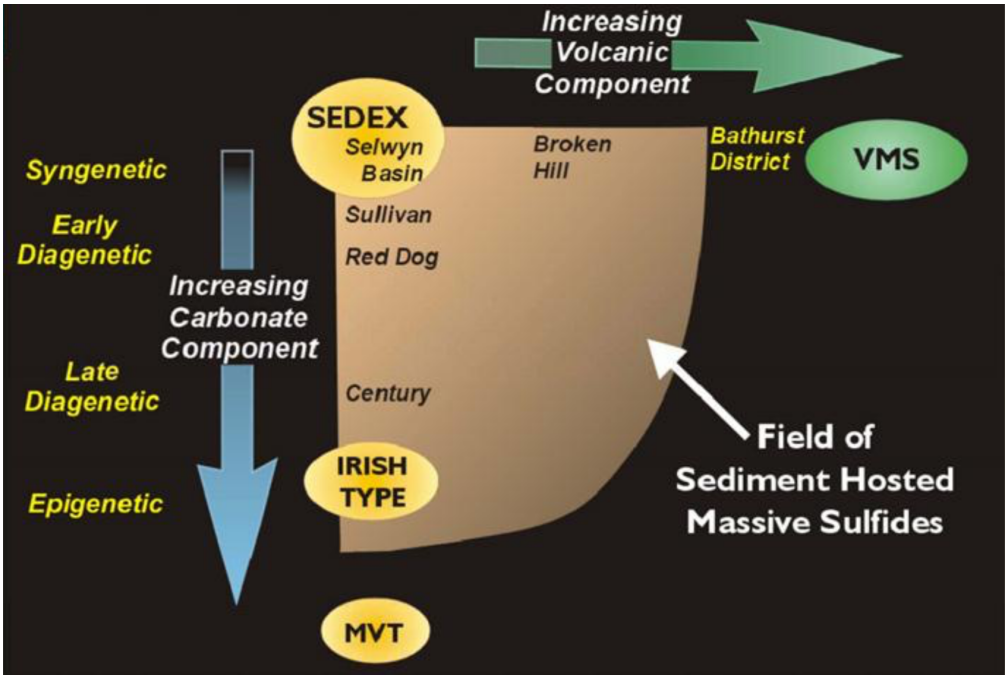


Figure 1: Illustration of the continuum of SHMS deposits from pure SEDEX, to VHMS deposits (also known as VMS) with increasing volcanic component, to MVT and more basinal and Carbonaceous components. (from Allen 2003[15])

Deposit	Age	Paleotectonic Setting	Host Rocks	Type	Tonnes (millions)	Zn %	Pb %	Ag g/t	Cu %	Barite (in ore)
Mount Isa, Australia	M. Proterozoic	Leichhardt R. Trough	dolomitic shales, siltstones	1B	88.6	6.3	6.9	149		low
McArthur River, Australia	M. Proterozoic	Batten Trough	dolomite, evaporitie, siltstone	1B	190.0	9.5	4.0	45		low
Sullivan, Canada	M. Proterozoic	Purcell Basin	basinal turbidites	1A	155.0	5.7	6.6	7		low
Faro, Canada	Cambrian-Ord.	Selwyn Basin	graphitic phyllites	1A	57.6	4.7	3.4	36		mod.
Howard's Pass, Canada	Early Silurian	Selwyn Basin	dolomitic shales, chert	1B	125.0	5.4	2.1			low
Jason, Canada	Late Devonian	Selwyn Basin	shales and turbidites	1A	15.5	6.6	7.1	79		mod.
Tom, Canada	Late Devonian	Selwyn Basin	silty and cherty argillite	1A	15.7	7.0	4.6	49		mod.
Cirque, Canada	Late Devonian	Kechika Trough	cherty argillite, shale	1A	32.2	7.9	2.1	48		high
Rammelsberg, Germany	Dev.-Miss.	Variscan Trough	shales, siltstones	1A	22.0	19.0	10.0	120	2.00	low
Meggen, Germany	Dev.-Miss	Variscan Trough	shales, siltstones	1A	50.0	10.0	1.3	3	0.02	mod.
Red Dog, Alaska	Mississippian	Kuna Formation	black sil. shale and cherts	1A	85.0	17.1	5.0	82		mod.
Tynagh, Ireland	Mississippian	Central Midlands Basin	shelf carbonates	1C	8.6	5.5	7.0	81	0.58	low
Silvermines, Ireland	Mississippian	Central Midlands Basin	shelf carbonates	1C	12.9	6.8	2.6	25		low
Navan, Ireland	Mississippian	Central Midlands Basin	shelf carbonates	1C	69.9	10.1	2.6	11		low

Deposit Types (this paper): 1A — clastic hosted, starved basin; 1B — mixed carbonate-clastic, shallow marine; 1C — shelf carbonate hosted

Figure 2: Table 3-1 from MacIntyre 1991[13]. Note that MacIntyre uses SEDEX as synonymous with this thesis's definition of the SEDEX-MVT suite. In the current nomenclature 1A would be a pure SEDEX deposit, 1B a mix, and 1C the Irish Type deposit approaching the MVT in figure 1

From a mineralogical aspect one can make a few preliminary observations. A higher Pb content points to a likely crustal source for the metals in the deposit, which has been confirmed in certain deposits by isotope analysis[16]. The prevalence of minor Cu content gives information about the local temperatures and gradients present at the time of deposition, as higher temperatures are needed for effective Copper-complexiation[17]. When compared to VHMS deposits, however, the temperatures present during SHMS deposition appear to be distinctly lower[17]. Finally, the existence of Ba seems to be confined primarily to SEDEX-type deposits and points to the existence of ponds of localized oxidizing conditions accommodating oxidized sulphur(SO_4^{2-}) in the body of water subject to exhalative venting[13]. This is in stark contrast to the other SHMS deposits that require reduced sulphur(S^{2-}) present for metal deposition[17].

2.3 Wilson Cycle

A useful concept in tectonics and resource geology when considering what type of deposit may be present in a region, or whether or not a given deposit seems to be likely, is called the Wilson Cycle[17]. The Wilson Cycle describes the cyclical opening and closing of ocean basins due to plate tectonics and is named after the man who first described it, John Tuzo Wilson[18].

This cycle is often referenced when discussing mineral forming processes as it has an inherent connection to relative sea level, geotherms, and depositional environments, all of which define preferential ore development, and includes the stages; fragmentation, dispersal, assembly, and stasis[17]. The stages, along with sea level and ore occurrences, are shown in figure 3.

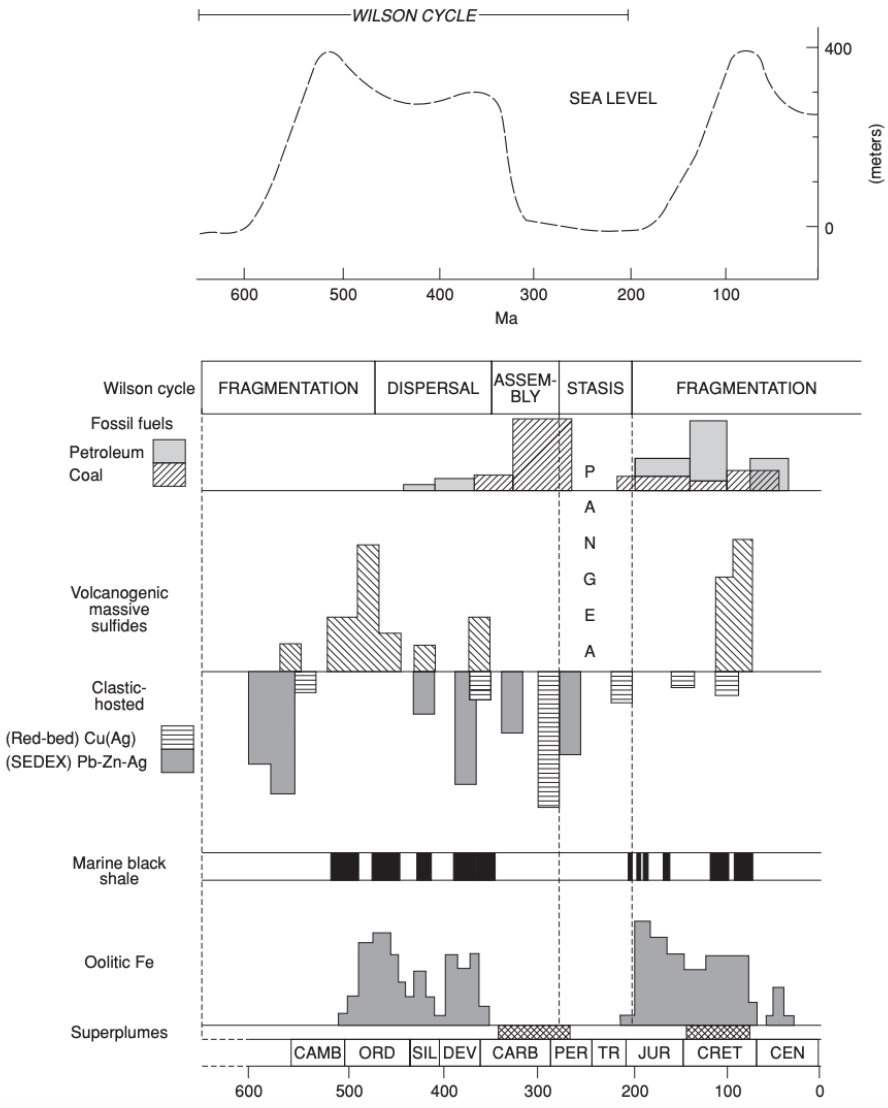


Figure 3: Illustration of the Wilson cycle(from Robb 2014[17])

During fragmentation and dispersal continental rifting occurs, the sea level is at a highstand, and oceanic crust production is enhanced. This appears to preferentially develop the SEDEX-VHMS suite of deposits and organic rich shales, as continental dispersal is accompanied by hydrothermal exhalation favouring VHMS deposits, again driving CO_2-production and global warming which increases organic production.

During assembly and stasis continents amalgamate to make super continents, sea level is at a lowstand, and collisions occur. This appears to preferentially develop the SEDEX-MVT suite of deposits, especially during maximum stasis, as collisions causing orogenesis drive fluid flow for MVT-deposits and intracratonic rifting creates the basinal setting needed for SEDEX-deposits.

2.4 Location and setting of the Barents Sea

The Norwegian continental shelf is large and heterogeneous, so to limit the scope of thesis, and due to the rationale given in section 1.1, this thesis focuses on the continental shelves contained within the Barents Sea.

As illustrated in Figure 4, the Barents Sea is situated in an epicontinental setting as a marginal sea located between Northern Norway, Svalbard(Norway), Franz Josef Land(Russia), Novaya Zemlya(Russia), and the Kola peninsula(Russia), as well as the Norwegian, Greenland, and Arctic seas.

Figure 4: Location of Barents Sea in relation to Norway, Greenland, and Russia. (by Norman Einstein[19])

As many areas in the Barents Sea are protected from georesource exploration, ice covered, or disputed, the scope can be restricted further by limiting the focus to the southern part of the Barents Sea that currently is opened by the Norwegian government for georesource exploration. This area is shown in figure 5 and has the added benefit of having available data surveys with EM and seismic data from petroleum exploration.

The red dots in the figure show petroleum discoveries, and the orange fields show the currently active production licenses as of May 2017, but there may be large additions to this after the 24th licensing round in 2018 with 93 possible blocks in the Barents Sea[20].

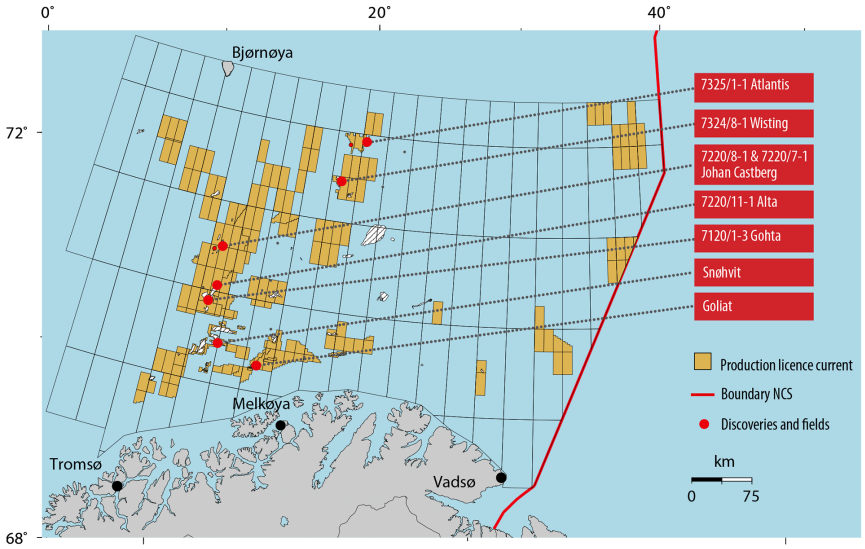


Figure 5: Areas of Barents Sea South opened for petroleum activities, fields on stream (Snøhvit and Goliat), discoveries in the evaluation phase, and the northernmost exploration wells. (Map by NPD[21])

2.5 Machine Learning (ML) basics

ML is considered by many as one the most exciting fields in computer science, has the potential to automate various tasks done by humans today, and will be the main area of focus in this thesis when discussing automation. In this section the semantics of ML and leading ideas will be covered.

2.5.1 Essentials of a Machine Learning (ML) algorithm

ML is very simply put, the ability of an algorithm to acquire knowledge by extracting patterns from raw data, but how is this normally done? Goodfellow et al., 2016, discuss a large number of popular ML algorithms, and show how they all can be decomposed into 3 fundamental parts that essentially build up any ML algorithm[8]:

- A task - Some problem that the algorithm is trying to solve.
- A performance measure - Some way to evaluate how well the algorithm is currently doing.
- Experience - Improvement with increased training.

An example of this is the computer vision algorithm, *AlexNet*, by Krizhevsky et. al. [22]. The **task** is to do image classification on a large image data set called ImageNet with over 22,000 image categories and 15 million labeled images[23]. The **performance measure** is how many percent of images it correctly classifies using deep convolutional networks on a given subset of the image data (discussed in more detail in section 4.1). The **experience** comes from iteratively updating the convolutional network to correctly classify more of the images using a form of statistical optimization (Stochastic Gradient Descent (SGD) with momentum and weight decay, see Goodfellow et al.[8] or Krizhevsky et al.[22] for further information).

The task in this thesis is the classification of geophysical data based on how it effects the favourability of a given deposit. The performance measure and experience depends on the method to be chosen, which is discussed in more detail in section 4.1.

2.5.2 Supervised and unsupervised methods

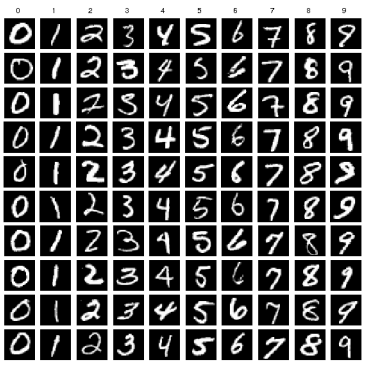


Figure 6: Example images taken from the MNIST dataset. (by Lim et al. [24])

A major division between different ML algorithms is whether they are supervised or unsupervised, and is dependent on how the performance measure is set up.

A **supervised learning algorithm** takes input data with a *known* label, or output value, that the input should be mapped to. The performance measure can then be formalized as finding the function(f) that *best* maps the input data($\mathbb{X}_i$) to the correct output(y_i)($f : \mathbb{X} \rightarrow y$). A famous example of a labeled data set is the MNIST data set, shown in figure 6, where the labels are a number between 0 and 9 shown on the top, and the input is a 20x20 pixel image that displays what digit it is. The dataset is designed to help train algorithms that can classify handwritten digits to automatically read area codes in the postal industry. The task of a supervised learning algorithm in this case is to map any given image to the correct label(0-9).

An **unsupervised learning algorithm** works quite differently, as it involves the algorithm getting input data that is unlabeled, or with an *unknown* output value, and having to make its own categories of the data. Having an overall performance measure is in this case often difficult, but it's usually done by figuring out which data points are most similar and dissimilar in some parameter space and then returning categories based on a segmentation of this parameter space, often along with the transformation function to this parameter space or even a statistical distribution over parameter spaces. A way to think about this is as a form of non-linear, intelligent clustering, where the famous linear equivalent would be a Singular Value Decomposition (SVD) or Principal Component Analysis (PCA)[8]. Mathematically this can be formalized as making the most natural or logical, disjoint and cumulatively exhaustive sub sets($\mathbb{A}_i$) from the input data($\mathbb{X}$), i.e. $(\mathbb{A}_1 \cup \mathbb{A}_2 \cup ... \cup \mathbb{A}_N) = \mathbb{X}$, and $\mathbb{A}_i \cap \mathbb{A}_j = \emptyset \quad \forall \quad i \neq j$. If it's impossible to achieve this whilst keeping the sets disjoint, the goal is then generally to make the intersection of the sub sets $(\mathbb{A}_i \cap \mathbb{A}_j)$ as small as possible.

2.5.3 Training and testing

For an algorithm to gain experience it needs data to train on, and at the same time it needs data to evaluate its performance measure. Normally this data comes from two separate datasets; one known as the training data, that is used to gain experience, and the other known as test data, which is used to evaluate the performance of the algorithm independently. The algorithm's ability to correctly predict on data it hasn't trained on is often referred to as the algorithms ability to **generalize**. A way to think about this concept is an examiner testing students in a university subject. The students are given relevant material in advance to prepare, but *not* the exam itself. They then have to extract the knowledge and generalize from the material given to perform well, and their skill in the actual subject will be tested, not just their ability to memorize a particular exam.

An illustration of the training and testing processes in an algorithm are shown in figure 7. The training operations are shown as blue arrows, the testing operations are shown as green arrows. The steps that normally require storage or segmenting of data are shown as blue parallelograms, steps that normally require more complex mathematical operations are shown as orange squares, and decision nodes are shown as green diamonds.

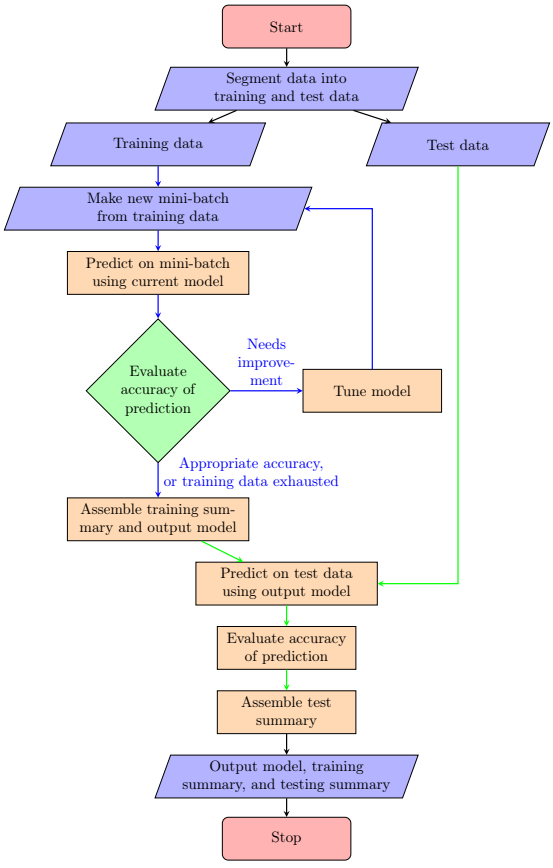


Figure 7: Flowchart of a supervised training epoch, which is often repeated for improved results.

Other important concepts introduced in this illustration are *mini-batches* and *epochs*. A **mini-batch** is a sub set of the data that is easier to run operations on and store in the memory of a computer, as the training and test data sets normally are extremely large in machine learning problems and are impractical to handle collectively (e.g. the ImageNet data set of over 15 million images[23]). An **epoch** is the phrase used for one full training cycle over the entire training set, and is often repeated several hundreds of times allowing the algorithm to improve. When repeating an epoch, it is important to not tune the model based on the results from the test data, and that the segmentation is done in the exact same way, to ensure the algorithm stays unbiased. This can be thought of as a student *forgetting* the content of the previous exam, and making sure that the material for the next exam has not been given out in the relevant material for a previous exam.

Two parameters normally used to evaluate the success of an algorithm is *training error* and *test error*(also known as the validation error or generalization error). The **training error** is a measure on how well the algorithm is doing in making a model that actually fits the training data. For example if the goal is to classify images, the training error is the percentage of images in a mini-batch that the algorithm incorrectly classifies. This is continuously updated as the model is tuned using more and more images to correctly classify the data set. The **test error** is a measure of how well the algorithm can generalize. In the example of image classification, it is the percentage of images in the test data that are incorrectly classified. The model does *not* tune the model based on the test error.

2.5.4 Regularization, overfitting and underfitting

Related to the the training and testing errors are the terms **underfitting**, **overfitting** and **regularization**. If the model is underfit it will lack the appropriate complexity to properly estimate the underlying function or distribution. This can be simply imagined as a linear function approximating a cubic polynomial; no matter how much tuning is done on the parameters of the linear function it will never accurately estimate the polynomial.

On the other hand, if the model is overly complicated, the algorithm is overfitting, and will either, in the absence of noise, spend excessive amounts of time training, or, in the presence of noise, overadapt to the training data and predict based on characteristics that are not in the underlying function. This will result in a very low training error, but a high test error, and the algorithm generalizes poorly[8].

Regularization is simply any measure implemented to improve the algorithms ability to generalize to the test data, often being at the expense of a higher training error. This is widely used in many ML algorithms, and can either be implemented directly in the structure of the model, like having dropout layers that make some of the data disappear during training, in the model tuning step by adding a random disturbance, or by augmentation of the training data[8].

2.5.5 Prediction

The prediction process is normally as simple as utilizing the trained algorithm to perform the desired task. In terms of the MNIST dataset, that would be feeding in packages and read out the area codes. In this thesis that would be feeding in the geophysical data and receiving a prediction of favourability for a given deposit.

As shown in figure 8, this process is very straight forward to execute and implements no model adjustments. The important factors in this step of the process is normally the speed and accuracy of predictions, and the ease of visualization of the results.

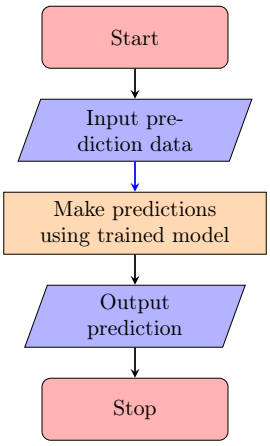


Figure 8: Flowchart for prediction process.

2.5.6 Deep Learning

Deep learning is the term used for an ML algorithm that makes use of a combination of simple concepts to understand increasingly complex concepts, or as described in *Deep Learning* by Goodfellow et. al., 2016:

...allow[ing] computers to learn from experience and understand the world in terms of a hierarchy of concepts, with each concept defined in terms of its relation to simpler concepts. ... If we draw a graph showing how these concepts are built on top of each other, the graph is deep, with many layers. For this reason, we call this approach to Artificial Intelligence (AI) deep learning.[8]

2.5.7 Hyperparameters

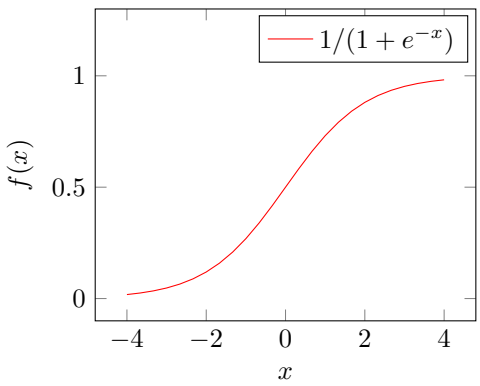
Hyperparameters in ML refers to the parameters that define the size and shape of a model. This is in contrast to model parameters that are tuned by the algorithm. Examples of hyperparameters are number of neurons in a neural network, number of hidden layers in a deep network, or the choice of epochs to train an algorithm over.

2.5.8 Nonlinearities

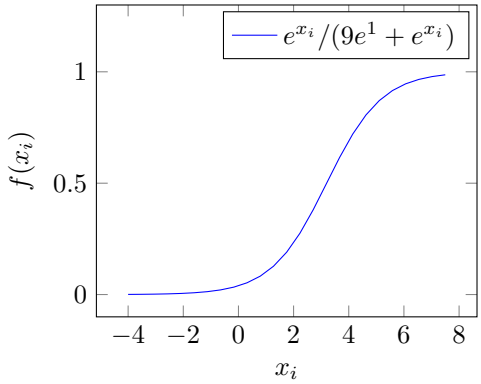
The real world is chaotic and nonlinear. This has been studied extensively, including famous fields like the study of turbulent flow in fluid dynamics, or chaos theory in mathematics[25], but is also highly influential in fields normally studied in ML, like computer vision, or human language processing. This means that the underlying function or distribution that maps the input(e.g. image data) to the output(e.g. classification label), can often be strongly nonlinear. For an algorithm to be able to make correct predictions in this case it needs to have implemented nonlinearities inside the model.

The way this is normally done is a very powerful technique where derived linear features of the input data are extracted first, and then combined in a series of nonlinear ways to get an output[26](p.389). In addition to this, in order to map a complex nonlinear relationship between input and output, it is common to use a combination of several less complex, or *softer*, nonlinearities, in stead of fewer more complex, or *harder*, ones. The softer nonlinearities are then faster to calculate, easier to analyze after training, and has shown to yield equally good results for prediction[27].

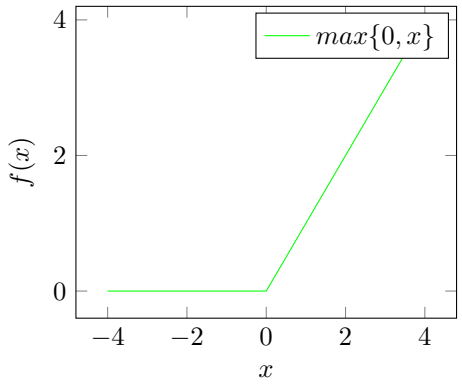
Some examples of commonly used *soft* nonlinearities used in machine learning are given in figure 9. A common characteristic of a soft nonlinearity, like the ones shown is that they are linearly increasing in one domain, often around 0, and approximately constant almost everywhere else, with a small transition domain in between.



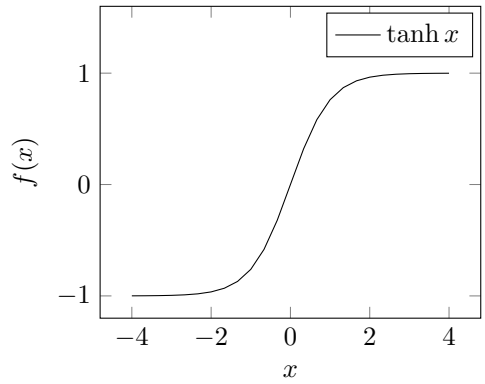
(a) Sigmoid function



(b) Softmax function with $x_i = 1$ for $i \neq j$ and $K = 10$



(c) Rectified Linear Unit (ReLU(x))



(d) hyperbolic tangent function (tanh(x))

Figure 9: Common soft nonlinearities used in ML algorithms, on the domain [-4,4]. with the exception of the softmax function(in blue) plotted on the domain [-4,8] to properly show flattening in the high end of the domain.

Note that the softmax function is dependent on the other values in a set and is normally used

to convert from absolute values to probabilities. The function is defined as[8]:

$$\boldsymbol{\sigma}(x)_i \triangleq \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}} \qquad \text{for} \quad j = 1, ..., K.$$

2.5.9 Machine Learning vs. conventional optimization problems

To finalize the semantics of ML a final question is posed. In the previous section the performance criteria for the two types of learning algorithms was discussed and attempted formalized mathematically, and so one is inclined to ask; what is the difference between ML and conventional, pure mathematical optimization?

The key difference is that in ML, either the datasets are intractable, the performance measure is very difficult to optimize directly, or both. This means that the performance measure is indirectly optimized by solving a different, hopefully representative problem that simultaneously optimizes the performance measure. This can often be very difficult, and entails finding a representative problem that is solvable for the algorithm. This being said, ML algorithms use pure optimization to tune its models, i.e. working to solve the representative problem, but this is not directly optimizing the performance measure[8].

3 Regional Geology

In the following chapter the regional geology for the Norwegian part of the Barents Sea will be described. This includes a description of the structural evolution of the Barents Sea, a delineation of the ROI, its main structural elements, and relevant geological formations. The work in section 3.1 is rephrased from work done in the project paper from Fall 2017[2].

3.1 Geology and structural evolution of the Barents Sea

The Barents Sea has undergone phases of collision, statis, extension, and rifting, and clearly shows several, if not all, of the stages of the Wilson Cycle, having the potential to host many different mineral deposits[11]. In the following section the structural evolution of the Barents will be described chronologically with focus on the Wilson Cycle and the processes that are significant with respect to the genesis of SHMS deposits.

The initial creation of the Barents Sea was caused by two major continental collisions, first creating the western boundary in the late Ordovician(450 Ma) followed by the eastern boundary in the late Carboniferous(300 Ma). The western boundary was created as part of the closing of the Iapetus ocean and acted as the northern edge of the Caledonian orogeny when Laurentia from the West collided with Baltica in the East to create Laurasia. The eastern boundary was created when the Laurasian super continent collided with the Western Siberian plate from the north-east causing the Ural orogeny[28]. This is considered the point of assembly in the Wilson cycle, and the compression persisted well into the Carboniferous-Permian, during which the Laurasian supercontinent collided with Gondwana in the south to create the supercontinent Pangaea.

In the **Carboniferous-Permian** the Caledonian orogeny went into its final stage, known as the Svalbardian phase, and what is now the Barents sea was part of a large continuous inland area between current Northern Greenland and Novaya Zamlya. In the Wilson cycle this is termed the point of statis, as the land mass at this point is relatively stationary, or static. By the **Late Permian**, as the tectonic forces in the area went from compressive to extensive, the Barents Sea started moving into the later stages of stasis in the Wilson cycle. At this point several basins formed in the interior of the area bounded by vast platforms and narrow seismic zones. These basins underwent rapid subduction, as is common in epicontinental rift settings, and the majority of the basins were formed on the western edge of the Barents Sea, in proximity to the reactivated Hornsund Fault Zone and the Senja Fracture Zone, shown in figure 10.

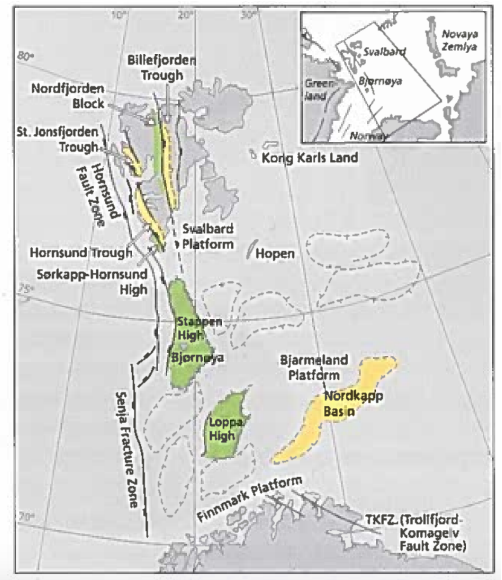


Figure 10: Structural elements in the Barents Sea during the Carboniferous-Permian. Elevated highs as green, active subsidence as yellow. (From *Making of a land*[28])

Throughout this period the area drifted northwards, starting from close to the equator, passing through many climatic zones, ending up halfway

between the equator and the pole. This is reflected in figure 11, showing the Carboniferous-Permian stratigraphy for the basins, by the highly contrasting basin deposits from the initial tropical deposits (green), overlain by arid desert sandstones and evaporites (pink), finally followed by a transgression and various carbonate deposits (blue).

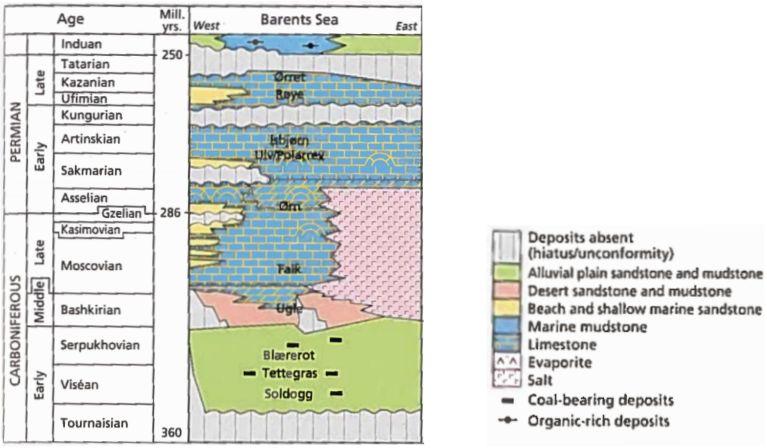


Figure 11: Stratigraphic column for the Barents shelf. These successions reflect a progressive change from arid to humid climate.(Adapted from *Making of a land*[28])

In the **Early Triassic**, what is now the Barents Sea was a part of a huge marine gulf in the northern part of Pangaea, also known as the Boreal Sea, shown in Figure 12. The depositional environment of the gulf changed rapidly due to highly varying global sea levels caused by lithospheric plate movements and global warming. In terms of the Wilson cycle, this can be termed the end of stasis and lead to the cyclical deposition of mud and sand. The **Early-Middle Triassic** is dominated by deep and shallow marine mud due to the rapid melting of the thick southern hemisphere ice caps, whilst the **Middle-Late Triassic** is dominated by sands from alluvial fans and coastal and shallow marine environments due to erosion of the Caledonides[28].

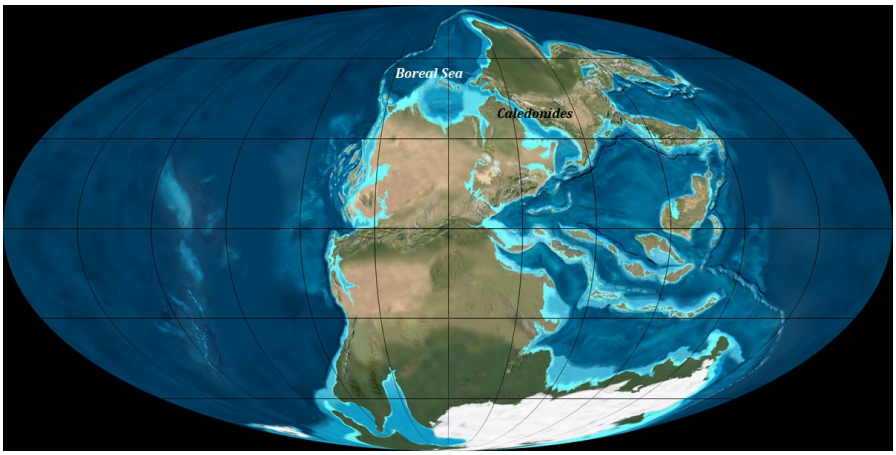


Figure 12: The supercontinent Pangaea during Late-Permian to Early Triassic. (From NAU Geology[29])

By the **Late Triassic**, the Barents Sea started moving from stasis into fragmentation in the Wilson cycle, and rifting occurred more extensively in the western part. Most of the Norwegian continental shelf was at this point covered by alluvial fans and primarily sandstone and mudstone was deposited in and around the rift basins. In the western part of the Barents Sea there are signs that the ocean actually encroached into a narrow strip along the larger rift zone between Norway and East Greenland, especially in times of sea level lowstand, shown in Figure 13. The intracratonic rifting, deposition of continentally derived sediments and rapid subsidence of what is now the Bjørnøya, Hammerfest, Nordkapp and Tromsø rift basins are strong indicators of potential SHMS deposits, discussed in more detail in section 4.1.

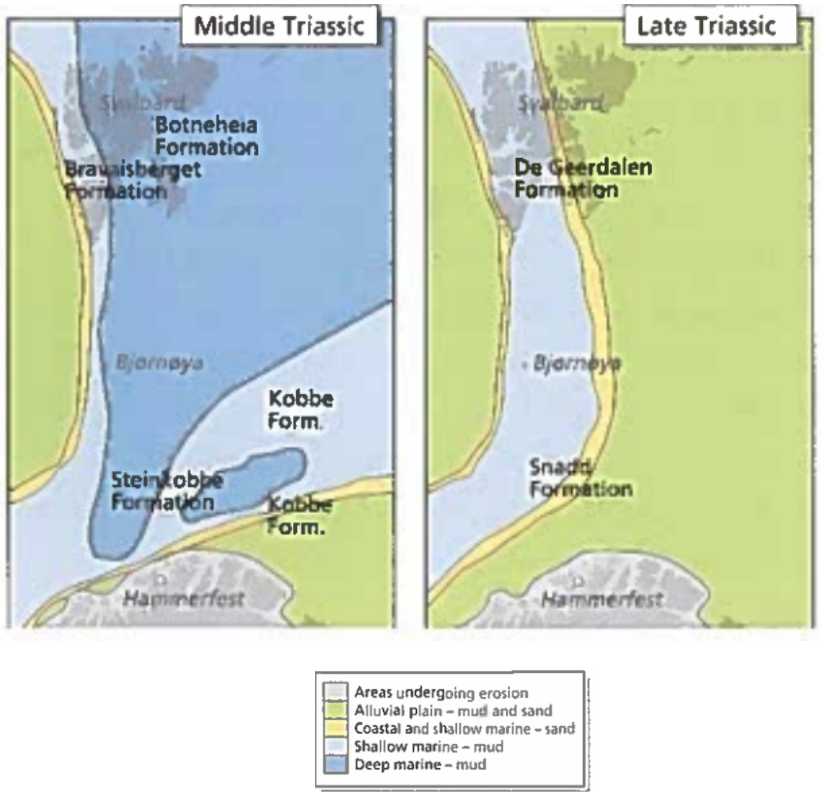
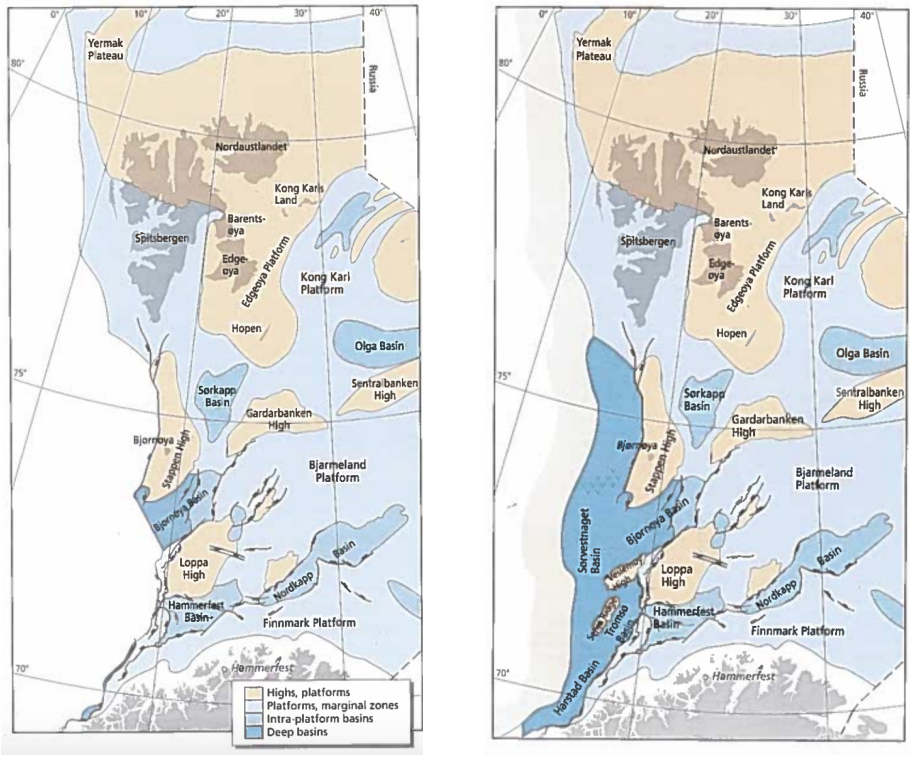


Figure 13: Paleogeography and depositional environment in the Barents sea during the Middle to Late Triassic. (from *Making of a land*[28])

In the **Early Jurassic** global sea levels continued to rise, the alluvial plains in the Barents Sea were flooded and due to the tropical climate were filled with swamplands and other vegetation. This lead to an extensive sand sheet being deposited on the shelf, overlying the fluvial deposits. In the **Early-Middle Jurassic** the Barents Sea was a relatively stable platform, likely accommodating high rates of fluid flow, as basin subsidence and sediment supply approximately offset each other. By the **Middle-Late Jurassic**, the Wilson cycles moved into its current state, full fragmentation, and high vertical tectonic movement sparked the further development of the Loppa High that divides the Barents shelf into the northern and southern groups of rift basins. This again lead to the deposits of large amounts of sediments into the adjacent basins, in particular the Hammerfest basin in the Middle-Late Jurassic, and the Bjørnøya basin in the Late Jurassic-Early Cretaceous, as shown in Figure 14a.

In the **Late-Jurassic** the Bjørnøya basin, Hammerfest basin, and a few other areas in the south-western Barents Sea were affected by rifting in the Norwegian Sea to the south. This rifting caused fault block rotation, further basin subsidence, and persisted until the Early-Cretaceous. By the end of the Jurassic the southern Barents shelf had become a continuous, open sea with excellent oxygen circulation and a plethora of life. This lead to deposition of mudstones that to different extents are calcareous and organic rich.



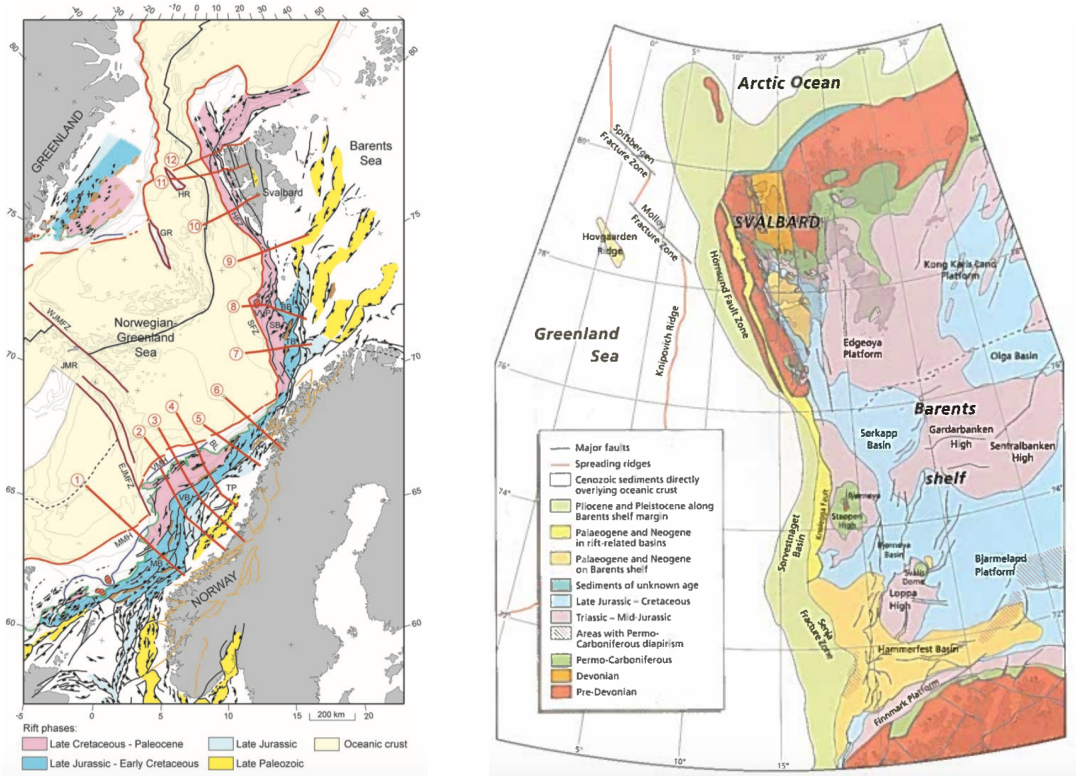
(a) Late-Jurassic rifting and crustal extension.

(b) Cretaceous rifting and crustal extension.

Figure 14: Major structural elements of the Barents Sea and Svalbard in the Late-Mesozoic. (Adapted from *Making of a land*[28])

The **Cretaceous** is dominated by a large rifting event in the west of the Barents Sea, namely the opening of the Atlantic ocean. It now becomes clear that the Hammerfest and Bjørnøya basins are divergent rift arms representing the Atlantic Ocean's failed attempt to move north-east through the massive sedimentary Barents Sea platform. The crustal extension and rifting associated with the Atlantic Ocean did however continue along the western margin of the Barents Sea, and created the deep Harstad and Tromsø basins. In addition it caused the further subsidence of the Hammerfest basin and Bjørnøya basins, and the western part of the Bjørnøya basin in particular, as shown in Figure 14b. The depositional environment remained fairly stable in the Cretaceous, as the major deposits continued to be sandstones in the proximity of highs and mudstones in the basins. However, as the basins continued to subside further, oxygen flow, and therefore life, diminished and the mudstone deposits become less calcareous and less organic rich in the deeper sections.

Finally, in the **Cenozoic** the only deposits found are thin bands in certain regions like the Hammerfest and Bjørnøya basins as the whole Barents Sea region was uplifted and the formations eroded away. The uplift and erosion is believed to have happened in the Neogene and is believed to be up to several thousands of meters and has removed nearly all of the Paleogene deposits. One does, however, find signs of a collision in the Paleogene between Svalbard and northern Greenland. This created the mountain belts seen on Svalbard and reactivated certain fault zones further south in the western Barents Sea, as shown in figure 15a. This was initially thought to be caused by a direct collision event, but is today believed to be the result of lateral displacement and transpressional movements between Svalbard and Greenland, similar to what are seen today in the San Andreas fault[30]. A summary of the rift phases in the Barents Sea and an overview of the deposits on the seafloor and continental Norway and Svalbard is shown in Figure 15, where the red lines in 15a are cross sections attached in appendix A.



(a) Rifting phases in the Norwegian shelf. (Faleide et al., 2008)[11]

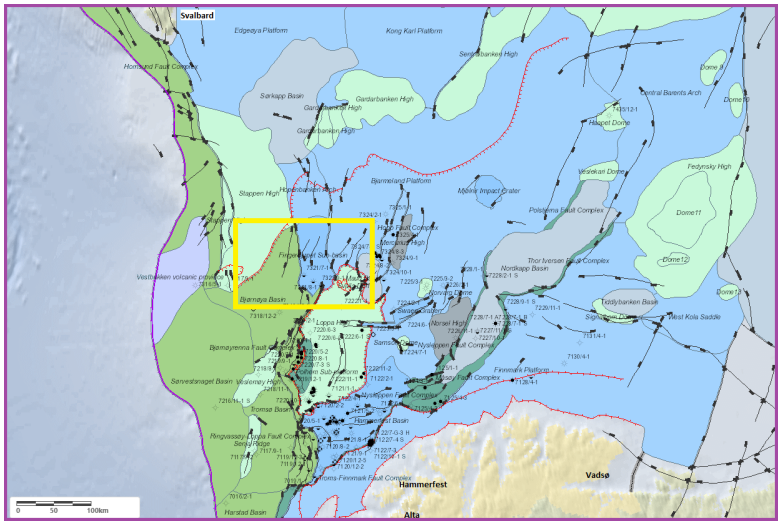
(b) Sedimentary deposits on the seafloor and continental Norway and Svalbard. (Adapted from *Making of a land*[28])

Figure 15: Barents Sea stratigraphy, basins, and rifting episodes.

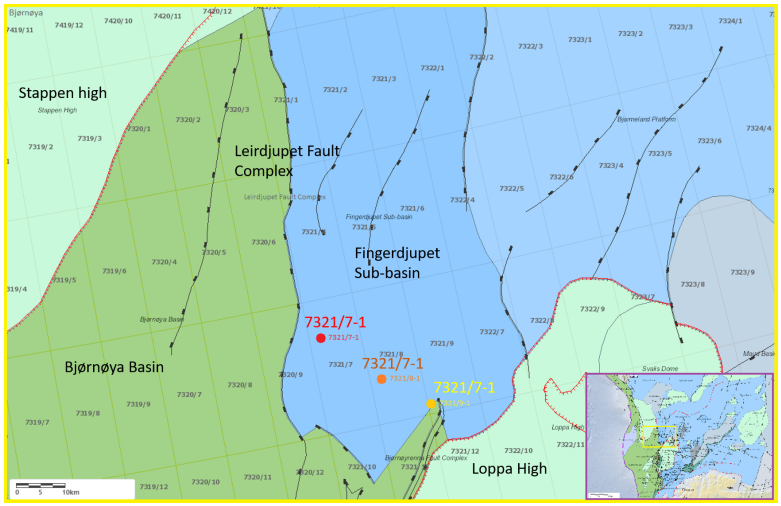
Based on the information presented above the most likely SHMS deposit type seems to be the SEDEX-MVT suite. This is because there is little evidence for volcanic activity, but high rates of fluid, varying pH and Eh conditions in the sediments, and therefore strong potential for diagenetic and epigenetic deposit development and enrichment. The most promising locations in the region appears to be on shoulders of the Bjørnøya basin and the Hammerfest basin. This is because both basins are intracratonic rift basins with historically high influx rates of continentally derived sediments.

3.2 Delineation of the Region of Interest (ROI)

Available data and the observations made above favour the Bjørnøya basin as a specific ROI. This region shows long-lived fault zones with cross-cutting secondary and tertiary faults, similar to those observed in the deposits of the PAOB[14], that serve as fluid pathways. In addition the Bjørnøya basin has been subject to more rifting and subsidence than the Hammerfest basin to the south. The location of the ROI and the surrounding region is shown in figure 16.



(a) Location of the ROI in relation to the rest of the Barents Sea, Svalbard and Norway. (Adapted from NPD[31])



(b) Location of the Stappen high, Bjørnøya basin, Leirdjupet fault complex, Fingerdjupet sub-basin, Loppa high, and petroleum exploration activity. (Adapted from NPD[31])

Figure 16: Location of the ROI.

The large scale depositional trends and geodynamic events for the ROI is shown in figure 17, stretching from Bjørnøya in the North-West to the Loppa high in the South-East.

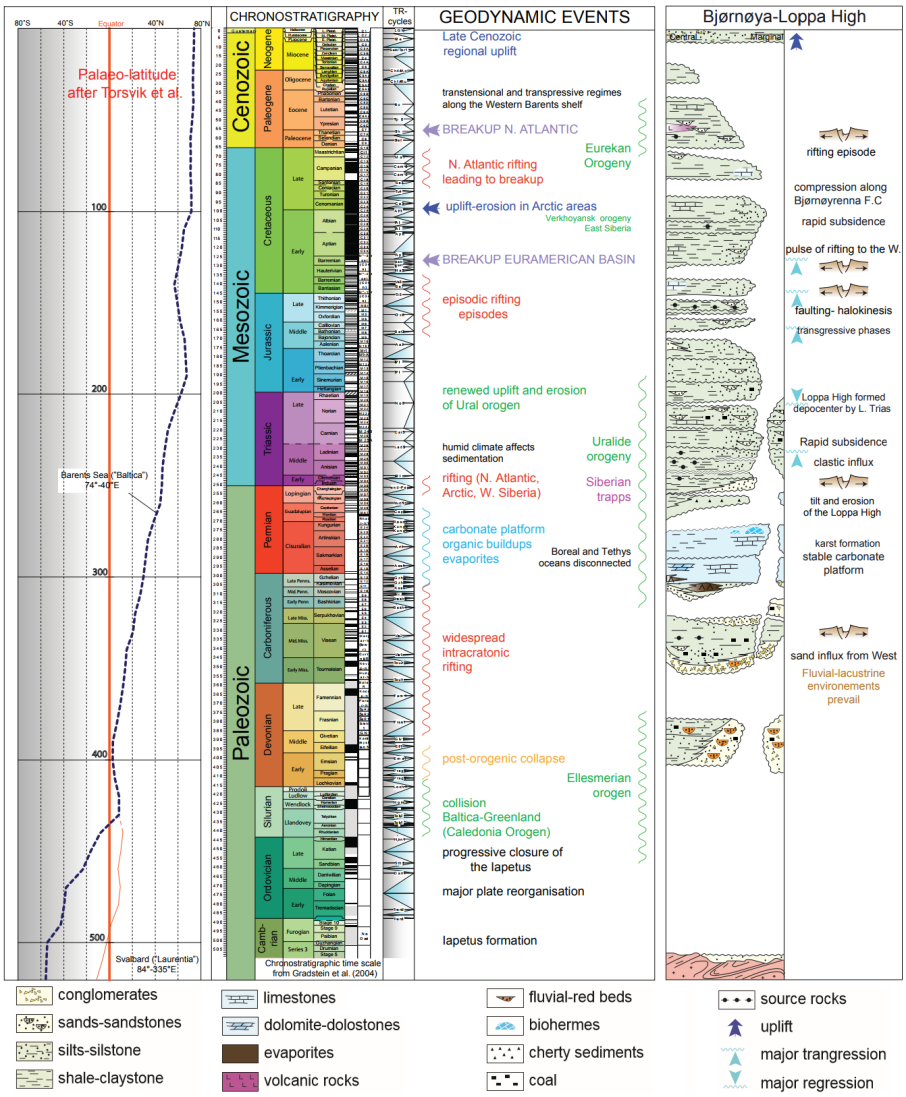


Figure 17: Chronostratigraphy, transgression-regression (T-R) cycles, lithostratigraphy, and geodynamic events for the Bjørnøya-Loppa high. (Adapted from The Geological Survey of Norway (NGU) ATLAS: Geological history of the Barents Sea[32])

3.3 Main structural elements of the Region of Interest (ROI)

The ROI includes several geological structures, namely the Bjørnøya basin, the Leirdjupet fault complex and the Fingerdjupet sub-basin, and is influenced by additional structures surrounding it; the Stappen high and Loppa high, as shown in figure 16. The evolution of these structures are briefly mentioned in section 3.1, but will be described in more detail in this section.

3.3.1 Stappen high

On the western edge of the ROI is the Stappen high with its highest point being the Bjørnøya island in the North-western part of the high. The Stappen high was initially formed as a high in the Permian and underwent several phases of complex tectonic activity in the late Permian and late Jurassic including faulting, tilting, and uplift[28]. This lead the Stappen high to be a source of sediments for the surrounding basins from the early Triassic, and continuing through the Mesozoic.

Through the Jurassic and Cretaceous the Stappen high was subject to rapid subduction, but due to fault reactivation and phases of Cenozoic uplift only late Permian and early Triassic deposits are shown in high variation today.

3.3.2 Bjørnøya basin

South-East of the Stappen high is the Bjørnøya basin, a sedimentary basin formed primarily by subsidence due to the rifting through the Jurassic and Cretaceous. This basin stretches to the Loppa high and Bjarmeland platform in the east and is bounded by the Bjørnøyrenna fault complex in the south. The tectonic activity in the Cenozoic has also caused the basin to tilt, showing characteristics of a large-scale half-graben. [33]

The western part of the basin is deeper and divided from the shallower, eastern part called the Fingerdjupet sub-basin by the Leirdjupet fault complex. The majority of deposits in the basin are likely of Cretaceous age, particularly thick in the western part, and likely come from the surrounding Stappen and Loppa highs[34]. The Fingerdjupet sub-basin in the east has earlier Mesozoic sediments, as shown in figure 18, and will be discussed in more detail later in this chapter.

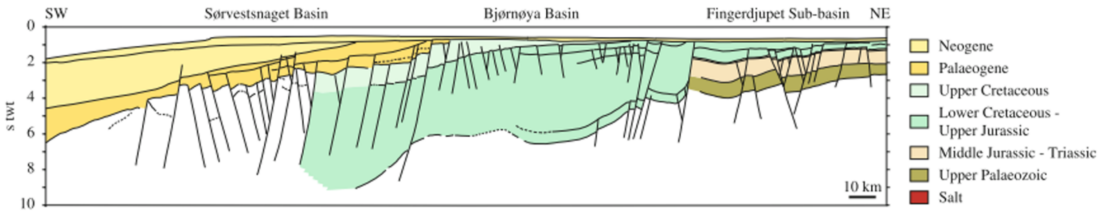


Figure 18: Seismic line crossing the Sørvestnaget and Bjørnøya basin, showing assumed ages of sedimentary deposits[35] (Line shown in figure 16).

3.3.3 Leirdjupet fault complex

Dividing the Bjørnøya basin into its mainly Cretaceous deeper western part and mixed Mesozoic shallower eastern part known as the Fingerdjupet sub-basin is the Leirdjupet fault complex. The fault complex changes from a single fault with significant western trough in the south, to a set of faults with normal troughs in the north[34]. The majority of movement in the fault complex occurred simultaneously with the major subsidence episodes of the Bjørnøya basin, dated to the Middle Jurassic, Early Cretaceous, and the Cenozoic, and has possibly functioned as a conduit, stockwork zone, or venting duct. It likely made the pronounced division of the basin sometime in the Early Cretaceous.

3.3.4 Fingerdjupet sub-basin

Situated in the Eastern part of the Bjørnya basin, the Fingerdjupet sub-basin is shallower than the western part and shows a horst and graben pattern outlined by a system om NNE-SSW trending fault blocks[34]. From middle Triassic to middle Jurassic the sub-basin is thought to have been a part of the regional platform, leaving it with a similar succession to the Stappen and Loppa high. The main fault trends in the basin were created during the late Jurassic, and the sub-basin itself separated from the western part when it was not subjected to large scale subsidence during the Early Cretaceous extensional tectonics. This sub-basin retains more of its Mesozoic deposits, likely coming from alluvial fans and deltas in the east[28]. As with the rest of the region, late Ceonozoic uplift caused erosion of all Cenozoic sediments, and so the only deposits remaining are the Upper Paleozoic and Mesozoic as shown in figure 18

3.3.5 Loppa high

South-East of the Bjørnøya basin, bounded by the Bjørnøyrenna fault complex to the west and the Bjarmeland platform in the North is the Loppa high. The high became pronounced during two periods of tectonism, Late Jurassic to Early Cretaceous and Late Cretaceous to early Cenozoic, but the western edge of the high shows signs of reactivation up to four times. From middle Triassic to middle Jurassic the high was part of a regional cratonic platform also underlying the Fingerdjupet sub-basin[34]. During the Cretaceous the Loppa high was an island with deep canyons cutting through its sides penetrating into Triassic sediments, supplying the surrounding basins with sediments. Finally, all late Cretaceous and Cenozoic shales deposited on the high were eroded away by late Cenozoic uplift, and high magnetic and gravity anomalies indicate a shallow, metamorphic Caledonian basement, particularly shallow in the western part.

3.4 Outline of relevant stratigraphy in the Region of Interest (ROI)

The formations from the late Permian to the middle Jurassic are described in more detail in the following section as these are the most promising in terms of Wilson cyle, sediment influx, and tectonic activity. A more comprehensive lithostratigraphy, including formation names and megasequences is also shown in figure 19. The formations have been determined by well cores, seismic data, and analogues on Svalbard[36].

3.4.1 The Tempelfjorden group

Deposited from mid until late Permian, this group is composed of fine-grained siliciclastics. This includes siltstones, sandstones, marls, silicified limestones, calcerous claystones, and shales. The group represents deposition in cool water, basinal and temperate shelf environments[38].

The group is made up of the Røye formation overlain by the Ørret formation, both found in well cores from the Loppa high in our ROI.

The **Røye formation** is composed of silicified sediments. The lower section is dark colored silicified calcerous claystone and mudstone, whilst the middle to upper section shows signs higher degree of calcification with marls, spiculites, cherts, limestones and mudstones in addition to the claystones found in the lower part. The depositional environment is determined to be distal marine in the lower part transitioning into distal marine, moderate to deep shelf setting in the middle and upper part.

The **Ørret formation** is composed of sandtones, siltstones and shales. The sandstones occur as fine-grained, isolated thin beds, separated by shales. The depositional environment is determined to be deep shelf environments and deltaic, coastal plains, with anoxic conditions in the eastern part of the Barents shelf.

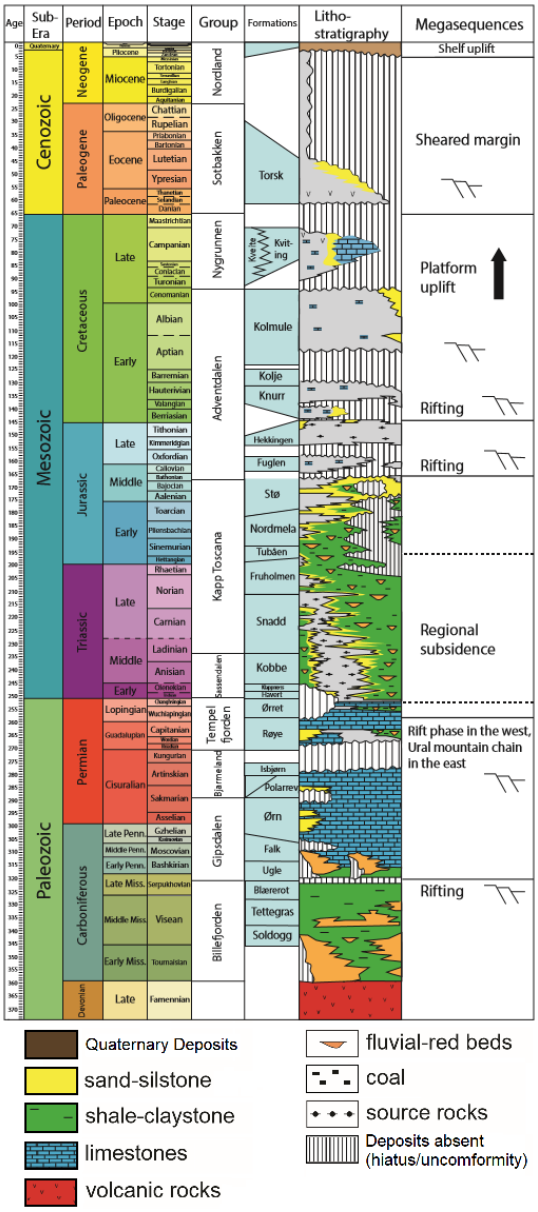


Figure 19: Map illustrating major structural elements on the Barents shelf from the Late Devonian to the Late Neogene. (Adapted from Glørstad-Clark et. al. 2010[36], and Norkus 2015[37])

3.4.2 The Sassendalen group

Deposited from early to middle Triassic, this group is composed of dark colored claystones and shales, with siltstones and sandstones in the upper part [39].

The group is made up of the Havert formation overlain by the Klappmyss formation, again overlain by the Kobbe formation, found in cores primarily drilled south of the ROI, closer to the Snøhvit gas field.

The **Havert formation** is composed of dark colored shales with interbedding siltstones and sandstones. The depositional environment is determined to be marginal to open marine.

The **Klappmyss formation** is composed of dark colored shales with interbedding siltstones and sandstones in the upper part. The depositional environment is determined to be marginal to open marine.

The **Kobbe formation** is composed of a thick shale unit in the lower section overlain by interbedded carbonate semented siltstone, sandstone and shales. The depositional environment is determined to be marginal marine.

3.4.3 The Kapp Toscana group

Deposited from late middle Triassic to middle Jurassic, the group is composed of shales with thin layers of coal in the lower part and sandstones in the upper part[39]. The group is made up of the Snadd formation, overlain by the Fruholmen formation, overlain by the Tubåen formation, overlain by the Normela formation, again overlain by the Stø formation.

The Snadd formation is found as a thick layer (>1000 meters) in cores on the Loppa high, whilst the rest of the group is found in thinner layers in cores from the Bjarmeland platform to the east of the ROI.[40]

The **Snadd formation** is composed of shales, limestones and calcerous beds in the lower part and interbedded siltstones and sandstones with coaly lenses in the upper part. The depositional environment is determined to be distal marine and a large prograding deltaic system, with storm derived sand and silts.

The **Fruholmen formation** is composed of interbedded shales and coal lenses in the lower part, sands in the middle part, and shales with sandstone lenses in the upper part. The depositional environment is open marine for the shales and fluvial or coastal for the sandstones.

The **Tubåen formation** is composed of sandstones with subordinate shales and coal. The upper and lower parts are dominated by sandstones, whilst the middle section is mainly shale. The depositional environment is determined to be stacked fluviodeltaic deposits for the sandstones and distal marine in the north-west for the shales with coals deposited in protected backbarrier lagoonal environments in the south-east.

The **Normela formation** is composed of interbedded siltstones, sandstones, shales, and mudstones with minor coal, with sandstones thickening towards the top of the formation. The depositional environment is determined to be tidal flat to flood plain environment, where individual sandstones represent tidal channels. This formation shows clear thickening to the west, likely due to the subsidence of the Bjørnøya and Tromsø basins.

The **Stø formation** is composed of sandstones, moderately sorted, with thin layers of shale and siltstone. The depositional environment is determined to be a preograding coastal regime for the sands, with regional transgressive pulses causing the shales and silts.

4 Theory

In the following chapter a literature review will be completed on genetic modeling of SHMS-deposits, exploration models and its limitations for SHMS-deposits and other resources in the ROI. In addition, existing work on automation in geophysics using machine learning, and automation in medical imaging as a proxy will be investigated. The work in this section is heavily influenced and inspired by the work done in the project paper from Fall 2017[2].

4.1 Genetic Modelling

Based on the geology of the Barents Sea presented in chapter 3, the SEDEX-MVT suite seems more likely due to the lack of intrusive magmatism, and prevalence of carbonates. The SEDEX-VHMS suite should however not be completely ruled out, and may even be present in sedimentary sequences in the same areas as the SEDEX-MVT continuum. This can also be seen from cross-section 8 in Figure 50b, where there are extrusives present to the west of to the Bjørnøya basin further pointing to the possible, albeit unlikely, existence of SEDEX-VHMS continuum deposits in the ROI. Therefore genetic models for both suites are presented below.

4.1.1 Genetic models for SEDEX-VHMS deposits

For all SHMS deposits there needs to be a heightened geotherm causing fluid circulation, a conduit system, and a chemical or physical barrier that causes metal precipitation. According to Robb, 2014, the tectonic setting for the SEDEX-VHMS continuum is an intracratonic rift setting, creating an inland ocean with a large influx of continental sediments from its margins[17]. This setting also holds heightened geotherms caused by crustal thinning, again driving fluid circulation and exhalative vents on the sea floor. When the fluids circulate through deeper anoxic sediments or magmatic rock the fluids become reducing and acidic and leach the overlying pelagic or marine sedimentary pile or even the crystalline basement rocks for metals. When the fluid is exhaled onto the sea floor, the sudden change in temperature, oxygen fugacity and pH induces metal precipitation and, in the presence of reduced sulphur, will deposit stratified Sphalerite(ZnS) and Galena(PbS) or, in the presence of oxidized sulphur, deposit stratified Barite($BaSO_4$) in the siliciclastic bottom sediments, described by MacIntyre, 1991[13].

The shape of the resulting deposits are largely dependant on the temperature and density of the fluid at the time of exhalation, as a warmer, less dense fluid disperses further and is less constrained to proximal depressions in the sea floor[17].

A given deposit may be more volcanic depending on the influence of intrusive magmatism and black smokers in the area, i.e. similar to the conventional VHMS deposits. As described by Walters, 1998, this can either be syngenetic with direct influence from rift magmatism, or epigenetic, as is the case with the Broken Hill Type (BHT) deposits mentioned in figure 1, where later magmatic events have caused metasomatism of the original deposits[41].

MacIntyre, 1991, describes several SHMS deposits in the Canadian cordillera showing similar genetic trends and mineralizations, further supporting the model by Robb, 2014[13]. Additionally, a modern example of an ongoing SEDEX-VHMS process showing the same trends is described; the deposition of stratified Sphalerite and Galena deposits in brine pools in the Atlantis II deep of the Red Sea between the Arabian peninsula and Africa.

4.1.2 Genetic models for SEDEX-MVT deposits

The main difference from the SEDEX-VHMS continuum is the presence of carbonates and absence of volcanics in the host rock of the deposit, and when it comes to tectonic regimes there are two major settings that are prevalent in the literature.

In the setting described for the SEDEX-VHMS deposits above there can be a scenario where the rising fluid precipitates its metals below the sea floor. According to Robb, 2014, this may either be triggered by boiling during ascent due to rapid decompression, or by coming into contact with a shallow reservoir of sulphur-rich fluid. In the case of boiling the release of energy causes brecciacion and metal precipitation in a very massive Sphalerite(ZnS) and Galena(PbS) dominated deposit, similar to the deposits near the Juan de Fuca Ridge[17]. If the metalliferous brine comes into contact with a shallow fluid reservoir hosted by a very carbonate rich wall rock the deposit is referred to as an IVT which account for some large deposits, like the Navan deposit in Ireland, the largest Zinc resource in Europe, described in detail by Wilkinson, 2005[42].

In addition to this, Goodfellow, 2004, describes deposits that are hosted in extensional and compressional settings on the shoulders of fault-bounded grabens in reactivated epicratonic rifts[43]. The fluid flow is here driven by several additional factors to the crustal thinning mentioned above, namely; diagenetic dewatering, metamorphic dehydration, magmatic intrusions, and in some cases even high levels of radioactive heat from the basement rocks[16]. A famous example of such a deposit is the Red Dog formation in Alaska. According to the study by MacIntyre, 1991[13], this deposit is formed during tectonic extension in the Devon-Carboniferous with heightened geotherms and faulting driving fluid flow, and deposition occurring from both exhalation in a narrow intracratonic basin and by replacement in the subsurface. Robb, 2014, notes that during rifting the graben normally subsides rapidly and siliciclastic sediments, often carbon rich shales, are deposited. Then, during reactivation, the graben is thrusted upwards and evaporites and/or carbonates are deposited giving large Eh and pH contrasts[17]. Borg, 2000, observes that due to reactivation, faults also act as long lived pathways for fluids to flow through and allows for epigenetic deposition of metals when passing contrasting sedimentary layers, as well as on the sea floor through vents. Visually the deposits tend to resemble IVT deposits, but are somewhat less stratified, include stockwork zones from the feeder systems and can also give rise to Karst formations[14]. Geochemically these deposits are similar to the deposits mentioned above, but from isotope studies, like the one by Misi et. al., 2005, it is shown that metals in certain deposits can be derived from even deeper sources than the basement rocks, like the mantle[16].

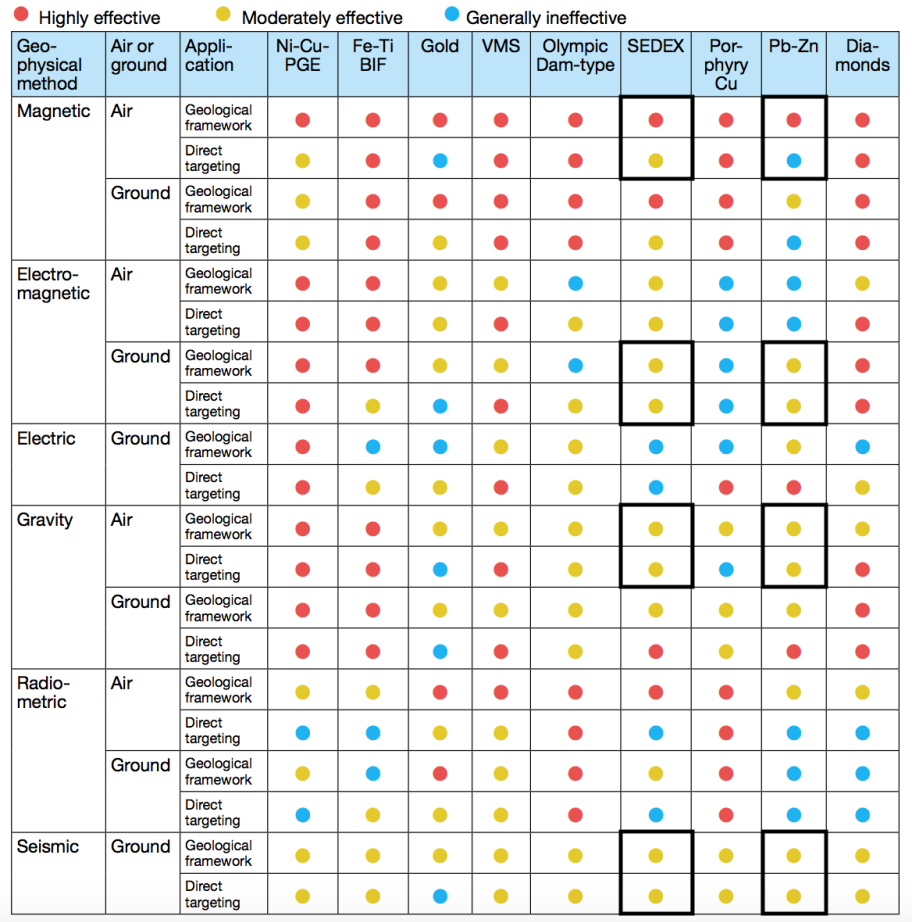
MVT deposits are normally not considered SHMS deposits, but often form in a similar sedimentary sequence, and in areas close to the deposits mentioned above[14]. Explained by Robb, 2014, the fluid flow is in this case driven by topographic elevation from orogeny due to compression, and the deposition occurs when an acidic, saturated brine epigenetically releases its metals when coming into contact with a contrasting sedimentary wall rock in the foreland basin. The release of metals can be driven by a sulphide rich fluid reservoir in the wall rock, organic matter driving pH and Eh changes, or with a carbonate rich wall rock driving pH changes and brecciacion[17].

Finally, as mentioned in section 2.3, these ores have been observed to preferentially develop at periods of maximum continental amalgamation and stasis, namely Pangaea in the Permian-Triassic, and Gondwana in the Precambrian-Cambrian[17]. These are not just periods of intracratonic rifting, but also sea level low stands, global warming and high points of organic matter. As plants can help create the necessary pH and Eh conditions, or even supply necessary Sulfide ions for precipitation, several studies have shown that hydrocarbon systems and SEDEX-MVT deposits can be found in the same regions and be part of the same geological systems[44].

4.2 Exploration models

As explained in section 2.1, an exploration model is a model of how a given deposit and geologic framework are observed in geophysical data. Now that genetic models have been defined for the two suites of SHMS deposits, the different ways they may present themselves will be investigated.

M. L. Airo, 2015, explores the geophysical responses and geological characteristics of some of the more common mineral deposits in Finland and their minerals[45]. It summarizes, from a previous study, geophysical methods used in on-shore Canada, and which have been the most efficient or useful for different deposit types found in Finland, modified and shown in Figure 20. Here the label Pb-Zn deposits is given to deposits on the SEDEX-MVT suite, with most of them being towards the MVT end. These indications are primarily made for Finland in an on-shore settings, but are expected to be similar in overall trends and to be indicative of exploration on the continental shelf. Therefore the following section will focus on, and further analyze using existing literature, the methods presented in this diagram, and what is being measured.



Geo-physical method	Air or ground	Appli-cation	Ni-Cu-PGE	Fe-Ti BIF	Gold	VMS	Olympic Dam-type	SEDEX	Por-phyry Cu	Pb-Zn	Dia-monds
Magnetic	Air	Geological framework	●	●	●	●	●	●	●	●	●
		Direct targeting	●	●	●	●	●	●	●	●	●
	Ground	Geological framework	●	●	●	●	●	●	●	●	●
		Direct targeting	●	●	●	●	●	●	●	●	●
Electro-magnetic	Air	Geological framework	●	●	●	●	●	●	●	●	●
		Direct targeting	●	●	●	●	●	●	●	●	●
	Ground	Geological framework	●	●	●	●	●	●	●	●	●
		Direct targeting	●	●	●	●	●	●	●	●	●
Electric	Ground	Geological framework	●	●	●	●	●	●	●	●	●
		Direct targeting	●	●	●	●	●	●	●	●	●
Gravity	Air	Geological framework	●	●	●	●	●	●	●	●	●
		Direct targeting	●	●	●	●	●	●	●	●	●
	Ground	Geological framework	●	●	●	●	●	●	●	●	●
		Direct targeting	●	●	●	●	●	●	●	●	●
Radio-metric	Air	Geological framework	●	●	●	●	●	●	●	●	●
		Direct targeting	●	●	●	●	●	●	●	●	●
	Ground	Geological framework	●	●	●	●	●	●	●	●	●
		Direct targeting	●	●	●	●	●	●	●	●	●
Seismic	Ground	Geological framework	●	●	●	●	●	●	●	●	●
		Direct targeting	●	●	●	●	●	●	●	●	●

Figure 20: Applicability of different geophysical methods in the exploration of various mineral systems in on-shore Canada (Adapted from Airo 2015(Table 1)[45]). Relevant methods and deposit types marked in bold.

4.2.1 Magnetic

Magnetic methods measure the remnant magnetism in the minerals of a rock mass, with an example for the Barents Sea shown in figure 21.

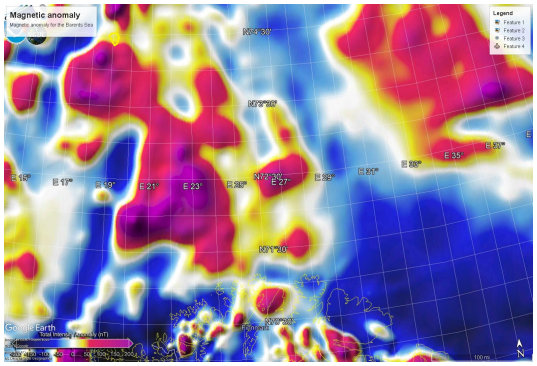


Figure 21: Global Earth Magnetic Anomaly Grid (EMAG) compiled from satellite, ship, and airborne magnetic measurements for the Barents sea, on a 2-minute grid, 4 km. above the geoid. (Compiled by The National Oceanic and Atmospheric Administration (NOAA)[46])

In 1996 the the United States Geological Survey (USGS) composed a compilation of mineral deposit models and what exploration techniques to use when prospecting these, based on a series of research articles[47]. This book covers in detail SEDEX and MVT deposits, and conventional geophysical exploration methods used for these, including magnetic and EM methods, gravimetric methods, and electric methods. The sulphide mineralizations have shown to give characteristic responses with EM and magnetic methods, especially pyrite and pyrrhotite-rich deposits.

M. L. Airo, 2015, also notes that pyrite and pyrrhotite-rich deposits will spark high anomalies, as these minerals hold remnant magnetism, and that the method is not disturbed by lakes, soil or waterways that may cover the bedrock in Finland. Unfortunately, the method only measures lateral contrasts, and the resolution of the method gets hindered by the fact that the data has to be collected by airplane[45].

4.2.2 Controlled Source ElectroMagnetic (CSEM)

Controlled Source ElectroMagnetic (CSEM) methods measures the resistivity of a rock mass.

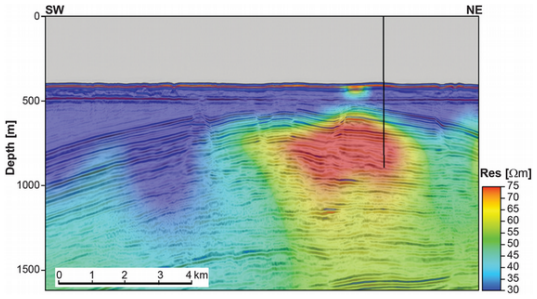


Figure 22: The Wisting oil discovery in the Barents Sea as imaged by 3D CSEM inversion (vertical resistivity), displayed on a seismic line. (from Fanavoll et al., 2014[48])

In the compilation by the USGS this is described in a similar manor as for magnetic methods, where sulphide minerals cause lower resistivities, primarily pyrite and pyrrhotite[47].

M. L. Airo, 2015, notes that since CSEM methods measure electrical resistivity, all sulphide minerals, except sphalerite, should display a low resistivity and make the rock mass display anomalous responses. The issue with this method, however, is that it requires extensive processing and is vulnerable to non-geological noise[45].

Finally, CSEM methods are widespread in hydrocarbon exploration, as petroleum has a high resistivity. In addition, by installing the equipment on the sea bed, this gives much higher resolution than magnetic methods, as shown in figure 22

4.2.3 Induced Polarization (IP)

Both the USGS and M. L. Airo report some success with an electric method called Induced Polarization (IP), that measures electrical features like resistivity and phase angle changes induced when passing through a deposit[47][45]. However this method is not feasible off-shore as it bases itself on sending currents through the rock mass.

4.2.4 Gravimetric

Gravimetric methods measure the density of a rock mass, with an example for the Barents Sea shown in figure 23.

In the compilation by the USGS the SEDEX deposits, and the extent of barite-rich zones in particular within these, can be indicated by gravity anomalies, as barite is an exceptionally heavy mineral[47].

In addition to barite, M. L. Airo, 2015, discusses the higher density of iron sulfides, like pyrite and pyrrhotite, and their likely appearance as anomalies on a gravity map. As for magnetic methods, a limitation of the method is that it only measures lateral contrasts and is limited in resolution as the data is collected by airplane.

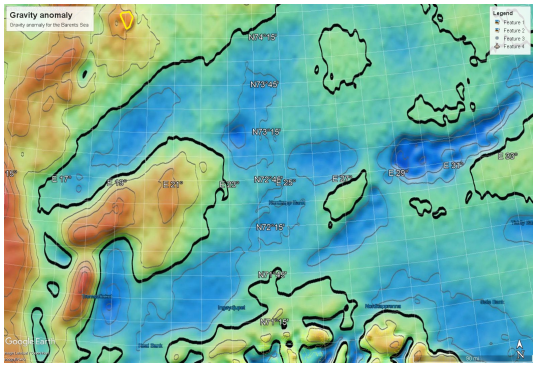


Figure 23: Satellite derived gravity anomaly map for the Barents sea, with contour lines every 20 mGal, on a 1-minute grid. (Compiled by the University of California, San Diego (UCSD)[49])

4.2.5 Well logs (Radiometric)

As with the electrical methods, conventional radiometeric methods are not feasible in an off-shore setting, due to physical limitations of the method[45].

However, Clavier et al., 1976, report the occurrence of pyrite in drill cores from hydrocarbon fields causing anomalous radioactivity logs[50]. This attests to the relationship between sulphide mineralization and hydrocarbon systems, and the potential usefulness of cores from petroleum activities. It is in fact quite well documented that the reduced sulphur essential in SHMS formation can come from organic and biological sources, even directly from bitumen[17].

This means that radioactivity logs collected from petroleum activity could prove useful, but this is very localized, and not true for conventional radiometric methods.

4.2.6 Seismic

Wu et al., 2013, present a possible genetic model for the relationship between the hydrocarbon systems and MVT deposits in south-western China [51]. Here a mechanism is proposed where the formation of MVT deposits coincides with the degradation of hydrocarbon reservoirs due to migration of basinal fluids. This leads to the proposition that seismic methods, widely used for

hydrocarbon exploration to map stratigraphic sections, could be useful in mapping the geological framework for sediment hosted deposits as well.

E. L'Heureux, B. Milkereit and E. Adam, 2005, discuss how seismic methods could be used in massive sulphide exploration[52]. The main takeaway seems to be that sulphide ore would make a strong seismic reflector. They also conclude that for seismic to be useful for direct targeting, there needs to be a comprehensive 3D survey. However the energy will likely be converted to shear waves and dispersed in dip-dependent directions.

M. L. Airo, 2015, reviews the usefulness of seismic methods and concludes that the ability to detect vertical as well as lateral changes is extremely useful, however as it doesn't detect deposits directly it is primarily useful for indirect or framework detection[45].

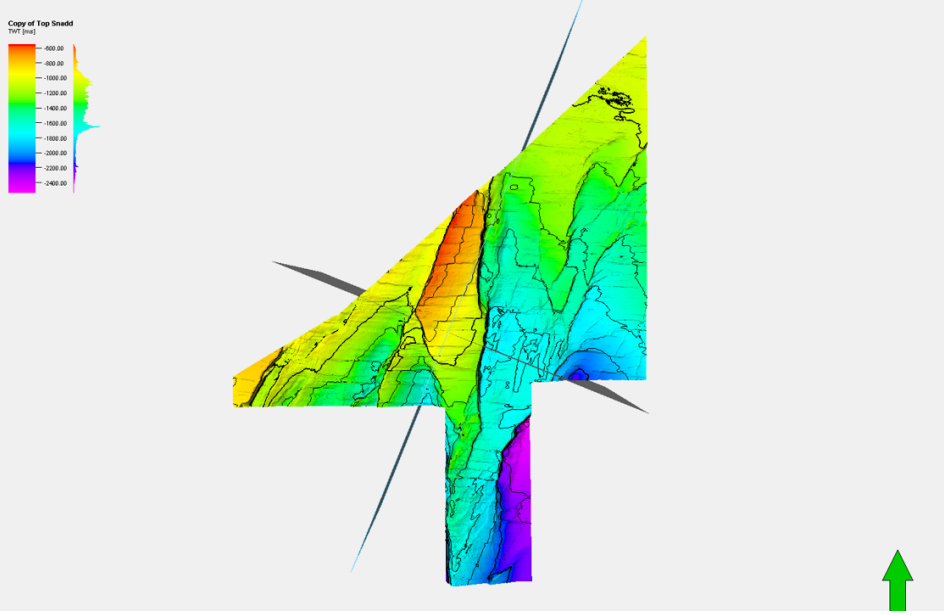


Figure 24: Example of a seismic horizon visualized in Petrel. (Top Snadd from seismic dataset provided by TGS-NOPEC Geophysical Company ASA (TGS))

The relative efficiencies of the methods is primarily derived from the response created by the respective mineralogies. It seems like sulphide mineralizations generally have a higher acoustic impedance and specific gravity due to their high density(Galena and Pyrite)[52], but most significantly give a strong magnetic anomaly due to the magnetic character of Iron sulfides. In addition it seems like a magnetic survey could help determine the sub-surface structures related to a deposit, but that a combination of several methods would be necessary to model the geological framework of the system accurately.

4.3 Automation using Machine Learning (ML)

Automation in geophysical interpretation using ML is in its early stages, and other fields of research have shown great advancements using such techniques to interpret data. Therefore, to get a wider range of literature and more information about the field, studies on using machine learning for medical imaging have also been examined. The main points of these studies are not necessarily always applicable, but if studied critically can provide important insights as they also deal with the automatic inverse 3D-tomography of sound wave data (Ultrasound), magnetic data (MRI), and EM data (CAT, Positron Emission Tomography–Computed Tomography (PET-CT), and X-ray).

Zhang et al., 2016, attempted to automate geophysical feature detection using deep learning[53]. The method used with some success appears to be a straight forward Deep Neural Network (DNN) in 2-dimensions. The results seemed promising for both fault detection and salt body detection.

Waldeland and Solberg, 2017, showed that salt body detection could also be done with accuracy using 3D-CNNs[10]. This work was also expanded upon by Ildstad and Bormann to be able to detect sedimentary facies and faults[54].

Xie et al., 2018, compared different machine learning methods for formation lithology identification comparing tuning processes and performance[55]. The performance was tested on well-log data, not explicitly defined in the paper, and evaluated by cross-referencing. Ensemble methods turned out to be the best on these data, however the authors note that the number of features in the data was limited, perhaps rendering the CNN too complicated, as these methods perform better in high-dimensional spaces resulting in overfitting.

Cracknell and Reading, 2014, also compared five different machine learning algorithms and their ability to do geological mapping using remote sensing data[56]. The methods included Support Vector Machines (SVM), Random Forests (RF), and a neural network, among others. The results were based on several factors, like stability, ease of use, processing time and prediction accuracy, and RF, an ensemble method, came out as the best. It is worth noting that the neural network used here was a very simple DNN, namely a multilayer perceptron.

Roden and Sacrey, 2016, use Self Organizing Maps (SOM), a type of unsupervised neural nets, to do seismic interpretation[57]. This is after first doing feature selection by experience, or PCA, and having the algorithm successfully distill numerous attributes to identify geologic objectives.

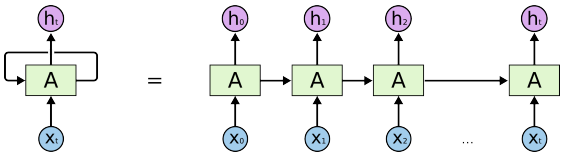


Figure 25: Basic structure of a simple Recurrent Neural Network. (by Christopher Olah[58])

In the 2017 publication by Springer publishing, a number of papers on new methods in medical imaging using machine learning is collected[9]. This includes some studies using ensemble methods, in particular RF and Gradient Boosted Trees (GBT), but the majority of the works is using a form of CNN, primarily in 3D. The most notable architectures implement state of the art concepts successfully to get faster and more accurate imaging, including cascading structures, U-nets, and Recurrent Neural Network (RNN)s. A cascading structure simply means that there are more than one CNN to process information, RNNs send outputs from nodes back into themselves to reuse information(figure 25), and U-nets use a reverse version of the trained neural network to create more training data and use the initial samples more efficiently[59].

There are a plethora of other ML methods that have been shown to be useful, but to limit the scope of this thesis only one will be investigated in more detail. Based on the literature the most promising appears to be Deep Learning, and DNNs in particular, and is the type of method reviewed further.

4.3.1 Deep Neural Networks (DNN)

Goodfellow et. al., 2016, describe a series of Deep Learning methods, with the simplest DNN method, shown in figure 26, called a multilayer perceptron[8]. It is called this because it emulates the biological neural network that constitutes animal and human brains, and implements hidden layers giving depth to the model. Between each layer there is a soft non-linearity, differentiating this method from simple linear interpolation. Shown in Waldeland et. al., 2017, this allows the network to model more complex non-linear phenomena like the relation between 3D seismic amplitude profiles and a given classification label[10].

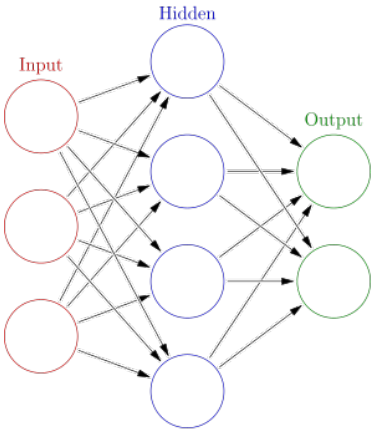


Figure 26: A multilayer perceptron with one hidden layer. Each circular node represents an artificial neuron and an arrow represents a connection from the output of one neuron to the input of another. (by Glosser.ca [60])

In the book *Statistical Learning* by Hastie et. al., 2009, it is found that DNNs solving tasks involving local features, or more generally looking at regionalized variables, can be made more efficient by implementing **convolutional layers**. As illustrated in Figure 27, a convolutional layer is a layer that is not fully connected in its input and/or output, and so only looks at values in a certain region of the image, also known as its receptive field. This is normally quite useful as the internal relationship between values in the receptive field is more informative than the relationship between values that are spatially further away from each other. This is naturally useful when classifying regionalized variables, and in addition to greater efficiency there are several other advantages to convolutional layers, like more complex feature extraction and increased parallelizability[26].

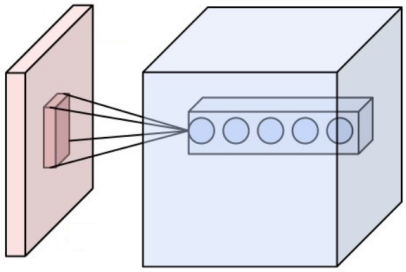


Figure 27: Neurons of a convolutional layer (dark blue), connected to their receptive field (dark red). (by Aphex34 [61])

To improve generalization of CNNs, Srivastava et al. 2014 implemented a technique called **dropout** and showed that it improved performance. Dropout is simply making some percent-

age of the nodes in the network inactive, or dropped, and varying which nodes to deactivate randomly during training. This makes the network avoid converging into local minima during optimization[62], and generalize better.

In the case of limited training data, Wong et. al. 2016, showed that **data augmentation** could be a useful technique to increase the training data space. This is simply implementing spatial, or other, transformations on the input data, like mirroring and rotations, to prevent over training[63].

Ioffe and Szegedy, 2015, show that training time and generalization can be further improved by implementing **batch normalization**[64]. This involves normalizing training data within a mini-batch to a set mean and standard deviation. This invokes 2 more parameters that can be tuned, the mean and standard deviation, makes the learning more robust, and enables higher learning rates.

Learning involves using a mathematical optimizer on a mini-batch of training data over the neural network, and naturally many papers have been produced attempting to find better optimizers[26]. An optimizer using both momentum, adaptable learning rates, and weight decay is the ADAM optimizer, described by Krizhevsky et. al., 2012[65]. This optimizer is frequently used in image recognition, and has performed very well on the ImageNet dataset.

Finally, DNNs is an extremely fast growing and popular field of research with new articles being published prolifically, and so there is surely additional research available not covered in this literature study.

5 Data

In this chapter an overview of available data for the ROI is presented. This includes a description of the data, acknowledgement to the supplier of the data, and an explanation if not visualized.

5.1 Magnetic

From the Barents Sea ATLAS by NGU there is a comprehensive magnetic anomaly map available with derivatives, shown in figure 28. The combined data set is interpolated to a square cell of five-kilometres using a minimum curvature method[32].

5.2 Controlled source electromagnetic (CSEM)

CSEM data was supplied by Schlumberger for the ROI, but due to its proprietary nature can not be visualized in this thesis and only be used for *sanity checking* other findings.

5.3 Gravimetric

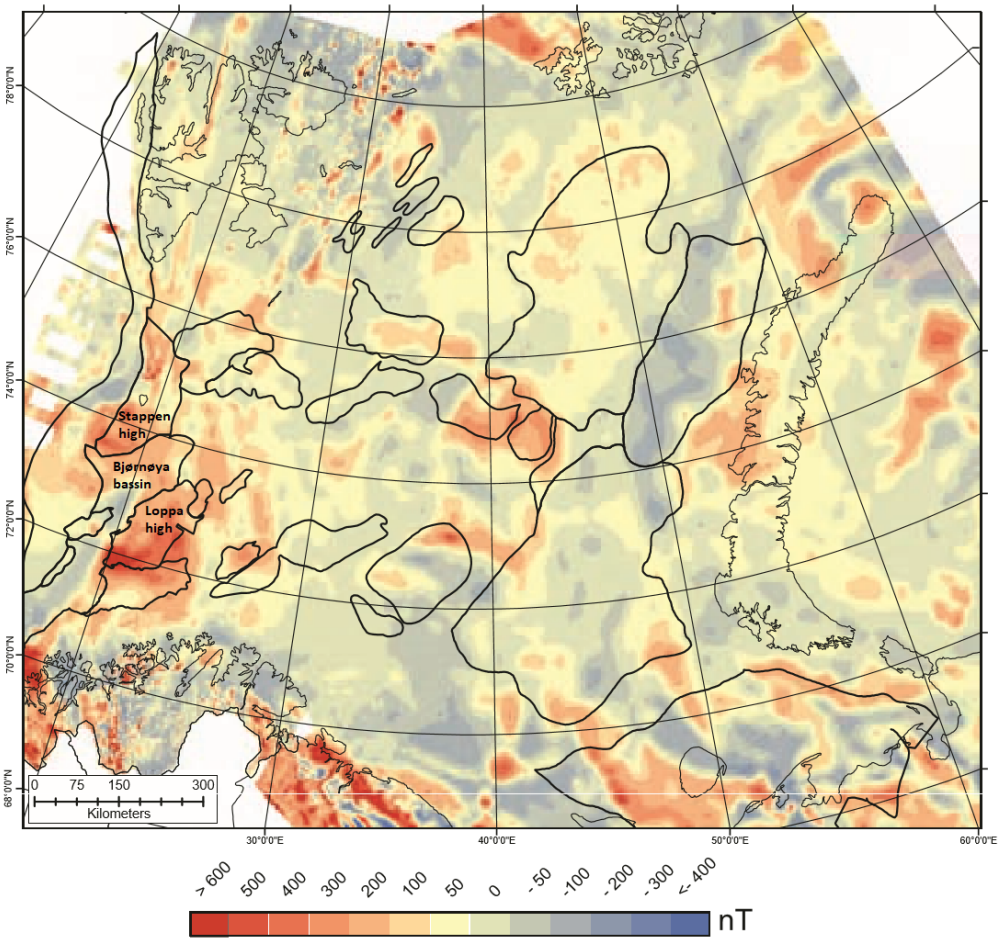
From the Barents Sea ATLAS by NGU there is also a comprehensive gravity anomaly map available with derivatives, shown in figure 29. On land, the combined data sets consist of terrain-corrected, Bouger anomaly values computed using a rock density of 2670 kg/m^3. For the oceanic area, the free air anomaly is retained[32].

5.4 Well logs

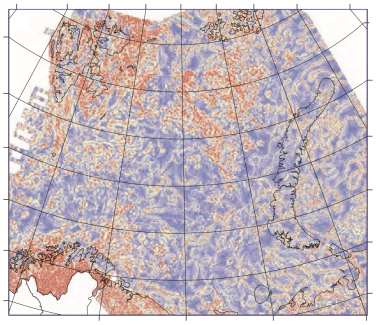
From the NPD fact pages, there are 3 openly available well logs that contain the relevant lithostratigraphy, shown in figure 16. These wells are described in table 1 along with lithostratigraphy, and can be used to tie seismic reflectors in the ROI to formation horizons. It is important to note that the lithology changes when moving west, especially crossing the Leirdjupet fault complex, where only a subset of the reflectors remain. This will be discussed in more detail in section 8.

Era	Period	Group	Formation	Wells		
				7321/7-1[66]	7321/8-1[67]	7321/9-1[68]
Mesozoic	Jurassic	Kapp Toscana	Stø	1999	1437	1379
			Nordmela	2022	1455	1417
	Triassic		Fruholmen	2039	1467	1424
			Snadd	2207	1626	1572
		Sassendalen	Kobbe	-	-	-
			Klappnyss	-	-	-
			Havert	-	(3362)	-
Paleozoic	Permian	Tempelfjorden	Ørret	-	-	-
			Røye	-	3398	-
Total well depth				3550	3470	1800

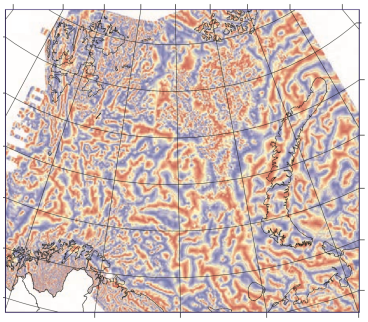
Table 1: Formation depths based on collected well data. (Adapted from Norkus, 2015[37])



(a) Magnetic anomaly map.

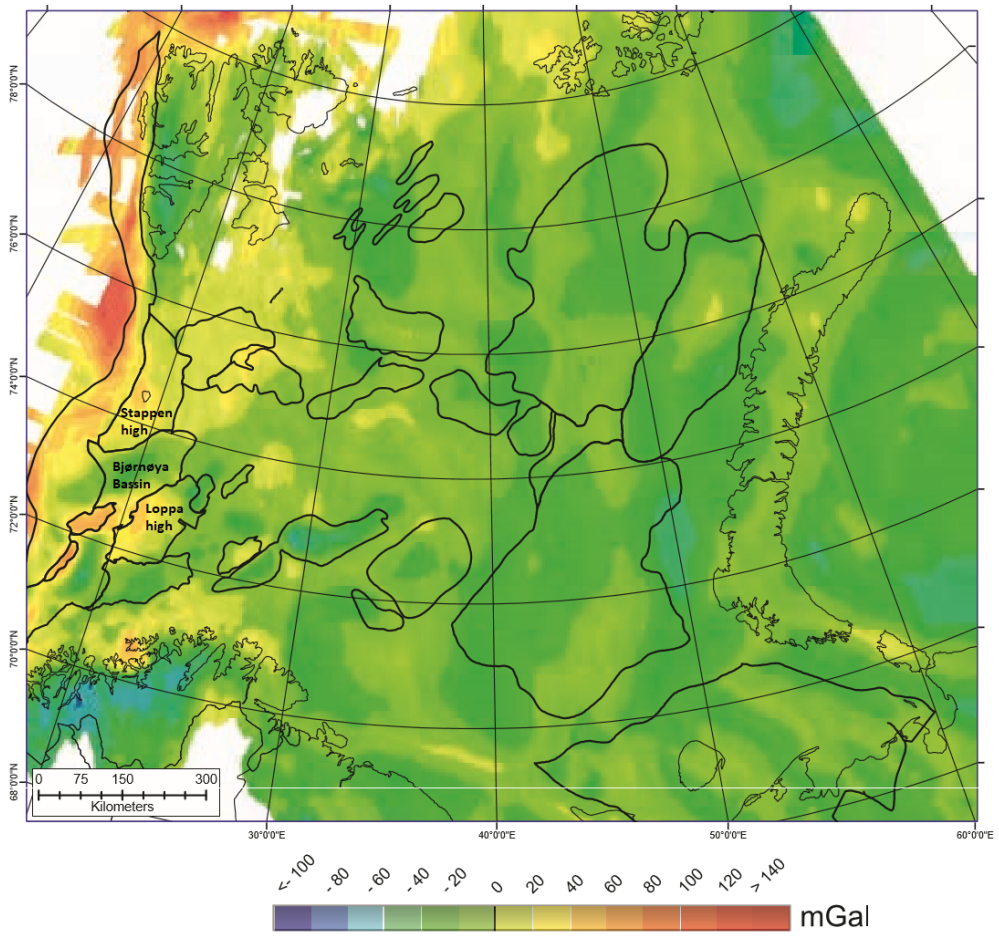


(b) Horizontal tilt derivative of the magnetic anomaly field.

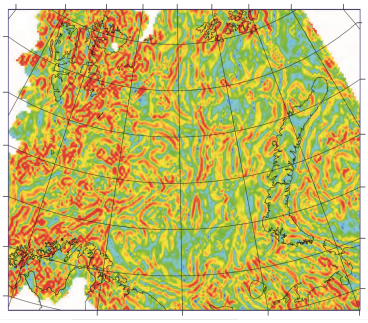


(c) Vertical tilt derivative of the magnetic anomaly field.

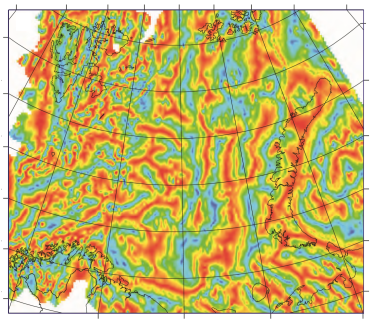
Figure 28: Magnetic data maps of the Barents and Kara Seas. (Presented by NGU in the Barents Sea ATLAS[32])



(a) Gravimetric anomaly map



(b) Horizontal tilt derivative of the gravitational anomaly field.



(c) Vertical tilt derivative of the gravitational anomaly field.

Figure 29: Gravimetric data maps of the Barents and Kara Seas. (Presented by NGU in the Barents Sea ATLAS[32])

5.5 Seismic

The seismic data is supplied by TGS, and is covering the region shown as a coloured surface in figure 30. The data is processed time migrated 3D-seismic covering approximately 3000 in-lines, 7000 cross-lines, and 4000 milliseconds of time data, sampled at 1 in-line, 2 cross-line and 4 ms intervals.

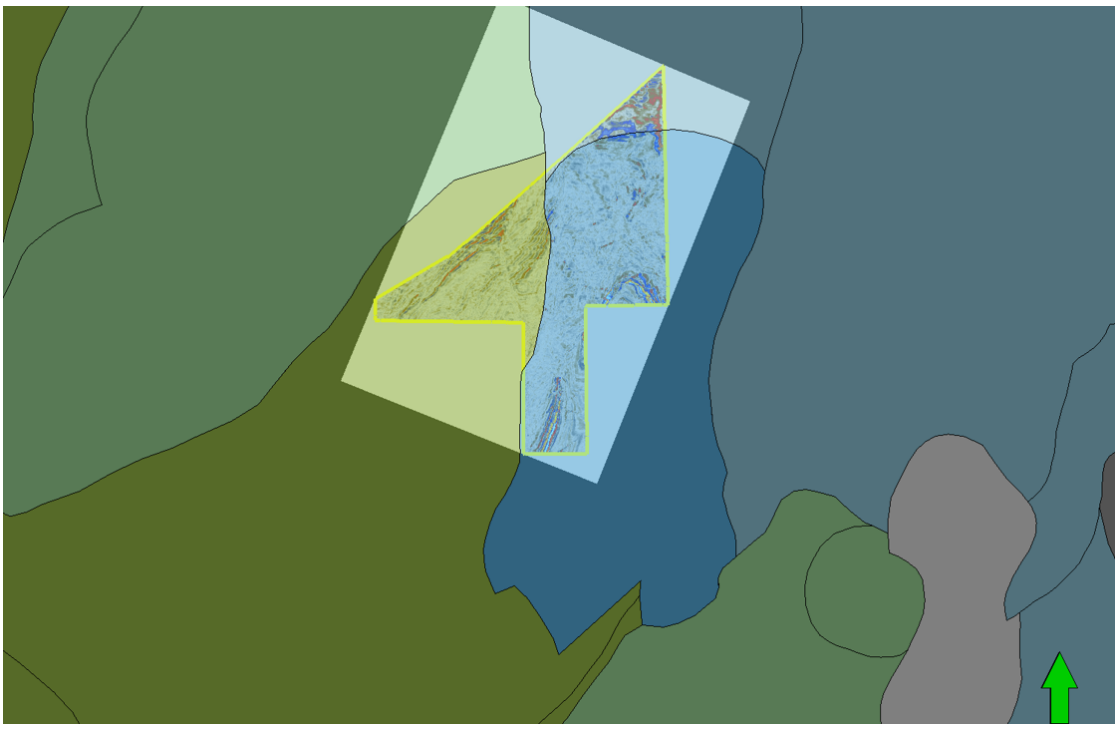


Figure 30: Location of the seismic survey provided by TGS within the greater ROI (Visualized in Petrel).

6 Method

In the following chapter a description and explanation of the selected exploration strategy and machine learning algorithm used for automation will be described. The work in section 6.1.1 and 6.2 is rephrased from work done in the project paper from Fall 2017[2].

6.1 Exploration strategy

From section 2.1, an exploration strategy is defined as needing 3 main steps. What to look for, where to look for it, and how to find it. A type of commodity that is favoured to exist in the Barents based on the considerations in sections 2.1, and 3.1 is the SHMS-MVT suite of deposits, with the major commodity being Galena(Pb) and Sphalerite(Zn). The ROI has been extensively described in section 3, so to fully assemble an exploration strategy the remaining step is to explicate the genetic model and exploration model.

6.1.1 Genetic model

Based on the genetic models described in section 4.1 the genetic model for potential deposits in the Bjørnøya Basin are of similar character to the large, Red Dog Pb-Zn deposit in Alaska[13], shown in figure 31. Like the Bjørnøya basin, there is a rapidly subsiding basin in a continental rifting setting. The Leirdjupet fault complex could have functioned as a conduit system hosting active hydrothermal circulation with feeding and venting in a deep and narrow sedimentary basin. The highly varying depositional environment of the formations in the basin, along with several phases of fault reactivation, could also cause secondary alteration mineralization in the deeper formations, like what is seen in the IVT deposits[42], also shown in figure 31.

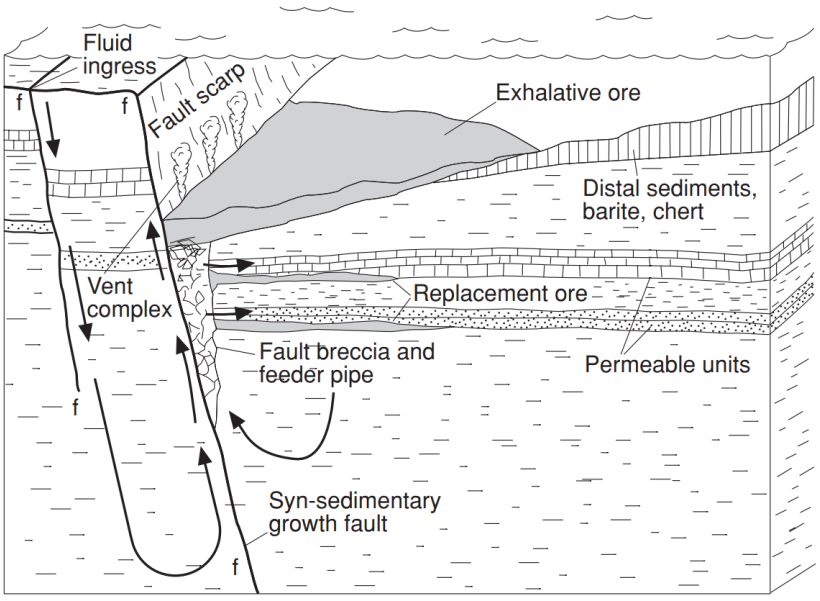


Figure 31: Block model for the Red Dog deposit in Alaska. (Adapted from Robb, 2014 [17])

The specific location for deposits is here likely to the west of the Leirdjupet fault complex, in the starved deep basin as a massive SEDEX type deposit in siliciclastic host rock, or as a replacement IVT deposits in carbonates. This was a rift zone between Greenland in the west and Norway in the East from the Triassic, and has high rates of influx of continental sediments. Here the absence of oxygen is evident, as it is a rich source rock throughout the Triassic, and coal lenses found in the entire region. In addition, the underlying Carboniferous-Permian layers are evaporites, desert sands, and carbonates, potentially giving very high salinity fluids, as well as high metal influx from continental sediments surrounding it. There is also chert on the edges of the Røye formation, like in the Red Dog deposit, pointing to hydrothermal fluid flow[13]. The specific formations of interest, based on their stratigraphy, thickness, and depositional period, are then the Snadd and Kobbe formations.

6.1.2 Exploration model

Taking the genetic model above as a basis, the main mineral constituents of the reservoir will be Sphalerite, Galena, and to a lesser extent Pyrite, Pyrrhotite and Barite. These minerals in the relevant setting are, as discussed in section 2.1, directly targeted using magnetic, gravimetric, EM, and to a lesser extent seismic methods. The deposits are normally also massive and stratified, making seismic methods useful for indirect detection, i.e. identification of the geological structures and framework.

Ideally, one would start with lateral delineation of the region of interest using gravimetric and magnetic methods. Secondly, seismic methods tied to relevant wells would be used to identify the geological framework and formations for the deposit, including the fault zone conduits. Finally, EM, and to a lesser extent seismic, methods would be used to directly target the sulphide deposits as low resistivity anomalies. This approach is illustrated as a list below, where methods directly registering a deposit is marked in **bold text**.

- **Further lateral delineation** - Gravimetric and magnetic data interpretation.
- Identify potential host rocks - Seismic-well tie, and Horizon tracking.
- Identify conduit system - Seismic data interpretation
- **Directly identify using seismic** - Seismic horizon amplitude analysis
- **Directly identify using CSEM data** - Negative anomaly probing

Unfortunately, due to limitations in availability of CSEM data and resolution in gravimetric and magnetic data, the approach is limited to direct and indirect targeting using seismic methods in this thesis, and will follow the process descrribed in the following list. In addition to the steps outlined, seismic interpretation requires significant work in familiarizing oneself with the data and region covered.

- Identify potential host rocks - Seismic-well tie, and Horizon tracking.
- Identify conduit system - Seismic data interpretation
- **Directly identify using seismic** - Seismic horizon amplitude analysis

6.2 Automation using Machine Learning (ML)

Based on the exploration model, and section 4.3, ML could be used to automate multiple, if not all, stages of the exploration process. The approach taken in this thesis, based on the available data, is limited to automating interpretation of seismic horizons and faults, using 3D-CNNs on processed seismic data.

Both algorithms for seismic data interpretation consists of the following 5 steps:

1. **Create and segment training and testing data** – Classify the relevant stratigraphy, or faults, one seismic cross-line to be used as training data. Then do the same on one seismic in-line to be used as independent test data.
2. **Prepare data** – Partition both dataset into sub-cubes surrounding every pixel in the original line.
3. **Batch and augment the training data** – Create mini-batches of 50 training samples and use basic spatial transformation as data augmentation.
4. **Training the algorithm** – Perform 100 training epochs over the chosen convolutional network (totalling over 10 000 000 training iterations) and validate performance on testing data
5. **Predicting new data** – Use the trained network to predict the rest of the seismic cube and inspect visually

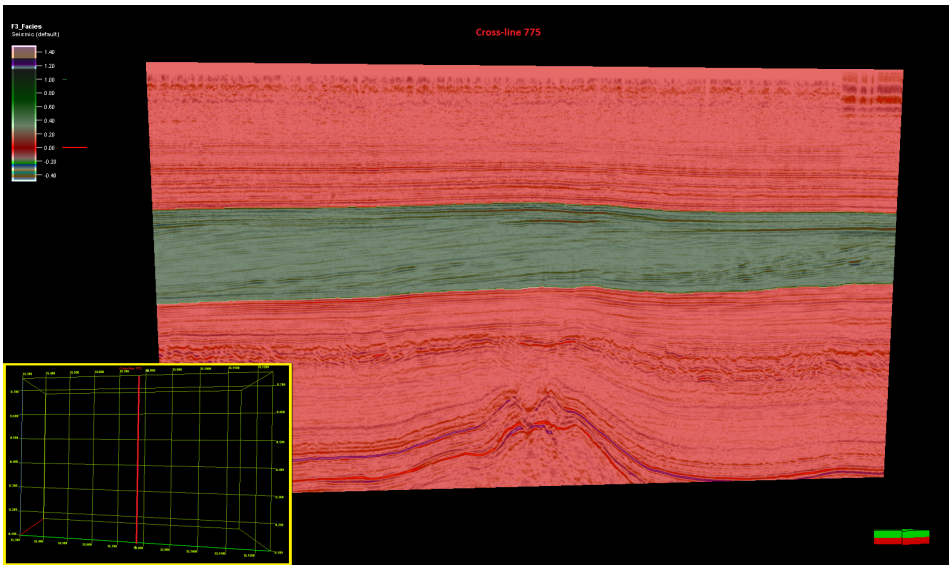
These steps are described in more detail in the rest of this chapter, along with descriptions of the two convolutional network architectures.

6.2.1 Data Engineering

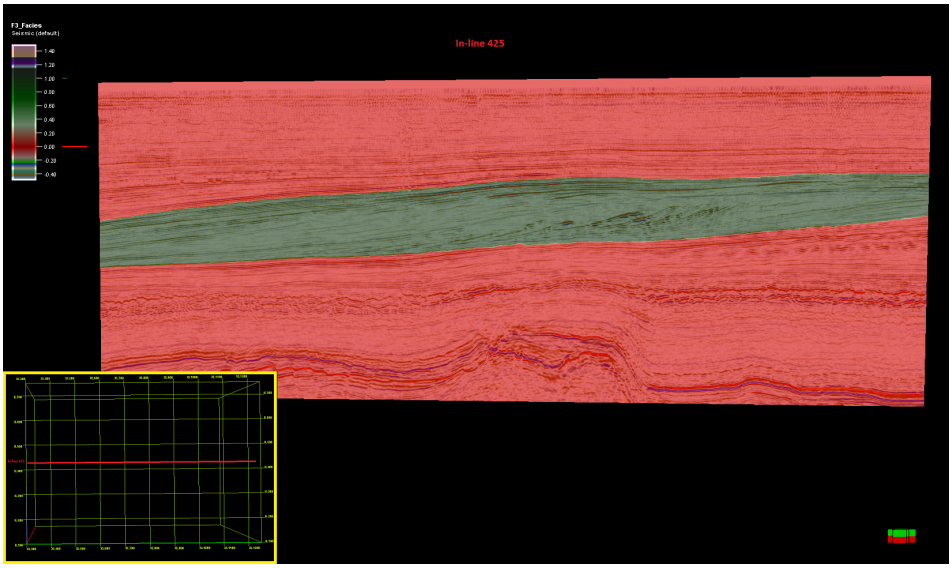
Data set creation and segmentation

The data is classified in advance into a positive and negative class, e.g. fault v. no-fault, or host formation v. not-host formation, and segmented into training and testing data. An example of such a classification is shown in figure 32 for seismic horizon tracking on data from the openly available F3 data set from the Netherlands. The green shaded area is the data labeled as a potential host rock, or *positive*, whilst the red shaded area is background, or *negative*, data.

As the training and testing data are spatially orthogonal to each other, they should be fairly independent, except for the data that are close to the intersection of the planes that show higher similarities. For the fault data, this same independence is achieved by segmenting the data into fault sticks and using separate sets of fault sticks for training and testing, preferably spatially orthogonal.



(a) Classified cross-line that is passed through the algorithm as training data.



(b) Classified in-line that is passed through the algorithm as testing data.

Figure 32: Input data to the machine learning network from the F3 data set.

Data set preparation

The input data is preparared by making a smaller sub-cube around a centre Volumetric Pixel (voxel) that should be classified. This enables the algorithm to use the data surrounding a voxel to determine the class. The sub-cube sizes were selected by empirical testing balancing enough input for classification and processing time. For seismic horizon tracking the sub-cube is 61x61x61 voxels, shown on the same F3 data set in figure 33, and for the fault classification is 61x61x125 voxels in order to target the longer vertical features of a fault.

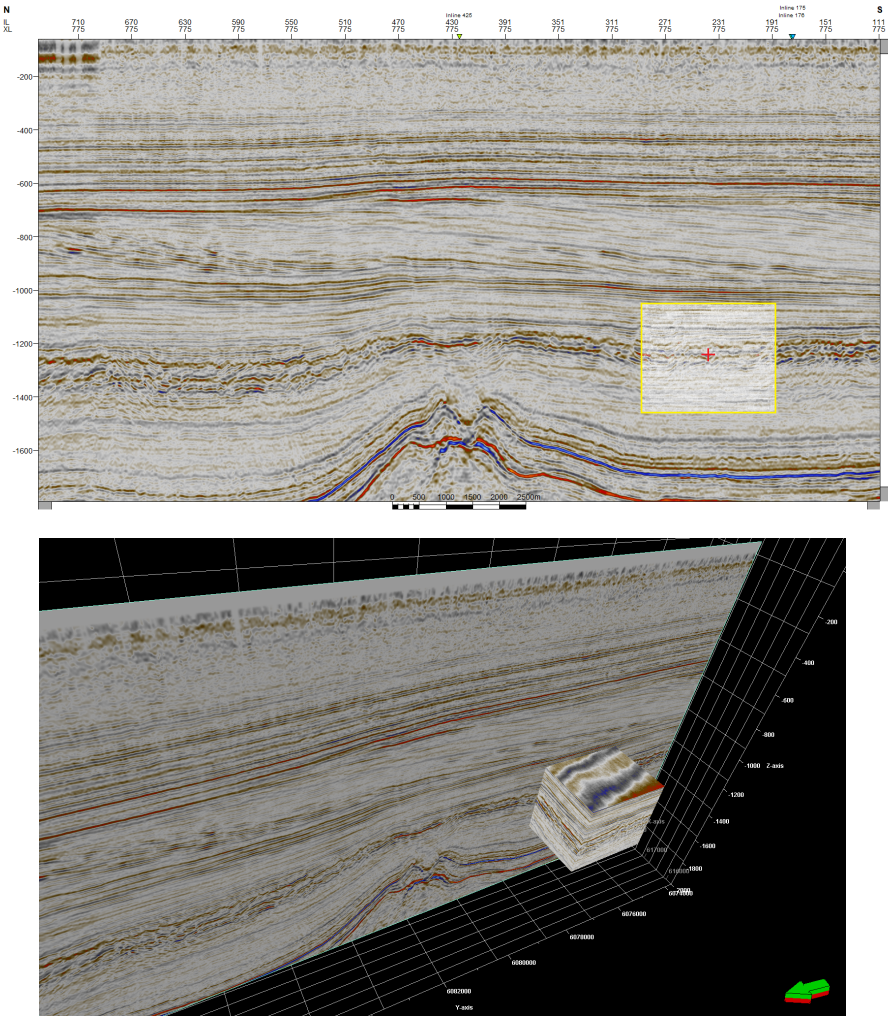
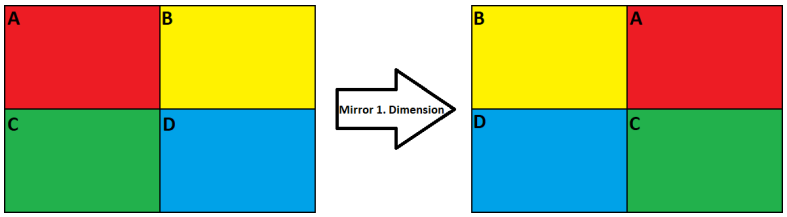


Figure 33: Visualization of the creation of a sub-cube from input data in the F3 data set.

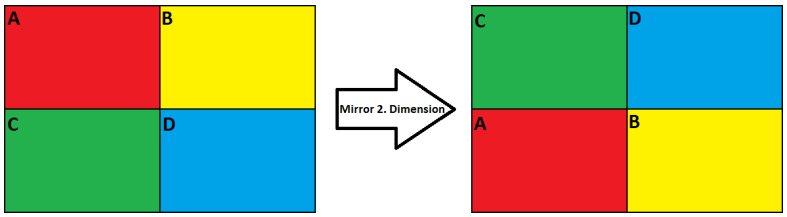
Batching and Augmentation

For the seismic horizon tracker, mini-batches are made by taking 50 sub-cubes without replacement, to make the training set as diverse as possible, and grouping them together. Each sub-cube individually within this mini batch is now, with a probability of 0.5, augmented with 90 degree horizontal rotation before the mini-batch is trained on.

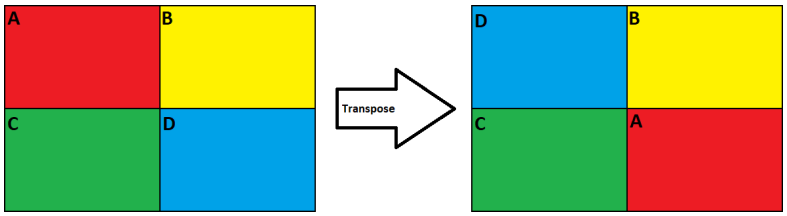
For the faults the batch creation is the same, but there is a more complex data augmentation scheme, as fault data sets are typically smaller than stratigraphy data sets. Here there are 3 operations that can be applied to the cube, each with a 0.5 probability of being implemented. The first two operations are horizontal mirroring operations in both spatial dimension, and the final is a transposition in the horizontal plane. All operations that are conceptually shown on a depth slice of a mini cube in figure 34, including the 90 degree rotation applied in the stratigraphy characterization. The final transformation, that is not shown, is simply mirroring in the depth direction and works as a simple upside-down flip, but was not implemented in the algorithms used as seismic data is not z-direction independent.



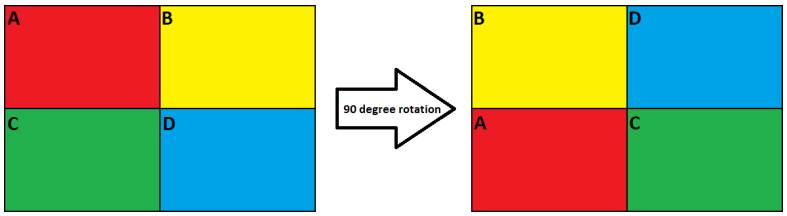
(a) In-line mirroring.



(b) cross-line mirroring.



(c) Horizontal transposition.



(d) 90 degree rotation.

Figure 34: Geometrical augmentation operations on a rectangle(depth slice of a sub-cube).

For the fault classification there may actually be more than one data augmentation operation applied to each sub-cube. This leads to more augmentation states, as shown in table 2, summarizing the possible combinations of 2 horizontal plane operations, with probabilities in parentheses.

This also holds true for the third operation, but is left to the reader to calculate.

	Mirror in-line (0.5)	Mirror cross-line (0.5)	Transposition (0.5)
Mirror in-line (0.5)		180°rotation (0.25)	-90°rotation (0.25)
Mirror cross-line (0.5)	180°rotation (0.25)		90°rotation (0.25)
Transposition (0.5)	-90°rotation (0.25)	90°rotation (0.25)	

Table 2: Combinations and probabilities for 2 steps of augmentation operations.

6.2.2 Architecture

The architecture of both networks are made up of distinct blocks combined to make a deep overall structure, with specialized hyperparameters for their particular purpose. For the seismic horizon tracker this is shown in figure 35 along with individual subcube dimensions. For the fault model it is slightly different, as it has another convolutional layer to accommodate the added size of input data in the depth direction, and 20 neurons in stead of 50 in the convolutional and dense blocks, as it is looking at a less complex feature. The hyper parameters, like the number of layers and number of neurons, in the networks are set using empirical testing and tuning in order to balance accuracy and processing time. The different blocks used in this thesis are shown in more detail in appendix B, and the full code is available on GitHub[69][70].

Training The networks were trained using the ADAM optimizer with standard parameters, and consistently converged on the validation data before completing 100 training epochs. This usually took a few hours.

Prediction After completing training, the networks predicted on a section of the seismic surveys, with results presented in section 7. This usually took about a day.

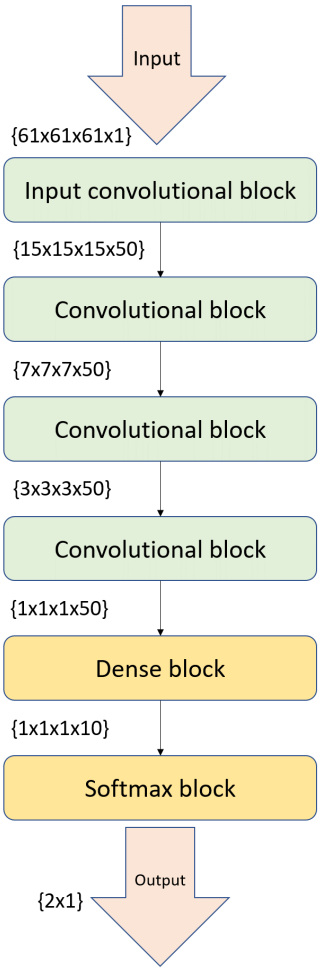


Figure 35: Network structure for seismic horizon tracking.

7 Results

In the following chapter the received results from the exploration strategy and machine learning algorithms will be presented. All visualization are made using the Schlumberger seismic visualization software Petrel.

7.1 Exploration Model

As described in section 2.1 the utilized exploration model consists of seismic well tie, horizon tracking, fault identification, and seismic amplitude analysis. The seismic data used in this thesis had already been tied to nearby wells and depth adjusted, and so the well tie step is omitted.

7.1.1 Horizon tracking

Based on available wells and interpretations from the area the relevant formations that could be tracked were the top of the Stø formation, the top of the Snadd formation, and the bottom of the Snadd formation(or bottom of the Kobbe formation), shown in figures 36 and 37. To be able to do this interpretation, significant time needed to be spent understanding the seismic survey and exploration wells from the area as well.

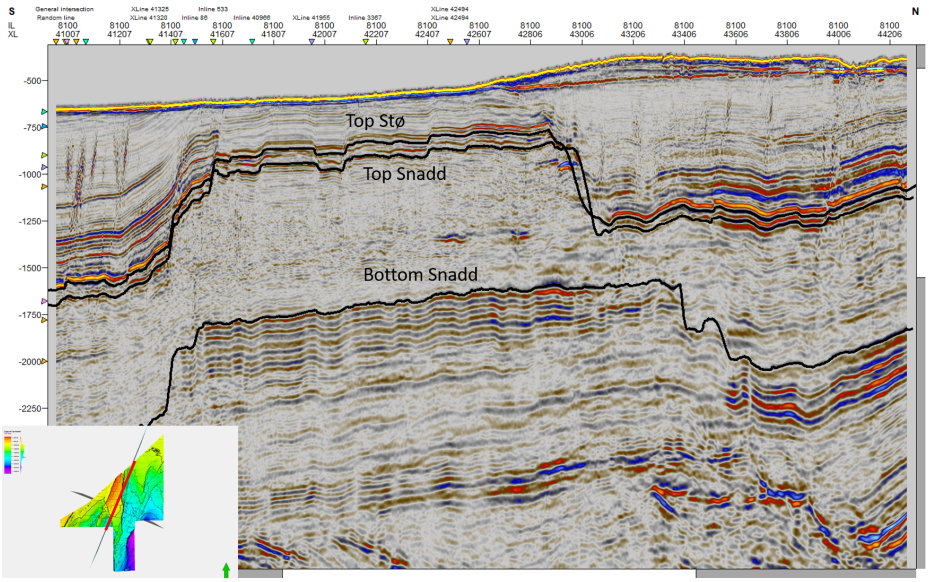


Figure 36: Inline number 8100 from the Sesmic survey provided by TGS. Horizons interpreted to be top Stø, top Snadd, and Bottom snadd are visualized

7.1.2 Fault identification

Faults were cutting through the formations were identified and studied in the region around the Leirdjupet fault complex, shown in figure 38.

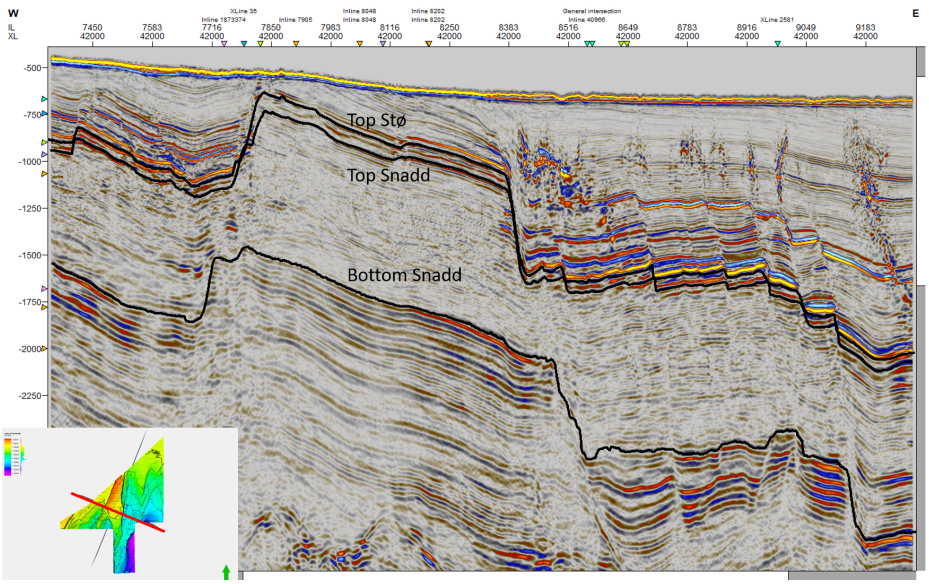


Figure 37: Inline number 42000 from the Sesmic survey provided by TGS. Horizons interpreted to be top Stø, top Snadd, and Bottom snadd are visualized

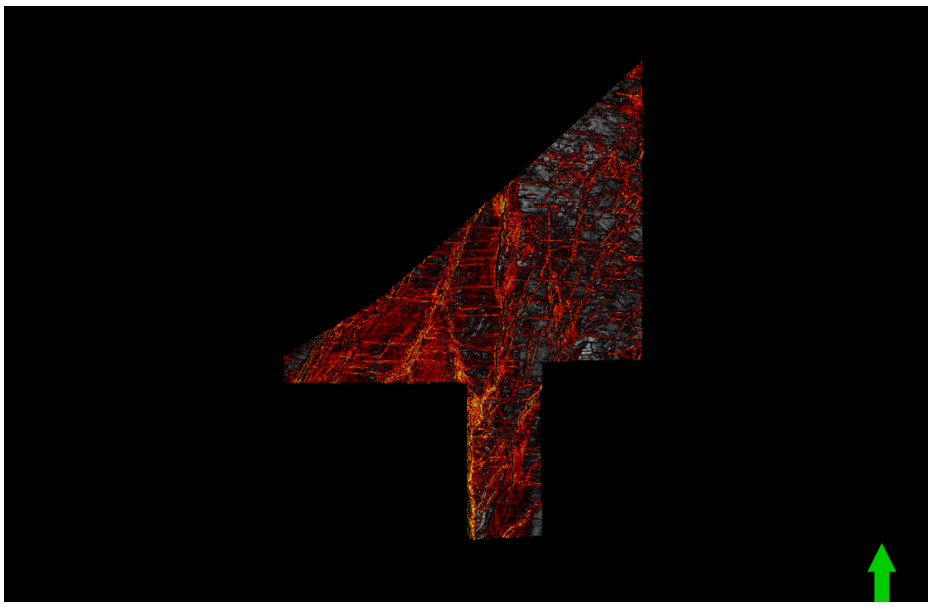
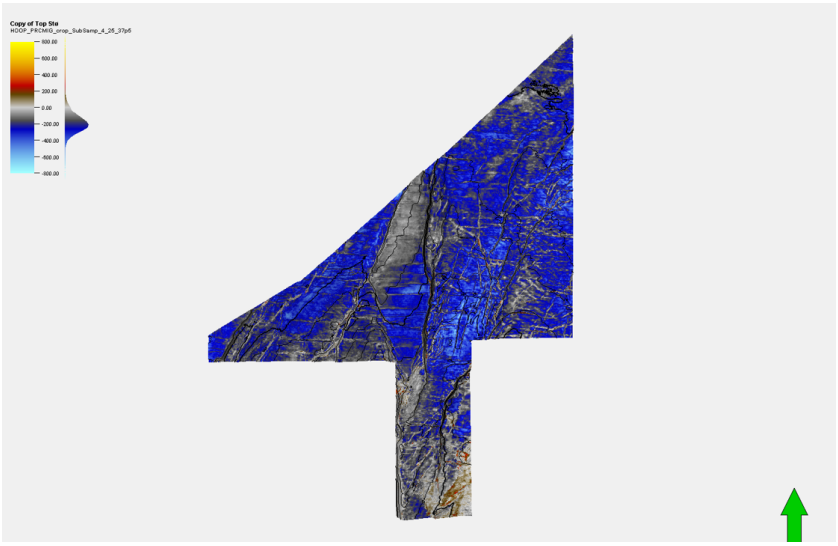


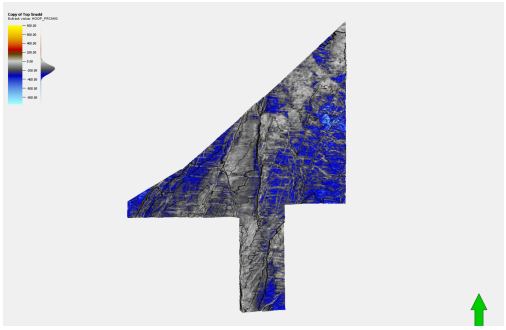
Figure 38: Visualization of a fault trends on the Snadd surfacein Petrel.

7.1.3 Amplitude analysis

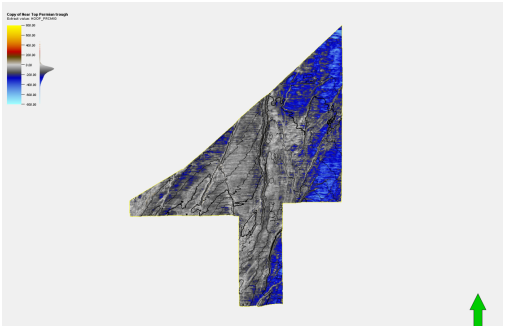
Once the top Stø, top Snadd, and bottom Snadd(or Kobbe) horizons were tracked on multiple inlines and cross-lines, a surface was interpolated and an amplitude map was made, shown in figure 39.



(a) Top Stø



(b) Top Snadd



(c) Bottom Snadd

Figure 39: Visualization of created amplitude maps on the interpreted surfaces in Petrel.

7.2 Automation using Machine Learning (ML)

To reduce processing times, the seismic survey was cropped to a representative region in the section surrounding the Leirdjupet fault complex, as shown in figure 40. By using the method described in section 6.2, the Snadd formation was attempted tracked using the seismic horizon tracker, and fault tracking was attempted using the fault identifier, using training data from the centre in-line and cross-line in the seismic survey.

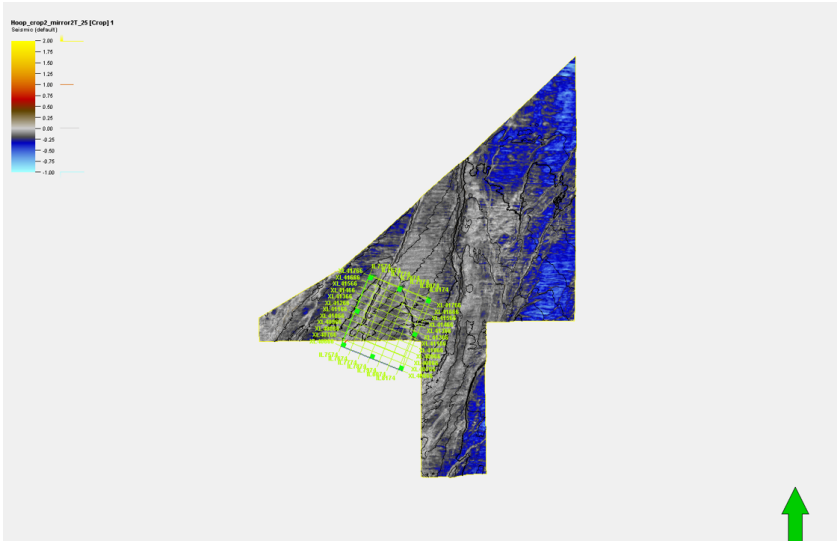


Figure 40: Visualization of cropped seismic survey region in green, imposed on the ful seismic survey.

7.2.1 Horizon tracking

Results from the seismic horizon tracker is shown in figures 41 and 42.

7.2.2 Fault identification

Results from the fault identifier is shown in figures 43 and 44.

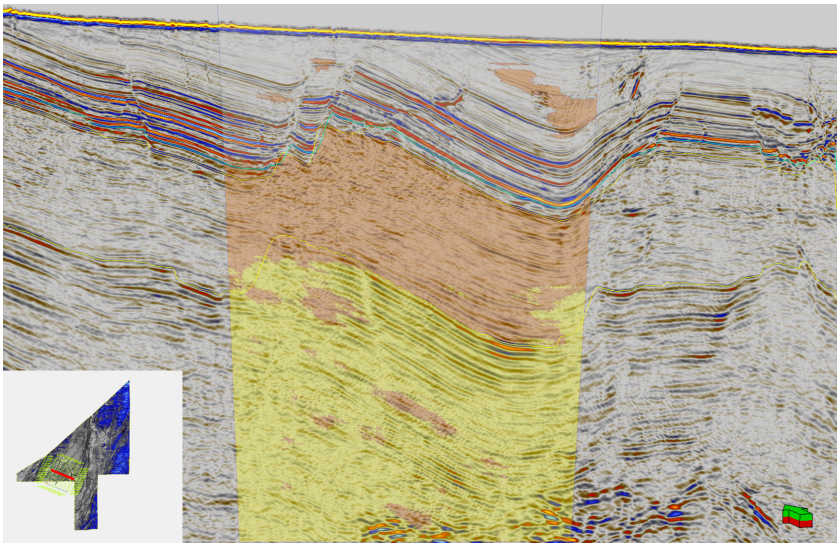


Figure 41: Automatic interpretation of an in-line, with the orange section being the Snadd formation. In-line shown in red in bottom left corner.

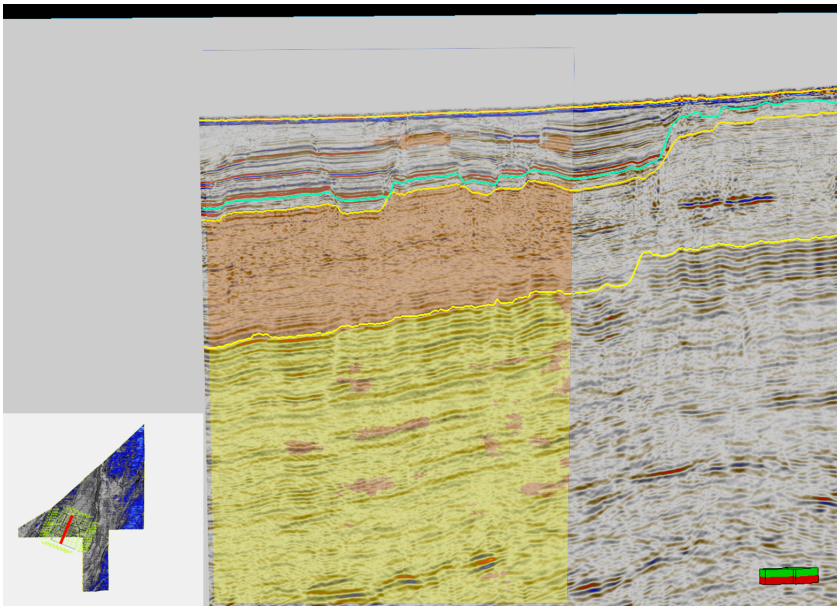


Figure 42: Automatic interpretation of a cross-line, with the blue sections illustrating faults. Cross-line shown in red in bottom left corner.

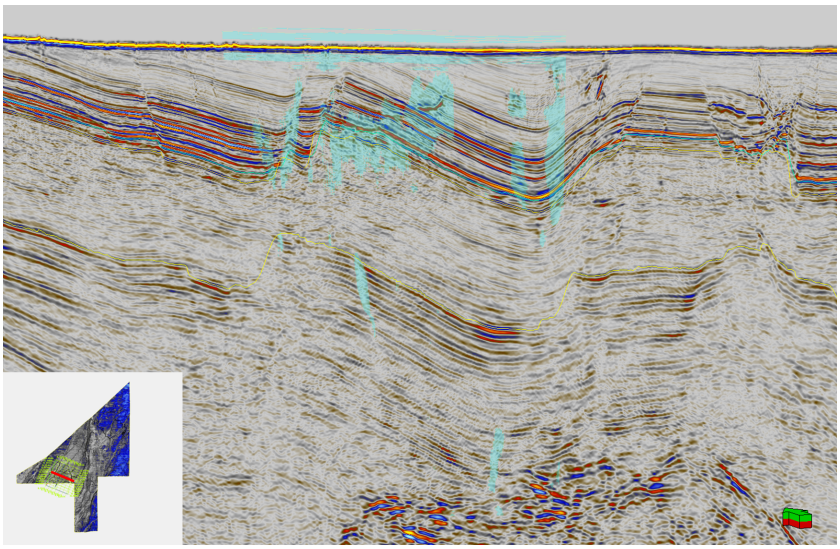


Figure 43: Caption

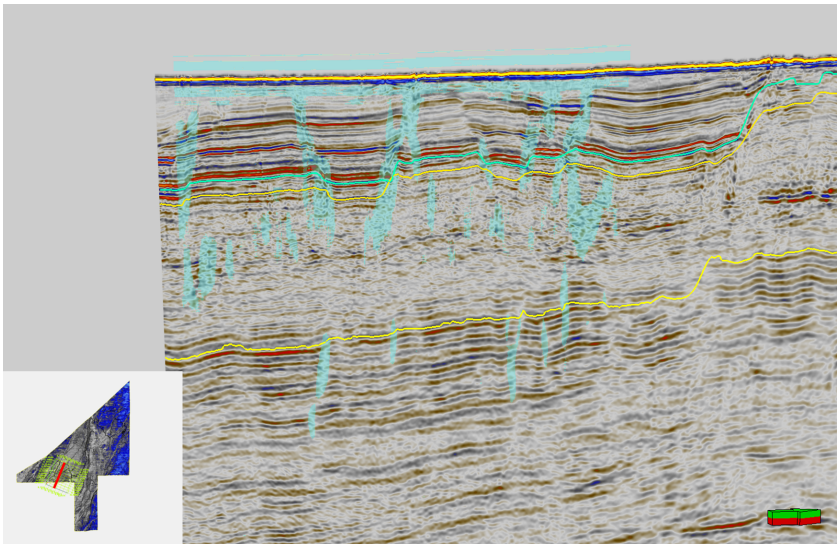


Figure 44: Automatic interpretation of a cross-line, with the blue sections illustrating faults. Cross-line shown in red in bottom left corner.

8 Discussion

In the following chapter an analysis of the received results, possible sources of error, and reflections around the approach taken are presented. This includes uncertainties of the exploration strategy, and inspection of the trained machine learning networks. The work in this chapter is heavily building on the discussion of the work done in the project paper from Fall 2017[2].

8.1 Exploration Strategy

Section 7.1 shows that the Snadd formation has a large vertical extension in the entire section, and is present in the basin on both sides of the Leirdjupet fault complex. There are various secondary and tertiary faults crossing the Leirdjupet, suggesting several phases of reactivation as hypothesized, and again points to the potential of conduit systems in the area. This long lived conduit would accommodate extended deposition of the SEDEX type on the west side of the faults, or more likely, IVT deposits stretching further laterally around the faults. From the amplitude analysis on the top Snadd surface there does not appear to be a high reflector on the west of the fault system, and is either due to a loss of energy, as is a problem with the method, or a lack of a deposit. On the other hand, there does appear to be some high amplitude sections, especially to the east of the faults, that could be IVT deposits, in accordance with the fault reactivation.

There are many uncertainties connected to designing an exploration strategy and analyzing its results, and the more pressing ones, along with the rational for assumptions made, are discussed below.

8.1.1 Deposit type

From section 4.1 it is known that an SHMS deposit needs a heightened geotherm causing fluid circulation, a conduit system, and a chemical or physical barrier that causes metal precipitation.

In the Barents Sea there is a long history of tectonic activity causing rifting and collisions, again giving rise to a heightened geotherm. Further, the depositional environment has changed drastically several times causing deposition of highly contrasting sediments and therefore chemical barriers in the subsurface, including pH(sandstone/shale-carbonate) and fO_2(oxidized sandstones-reduced shales). In addition, there have been periods when the overlying sea water is expected to have contained reduced sulphur, potentially accommodating sulphide mineralization. Finally, there has been severe rifting, faulting and reactivation in the region throughout giving rise to conduit systems for abyssal brines.

Based on this, it is clear that the Barents Sea has had the conditions needed to facilitate the creation of SHMS deposits of the SEDEX-MVT suite, making it the focus suite in the exploration strategy. However, as seen from cross-section 8 in Appendix A, there *is* evidence for magmatic activity to the west of the Bjørnøya Basin, and so the SEDEX-VHMS suite could have been investigated further as a potential type of deposit, giving different mineralogies and a different exploration model.

8.1.2 Region of interest

Moving forward with the SEDEX-VHMS suite, the most promising areas were determined to be proximally to the rift shoulders inside either the Bjørnøya basin or Hammerfest basins, and the Bjørnøya basin was chosen as it had more documented tectonic activity in the literature studied, and available seismic data from former Österreichische Mineralölverwaltung (OMV), TGS, and Schlumberger. This created a clear bias in this paper *against investigating* the Hammerfest basin in more detail, and it could prove to have been a more attractive ROI under a different study.

8.1.3 Genetic model

Based on the stratigraphic sequence, tectonic activity, and fault zones, the Red Dog deposit was determined a close proxy for a deposit in the Bjørnøya Basin. The Red Dog deposit is hosted by Black siliceous shale, like the ones found in the Snadd formation, but also contains *chert*. The chert points to hydrothermal fluid flow, and is only found in the Permian formation Røye in the Barents Sea. This does not exclude the existence of hydrothermal fluid circulation in the Leirdjupet fault complex, and the Red Dog deposit as a viable proxy for a deposit in the Bjørnøya basin, but it is an indication that there might be differences that could be accounted for moving forward, and that more formations should be considered as possible host rocks.

In addition there has been observed a clear correlation between the Red Dog deposit, and close by IVT deposits for shallow lagoonal environment), and MVT deposit for open spaces in carbonaceous strata. This leads to the argument that more emphasis could have been put on modelling the carbonate hosted deposits in the region, particularly in the upper Permian deposits. These deposit do however occur in less massive geometries, making them more difficult to detect, and leading to omission in the exploration strategy.

8.1.4 Exploration model

The genetic model above should yield a sulfide deposit that is dominated by Iron sulfides, Galena(Pb), Sphalerite(Zn), and mior Barite(Ba), with a large lateral extent. From the mineralogical standpoint it was assumed to be lead rich due to a high influx of crustal sediments, but if the sediment flux was lower than assumed, the deposit could have been dominated by more Iron and less Lead, and so could have had a different profile in geophysical data. Primarily this would make the seismic amplitude less, when considering direct targeting by seismic, but also make magnetic and EM methods more efficient as exploration tools.

Seismic data likely gives the best resolution and, together with validation from EM, is suggested as ideal for exploration in this thesis. This is in many ways the richest data types in the area, and due to petroleum exploration covers the basins of interest. However, the total data collected over the Barents Sea includes seismic, EM, gravimetric and aeromagnetic. Of these methods only magnetic, gravimetric and EM can detect sulphide mineralization directly, but their resolution is less than other methods, like seismic. Seismic methods on the other hand are very comprehensive and useful in detecting structures and systems that facilitate the deposits. This leads to another potential Exploration strategy where one could use seismic data as the primary data set to look for structures, and validate whether or not there is a deposit in these structures using aeromagnetic cross-correlation. This was not chosen in this thesis due to the lack of high resolution magnetic data.

Finally, there is the issue non-geological noise in data, e.g. measurement noise or human error in interpretation, causing a higher uncertainty when exploring a potential deposit. This should be kept in mind. and is particularly prevalent in seismic data, and despite extensive and complex seismic processing, effects the accuracy of the data, and so the interpretations and conclusions drawn from it. This noise is also present in other data sources used, like well logs, but will be less prevalent.

8.2 Automation using Machine Learning (ML)

Section 7.2 shows that the ML algorithm manages to follow the horizons to some extent, but bleeds a little across the horizons. This is likely because the data is overfit, or because the problem is more complex than what the current neural network can model. An observation that can be made, is that the horizon tracker performs well, even when crossing faults, which is not the case for conventional auto-trackers in Petrel. When it comes to the fault tracking, the performance is slightly worse, and only obvious faults are identified. This can again be attributed to overfitting, and likely will improve with larger input data sets, as the number of faults on a given line is normally quite limited. As discussed, above there are also inherent inaccuracies in the data, but the main reason for moderate performance of the algorithm is likely imperfections in approach and architecture rather than in the data.

The two studies covered comparing different machine learning methods used in geoscience both concluded that RF was the best method and that ensemble methods overall performed well. However the data types used in these studies were either poorly defined, as *well log data* with few features, or as *remote sensing data* likely being aeromagnetic and gravity data. Both studies also note that neural networks likely do better with more complicated tasks in higher dimensional spaces, like the one attempted in this thesis, and so the method applied should not be abandoned.

Another important aspect of CNNs is choosing the right hyperparameters, like number of layer, number of neurons in each layer, how they are connected, etc. This was done manually by empirical test in this thesis, but could also have been done using an optimizer, potentially improving the results, or even implementing representation changes, like auto-encoders[71].

All in all, the performance can likely be improved, which is further discussed in section 10, but the results are highly encouraging.

8.2.1 Feature visualization and Layer-wise Relevance Propagation (LRP)

A final issue with machine learning methods is that it often comes off as a *black box*, with questionable accuracy. To gain understanding of neural networks, several studies have been done on visualizing what happens in the process, and somewhat unwrapping this *black box*. Two of the most popular techniques involve visualizing the features trained in the network[72], and inverse engineering pictures to find out what triggers different nodes for classification[73]. This has been modified and implemented on the architectures used above to gain understanding of networks and their classification, with source code available on GitHub[69][70]. This is shown in figure 45, visualizing 8 of the final features used for classification in the horizon tracker and figure 46, showing what triggers a given class for the horizon tracker network.

Note that the filters are 3-dimesional, and so need to be sliced for 2 dimensional representation. This is shown on the right hand side of figure 45. From the feature visualization, one can see that the algorithm is indeed looking for smaller localized features, due to the waffle pattern, but that it is overfit as it has very similar features (at least based on the slices visualized).

When looking at what triggers a classification, it is obvious that it is based on actual geological features, as it triggers on the horizons moving through the sample image, supporting the statement that the method should not be abandoned. Again, here the input has been visualized, as in figure 46.

Additionally, a similar exercise was done for the fault identifier, shown in figure 47, showing that

it does focus on the centre of the pixel where a fault would be located, also giving faith in the efficacy of this network.

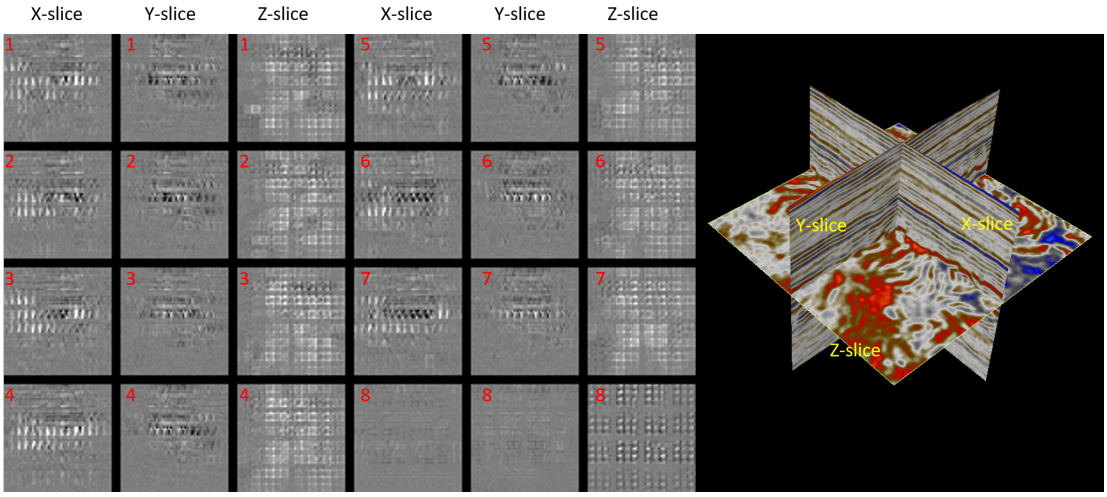


Figure 45: Visualization of 8 of the final filters in the seismick horizon tracker network. Conversion from 3D-filters to 2D-slices illustrated on the right hand side.

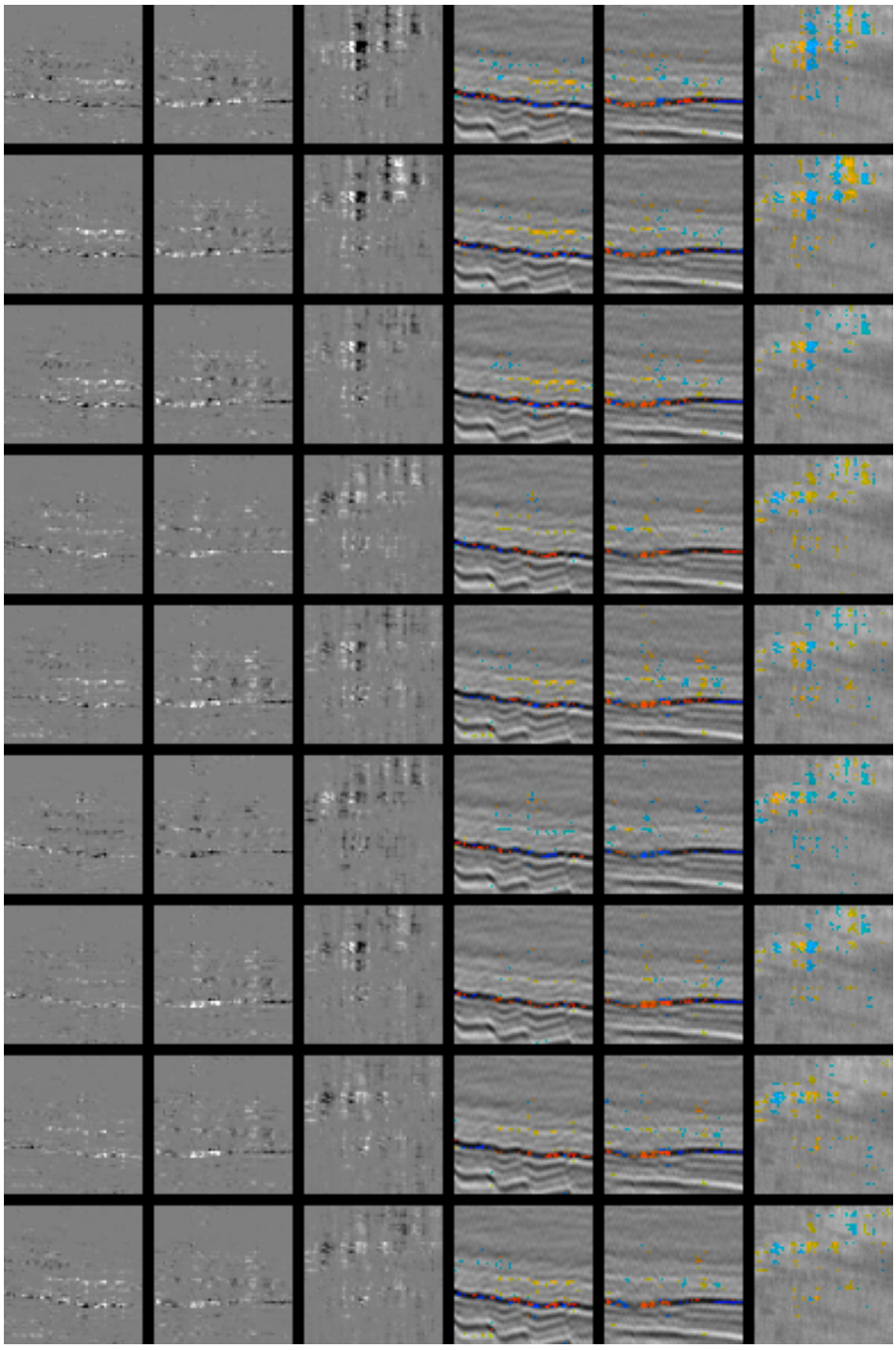


Figure 46: Left three images: Visualization of the triggering voxels from an arbitrary input sub-cube. Right three images: Overlay of the triggering pixels on the original input. The illustration is here done in the same way as shown in the figure above.

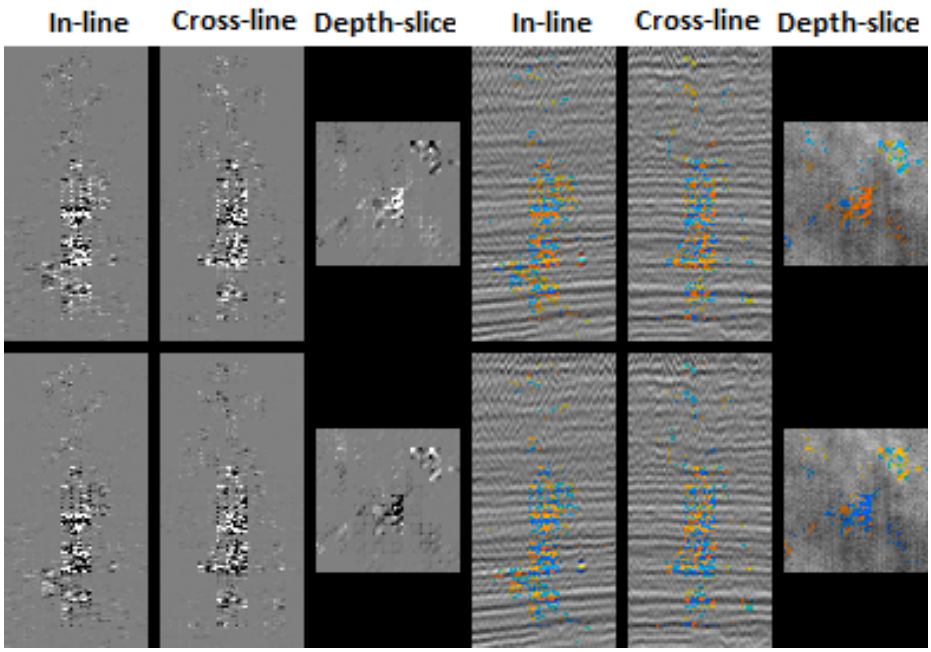


Figure 47: Left three images: Visualization of the triggering voxels from an arbitrary input sub-cube. Right three images: Overlay of the triggering pixels on the original input. The illustration is here done in the same way as shown in the figure above.

9 Conclusion

To conclude, this chapter will answer the research questions posed in the introduction.

- Choose a specific SHMS-deposit type anticipated to be prevalent on the Barents Sea.
- Define an appropriate ROI with the necessary mechanisms for these SHMS-deposits.
- Design an exploration strategy for this SHMS-deposit type in the ROI.
- Find a way to automate the steps in the exploration strategy by using machine learning.
- Apply the exploration strategy, with the automated steps where possible, in the ROI to create the final favourability map.

The Barents Sea contains the necessary genetic components for both suites of SHMS deposits, but based on the prevalence of carbonates, and scarcity of volcanics, the genesis of the SEDEX-MVT suite in particular is expected to be favoured.

The most likely location of such SEDEX-MVT deposits is proximally to the intracratonic rift basin shoulders of the Bjørnøya basin firstly, and the Hammerfest basin secondly.

The exploration strategy used was based solely on seismic data, and consisted of indirect detection using potential host formation tracking, fault detection, and direct detection by seismic amplitude analysis.

Both the tracking of the host formation and fault detection, was automated using 3D-CNNs, showing promising results.

Finally, the exploration method showed that the necessary geological framework to accommodate a deposit is in place, with the most likely type being an IVT, but further direct detection methods are needed for a conclusive answer. When superimposing the results, a favourability map can be created, and is shown in figure 48.

This shows clear primary and secondary faults in the Leirdjupet complex, and some local amplitude highs (shown in orange) that point towards the need for further studies of the area.

Figure 48: Favourability map for SEDEX-MVT suite deposits shown with surrounding the same features visualized as in the ROI delineation.

10 Further work

The further work is mainly occupied with obtaining more geophysical data for direct detection, and finding more applications for and improving on the machine learning algorithm.

10.1 Utilize more geophysical data

An alternative to this approach could be combining more than one data set in the training data, having the algorithm analyze, for example, both seismic and EM data at the same time, giving more accurate results. The downside to this being that it becomes incredibly data and computationally heavy. Additionally, magnetic or EM data with sufficient data can be hard to acquire.

10.2 Improve the Machine Learning (ML) algorithm

Results from peripheral fields, like medical imaging, suggest that more complex structures, like RNNs, U-nets, etc., can be used to an advantage[9].

Semantic segmentation is an alternative method of segmenting large datasets and has shown to be useful in increasing prediction speeds[74]. This could also be implemented to improve the network.

Alternatively one could imagine using ensemble methods as suggested in some of the literature comparing machine learning in geology to possibly obtain improved results results.

10.3 Infer the existence of SHMS on the Norwegian shelf

As the exploration strategy and algorithm in this thesis has been shown to be promising, there could be additional studies on deposit mapping on other sections of the Norwegian continental shelf.

10.4 Testing algorithm on other regionalized variables

Finally, one could imagine implementing the algorithm to train and predict on other regionalized variables, like soil parameters(pH/Eh-conditions), mapping of hydrocarbon systems, or even mapping of ground water pollution.

List of Figures

List of Abbreviations

AI Artificial Intelligence. 13

BHT Broken Hill Type. 28

CAT Computed Axial Tomography. 1, 34

CNN Convolutional Neural Network. i, 34, 35, 43, 57, 61

CSEM Controlled Source ElectroMagnetic. 31, 37, 42

DNN Deep Neural Network. 34–36

E&P Exploration and Production. 1

EM Electromagnetic. 1, 9, 31, 34, 42, 55, 63

EMAG Earth Magnetic Anomaly Grid. 31, 64

GBT Gradient Boosted Trees. 34

IGP The Department of Geoscience and Petroleum. i

IP Induced Polarization. 32

IVT Irish Valley Type. 4, 29, 41, 42, 54, 55, 61

MIT The Massachusetts Institute of Technology. i, ii

ML Machine Learning - Ability to acquire knowledge by extracting patterns from raw data[8]. i, 1–3, 10, 12–15, 34, 35, 43, 57, 64

MOR Mid-Ocean Ridge. 1

MRI Magnetic Resonance Imaging. 1, 34

MVT Mississippi Valley Type. 4, 5, 8, 20, 28–32, 41, 54, 61, 62, 65

NGU The Geological Survey of Norway. 22, 37–39, 64, 65

NOAA The National Oceanic and Atmospheric Administration. 31, 64

NPD The Norwegian Petroleum Directorate. i, 9, 21, 37, 64

NTNU The Norwegian University of Science and Technology. i, ii, 1

OMV former Österreichische Mineralölverwaltung. i, 55

PAOB Pan-African Orogenic Belt. 4, 21

PCA Principal Component Analysis. 11, 34

PET-CT Positron Emission Tomography–Computed Tomography. 34

PGE Platinum Group Elements. 4

ReLU(x) Rectified Linear Unit. 14

RF Random Forests. 34, 57

RNN Recurrent Neural Network. 34, 63

ROI Region of Interest. i, 1–3, 16, 21–23, 25, 26, 28, 37, 40, 41, 55, 61, 62, 64, 65

SEDEX Sedimentary Exhalative. 4–6, 8, 20, 28–32, 42, 54, 55, 61, 62, 64, 65

SGD Stochastic Gradient Descent. 10

SHMS Sediment Hosted Massive Sulfide. i, 1–6, 16, 18, 20, 28–30, 32, 41, 54, 61, 64

SOM Self Organizing Maps. 34

SVD Singular Value Decomposition. 11

SVM Support Vector Machines. 34

tanh(x) hyperbolic tangent function. 14

TGS TGS-NOPEC Geophysical Company ASA. i, 33, 40, 55, 65

UCSD the University of California, San Diego. 32, 64

USGS the United States Geological Survey. 31, 32

VHMS Volcanic Hosted Massive Sulfide. 4–6, 8, 28, 29, 54, 55, 64

VMS Volcanogenic Massive Sulfide. 4, 5, 64

voxel Volumetric Pixel. 45

Appendix A Cross sections for the Norwegian continental margin and the Barents Sea

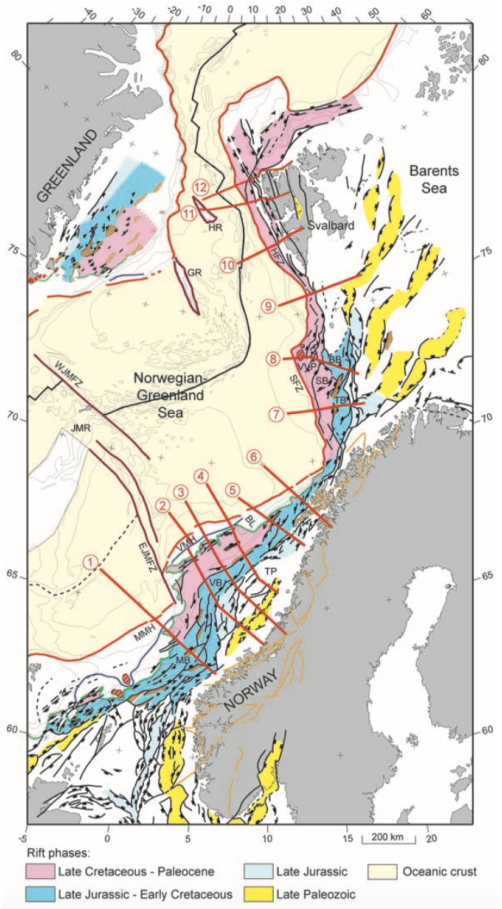
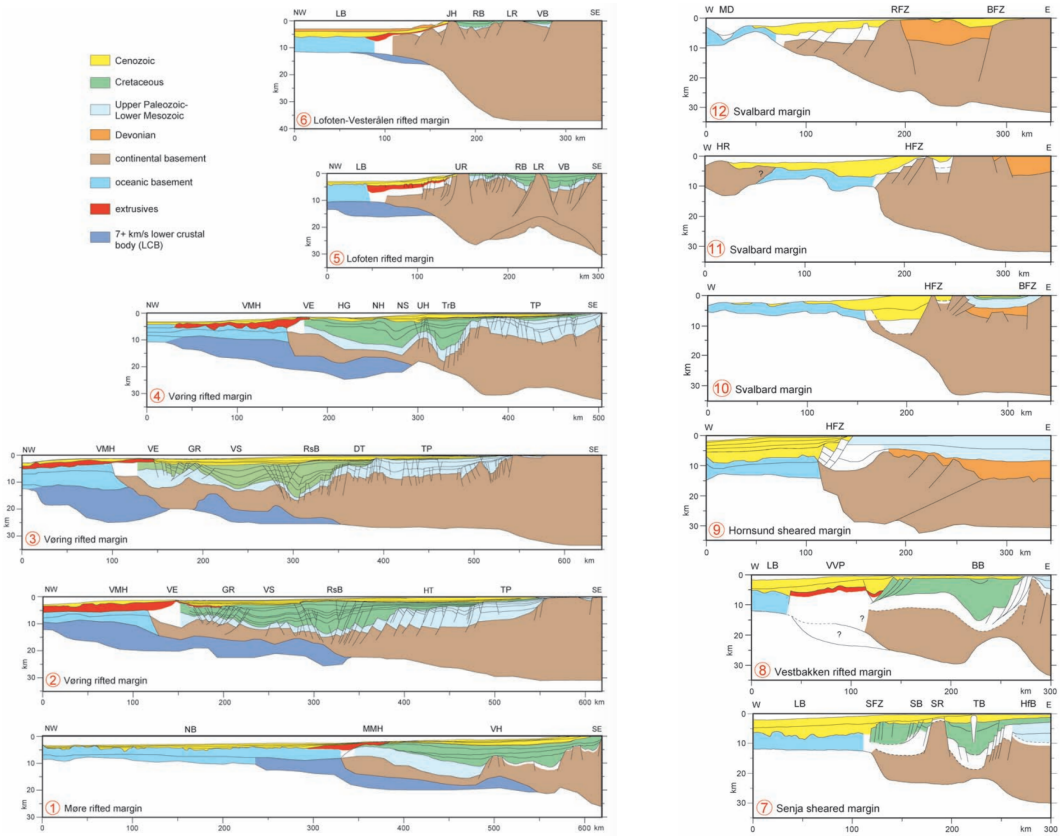


Figure 49: Regional structural map showing structural elements related to different rift phases affecting the NE Atlantic region. Location of cross sections in Figure 50 are also shown.

BB: Bjørnøya Basin, BL: Bivrost Lineament, EJMFZ: East Jan Mayen Fracture Zone, GR: Greenland Ridge, HFZ: Hornsund Fault Zone, HR: Hovgård Ridge, JMR: Jan Mayen Ridge, MB: Møre Basin, MMH: Møre Marginal High, SB: Sørvestsnaget Basin, SFZ: Senja Fracture Zone, TB: Tromsø Basin, TP: Trøndelag Platform, VB: Vøring Basin, VMH: Vøring Marginal High, VVP: Vestbakken Volcanic Province, WJMFZ: West Jan Mayen Fracture Zone.

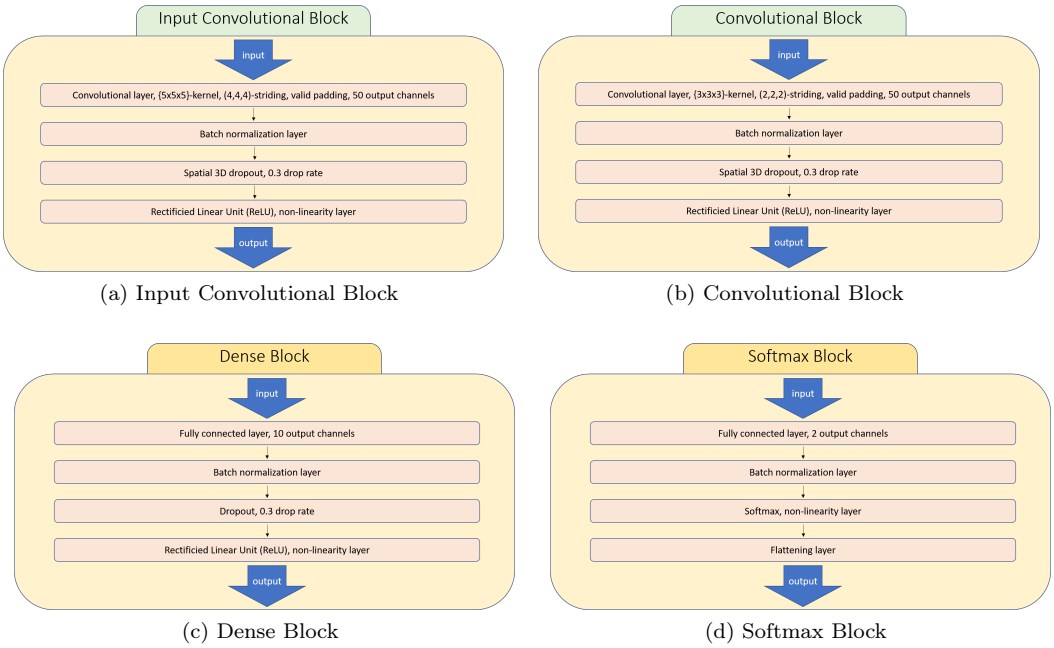


(a) DT: Dønna Terrace, GR: Gjallar Ridge, HT: Halten Terrace, HG: Hel Graben, JH: Jennegga High, LB: Lofoten Basin, LR: Lofoten Ridge, MMH: Møre Marginal High, NB: Norway Basin, NH: Nyk High, NS: Någrind Syncline, RB: Ribban Basin, RsB: Rås Basin, TrB: Træna Basin, TP: Trøndelag Platform, UH: Utgard High, UR: Utrøst Ridge, VB: Vestfjorden Basin, VE: Vøring Escarpment, VH: Vigra High, VMH: Vøring Marginal High, VS: Vigrid Syncline.

(b) BB: Bjørnøya Basin, BFZ: Billefjorden Fault Zone, HfB: Hammerfest Basin, HFZ: Hornsund Fault Zone, HR: Hovgård Ridge, LB: Lofoten Basin, MD: Molloy Deep, RFZ: Raudfjorden Fault Zone, SB: Sørvestsnaget Basin, SFZ: Senja Fracture Zone, SR: Senja Ridge, TB: Tromsø Basin, VVP: Vestbakken Volcanic Province.

Figure 50: Cross sections across the rifted continental margin off mid-Norway(a), and across the mainly sheared western Barents Sea-Svalbard margin(b). (Faleide et al., 2008[11])

Appendix B Machine learning blocks for seismic horizon tracking



(a) Input Convolutional Block

(b) Convolutional Block

(c) Dense Block

(d) Softmax Block

Figure 51: Block structures for the Machine Learning architecture

Appendix C Source code for flow charts in LaTeX

Training algorithm chart:

```
%%% Make a flow chart for the training process
\begin{tikzpicture}[node distance=2cm, scale=0.5]

%%% Define the nodes
% Section 1
\node (start) [startstop] {Start training epoch};
\node (in1) [io, below of=start, yshift=0.55cm, align=center, text width=5cm] {
    Segment data into training and test data};
\node (in2a) [io, below of=in1, xshift=-3cm, yshift=0.7cm] {Training data};
\node (in2b) [io, below of=in1, xshift=4cm, yshift=0.7cm] {Testing data};
\node (in3a) [io, below of=in2a, xshift=-0cm, yshift=0.5cm, align=center, text
    width=6cm] {Make new mini-batch from training data};
\node (pro1) [process, below of=in3a, xshift=-0cm, yshift=0.5cm, text width = 4cm]
     {Predict on mini-batch using current model};
\node (dec1) [decision, below of=pro1, xshift=-0cm, yshift=-0.75cm, align=center]
    {Evaluate\\accuracy of\\prediction};
\node (pro2a) [process, right of=dec1, xshift=3cm, yshift=0cm] {Tune prediction
    model};

% Section 2
\node (pro2b) [process, below of=dec1, xshift=-0cm, yshift=-1cm, align=center,
    text width=4cm] {Assemble training summary and output model};
\node (pro3) [process, below of=pro2b, xshift=3cm, yshift=0.5cm, text width=4cm] {
    Predict on testing data using output model};
\node (pro4) [process, below of=pro3, xshift=-0cm, yshift=0.5cm] {Evaluate
    accuracy of prediction};
\node (pro5) [process, below of=pro4, xshift=-0cm, yshift=0.5cm] {Assemble testing
     summary};
\node (out1) [io, below of=pro5, xshift=-0cm, yshift=0.5cm, align=center, text
    width=5cm] {Output model, training summary, and testing summary};
\node (stop) [startstop, below of=out1, yshift=0.55cm] {Stop};

%%% Draw the arrows between the nodes
% Section 1
\draw [arrow] (start) -- (in1);
\draw [arrow,green] (in1) -- (in2a);
\draw [arrow,green] (in1) -- (in2b);
\draw [arrow,green] (in2a) -- (in3a);
\draw [arrow,blue] (in3a) -- (pro1);
\draw [arrow] (pro1) -- (dec1);
\draw [arrow] (dec1) -- node[anchor=south, align=center] {Needs\\improve-\\ment} (
    pro2a);
\draw [arrow] (pro2a) |- (in3a);

% Section 2
```

```
\draw [arrow] (dec1) -- node[anchor=west, align=center] {Appropriate accuracy,\\or
     training data exhausted\vspace{1em}} (pro2b);
\draw [arrow] (pro2b) -- (pro3);
\draw [arrow,blue] (in2b) |- (pro3);
\draw [arrow] (pro3) -- (pro4);
\draw [arrow] (pro4) -- (pro5);
\draw [arrow] (pro5) -- (out1);
\draw [arrow] (out1) -- (stop);

\end{tikzpicture}
```

Prediction algorithm chart:

```
%%% Make a flow chart for the prediction process
\begin{tikzpicture}[node distance=2cm]

%%% Define the nodes
\node (start) [startstop] {Start prediction};
\node (in1) [io, below of=start, xshift=0cm, yshift=0cm, align=center, text width
    =2.5cm] {Input prediction data};
\node (pro1) [process, below of=in1, xshift=0cm, yshift=0cm, align=center, text
    width=4cm] {Make predictions using trained model};
\node (out1) [io, below of=pro1, xshift=0cm, yshift=0cm, align=center, text width
    =2.5cm] {Output prediction};
\node (stop) [startstop, below of=out1, xshift=0cm, yshift=0cm, align=center] {
    Stop};

%%% Draw the arrows between the nodes
\draw [arrow] (start) -- (in1);
\draw [arrow,blue] (in1) -- (pro1);
\draw [arrow] (pro1) -- (out1);
\draw [arrow] (out1) -- (stop);

\end{tikzpicture}
```

Appendix D Source code for Machine Learning Software

The machine learning software written in the course of this thesis is openly available on GitHub[69][70].

For the seismic horizon tracker, see the directory faces_net: `https://github.com/crild/facies_net`

For the fault identifier, see the directory fault_net: `https://github.com/crild/fault_net`

References

[1] Peter Bormann & Charles Rutherford Ildstad. Malenov, 2017. [Online; accessed January 15th, 2018], https://drive.google.com/drive/folders/0B7brcf-eGK8CRUhfRW9rSG91bW8 / https://github.com/bolgebrygg.

[2] Charles Rutherford Ildstad. Mapping and prediction of sediment hosted massive sulfide (shms) deposits on the norwegian continental shelf using machine learning (ml) - a feasibility study. *NTNU*, 2017.

[3] Mike Davidson. Introduction to this special section: Data analytics and machine learning. *SEG, The Leading Edge*, 36(3):206, 2017. https://library.seg.org/doi/abs/10.1190/tle36030206.1.

[4] Stephen E. Kesler. Mineral supply and demand into the 21st century. *USGS. Workshop on Deposit Modeling, Mineral Resource Assessment, and Sustainable Development*, page 55–62, 2007. https://pubs.usgs.gov/circ/2007/1294/reports/paper9.pdf.

[5] Steinar Løve Ellefmo. Ntnu oeans - pilot programme on deep sea mining, 2017. [Online; accessed May 12th, 2018], https://www.ntnu.edu/oceans/deep-sea-mining.

[6] Dr. Jort van Wijk. Blue mining: breakthrough solutions for sustainable deep sea mining, 2018. [Online; accessed May 12th, 2018], http://www.bluemining.eu/.

[7] Mark Hannington, Sven Petersen & Anna Krätschell. Subsea mining moves closer to shore. *NATURE GEOSCIENCE*, 10:158–159, February 2017. https://www.nature.com/articles/ngeo2897.

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

[9] Qian Wang, Yinghuan Shi, Heung-Il Suk, & Kenji Suzuki(Eds.). *Machine Learning in Medical Imaging*. Springer Publishing, 2017. https://books.google.no/books?id=DYQODwAAQBAJ&lpg=PA93&dq=SEG%20machine%20learning&hl=no&pg=PR2#v=onepage&q=SEG%20machine%20learning&f=false.

[10] A.U. Waldeland & A.H.S.S. Solberg. Salt classification using deep learning. *79th EAGE Conference and Exhibition 2017*, Conference:Seismic Interpretation – Analytics and Machine Learning for Interpretation, June 2017. http://www.earthdoc.org/publication/publicationdetails/?publication=88635.

[11] Faleide et al. Structure and evolution of the continental margin off norway and the barents sea. *Episodes*, 31:82–91, March 2008. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.475.9034&rep=rep1&type=pdf.

[12] David Michaud. Mining exploration process, 2018. [Online; accessed January 4th, 2019], https://www.911metallurgist.com/blog/mining-exploration-process.

[13] D.G. MacIntyre. Sedex - sedimentary exhalative deposits. *Ore Deposits, Tectonics and Metallogeny in the Canadian Cordillera*, 4:25–65, 1991. http://www.empr.gov.bc.ca/Mining/Geoscience/PublicationsCatalogue/Papers/Documents/P1991-04_chpt3.pdf.

[14] Gregor Borg. Regional controls on sediment-hosted pb-zn (ba-cu) occurences within the pan-african orogenic belts of namibia. *Communs Geological Survey Namibia*, 12:239–

249, 2000. http://www.mme.gov.na/files/publications/c7e_Borg_sediment-hosted%20Pb-Zn.pdf.

[15] Neal Reynolds & Peter Muhling. Diversity of zinc-lead metallogeny. AIG Base Metals Seminar, Perth, https://www.aig.org.au/wp-content/uploads/2012/12/Neal-Reynolds-Diversity-of-Zinc-Lead-Metallogeny-Implications-for-Targeting-and-Discovery.pdf, June 2011.

[16] Misi et al. Sediment hosted lead–zinc deposits of the neoproterozoic bambuí group and correlative sequences, são francisco craton, brazil: A review and a possible metallogenic evolution model. *Ore Geology Reviews*, 26:263–304, 2005. https://repositorio.ufba.br/ri/bitstream/ri/5266/1/S016913680500020X-main.pdf.

[17] Laurence Robb. *Introduction to Ore-Forming Processes*. Blackwell Publishing, 2014. http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0632063785.html.

[18] J. Tuzo Wilson. Did the atlantic close and then re-open? *Nature*, 211:676–681, August 1966.

[19] Wikipedia, the free encyclopedia, Norman Einstein assumed. Barents sea map, 2005. [Online; accessed January 15th, 2018], https://commons.wikimedia.org/w/index.php?curid=433786.

[20] Norwegian Petroleum Directorate. Solid applicants want to explore in the 24th round, 2017. [Online; accessed May 13th, 2018], http://www.npd.no/en/Licensing-rounds/Licensing-rounds/24th-Licencing-round/Solid-applicants-want-to-explore-in-the-24-round/.

[21] Norwegian Petroleum Directorate. Petroleum activity in the norwegian sector of the barents sea, 2017. [Online; accessed January 15th, 2018], http://www.npd.no/en/Publications/Reports/Geological-assessment-of-petroleum-resources---Barents-Sea-north-2017/Petroleum-activity-in-the-Norwegian-sector-of-the-Barents-Sea/.

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25*, pages 1097–1105, 2012. http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[23] Jia Deng, Wei Dong, Richard Socher, Li jia Li, Kai Li, and Li Fei-fei. Imagenet: A large-scale hierarchical image database. *In CVPR*, 2009. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.155.1729.

[24] Seung-Hwan Lim, Steven Young, and Robert Patton. An analysis of image storage systems for scalable training of deep neural networks. *The Seventh workshop on Big Data Benchmarks, Performance Optimization, and Emerging Hardware (in conjunction with ASPLOS'16)*, 04 2016.

[25] Edward N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20:130–141, 1963. https://doi.org/10.1175/1520-0469(1963)020%3C0130:DNF%3E2.0.CO;2.

[26] Trevor Hastie, Robert Tibshirani, Jerome Friedman. *The Elements of Statistical Learning*. Springer International Publishing, 2009. https://web.stanford.edu/~hastie/ElemStatLearn/.

[27] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition? *IEEE 12th International Conference on Computer Vision*, pages 2146–2153, Sept 2009. https://ieeexplore.ieee.org/document/5459469/.

[28] Ivar B. Ramberg, Inge Bryhni, Arvid Nøttvedt, Kristin Ragnes. *The Making of a Land*. Norsk geologisk forening (The Norwegian Geological Association), 2008. https://www.haugenbok.no/Fagboeker/Biologi/The-making-of-a-land/I9788292394427, figures used from pages(308, 311, 348, 413, 438, 473).

[29] Jerry Coffey. What is pangaea? *unpublished*, 2010. [Online; accessed January 15th, 2018], https://www.universetoday.com/73678/what-is-pangea/.

[30] J. I. Faleide & Alvar Braathen K. A. Leever, R. H. Gabrielsen. A transpressional origin for the west spitsbergen fold-and-thrust belt: Insight from analog modeling. *Tectonics*, 30, 2011. https://pdfs.semanticscholar.org/ae2d/6a846d3a1adc780fbdf21fa583a1e70a7953.pdf.

[31] Norwegian Petroleum Directorate. Npd faktakart. *Norwegian Petroleum Directorate*, 2018. [Online; accessed May 13th, 2018], http://gis.npd.no/factmaps/html_21/.

[32] Valeri A. Basov, Jörg Ebbing, Laurent Gernigon, Marianna V. Korchinskaya, Tatyana Koren, Natalia V. Kosteva, Galina V. Kotljar, Geir Birger Larssen, Tamara Litvinova, Oleg. B. Negrov, Odleiv Olesen, Christophe Pascal, Tatyana M. Pchelina, Oleg V. Petrov, Yugene O. Petrov, Hans-Ivar Sjulstad, Morten Smelror, Nikolay N. Sobolev, Victor Vasiliev, Stephanie C. Werner. *ATLAS: Geological History of the Barents Sea*. Norges geologiske undersøkelse (Geological Survey of Norway), 2009. http://issuu.com/ngu_/docs/atlas_-_geological_history_of_the_b/1?e=3609664/9026048, figures used from pages(48, 52).

[33] R.H. Gabrielsen. Long-lived fault zones and their influence on the tectonic development of the southwestern barents sea. *Journal of the Geological Society*, 141(4):651–662, 1984.

[34] L. N. Jensen R.H. Gabrielsen, R.B. Faerseth. Structural elements of the norwegian continental shelf. pt. 1. the barents sea region. *Norwegian Petroleum Directorate Bulletin*, .:., .

[35] R.H. Gabrielsen J. I. Faleide, K. Bjørlykke. *Geology of the Norwegian continental shelf*. Springer Petroleum Geoscience, 2010.

[36] E. Glørstad-Clark, J.I. Faleide, B.A. Lundschien, and J.P. Nystuen. Triassic seismic sequence stratigraphy and paleogeography of the western barents sea area. *Marine and Petroleum Geology*, 27(7):1448–1475, 2010. https://www.researchgate.net/publication/223101952_Triassic_seismic_sequence_stratigraphy_and_paleogeography_of_the_western_Barents_Sea_Area.

[37] Audrius Norkus. Pre-jurassic evolution of the fingerdjupet subbasin, sw barents sea. Master's thesis, University of Oslo (UiO), Norway, 2015.

[38] L. B. Henriksen S. Kristensen I. Nilsson T. Samuelsberg T. Svånå L. Stemmerik D. Worsley G. Larssen, G. Elvebakk. Upper palaeozoic lithostratigraphy of the southern norwegian barents sea. *Norwegian Petroleum Directorate Bulletin*, 9:76, 2002.

[39] K Ofstad A. Dalland, D. Worsley. A lithostratigraphic scheme for the mesozoic and cenozoic succession, offshore mid- and northern norway. *Norwegian Petroleum Directorate Bulletin*, 1988.

[40] Norwegian Petroleum Directorate. The kapp toscana group - realgrunnen subgroup. *Norwegian Petroleum Directorate*, 2018. [Online; accessed Jan 1th, 2019], http://www.npd.no/en/Publications/Reports/Compiled-CO2-atlas/6-The-Barents-Sea/61-Geology-of-the-Barents-Sea/The-Kapp-Toscana-Group-Realgrunnen-Subgroup/.

[41] S. G. Walters. Broken hill-type deposits. *AGSO Journal of Australian Geology and Geophysics*, 17(4):229–237, 1998. https://web.archive.org/web/20120731213610/http://www.ga.gov.au/image_cache/GA5412.pdf.

[42] J.J. Wilkinson et al. Intracratonic crustal seawater circulation and the genesis of subseafloor zinc-lead mineralization in the irish orefield. *Geology*, 33(10):805–808, 2005. http://portal-vip.cc.ic.ac.uk/pls/portallive/docs/1/30361696.PDF.

[43] M. Deb, W. D. Goodfellow. *Sediment Hosted Lead-Zinc Sulphide Deposits*. CRC Press, 2004. https://books.google.no/books?id=H64sxguvlpMC&printsec=frontcover&hl=no&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false.

[44] & Alexander P.Gysi Nicole C.Hurtig, Jacob J.Hanley. The role of hydrocarbons in ore formation at the pillara mississippi valley-type zn-pb deposit, canning basin, western australia. *Ore Geology Review*, 102:875–893, 2018. https://www.sciencedirect.com/science/article/pii/S0169136818304062.

[45] Meri-Liisa Airo. Geophysical signatures of mineral deposit types – synopsis. *Geophysical signatures of mineral deposit types in Finland, Geological Survey of Finland, Special Papers*, 58:9–70, 2015. http://tupa.gtk.fi/julkaisu/specialpaper/sp_058_pages_009_070.pdf.

[46] S. Maus, U. Barckhausen, H. Berkenbosch, N. Bournas, J. Brozena, V. Childers, F. Dostaler, J. D. Fairhead, C. Finn, R. R. B. von Frese, C. Gaina, S. Golynsky, R. Kucks, H. Lühr, P. Milligan, S. Mogren, D. Müller, O. Olesen, M. Pilkington, R. Saltus, B. Schreckenberger, E.Thébault, and F. Caratori Tontini. Emag2: A 2-arc-minute resolution earth magnetic anomaly grid compiled from satellite, airborne and marine magnetic measurements. *Geochem. Geophys. Geosyst., under review*, 2009. http://geomag.org/info/Smaus/Doc/emag2.pdf.

[47] Edward A. du Bray(Ed.). *Preliminary Compilation of Descriptive Geoenvironmental Mineral Deposit Models(Chap. 29 & 30)*. U.S. Geological Survey Open-File Book 95-831, 1996. https://pubs.usgs.gov/of/1995/ofr-95-0831/.

[48] Stein Fanavoll, Pål T. Gabrielsen, and Svein Ellingsrud. Csem as a tool for better exploration decisions: Case studies from the barents sea, norwegian continental shelf. *SEG: Interpretation (online version)*, 2(3):SH55–SH66, 2014. http://dx.doi.org/10.1190/INT-2013-0171.1.

[49] David T. Sandwell and Walter H. F. Smith. Global marine gravity from retracked geosat and ers-1 altimetry: Ridge segmentation versus spreading rate. *Journal of geophysical data*, 114:B01411(1–18), 2009. http://topex.ucsd.edu/sandwell/publications/122_JGR_gravity.pdf.

[50] C. Clavier, A. Heim & C. Scala. Effect of pyrite on resistivity and other logging measurements. *Society of Petrophysicists and Well-Log Analysts*, page 34, 1976. https://www.onepetro.org/conference-paper/SPWLA-1976-HH.

[51] Yue Wu, Changqing Zhang, Jingwen Mao, Hegeng Ouyang & Jia Sun. The genetic relationship between hydrocarbon systems and mississippi valley-type zn–pb deposits along the sw margin of sichuan basin, china. *International Geology Review*, 55(8):941–957, 2013. https://doi.org/10.1080/00206814.2012.753177.

[52] E. L'Heureux, B. Milkereit & E. Adam. 3d seismic exploration for mineral deposits in hardrock environments. *RECORDER, Canadian Society of Exploration Geophysicists*, 30:NO. 09, November 2005. https://csegrecorder.com/articles/view/3d-seismic-exploration-for-mineral-deposits-in-hardrock-environments.

[53] Chiyuan Zhang, Charlie Frogner, Tomaso Poggio, Mauricio Araya-Polo, Jan Limbeck & Detlef Hohl. Automated geophysical feature detection with deep learning. *unpublished*, April 2016. http://pluskid.org/slides/GTC2016-seismic.pdf.

[54] Ronny Setså (Charles Rutherford Ildstad & Peter Bormann). Ung, lovende, analytisk og lærevillig. Geoforskning.no, http://geoforskning.no/nyheter/olje-og-gass/1621-ung-lovende-analytisk-og-laerevillig, October 2017.

[55] Yunxin Xie, Chenyang Zhu, Wen Zhou, Zhongdong Li, Xuan Liu, Mei Tu. Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. *Journal of Petroleum Science and Engineering*, 160:182–193, January 2018. http://www.sciencedirect.com/science/article/pii/S0920410517308094.

[56] Matthew J. Cracknell & Anya M. Reading. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences*, 63:22–33, February 2014. http://www.sciencedirect.com/science/article/pii/S0098300413002720.

[57] Rocky Roden & Deborah Sacrey. Seismic interpretation with machine learning. *GeoExPro Magazine*, online:online, December 2016. https://static1.squarespace.com/static/558831bce4b04e6344650e69/t/58700b8ce3df284f0b4f80a2/1483738004643/Sesimic+Interpretation+%26+Machine+Learning%2C+GeoExPro+December+2016.pdf.

[58] Christopher Olah. Understanding lstm networks, 2015. [Online; accessed January 15th, 2018], http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

[59] O. Ronneberger, P.Fischer & T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a.

[60] Wikipedia, the free encyclopedia, Glosser.ca. Colored neural network, 2013. [Online; accessed January 15th, 2018], https://commons.wikimedia.org/w/index.php?curid=24913461.

[61] Wikipedia, the free encyclopedia, Aphex34. Conv layer, 2015. [Online; accessed January 15th, 2018], https://commons.wikimedia.org/w/index.php?curid=45659236.

[62] Alex Krizhevsky Ilya Sutskever Ruslan Salakhutdinov Nitish Srivastava, Georey Hinton. Dropout: A simple way to prevent neural networks from overtting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. https://www.cs.toronto.edu/~hinton/absps/JMLRdropout.pdf.

[63] Victor Stamatescu Mark D. McDonnell Sebastien C. Wong, Adam Gatt. Understanding data augmentation for classification: when to warp? *Computer vision and pattern recognition*, 2016. https://arxiv.org/abs/1609.08764.

[64] Christian Szegedy Sergey Ioffe. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Computer Science*, 2015. https://arxiv.org/abs/1502.03167.

[65] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *eprint arXiv:1412.6980*, Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015:1–15, 2014. https://arxiv.org/abs/1412.6980.

[66] Norwegian Petroleum Directorate (NPD). Factpages. *7321/7-1*, 1988. http://factpages.npd.no/FactPages/Default.aspx?nav1=wellbore&nav2=PageView|Exploration|All&nav3=1284&culture=en.

[67] Norwegian Petroleum Directorate (NPD). Factpages. *7321/8-1*, 1987. http://factpages.npd.no/FactPages/Default.aspx?nav1=wellbore&nav2=PageView|Exploration|All&nav3=1070&culture=en.

[68] Norwegian Petroleum Directorate (NPD). Factpages. *7321/9-1*, 1988. http://factpages.npd.no/FactPages/Default.aspx?nav1=wellbore&nav2=PageView|Exploration|All&nav3=1339&culture=en.

[69] Charles Rutherford Ildstad. Faciesnet, software for detection of given facies. *Unpublished*, 2018. https://github.com/crild/facies_net.

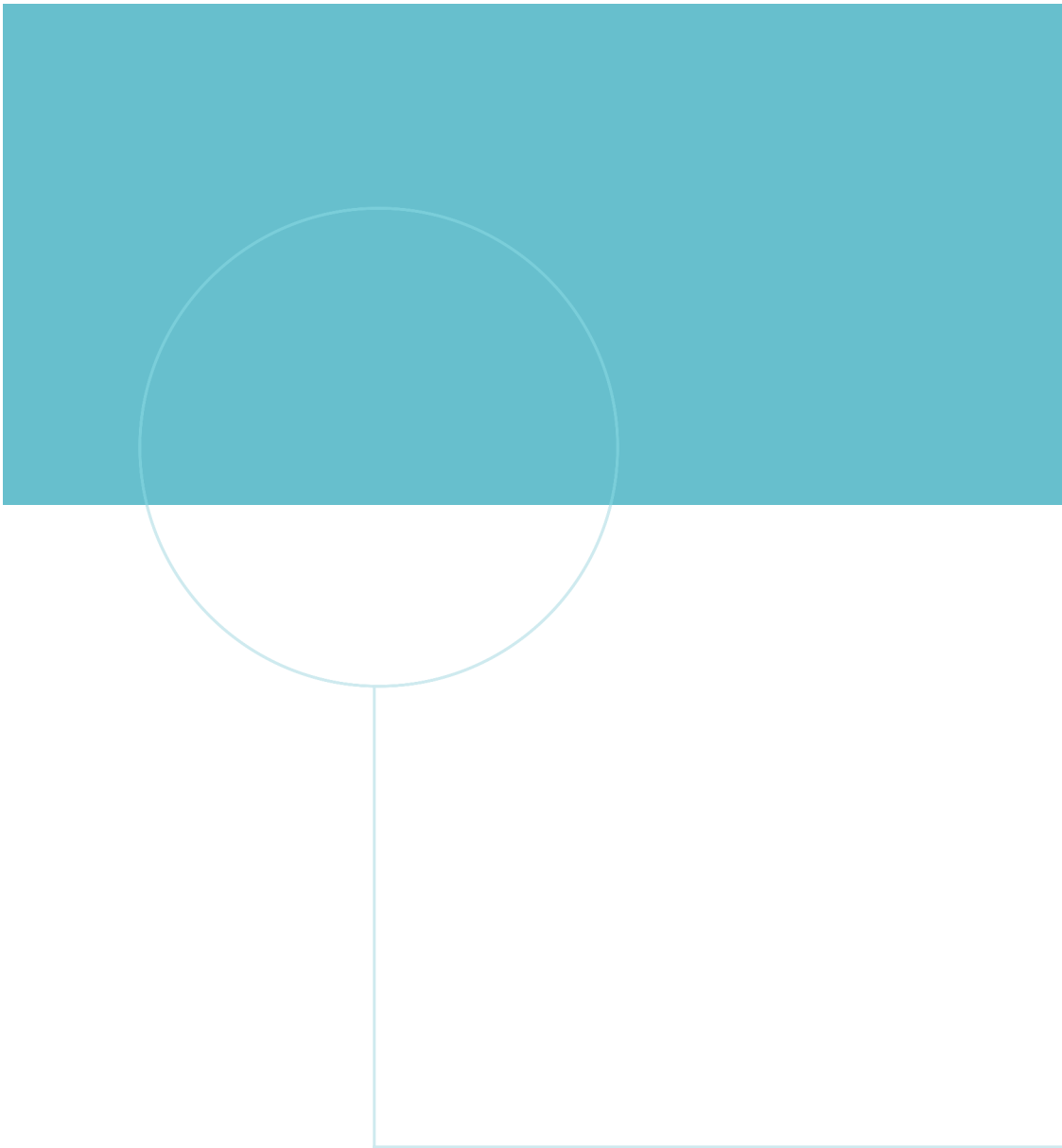
[70] Charles Rutherford Ildstad. Faultnet, fault detection software for seismic data. *Unpublished*, 2018. https://github.com/crild/fault_net.

[71] Diederik P Kingma & Max Welling. Auto-encoding variational bayes. *ArXiv e-prints*, May 2014. https://arxiv.org/abs/1312.6114.

[72] Rob Fergus Matthew D Zeiler. Visualizing and understanding convolutional networks. *Computer Vision and Pattern Recognition*, 2013. VisualizingandUnderstandingConvolutionalNetworks.

[73] Sebastian Bach Klaus-Robert Müller Wojciech Samek Alexander Binder, Grégoire Montavon. Layer-wise relevance propagation for neural networks with local renormalization layers. *Computer Vision and Pattern Recognition*, 2016. https://arxiv.org/abs/1604.00825.

[74] Trevor Darrell Jonathan Long, Evan Shelhamer. Fully convolutional networks for semantic segmentation. *Computer Vision and Pattern Recognition*, 2014. https://arxiv.org/abs/1411.4038.



NTNU
Norwegian University of
Science and Technology