Using Machine Learning to Understand Uncertainty in Subsurface Exploration

Keith Gray, Max Grossman, Anar Yusifov
BP plc

Challenge Science Domain: Geosciences

Data Set Name: Synthetic Seismic Realizations

Description of the Data Set

In the energy industry, an understanding of subsurface characteristics and structure is crucial to identifying and localizing untapped resources. At a high level, the process of taking an entirely unexplored region of earth and generating an actionable understanding of its structure includes:

1. Seismic data collection: Collect raw signals from the subsurface using techniques similar to sonograms used in hospitals.
2. Seismic data pre-processing: Quality check and clean the collected raw signals.
3. Seismic migration & velocity model construction: Use the raw signals and our understanding of the likely geology of the region to construct a 3D representation of the subsurface.
4. Seismic interpretation: Using the constructed 3D representation, interpret where faults, layers, and other important structural features are in the subsurface.

With each of these steps comes an amount of uncertainty from various sources of potential error: instrument error, human error, modeling error, and more. Despite this, the output of most seismic processing workflows is a single, gold standard, output image. An image which we know cannot possibly be 100% accurate!

It is crucial that future seismic processing workflows start to incorporate uncertainty when estimating the true subsurface structures. Rather than outputting a single interpretation, we should aim to emit a spectrum of possible realizations and an understanding of where uncertainty is high or low.

The dataset included in this data challenge serves as a starting point in exploring techniques for quantifying uncertainty in seismic processing workflows. In this dataset we are focused on quantifying and visualizing the uncertainty in our estimations of the density of the subsurface based on how varying those estimates impacts our output 3D volume. At a high level, this dataset consists of a set of synthetic but realistic models of the density of the subsurface, randomly generated based on a single, known, synthetic ground truth. This dataset also includes the final 3D realizations generated using those density models (also called velocity models). These files are stored in the industry standard SEGY format, and an example Jupyter notebook is provided to illustrate how to load and visualize them.

Challenge Questions

The end goal of this data challenge is to construct an uncertainty map for a given seismic survey, labeling each pixel in a final 2D seismic image with a value between 0.0 and 1.0 indicating how volatile the estimate for that pixel is.

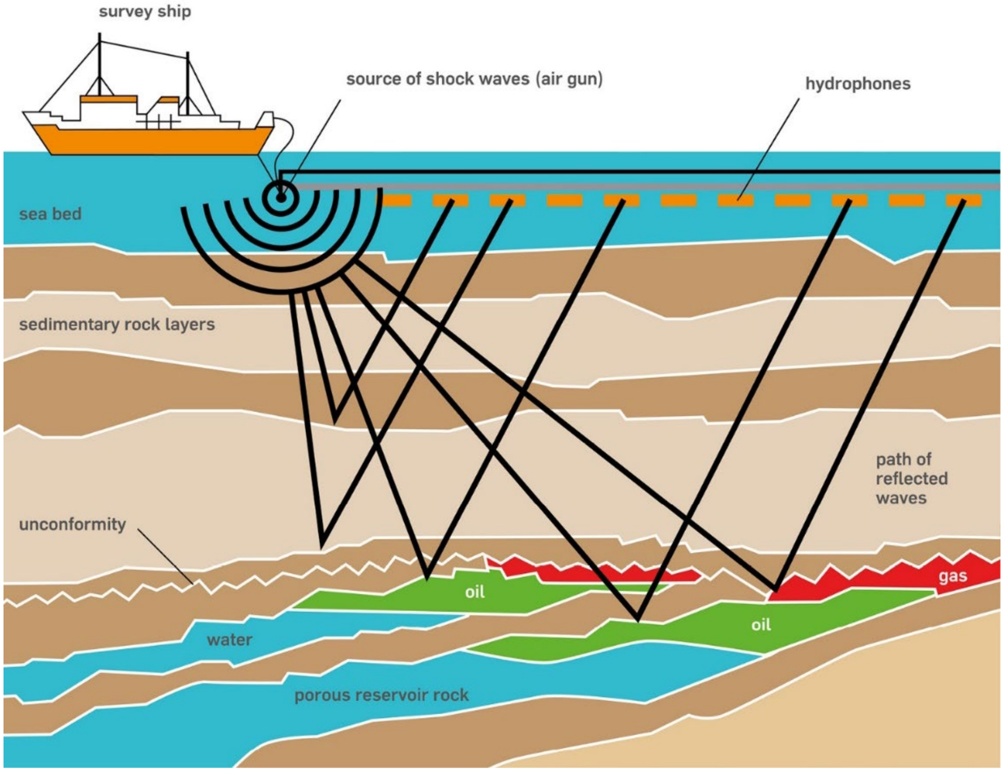
However, we also welcome submissions that include any intermediate work towards that end goal or answers to any of the below challenge questions. Even if you are unable to complete the entire challenge, any submissions that show progress towards this end goal and lay out ideas for how the challenge could eventually be completed will be considered.

- Given that geophysicists generally use horizontal lines in gathers as a good indicator of velocity model accuracy, build a model (analytical, mathematical, data-driven, or otherwise) to estimate the quality of each velocity model based on its associated gathers.
- Train a model to label each pixel with an uncertainty value between 0.0 and 1.0 indicating how uncertain any given realization of that part of the subsurface is.
- Generate a single uncertainty map given all of the velocity models, realizations, and gathers at hand.
- Generate some form of visualization of this uncertainty map of the subsurface.

For more background information, please see the following appendicies.

Appendix A: Seismic Data Collection

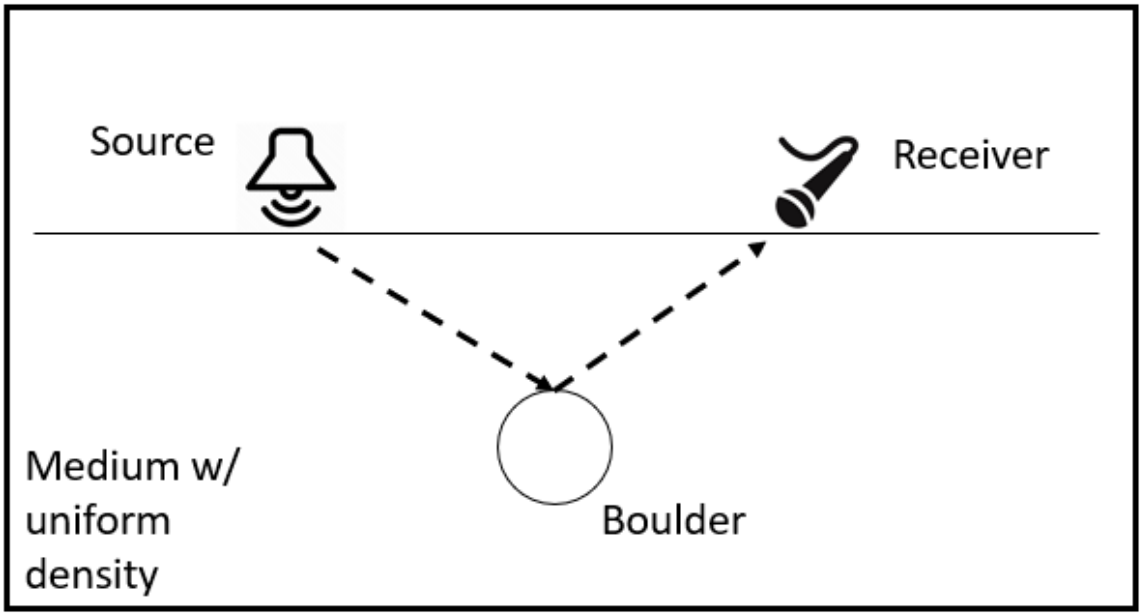
Seismic data collection (i.e., the process of conducting a seismic survey) involves transmitting powerful sound waves into the ground and then recording their echoes at the surface as they bounce off boundaries between layers in the Earth. This process parallels techniques used in x-ray and ultrasound imaging in the medical field to reconstruct structures inside the human body. The figure below depicts a typical offshore seismic survey setup, in which sound waves are transmitted from an air gun behind a survey ship and the return echoes are recorded by a line of hydrophones being towed behind the ship.



https://krisenergy.com/company/about-oil-and-gas/exploration/

During a seismic survey, one or more sources of sound energy are used to transmit waves into the ground. One or more receivers are used to record the reflection of that sound energy at the surface. The raw output generated from a seismic survey is a set of recorded waveforms at each receiver for each source. This recording stores the amplitude of the reflected sound wave at the surface as a function of the time it took to travel to the receiver.

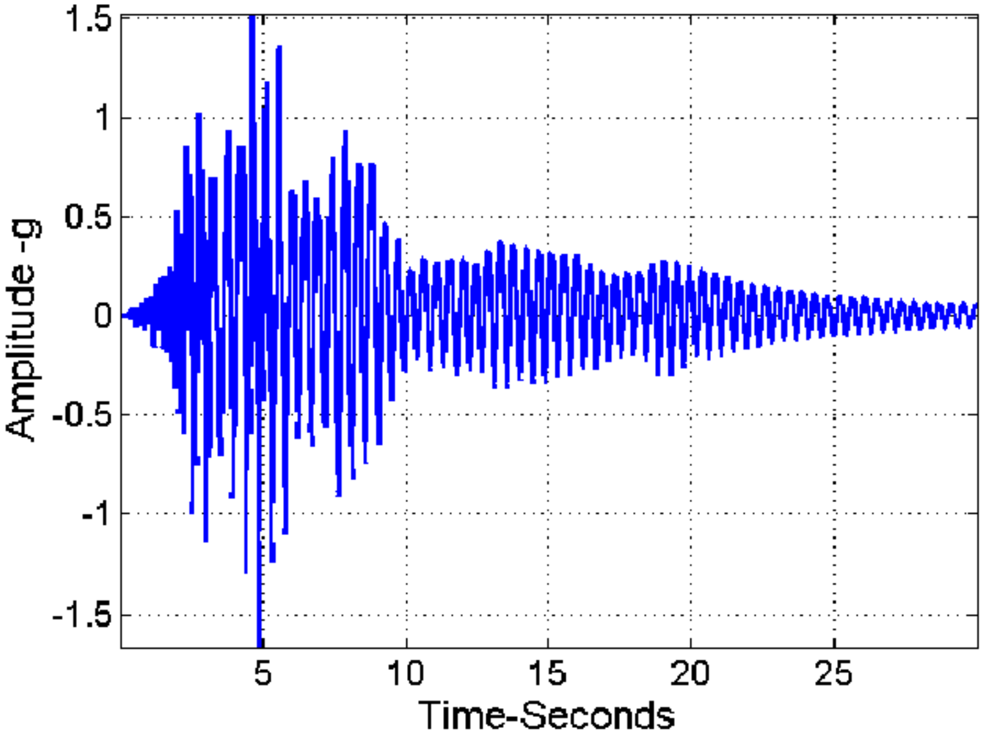
To start to see how this information could be useful in understanding subsurface structures, consider the trivial 2D example in the figure below, in which we have a single source; a single receiver; a single subsurface reflective object (e.g., a boulder); and a subsurface medium of perfectly uniform density (i.e., sound waves travel at a constant, known speed beneath the ground). With just the knowledge of (1) our source location, (2) our receiver location, and (3) an accurate measurement of how much time elapsed between the source transmitting and the receiver measuring the response, we can accurately compute the position of the boulder horizontally and vertically in the subsurface.



Appendix B: Seismic Data Preprocessing

Pre-processing of our raw seismic data can include a multitude of steps. Broadly, seismic preprocessing aims to clean up and strengthen signals in the seismic data while reducing noise, facilitating later stages of the seismic processing pipeline. While the example above would not require additional pre-processing, in the real world there are many sources of noise that make seismic data less than perfect.

The amplitude over time of the signal received at a given receiver is often referred to as a seismic trace. In a perfect world, the seismic trace would have a single spike for each boundary layer of the Earth that the transmitted signal bounced off. In reality, traces more often look like the figure below, with reverberations and other noise making a single peak in the signal difficult to pick out. Seismic preprocessing and denoising helps to reduce this noise.

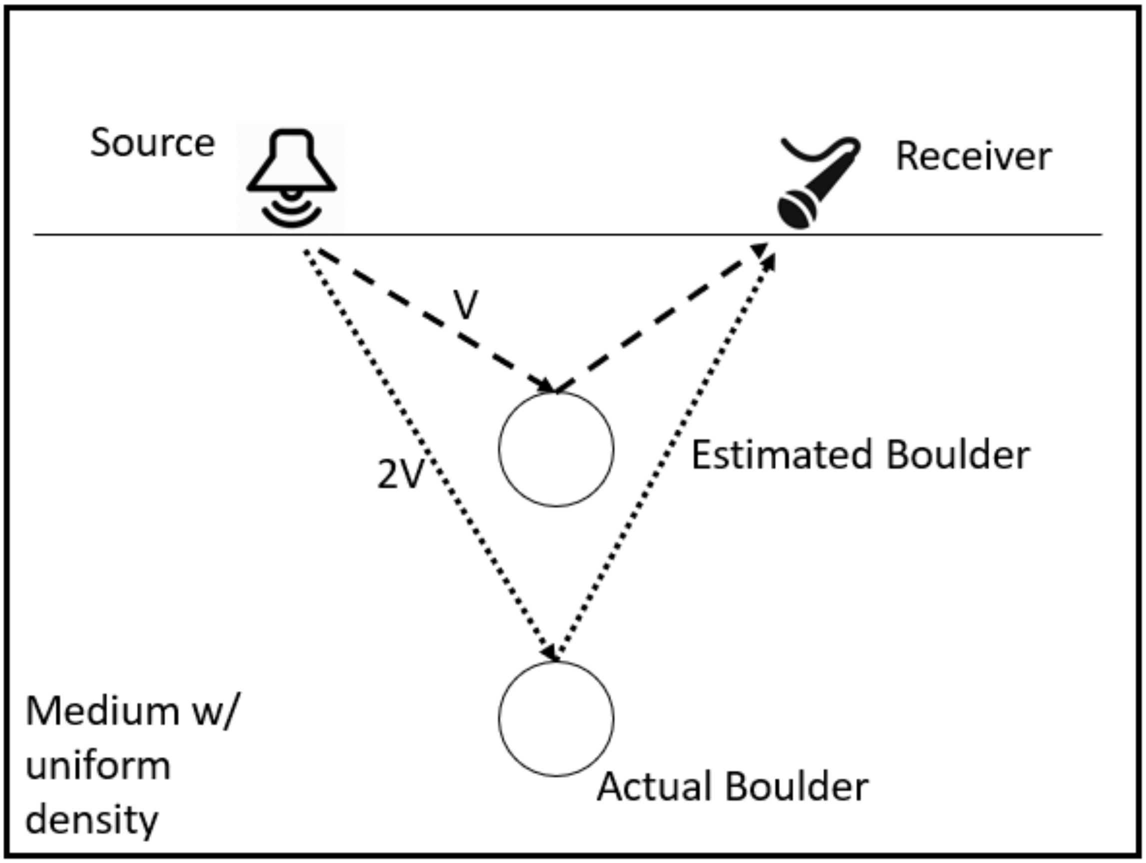


Appendix C: Seismic Migration and Velocity Model Construction

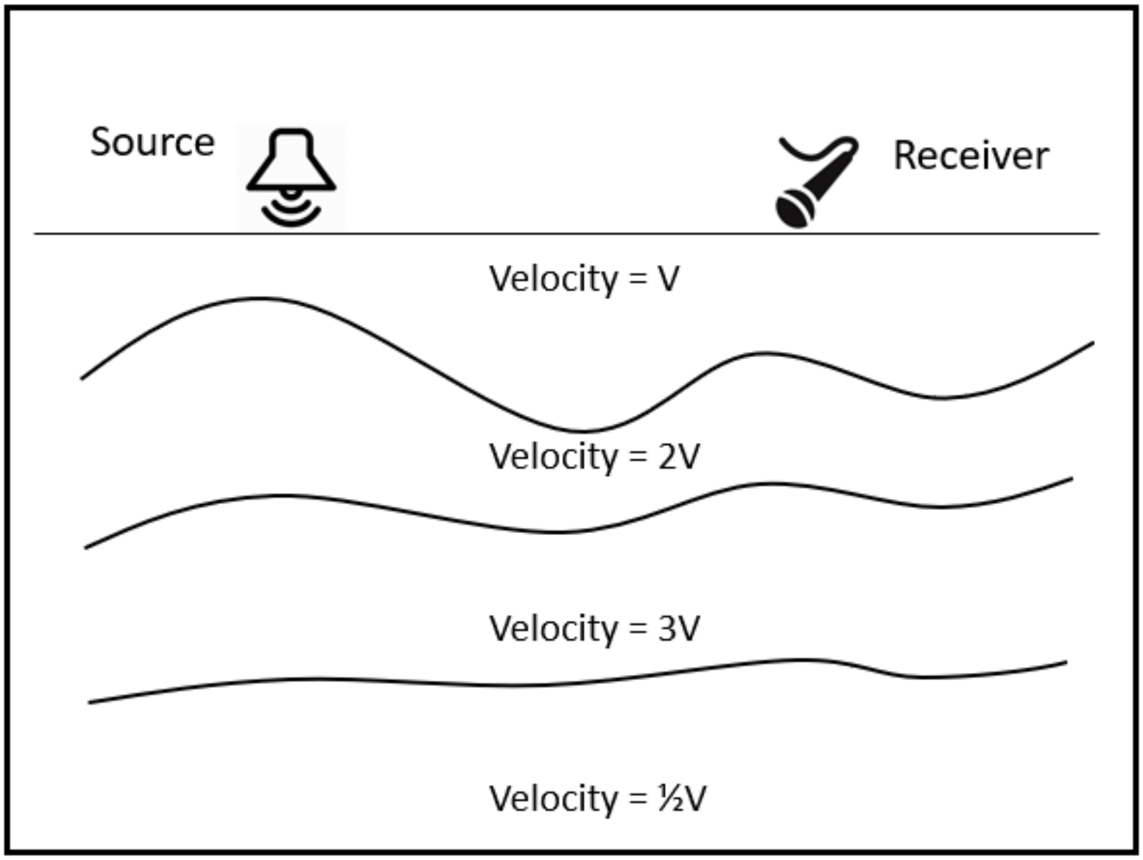
Seismic migration refers to the process by which the seismic waves received at receivers are backpropagated to the source through a simulated version of the seismic medium. Through knowledge of (1) the source location, (2) the receiver location, (3) the time/amplitude of the received signal, and (4) the medium through which the signal traveled, we can simulate in reverse the propagation of the signal through the subsurface, identify its reflection point, and thereby identify the location of a potential object/reflector of interest in the subsurface.

Note how crucial an accurate estimate of the subsurface velocity of sound waves is in this process. Without an accurate velocity estimate, it is impossible to accurately predict the distance traveled by sound waves in the subsurface in a certain period of time.

Take our simple example of a uniform subsurface medium and a single reflector (i.e., boulder) in the subsurface. Suppose that our geological understanding of the subsurface led us to estimate that sound travels with a velocity V in the subsurface, when in fact it travels with velocity 2V. Our incorrect velocity estimate leads us to calculate a much shorter distance traveled, thereby drastically underestimating the depth of the boulder (see below).

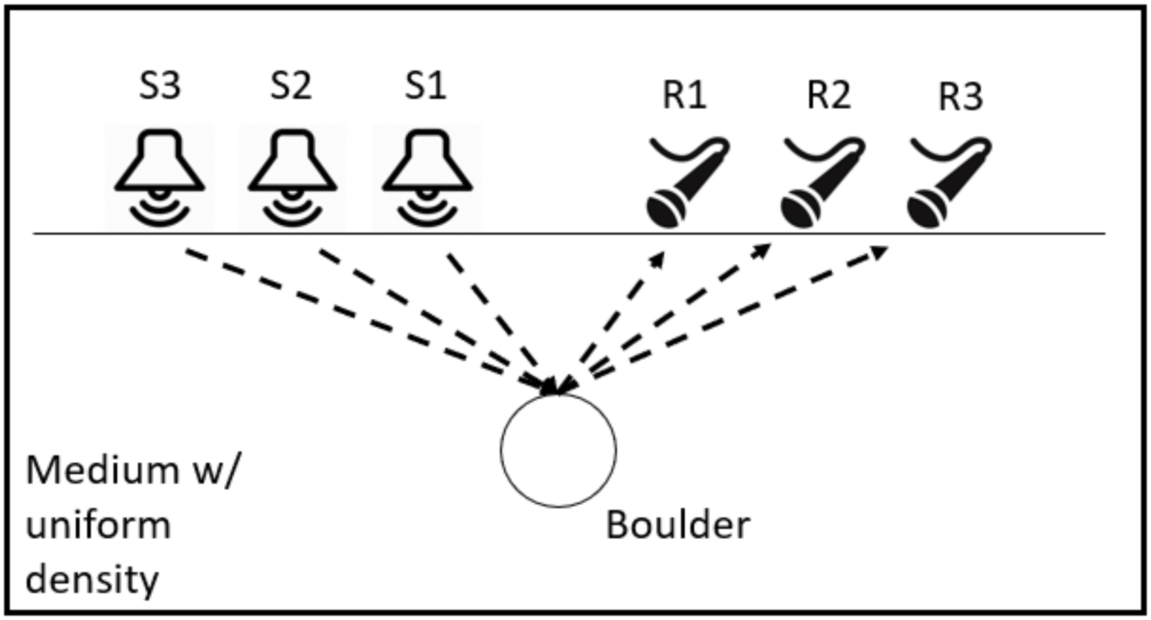


Of course, in the real world, we never have a subsurface medium of uniform density. The Earth has many layers of varying density, height, angle, and other attributes. These constantly changing attributes affect the velocity and attenuation of seismic waves as they travel through the subsurface. Our estimate of the true velocity of a seismic wave at every point in the subsurface is referred to as a velocity model. See below for an illustration of a velocity model in a simple layered environment.

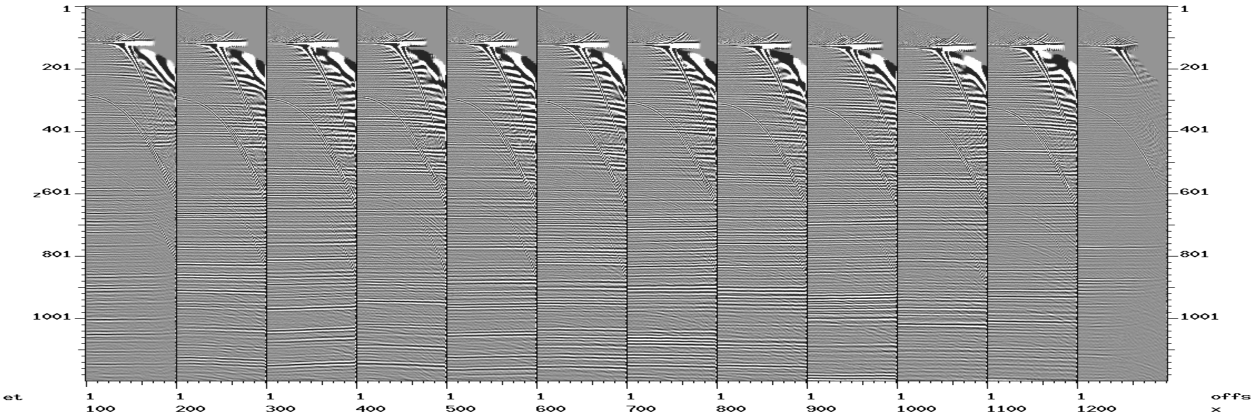


While building a velocity model is a critical component for accurate seismic reconstruction, a number of uncertainties are involved in the process. Simply asking two different geophysicists to perform velocity model construction on the same seismic traces can drastically change the selected velocity model. **Quantifying and visualizing this uncertainty in velocity models will be the prime focus of this data challenge.**

One common practice for checking the validity of a given velocity model is through offset pair gathers. Modern seismic surveys generally involve many sources and many receivers. As a result, many pairs of sources and receivers capture reflections off the same reflector in the subsurface (see below). This redundancy can be helpful in validating the quality of a velocity model, as an accurate velocity model is expected to produce similar/identical depth estimates for a given reflector no matter which offset pair a reflection is received from.

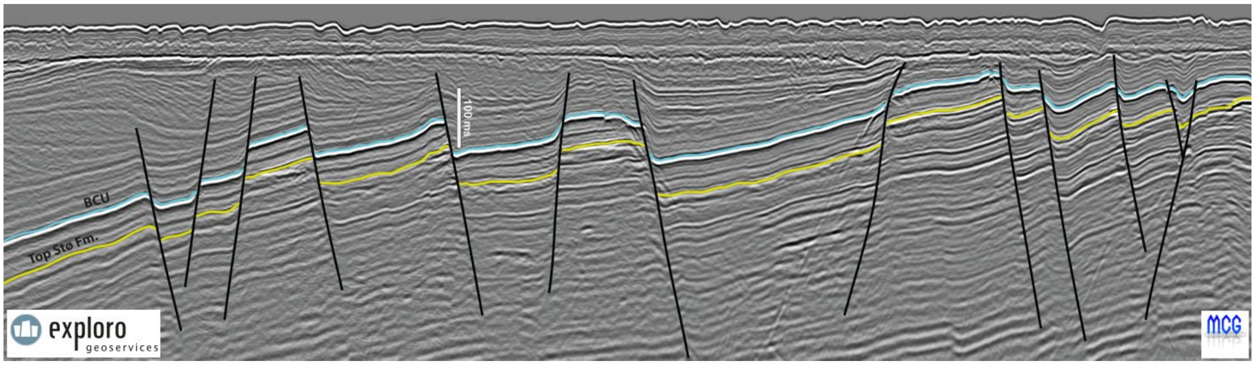


Gathers generally refer to collecting the depth estimate for a given reflector across many offset pairs and plotting them visually, with depth on the y axis and offset pairs along the x axis. In a gather of an accurate velocity model, geophysicists expect to mostly see horizontal lines, indicating that the depth estimate for a layer is the same across all offset pairs. See below for several examples of reasonable gathers, indicated by the prevalence of horizontal lines.



Appendix D: Seismic Interpretation

Once a final seismic image is rendered following seismic migration, seismic interpretation—the process of identifying faults, reservoirs, and other features of interest in the image—begins. This manual labeling is then used in field development and reservoir characterization. See below for an example seismic image with faults manually labeled and emphasized.



https://www.geoexpro.com/articles/2016/01/super-high-resolution-seismic-data-in-the-norwegian-barents-sea