DeepTrace:
A.I. for First-Break Picking

Branton Demoss
Front Range Geosciences
Boulder, CO

Lenovo™

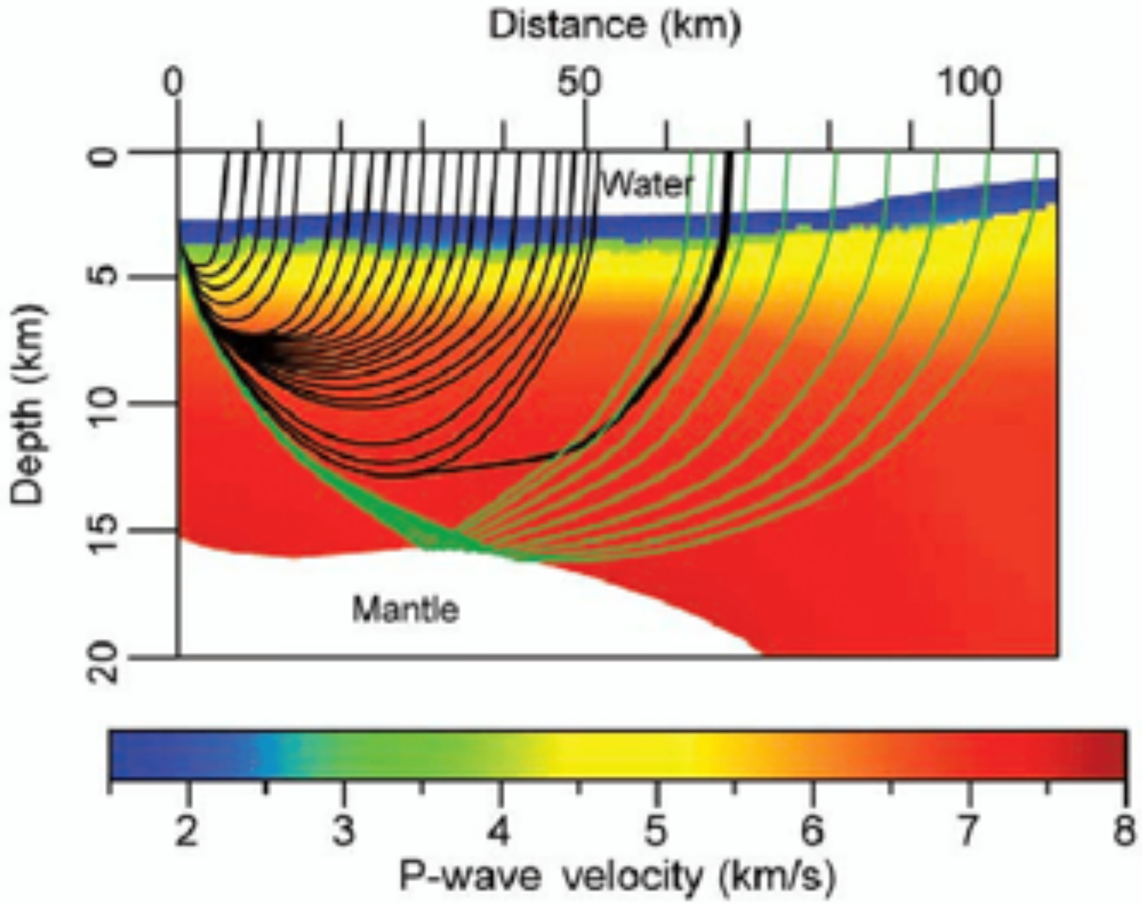
Roadmap

- Motivation and Problem
- History of Automatic First-Break Picking
- History of Machine Learning
- Hardware
- DeepTrace Results
- Phoenix
- Quantifying Confidence and Reliability

Background

why do we need first-break picks?

Near-Surface Modeling

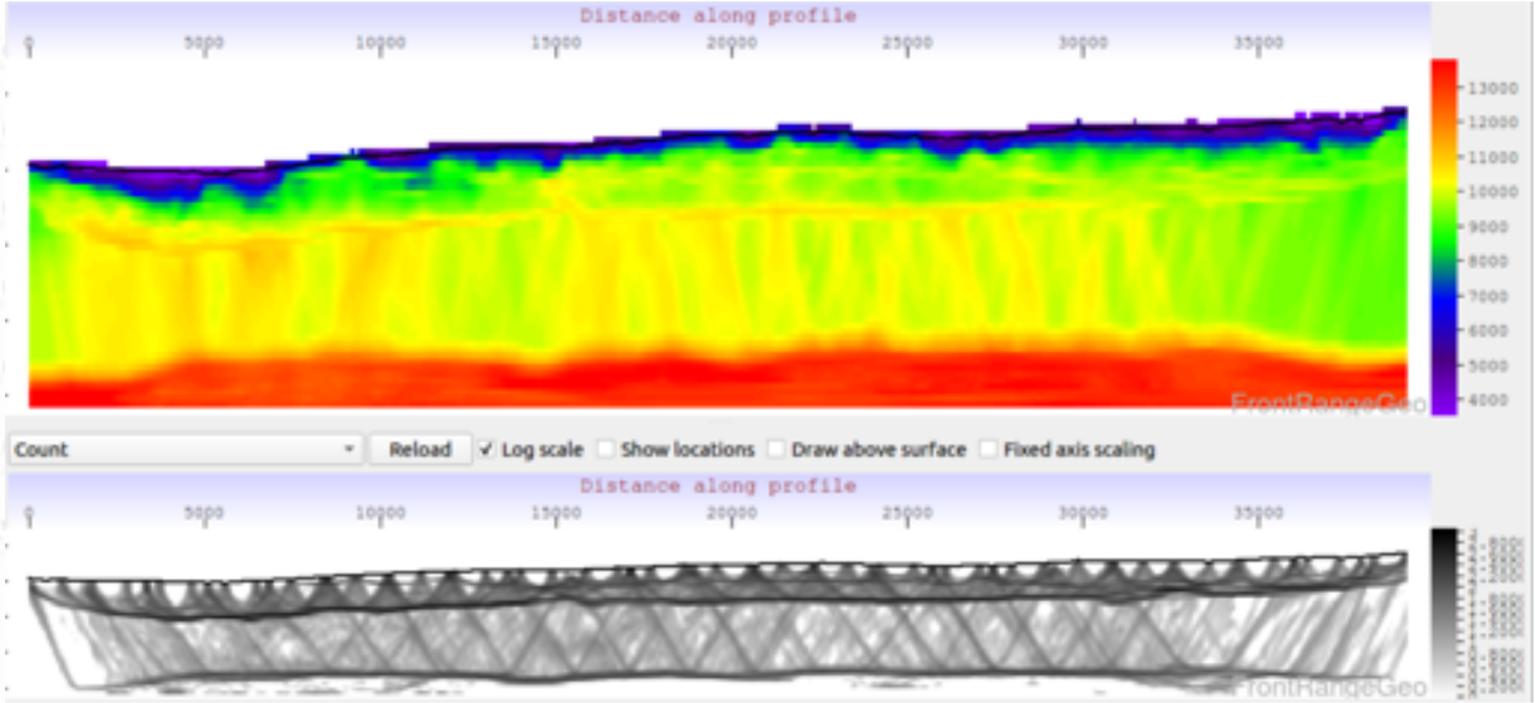


Seismic waves propagate through the earth, producing a seismic record (trace) at each receiver.

Pick the first arrival in seismic record, then invert along ray path to find a velocity model of the earth.

Lau et. al. *Low-frequency source for long-offset, sub-basalt and deep crustal penetration*. The Leading Edge. 2007

Near-Surface Modeling



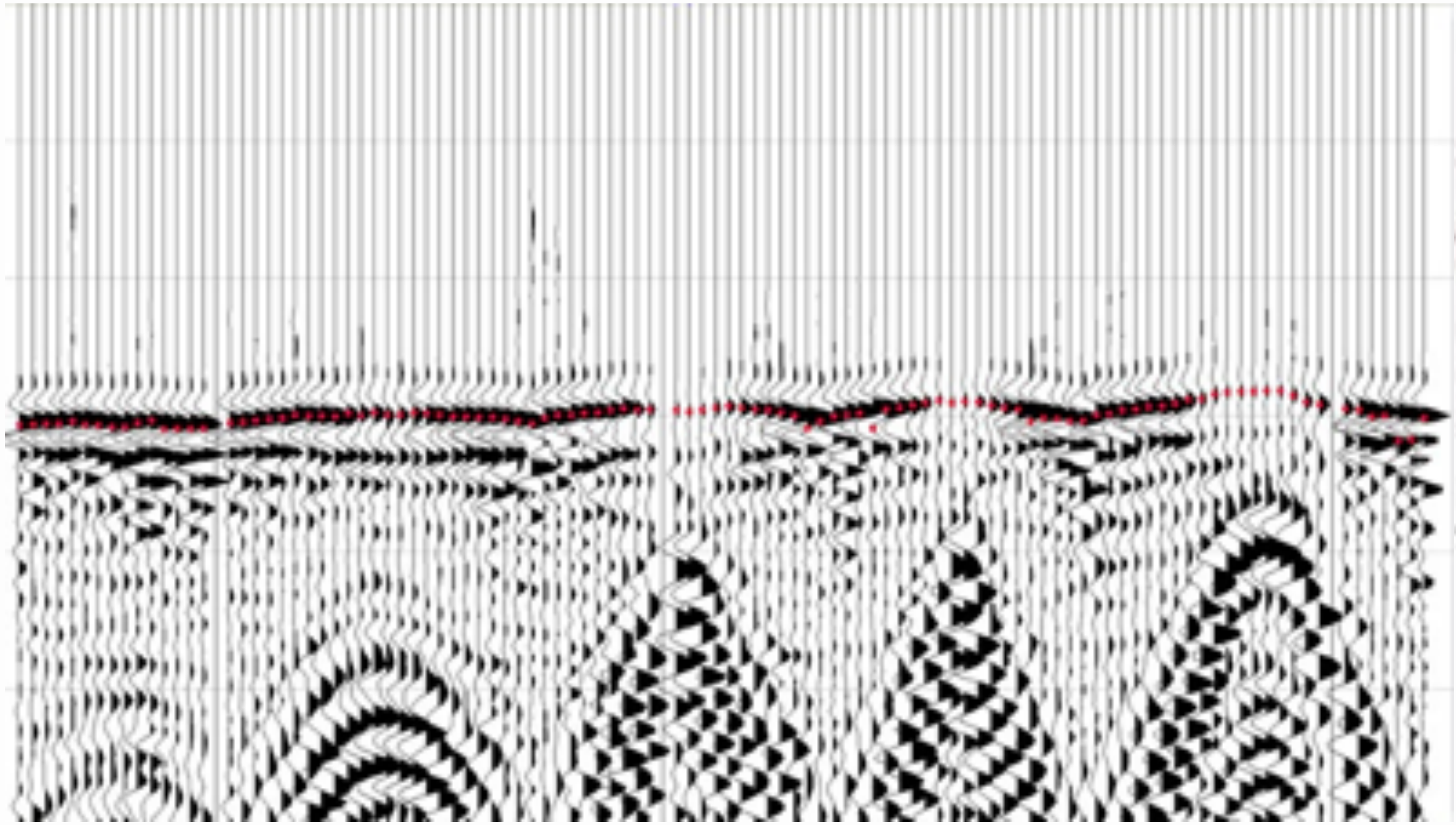
By comparing simulated travel times with picked arrival times, we can iteratively update a tomographic model.

Upper panel: profile of computed velocity field. Lower panel: simulated node hit counts (Teapot Dome Survey)

Static Correction

Final near-surface model used to correct for weathering layer.

Without static shift from near-surface, deeper imaging will have velocity anomalies.



Shot record with tomographic solution applied. Image courtesy XtremeGeo.

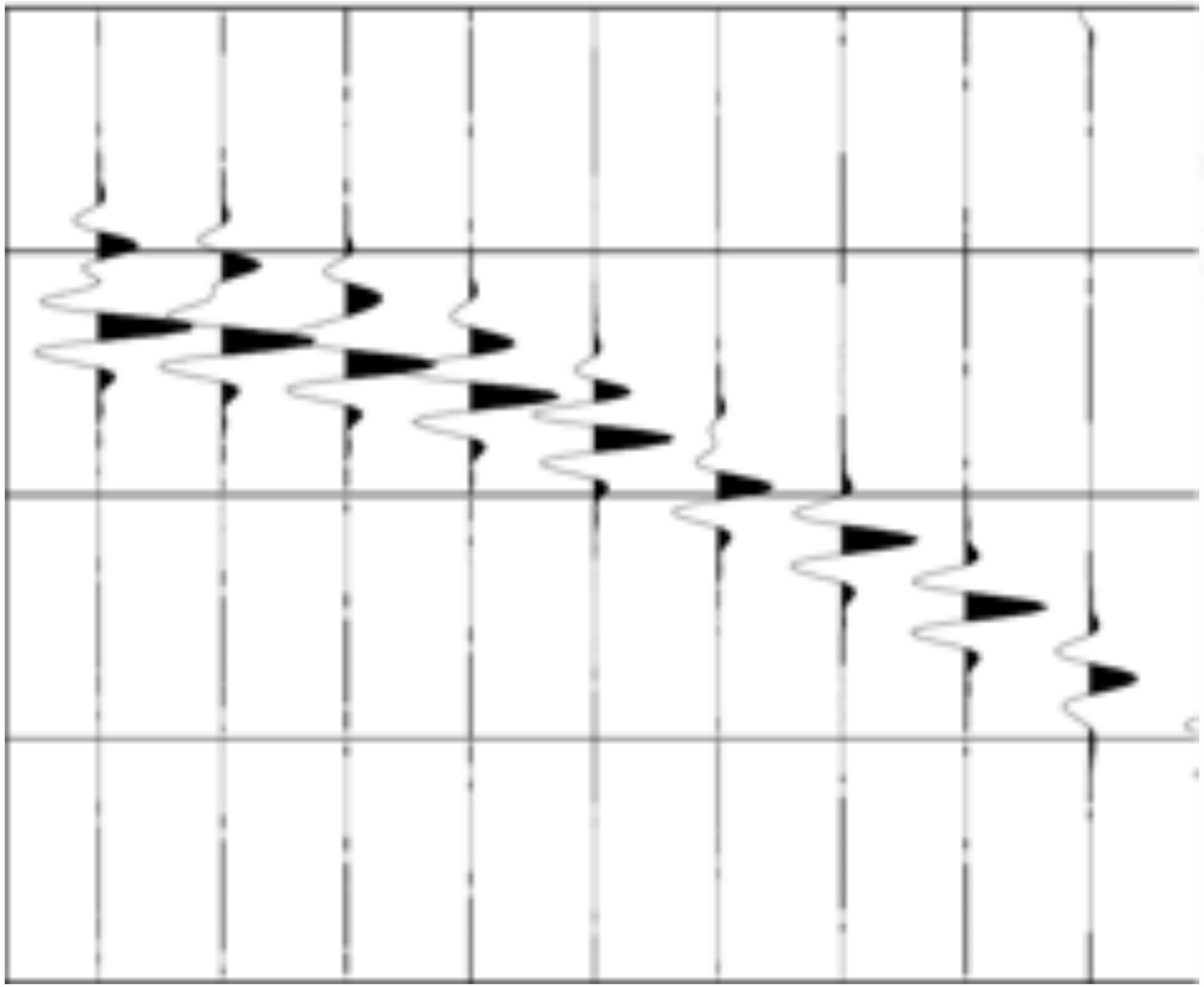
Seismic Traces

Clean traces have easy-to-spot arrival times.

Or do they? What is the arrival time of a wavelet? Is such a notion well-defined?

Peak, trough, zero crossing?

Consistency matters more than any specific event.

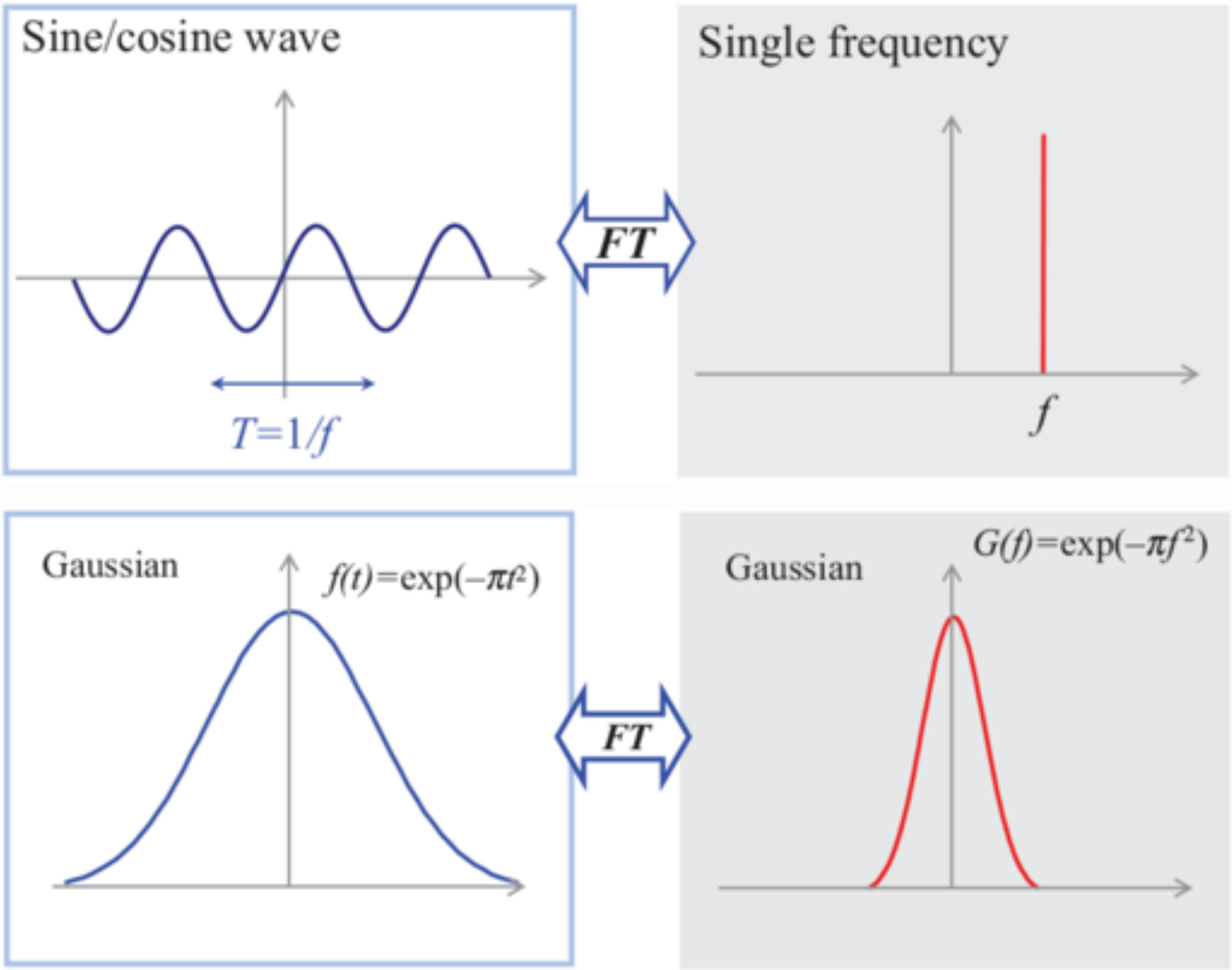


Waves

Seismic waves are spread out in time and frequency domains.

More correct model of arrival time is a *distribution.*

Different component frequencies of initial wavelet are dampened differently.



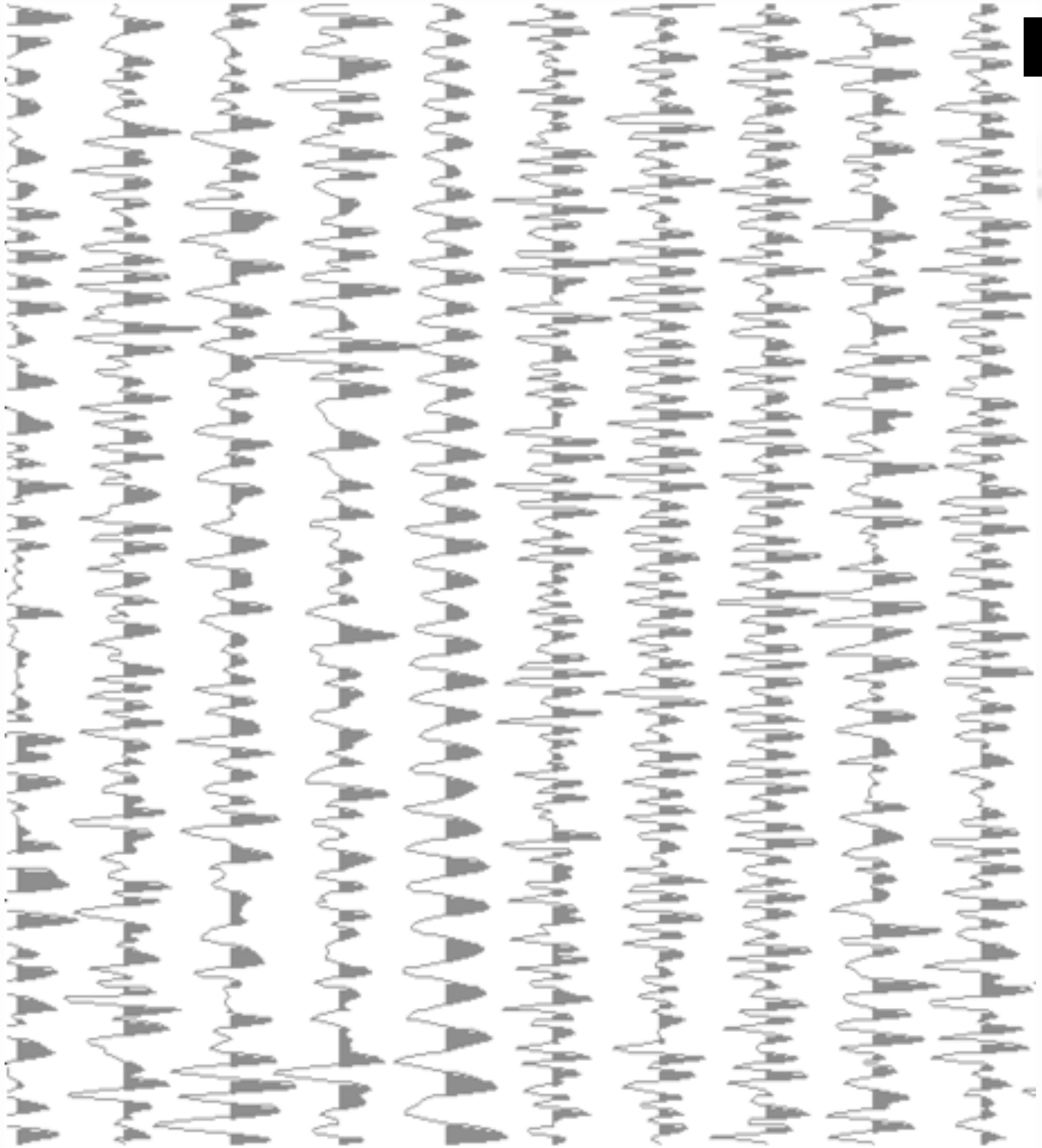
Noisy Seismic

The problem is even worse in the presence of noise.

Not only are wavelet arrival times inherently ambiguous, noise can completely obscure signal.

In difficult conditions, humans must label arrival times ‘by hand’.

This is incredibly time consuming.



Unedited real seismic data

Lenovo

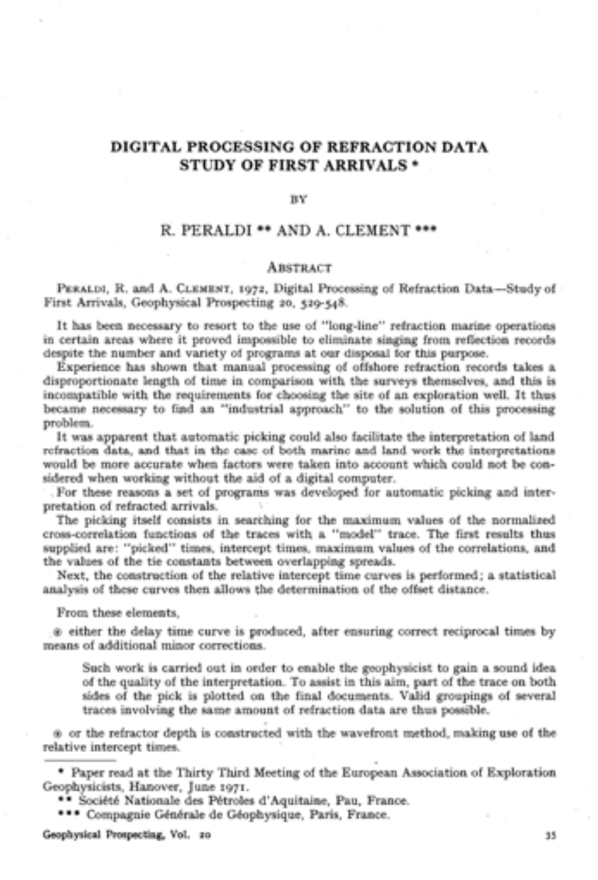
Automatic First-Break Picking

a brief history

Historical Perspective

“Experience has shown that manual processing of […] refraction records takes a disproportionate length of time in comparison with the surveys themselves, and this is incompatible with the requirements for choosing the site of an exploration well. It thus became necessary to find an ‘industrial approach’ to the solution of this processing problem.”

- Peraldi and Clement:
Digital Processing of Refraction Data - Study of First Arrivals.
Geophysical Prospecting. ***1972***



DIGITAL PROCESSING OF REFRACTION DATA
STUDY OF FIRST ARRIVALS *

BY

R. PERALDI ** AND A. CLEMENT ***

ABSTRACT

PERALDI, R. and A. CLEMENT, 1972, Digital Processing of Refraction Data—Study of First Arrivals, Geophysical Prospecting 20, 529-548.

It has been necessary to resort to the use of "long-line" refraction marine operations in certain areas where it proved impossible to eliminate singing from reflection records despite the number and variety of programs at our disposal for this purpose.

Experience has shown that manual processing of offshore refraction records takes a disproportionate length of time in comparison with the surveys themselves, and this is incompatible with the requirements for choosing the site of an exploration well. It thus became necessary to find an "industrial approach" to the solution of this processing problem.

It was apparent that automatic picking could also facilitate the interpretation of land refraction data, and that in the case of both marine and land work the interpretations would be more accurate when factors were taken into account which could not be considered when working without the aid of a digital computer.

For these reasons a set of programs was developed for automatic picking and interpretation of refracted arrivals.

The picking itself consists in searching for the maximum values of the normalized cross-correlation functions of the traces with a "model" trace. The first results thus supplied are: "picked" times, intercept times, maximum values of the correlations, and the values of the tie constants between overlapping spreads.

Next, the construction of the relative intercept time curves is performed; a statistical analysis of these curves then allows the determination of the offset distance.

From these elements,

- either the delay time curve is produced, after ensuring correct reciprocal times by means of additional minor corrections.

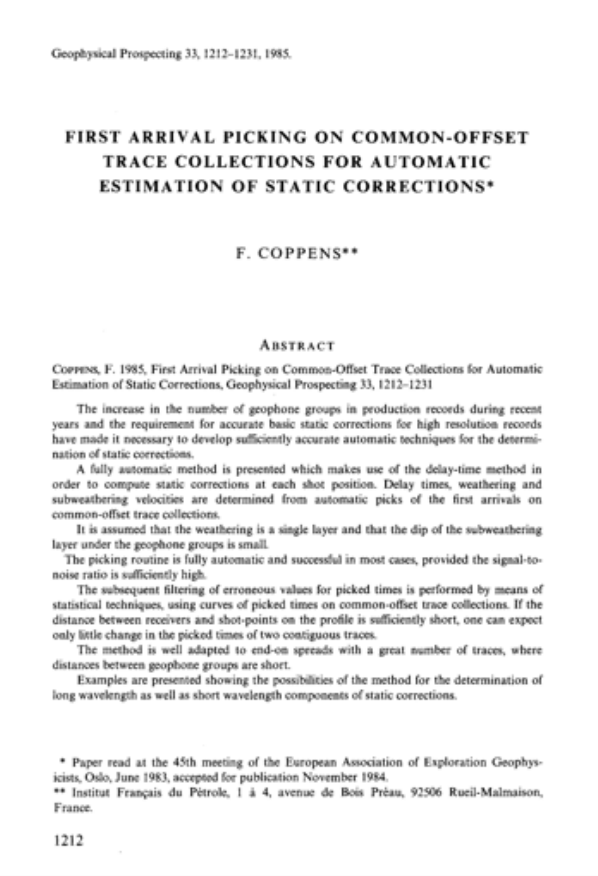
  Such work is carried out in order to enable the geophysicist to gain a sound idea of the quality of the interpretation. To assist in this aim, part of the trace on both sides of the pick is plotted on the final documents. Valid groupings of several traces involving the same amount of refraction data are thus possible.

- or the refractor depth is constructed with the wavefront method, making use of the relative intercept times.

* Paper read at the Thirty Third Meeting of the European Association of Exploration Geophysicists, Hanover, June 1971.
** Société Nationale des Pétroles d'Aquitaine, Pau, France.
*** Compagnie Générale de Géophysique, Paris, France.

Geophysical Prospecting, Vol. 20 35

1972 - Cross Correlations



Geophysical Prospecting 33, 1212–1231, 1985.

FIRST ARRIVAL PICKING ON COMMON-OFFSET
TRACE COLLECTIONS FOR AUTOMATIC
ESTIMATION OF STATIC CORRECTIONS*

F. COPPENS**

ABSTRACT

COPPENS, F. 1985, First Arrival Picking on Common-Offset Trace Collections for Automatic Estimation of Static Corrections, Geophysical Prospecting 33, 1212–1231

The increase in the number of geophone groups in production records during recent years and the requirement for accurate basic static corrections for high resolution records have made it necessary to develop sufficiently accurate automatic techniques for the determination of static corrections.

A fully automatic method is presented which makes use of the delay-time method in order to compute static corrections at each shot position. Delay times, weathering and subweathering velocities are determined from automatic picks of the first arrivals on common-offset trace collections.

It is assumed that the weathering is a single layer and that the dip of the subweathering layer under the geophone groups is small.

The picking routine is fully automatic and successful in most cases, provided the signal-to-noise ratio is sufficiently high.

The subsequent filtering of erroneous values for picked times is performed by means of statistical techniques, using curves of picked times on common-offset trace collections. If the distance between receivers and shot-points on the profile is sufficiently short, one can expect only little change in the picked times of two contiguous traces.

The method is well adapted to end-on spreads with a great number of traces, where distances between geophone groups are short.

Examples are presented showing the possibilities of the method for the determination of long wavelength as well as short wavelength components of static corrections.

* Paper read at the 45th meeting of the European Association of Exploration Geophysicists, Oslo, June 1983, accepted for publication November 1984.
** Institut Français du Pètrole, 1 à 4, avenue de Bois Préau, 92506 Rueil-Malmaison, France.

1212

1985 - Energy Ratio

GEOPHYSICS, VOL. 61, NO. 4 (JULY-AUGUST 1996); P. 1095–1102, 8 FIGS.

A fractal-based algorithm for detecting first arrivals on seismic traces

Fabio Boschetti*, Mike D. Dentith‡, and Ron D. List**

ABSTRACT

A new algorithm is proposed for the automatic picking of seismic first arrivals that detects the presence of a signal by analyzing the variation in fractal dimension along the trace. The "divider-method" is found to be the most suitable method for calculating the fractal dimension. A change in dimension is found to occur close to the transition from noise to signal plus noise, that is the first arrival. The nature of this change varies from trace to trace, but a detectable change is always found to occur. The algorithm has been tested on real data sets with varying S/N ratios and the results compared to those obtained using previously published algorithms. With an appropriate tuning of its parameters, the fractal-based algorithm proved more accurate than all these other algorithms, especially in the presence of significant noise. The fractal method proved able to tolerate noise up to 80% of the average signal amplitude. However, the fractal-based algorithm is considerably slower than the other methods and hence is intended for use only on data sets with low S/N ratios.

INTRODUCTION

The accurate determination of the traveltime of seismic energy from source to receiver is of fundamental importance in seismic surveying. This is particularly the case with seismic refraction and tomographic surveys where traveltimes, usually of first arrivals, are used to determine the seismic-velocity structure of the subsurface. To improve efficiency and speed of interpretation of such data it is common to use an automated technique for detecting seismic events, and several such algorithms have been published. As larger and larger data sets are now being used for such interpretations, these automatic methods of detecting seismic arrivals have become an essential part of the processing of seismic data.

Fundamentally, detection of first-arriving seismic data reduces to determining the time when the seismic trace ceases to be composed entirely of noise and also starts to include seismic signal. When such an operation is carried out manually, a subjective decision is made based on the change in the nature of the trace in terms of amplitude and/or frequency and/or phase both within the trace itself and also relative to its neighbors. However, what is a relatively simple operation for the human eye and brain is much more difficult to define mathematically and translate into an algorithm.

Several methods for locating a first break have been published (Coppens, 1985, Ervin et al., 1983, Gelchinsky and Shtivelman, 1983, Peraldi and Clement, 1972, Ramananantoandro and Bernitsas, 1987). Most of the methods are based on identifying a particular property of that part of the trace where the first arrival occurs. Some methods also rely on comparison of the trace with its immediate neighbors. The different methods proposed to detect first arrivals will give slightly different arrival times depending on exactly what property of the trace they are based on, but, in general, are extremely effective provided there is an adequate signal-to-noise (S/N) ratio. However, in a situation of very low S/N ratio, their accuracy may be affected seriously.

In this paper, we propose a new method of picking seismic first arrivals in noisy data sets based on the change in fractal dimension within the trace associated with the advent of the signal. Since fractal dimension can be thought of as measuring the "roughness", i.e., the overall shape, of the trace, the algorithm automatically simulates the way the human brain identifies the first arrival.

CALCULATION OF FRACTAL DIMENSION

Since its original introduction by Mandelbrot (1967) the concept of fractals and fractal dimension has found widespread applications in many fields including the earth sciences. For the definition and an extensive description of the concepts behind fractals the reader is referred to Feder (1988), Kaye (1989), Mandelbrot (1977, 1983) and Mandelbrot (1983), while their

Manuscript received by the Editor November 23, 1994; revised manuscript received August 28, 1995.
*Department of Geology and Geophysics and Department of Mathematics, University of Western Australia, Nedlands, Perth WA 6907.
‡Department of Geology and Geophysics, University of Western Australia, Nedlands, Perth WA 6907.
**Department of Mathematics, University of Western Australia, Nedlands, Perth WA 6907.
© 1996 Society of Exploration Geophysicists. All rights reserved.

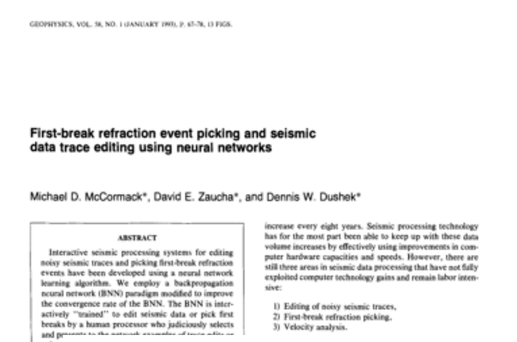
1095

1996 - Fractal Dimension

...

- A progression of more and more sophisticated methods.
- All claim method works as well as humans in noisy areas.
- Many note that their method 'naturally mimics the human eye'.

Neural Networks Appear



GEOPHYSICS, VOL. 58, NO. 1 (JANUARY 1993); P. 67-78, 13 FIGS.

First-break refraction event picking and seismic data trace editing using neural networks

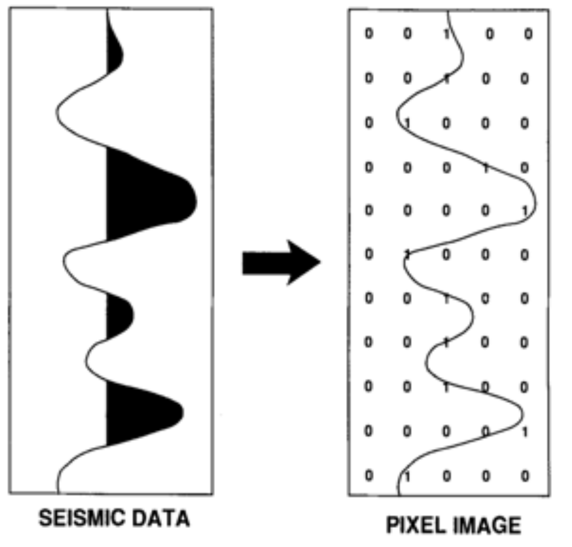
Michael D. McCormack*, David E. Zaucha*, and Dennis W. Dushek*

ABSTRACT

Interactive seismic processing systems for editing noisy seismic traces and picking first-break refraction events have been developed using a neural network learning algorithm. We employ a backpropagation neural network (BNN) paradigm modified to improve the convergence rate of the BNN. The BNN is interactively "trained" to edit seismic data or pick first breaks by a human processor who judiciously selects and pr...

increase every eight years. Seismic processing technology has for the most part been able to keep up with these data volume increases by effectively using improvements in computer hardware capacities and speeds. However, there are still three areas in seismic data processing that have not fully exploited computer technology gains and remain labor intensive:

1) Editing of noisy seismic traces,
2) First-break refraction picking,
3) Velocity analysis.



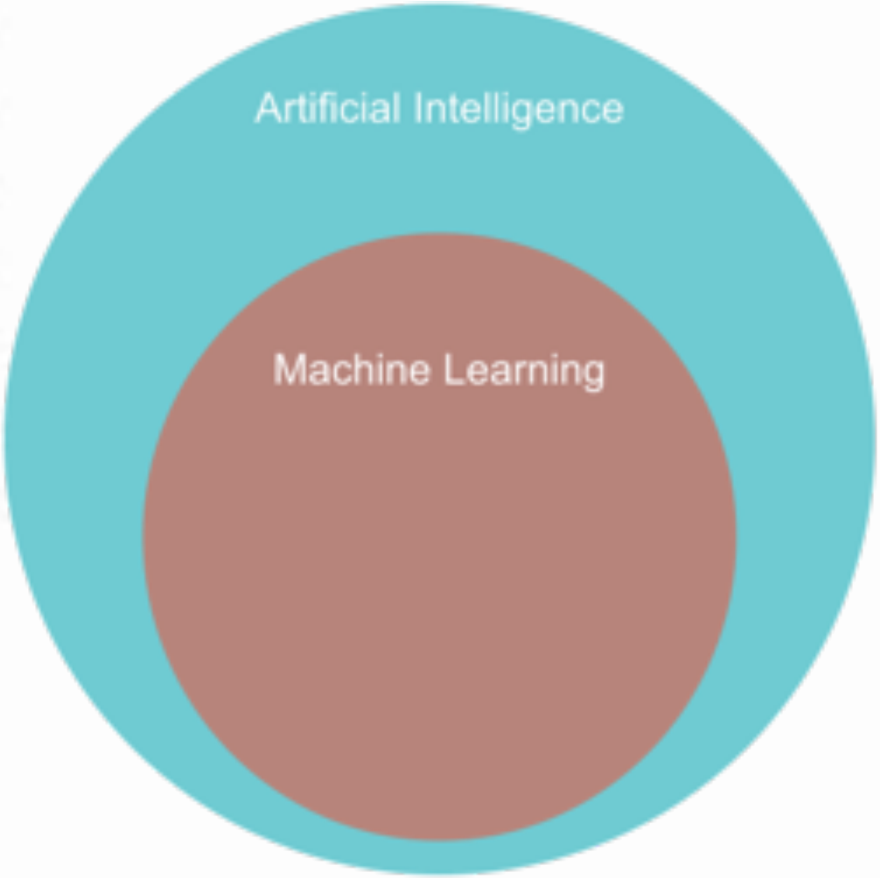
McCormack et al. (1993) describe the first implementation for a neural network to automate first-break picking.

Subsequent papers focus on the 'feature engineering' approach to neural network classification.

Machine Learning

a primer

A.I. vs Machine Learning

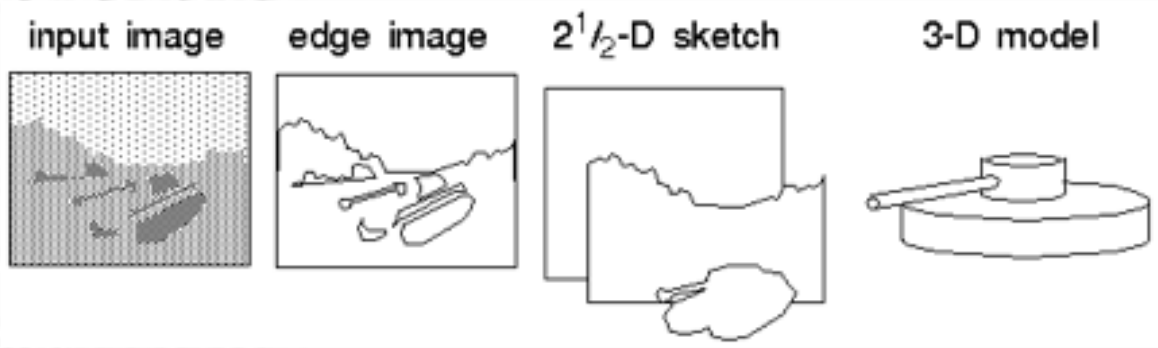


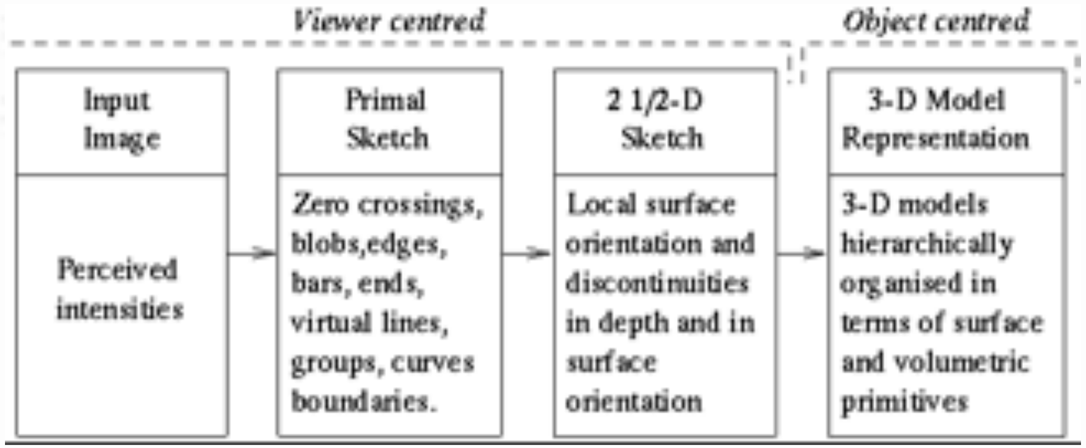
Working definitions:

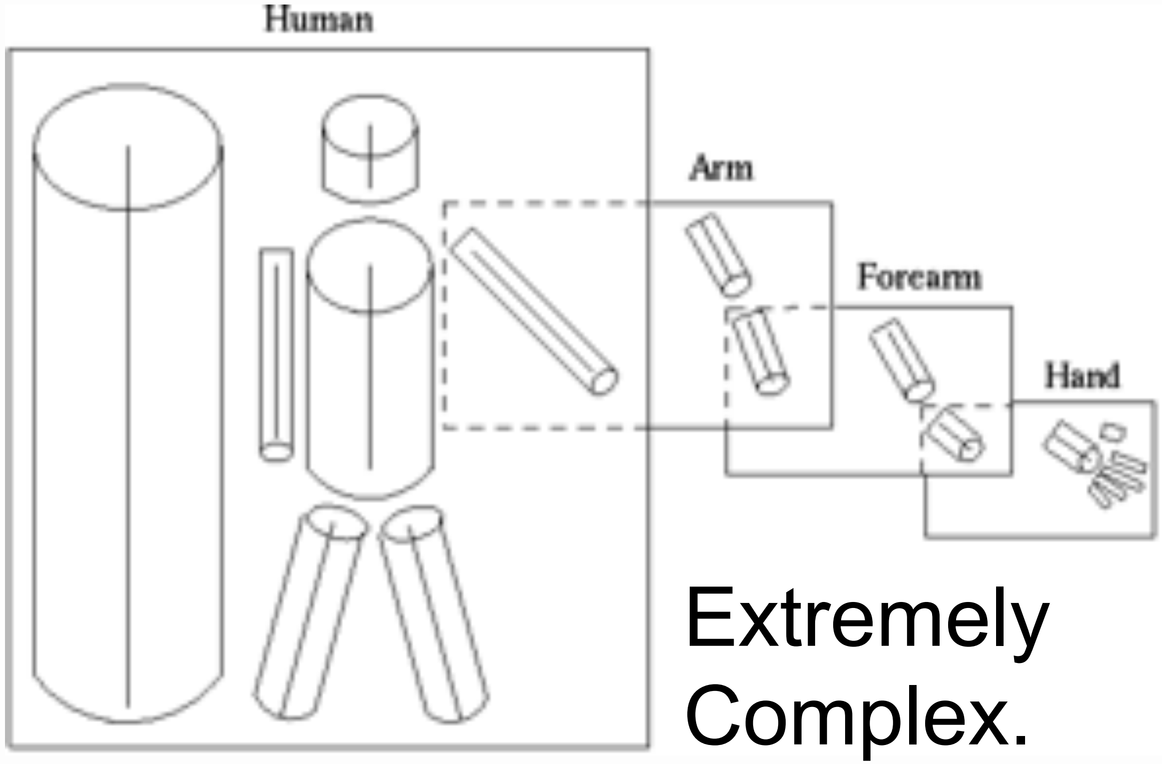
A.I. - Making computers perform tasks that humans traditionally perform well.

Machine Learning - An approach to A.I. where machines are given data, and come up with a mathematical model to fit the data “on their own”.

Old A.I.:
Classical Visual Recognition Stack



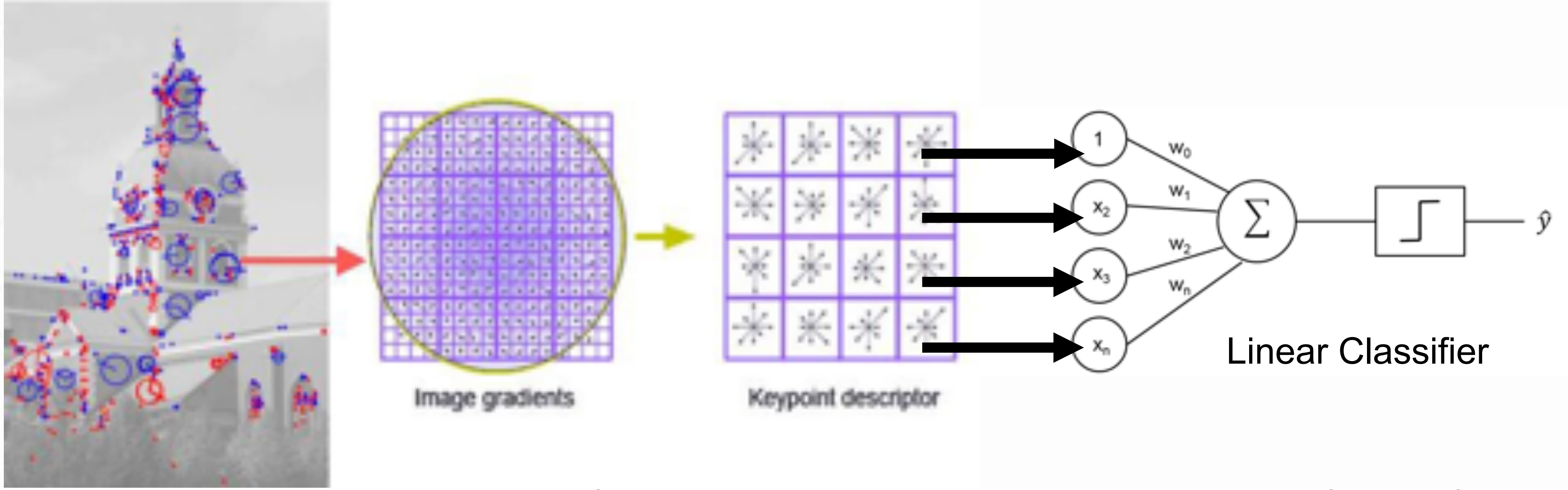




Extremely Complex.

Images from David Marr's *Vision*. 1982

Old A.I.: Feature Extraction

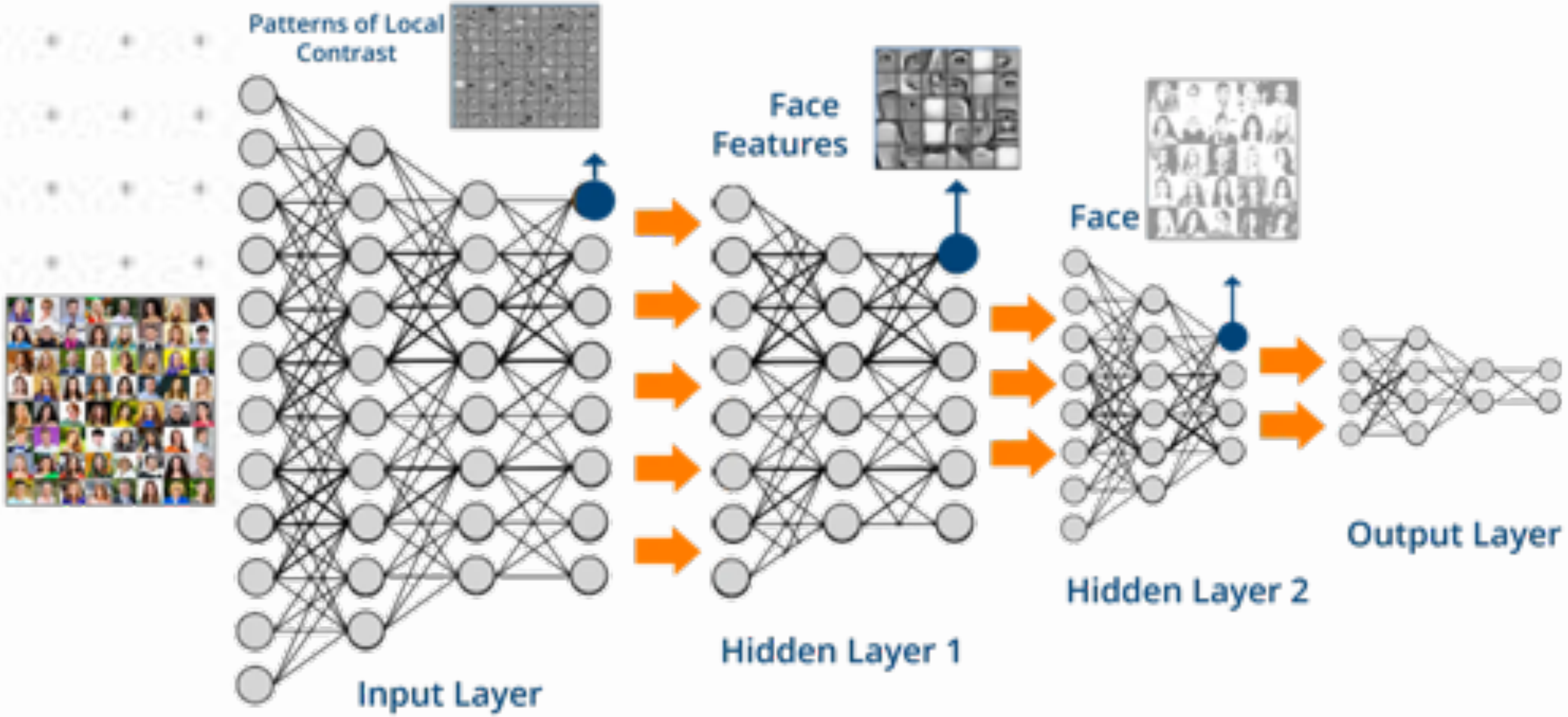


Instead of developing entire recognition stack, develop a few key features, and do statistics with them. Sometimes *many* features, extremely large state vectors.

Some recent first break picking papers focus on this (Hollander et.al 2018).

I tried this approach myself.

Modern A.I.: Deep Learning



Forgo features entirely, let A.I. learn its own representation.

Deeper layers tend to correspond to more abstract features.

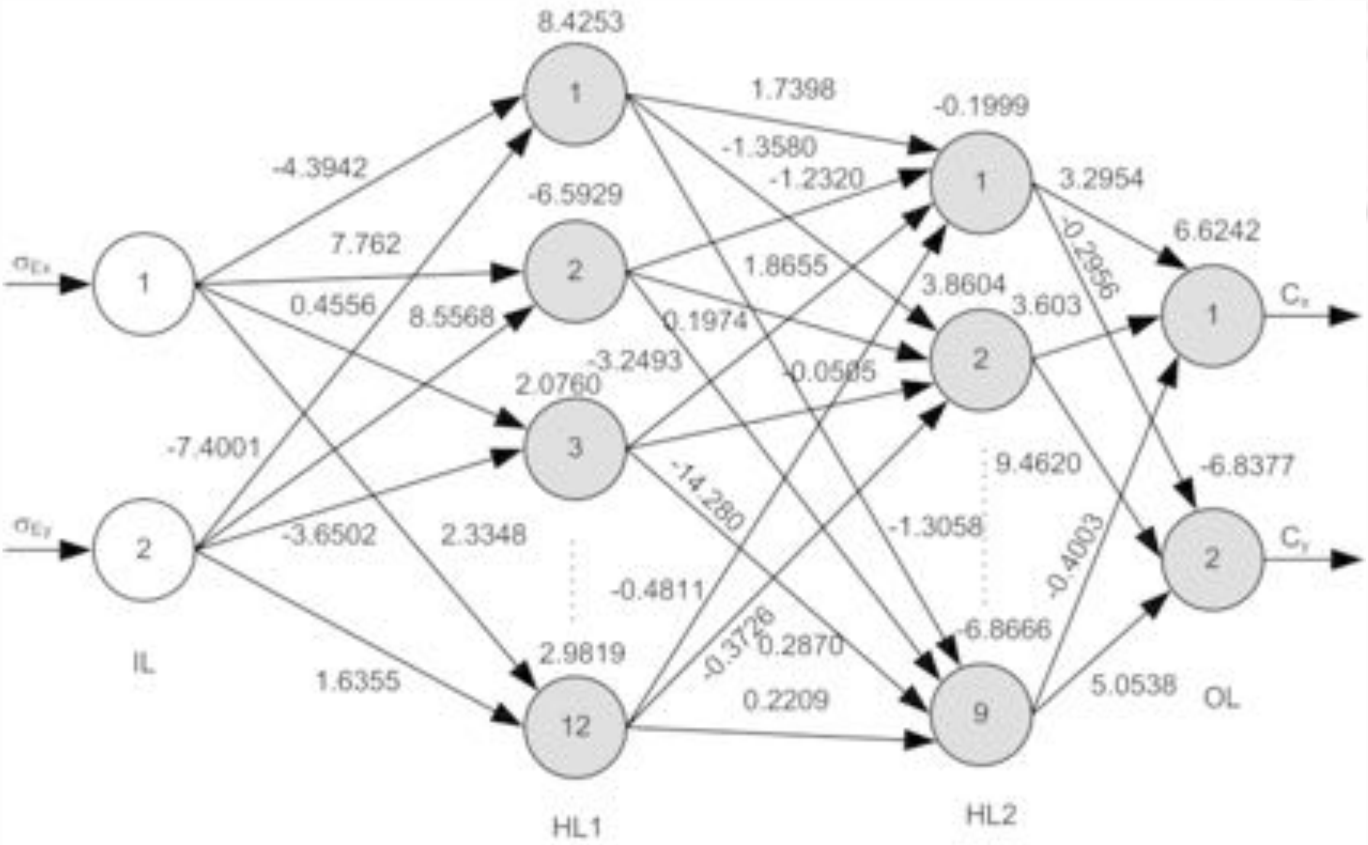
Implemented using a neural network.

Lenovo

Neural Networks

Neural Networks are:

- An example of a machine learning algorithm.
- Nonlinear functions (given an input, produce unique output).
- Randomly initialized, so start off knowing absolutely nothing.
- Adjusted via back-propagation and stochastic gradient descent (SGD).

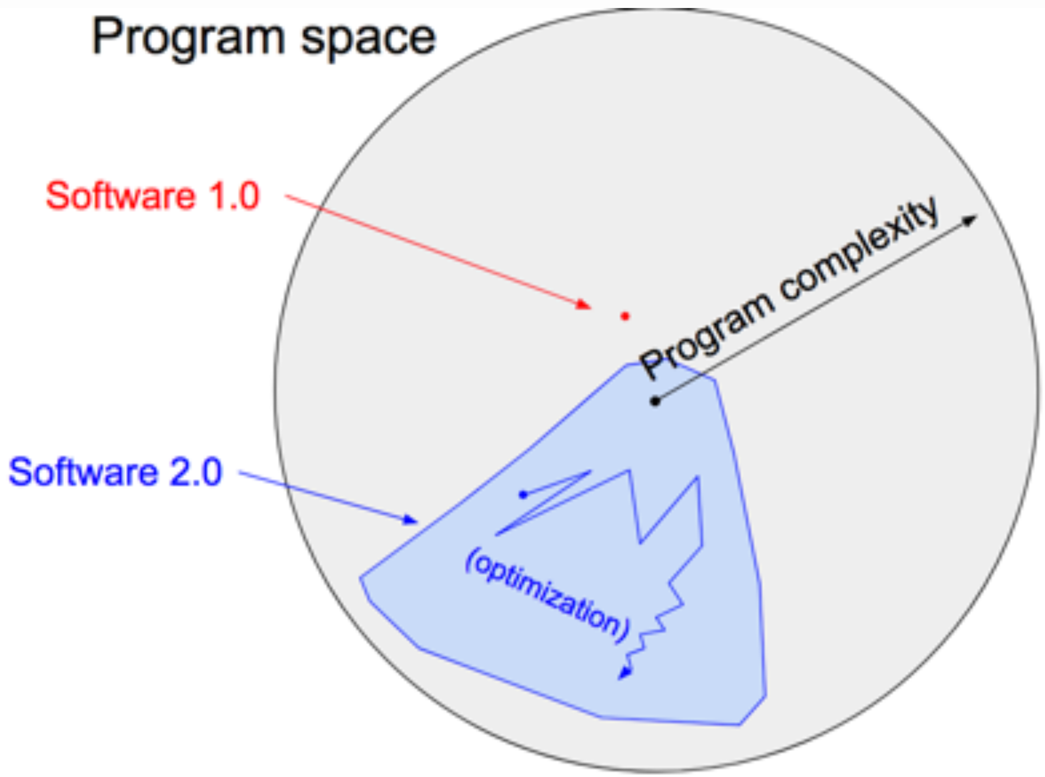


Andrej Karpathy. *Software 2.0.* Medium. 2017

Programming 2.0

In Programming 1.0 we identify a point in program space with desirable behavior by writing lines of code.

In 2.0 we specify the search space and let the optimizer find the best program (as represented by a neural network).



Andrej Karpathy. *Software 2.0*. Medium. 2017

Hardware

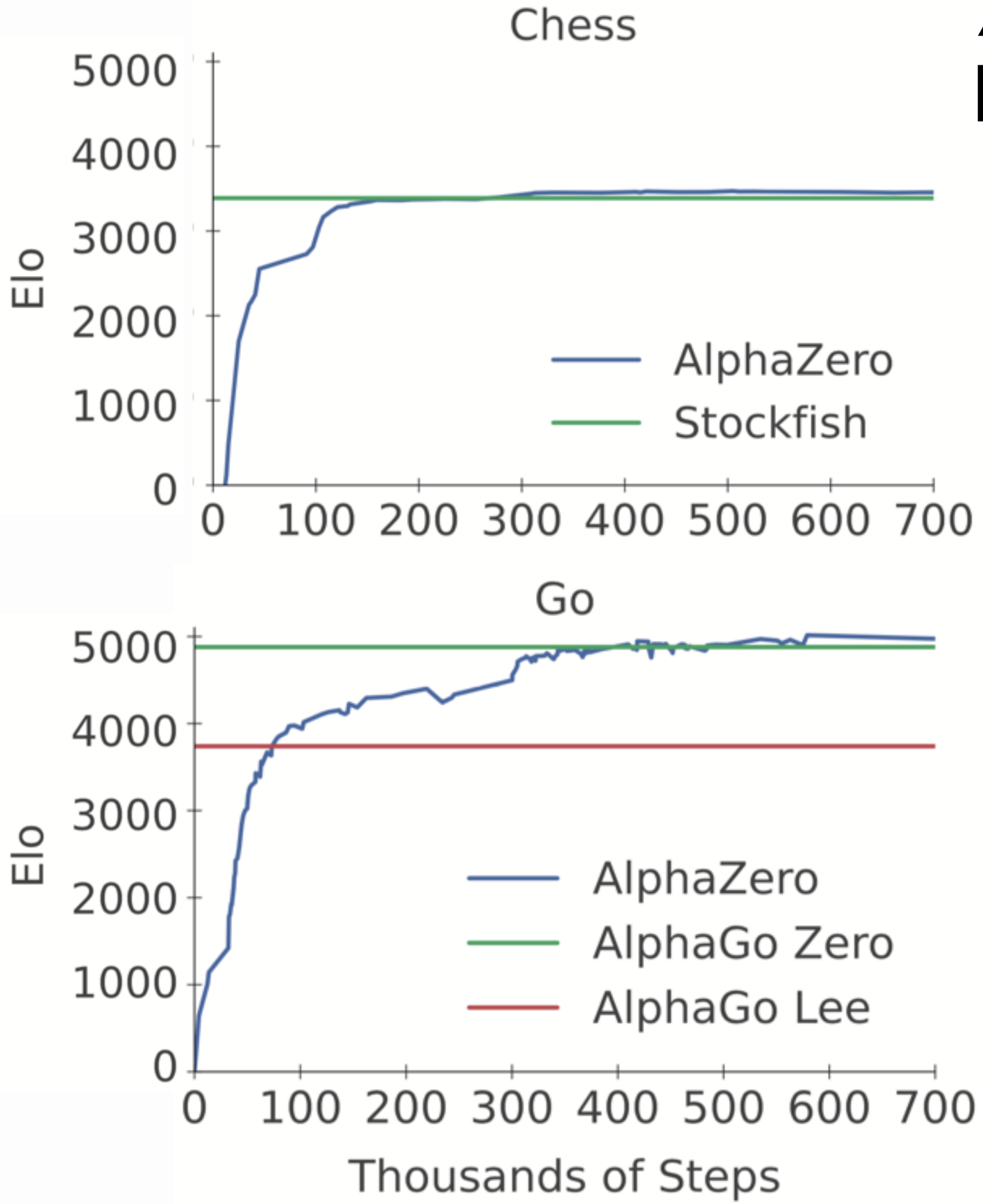
the fuel for modern machine learning

Performance Scales with Data

Generic deep learning result:

More data = higher accuracy and better generalization.

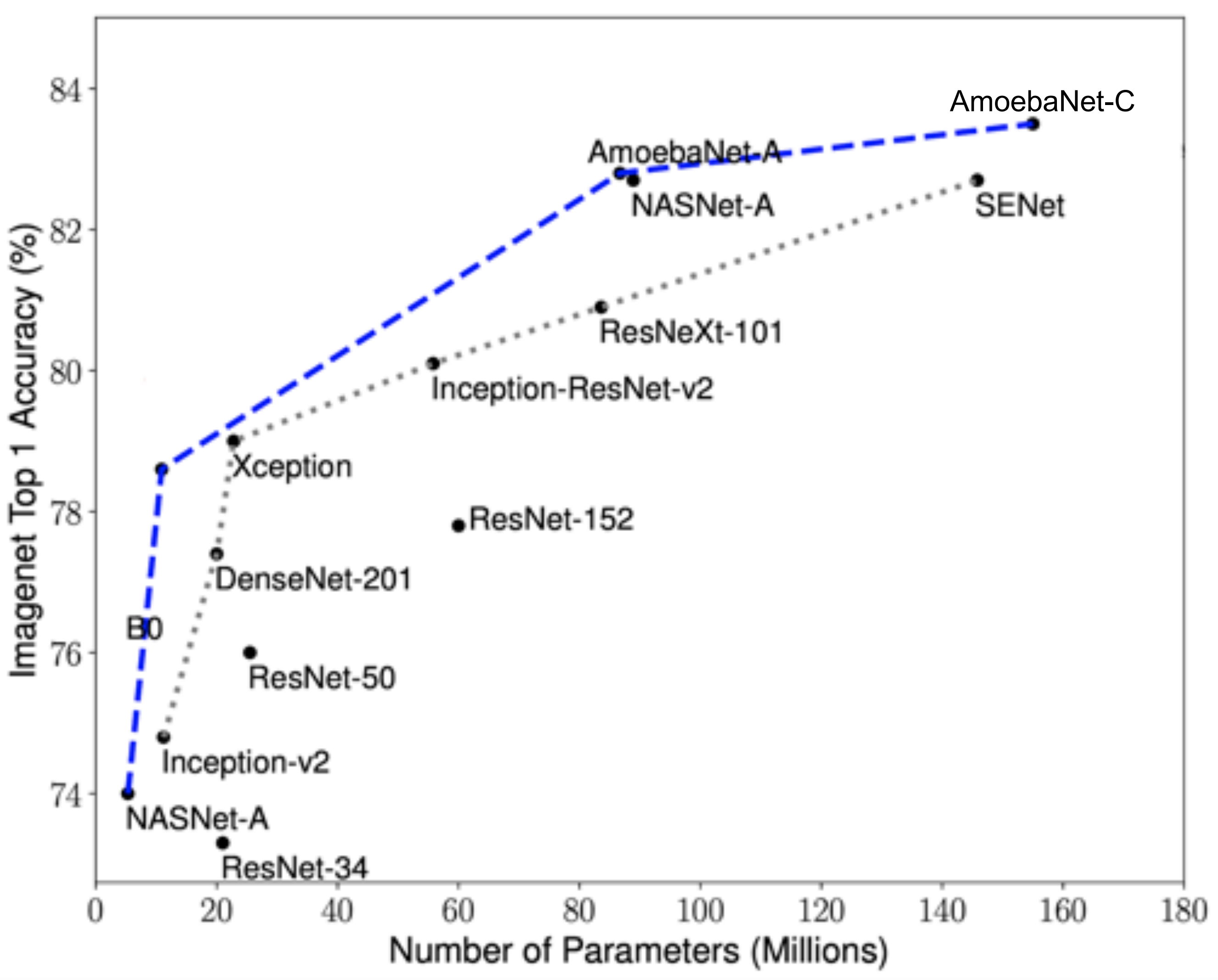
ImageNet, AlphaGo.



Performance Scales with Model Size

Neural nets aren’t new, we just haven’t had the computing and data scales to make them useful until recently.

The network size needed to get good image recognition results is around 10^7 neurons.

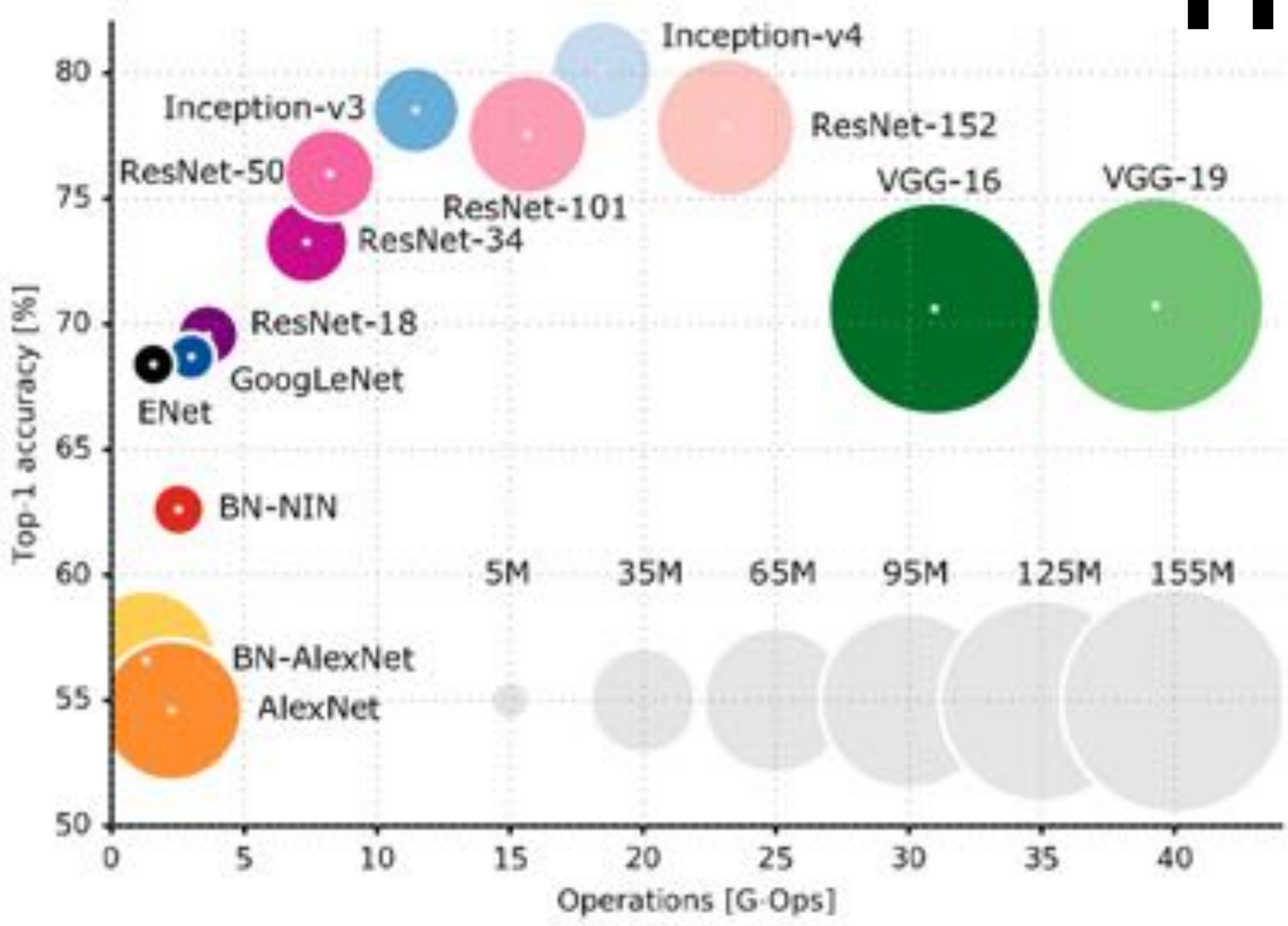


Canziani et. al. *An Analysis of Deep Neural Network Models for Practical Applications.* 2016

GPU Compute

GPUs can do many more operations in parallel than comparably priced CPUs.

Modern GPUs enable the FLOPS necessary to predict on modern large surveys in reasonable time, using deep neural networks.

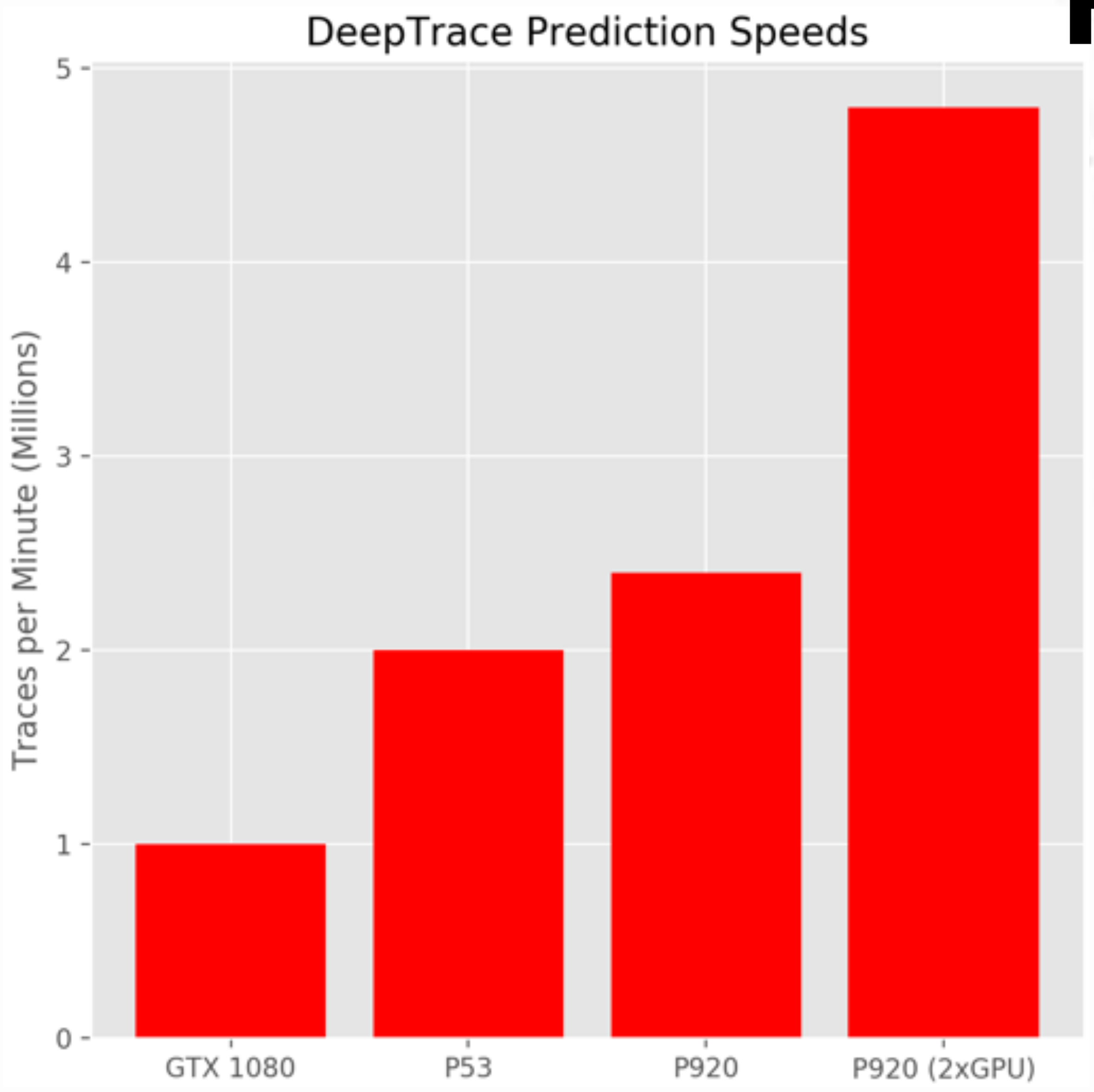


Canziani et. al. *An Analysis of Deep Neural Network Models for Practical Applications.* 2016

Lenovo Benchmarks

Running the fastest DeepTrace models, we can predict on ~2.5 million traces/minute.

With a modest 6 GPU cluster, we can get human-level predictions at nearly 1 billion traces per hour.

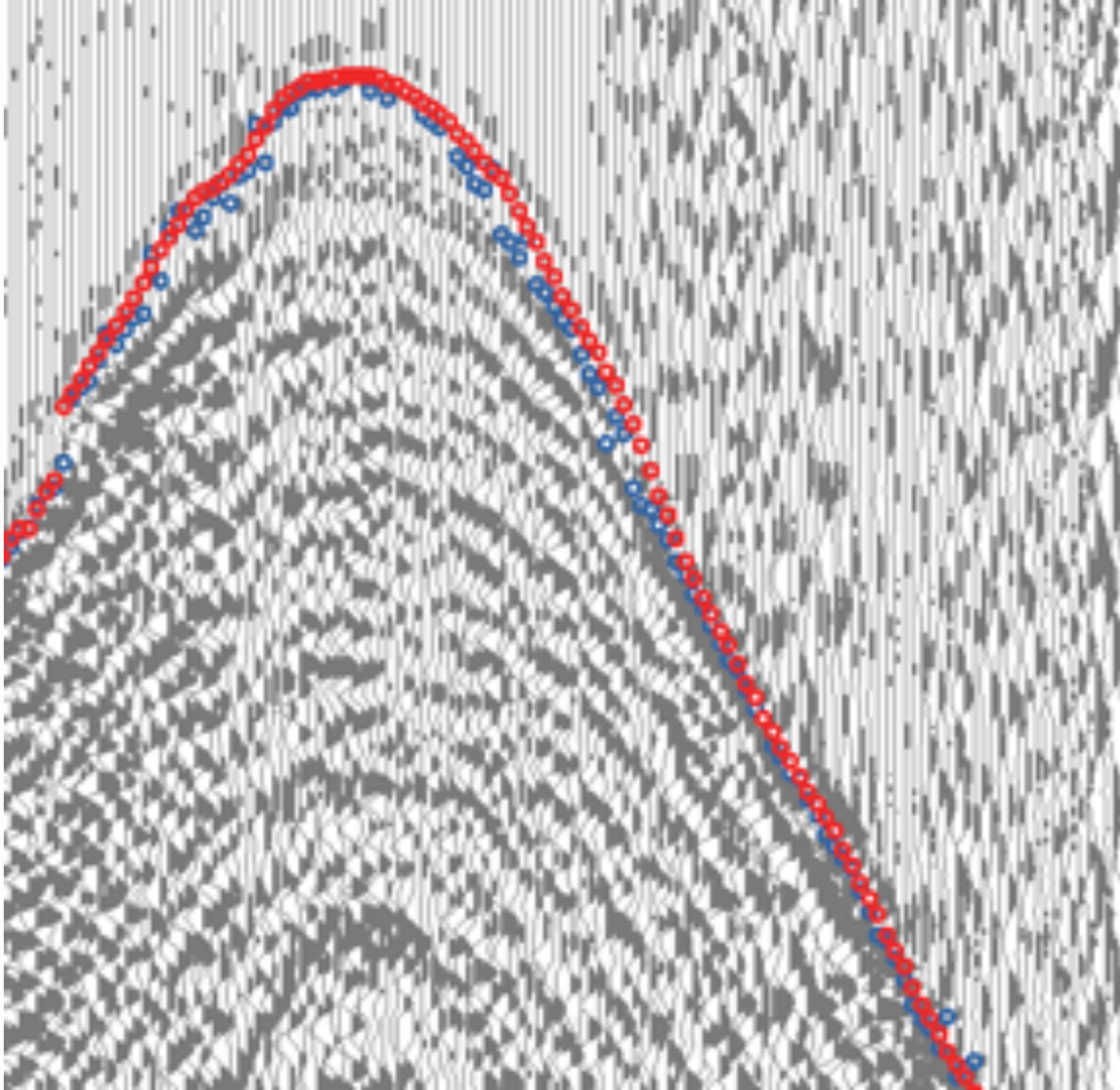


DeepTrace Results

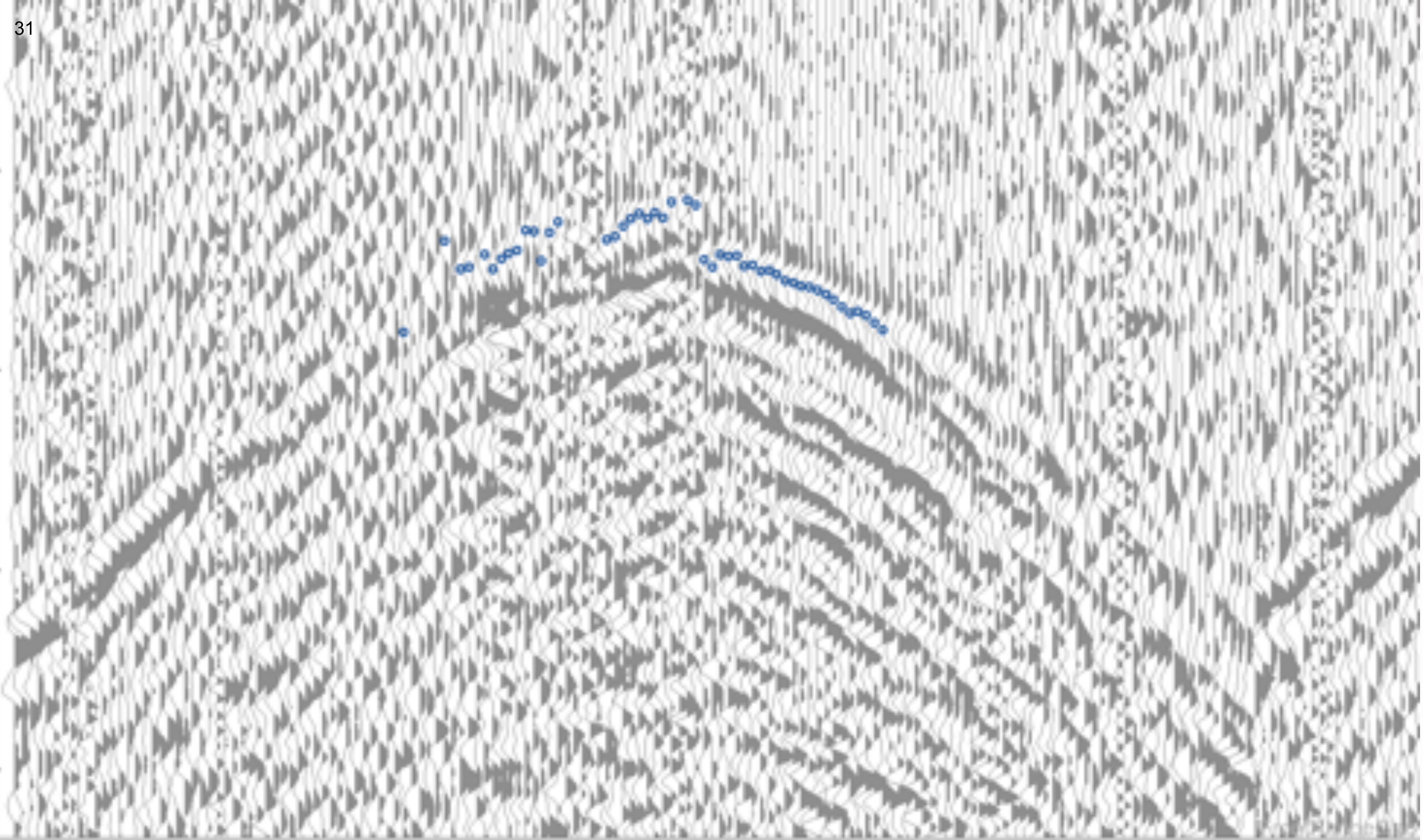
in collaboration with XtremeGeo

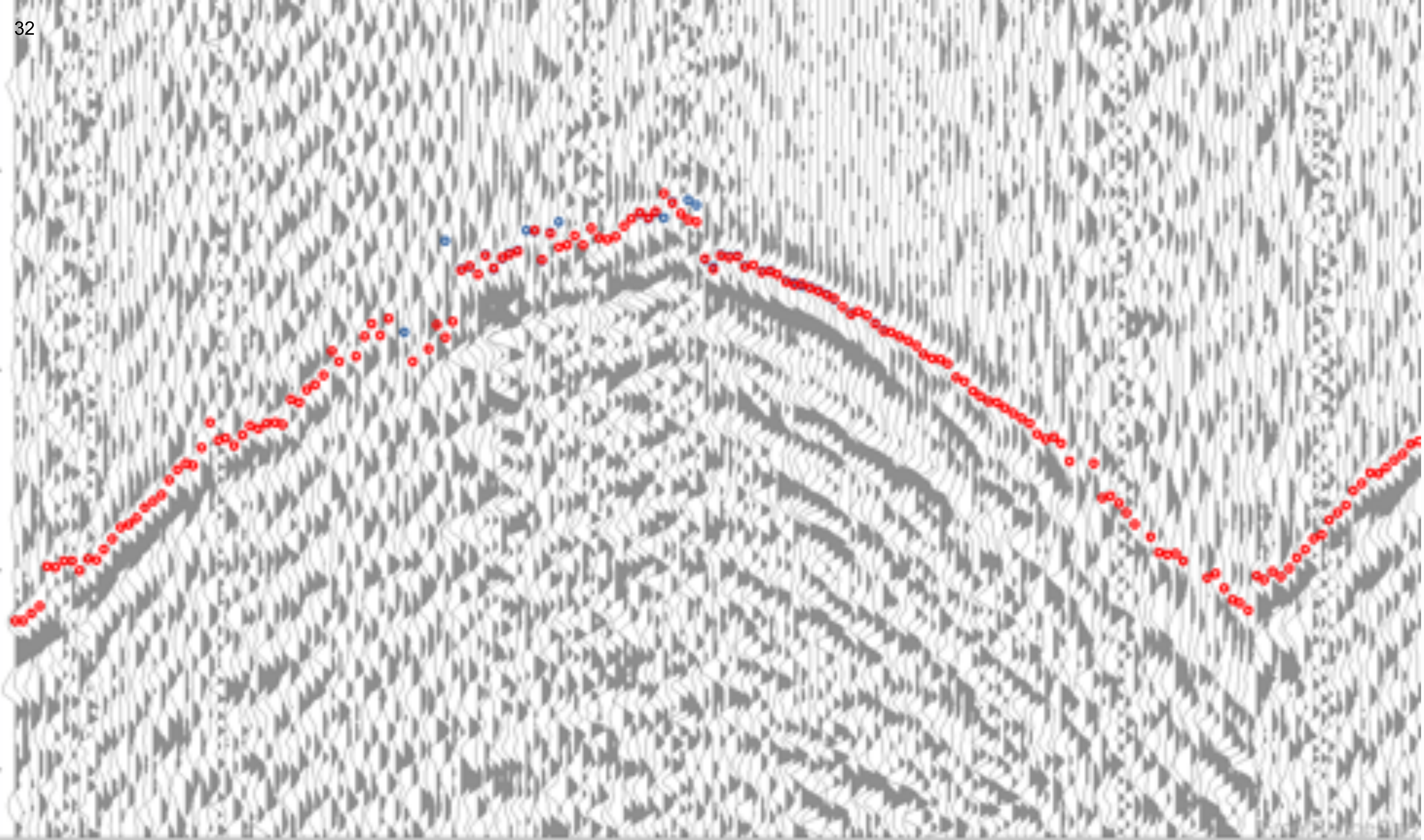
Noisy Seismic Data

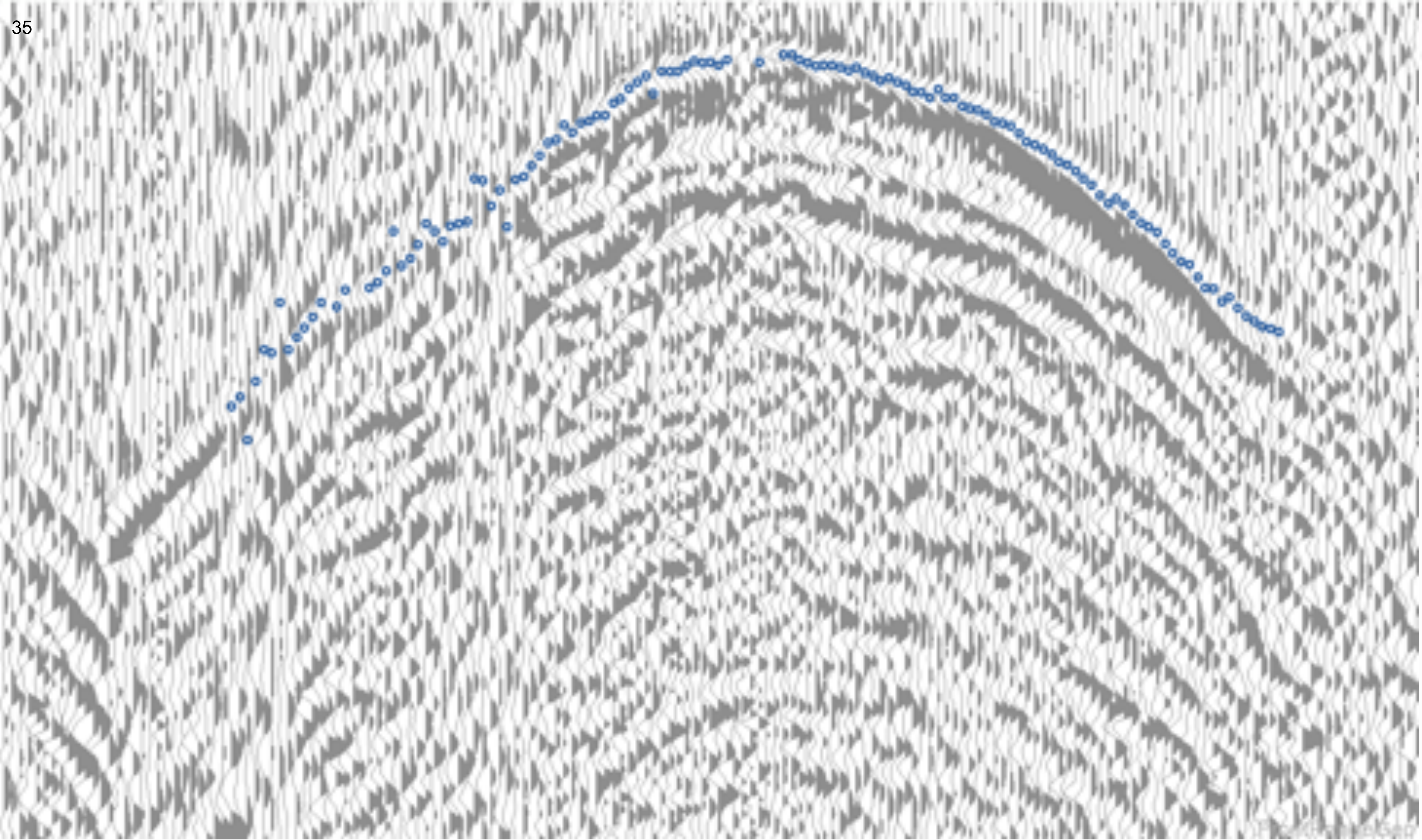
- Human
- DeepTrace

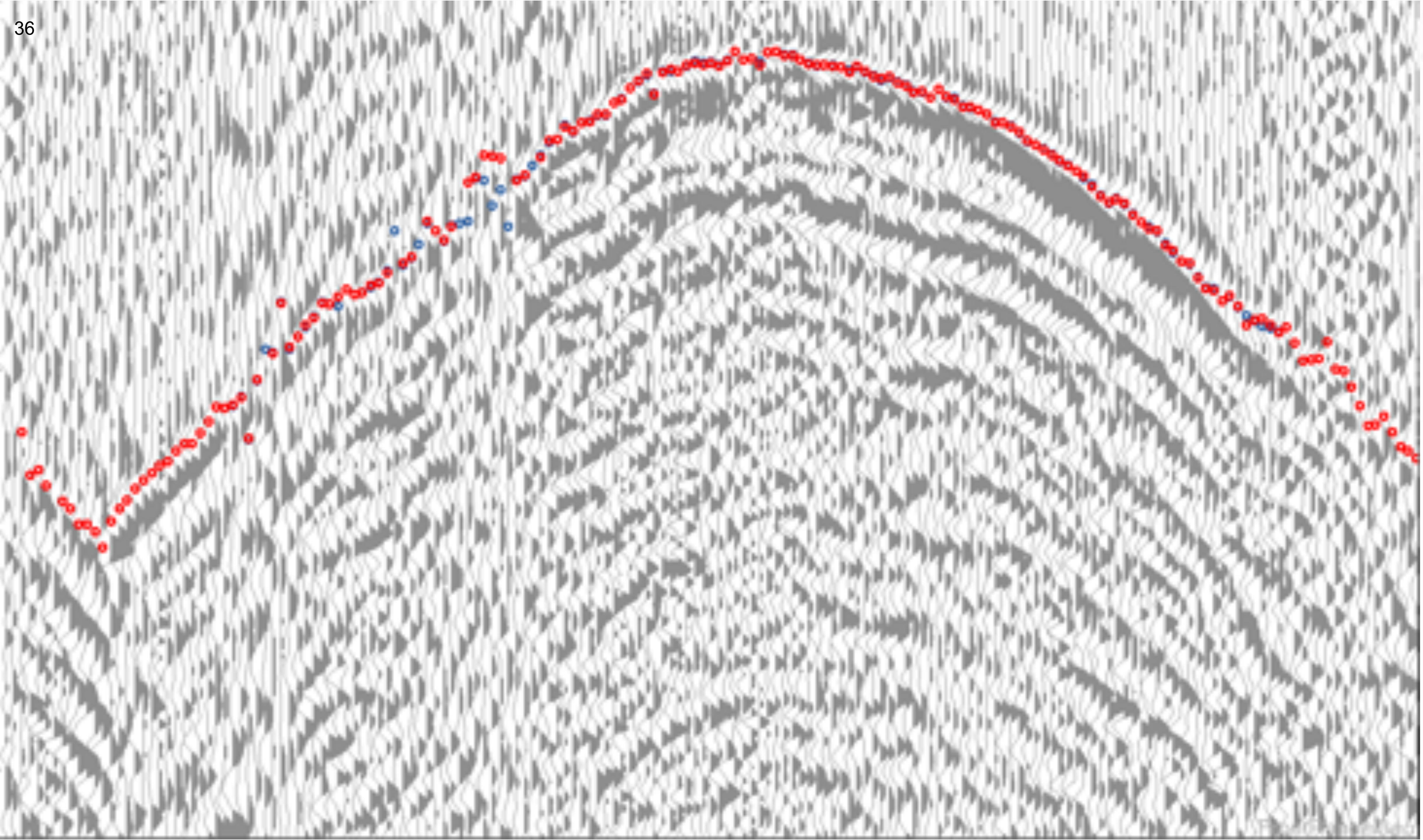


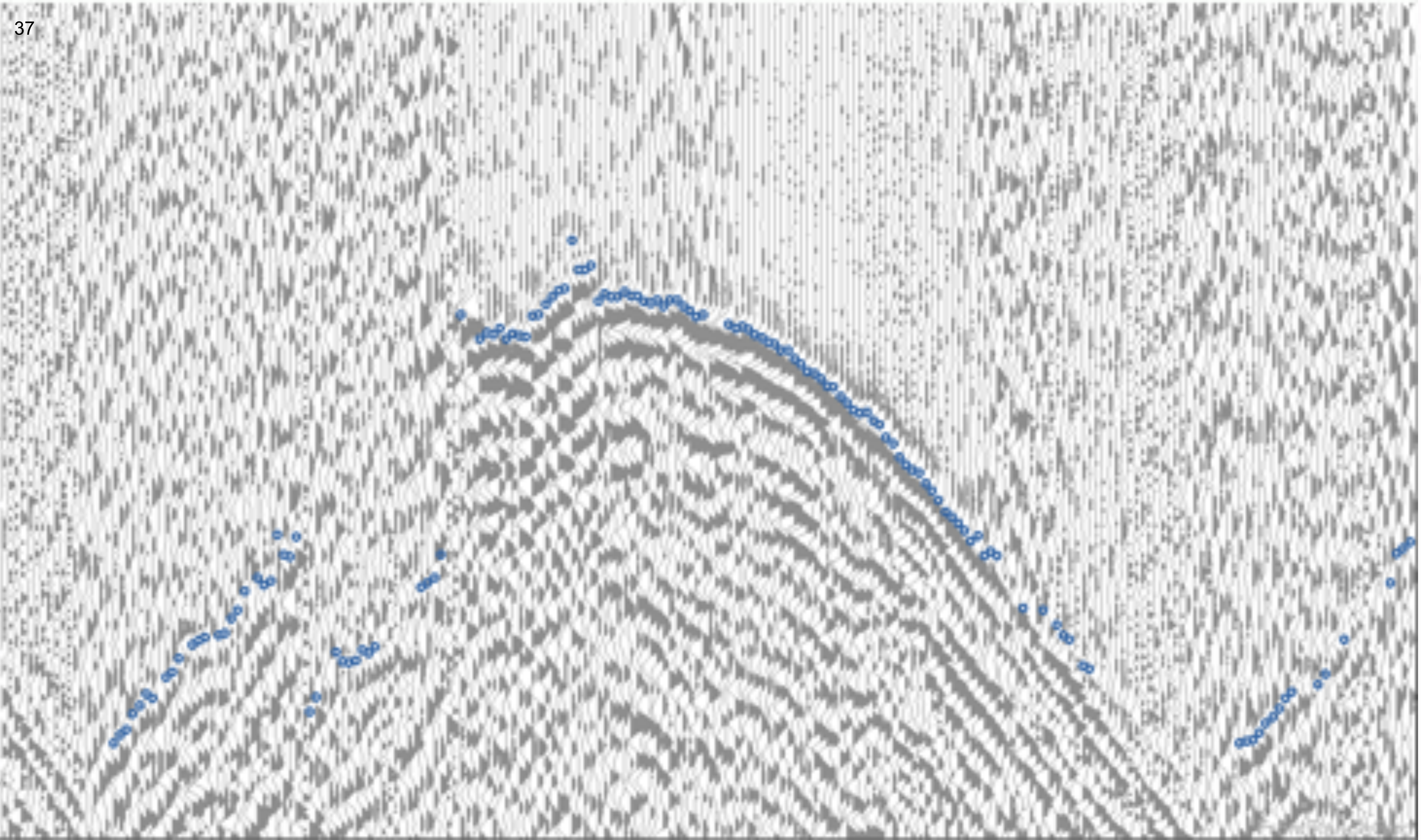
Data from the Permian Basin

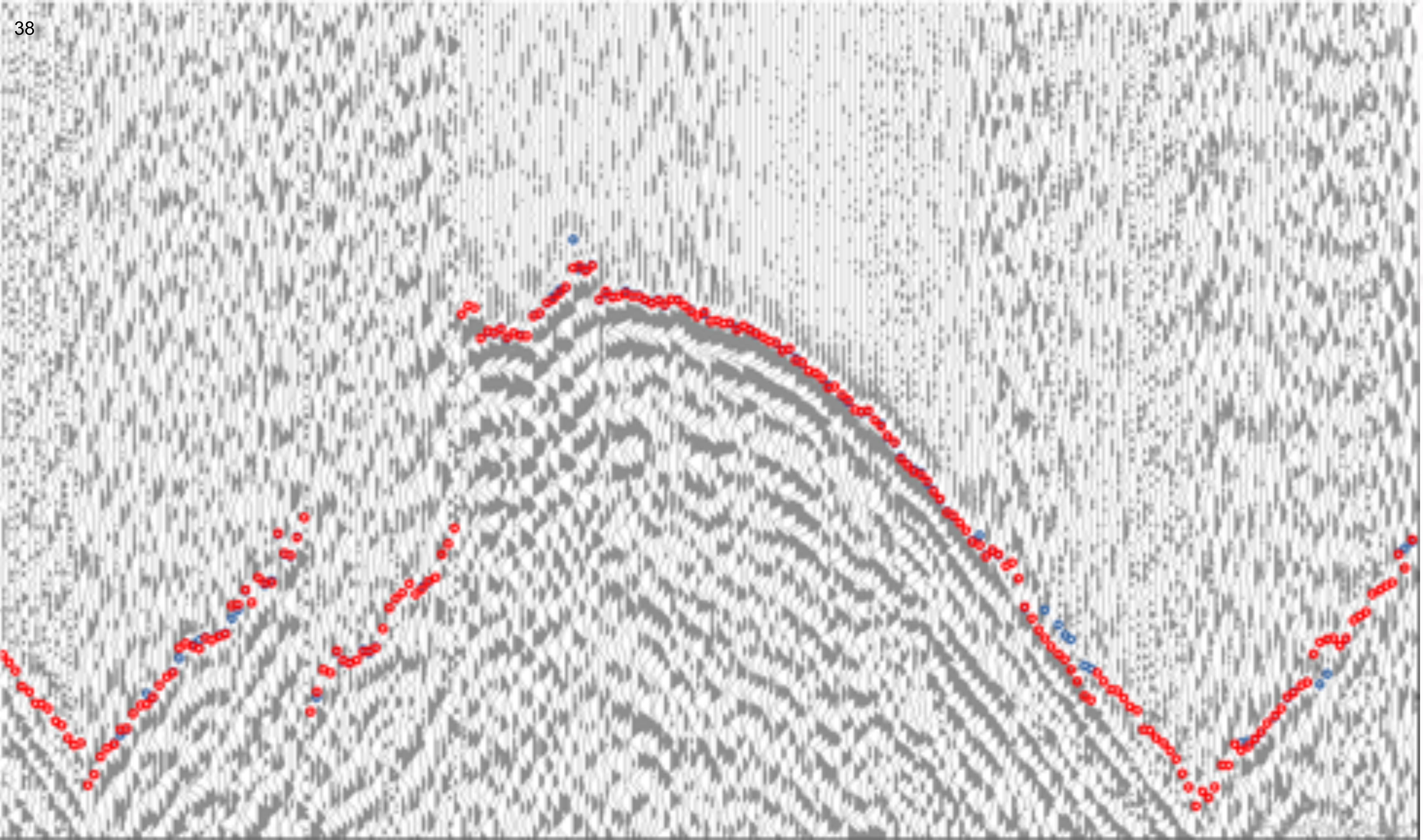


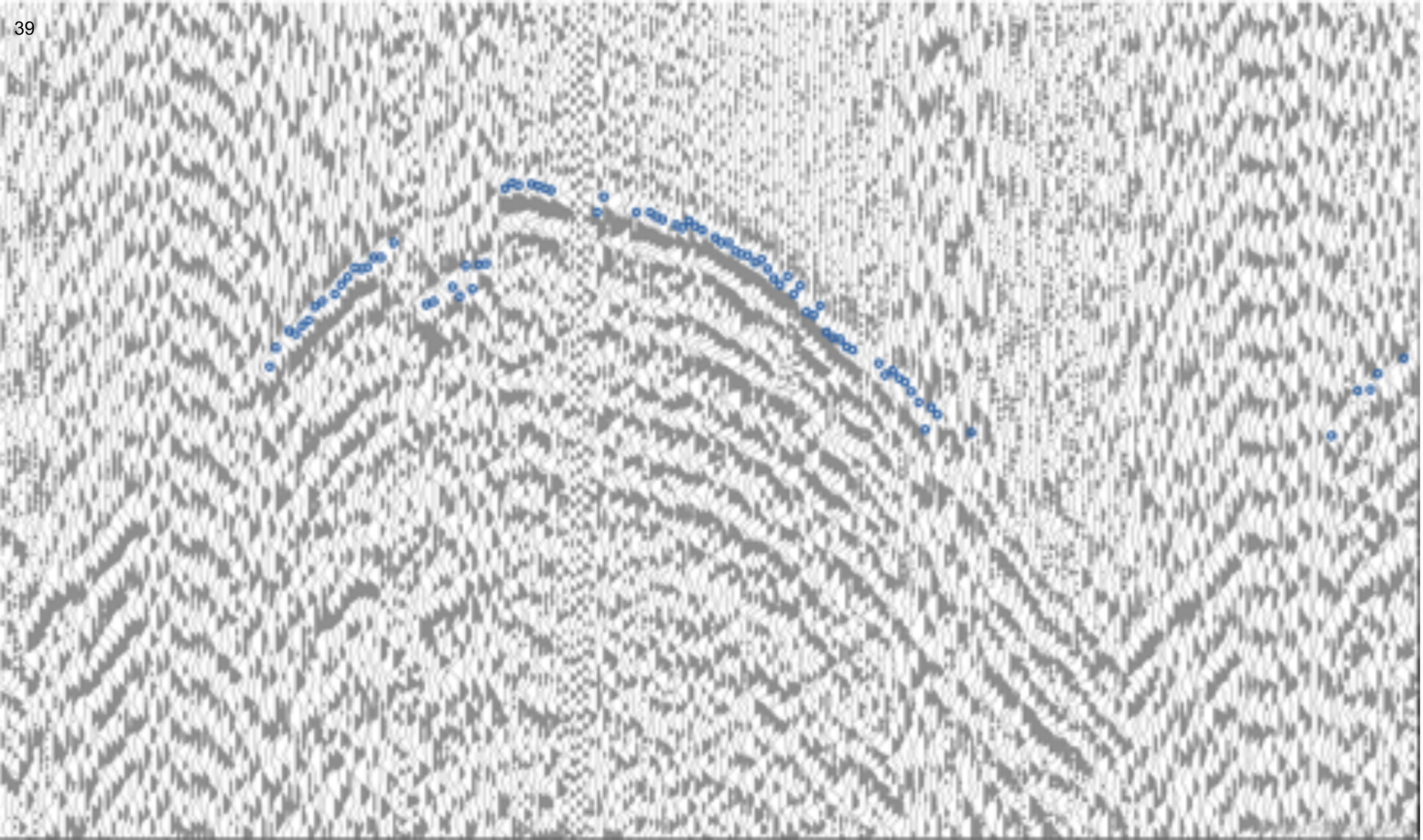


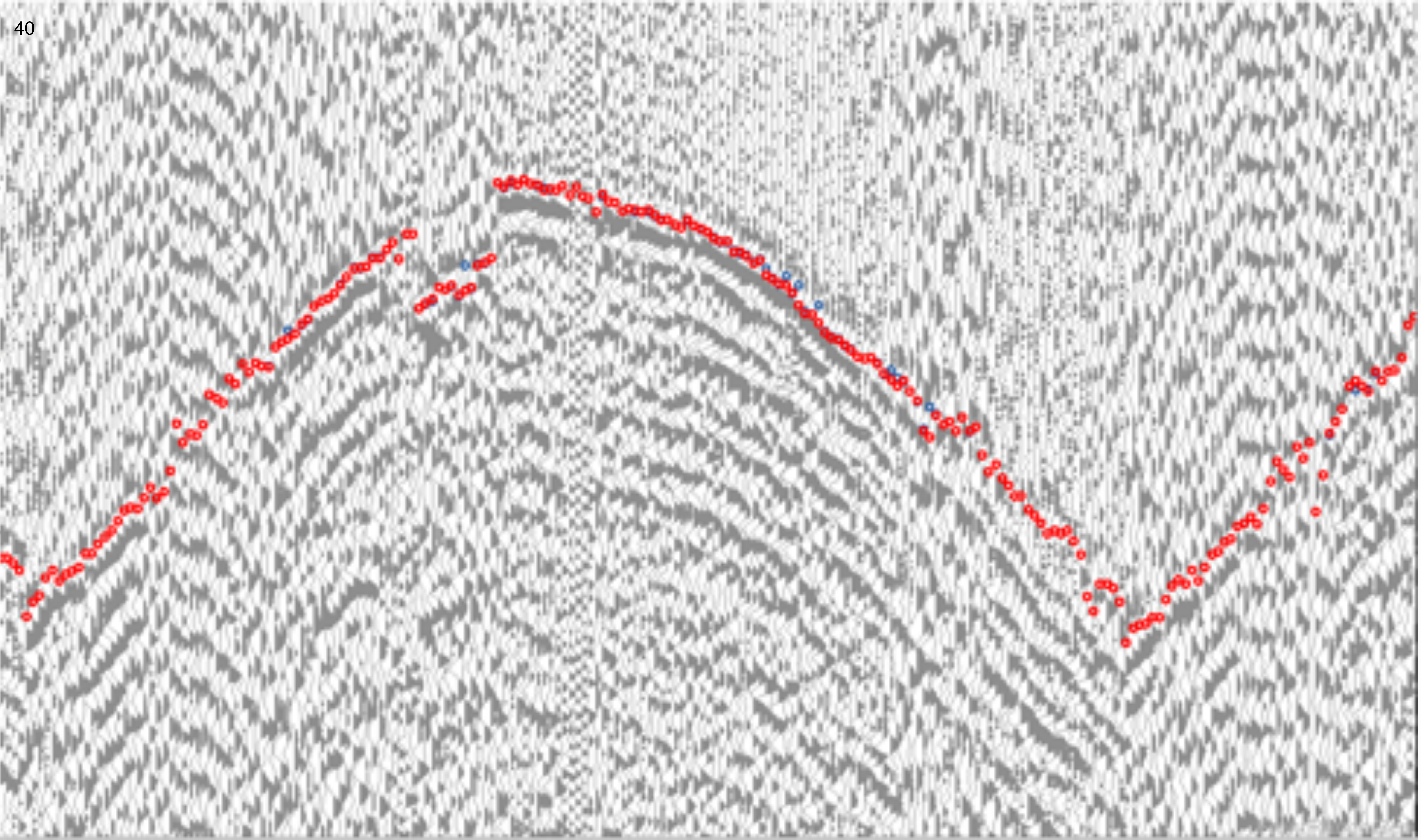


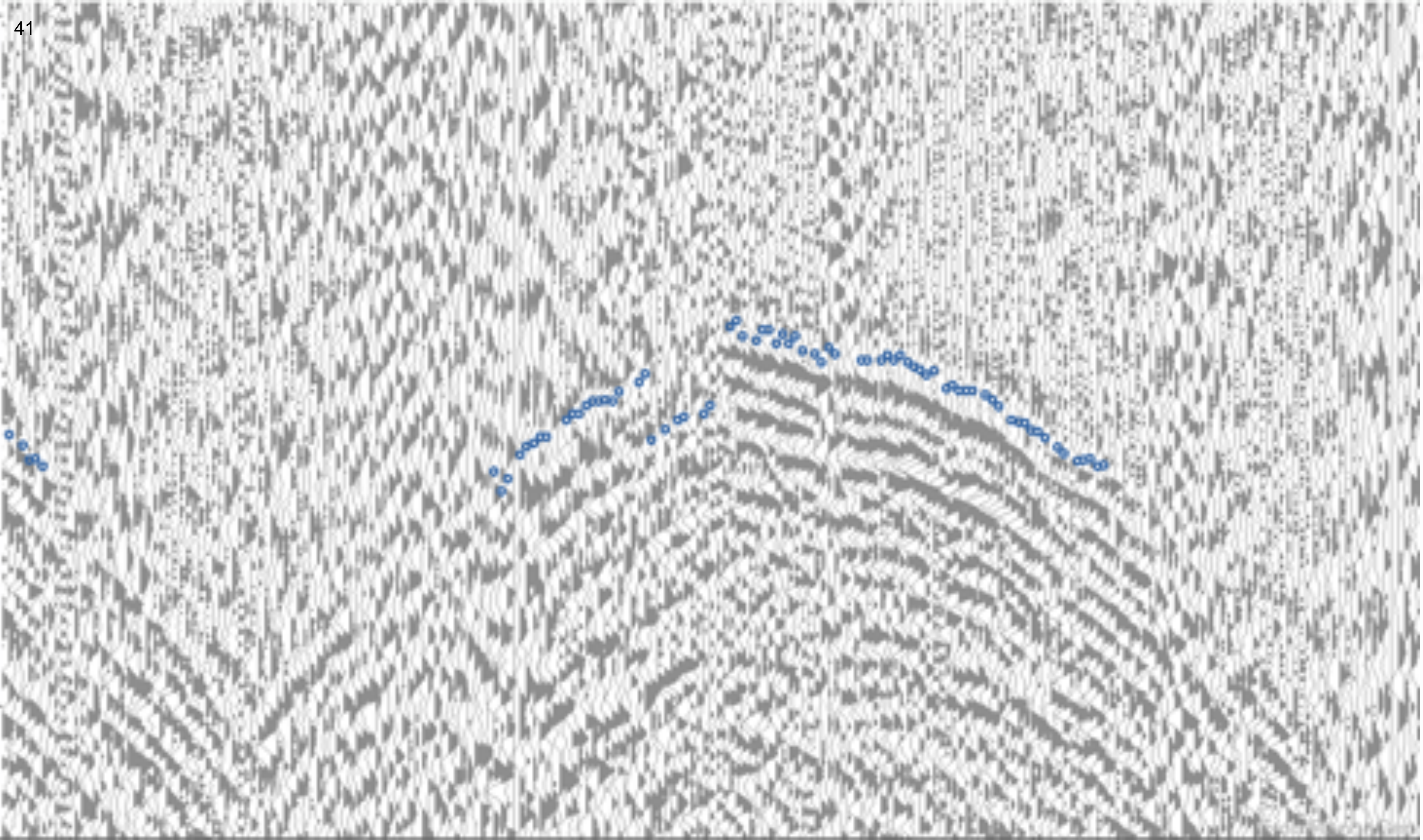


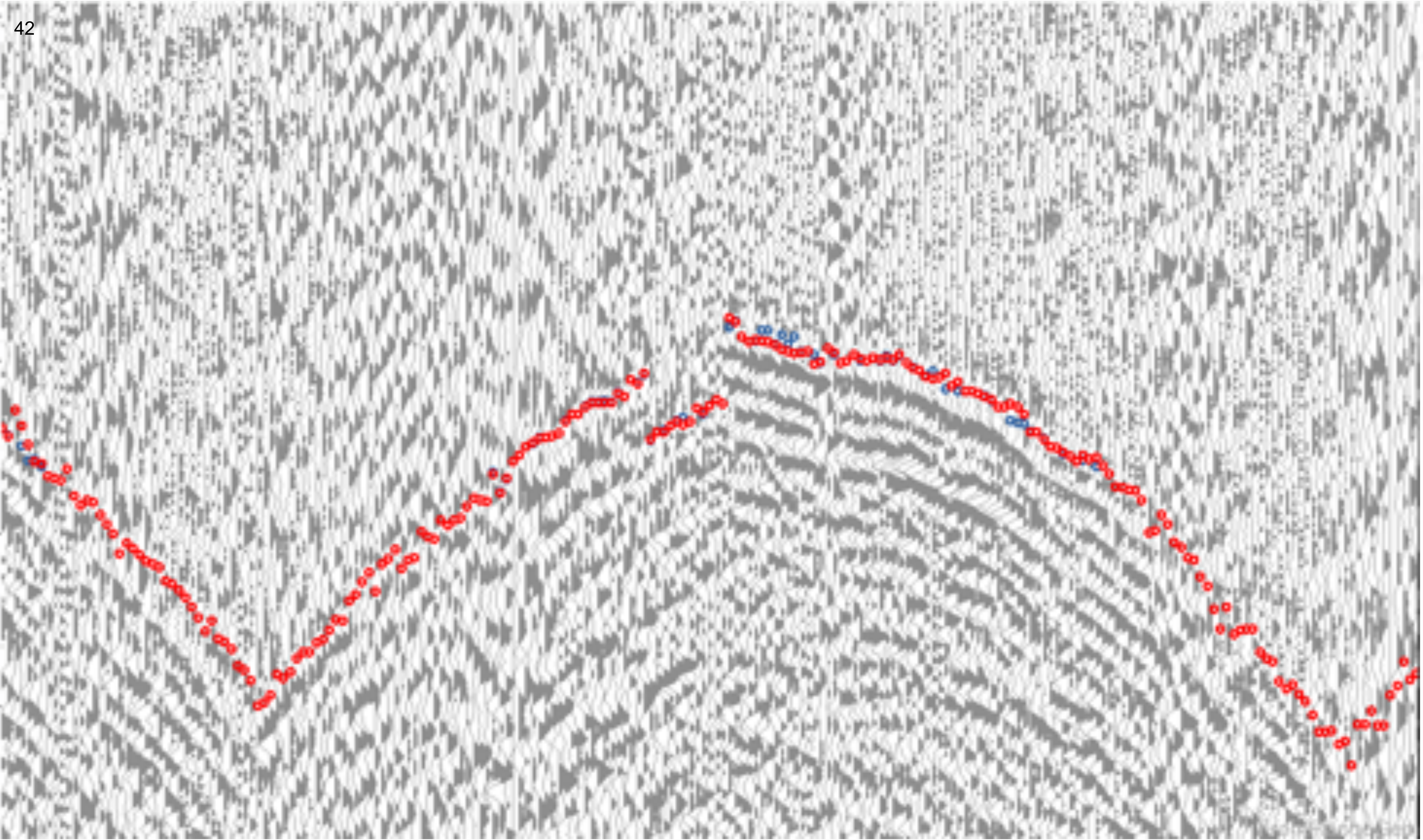


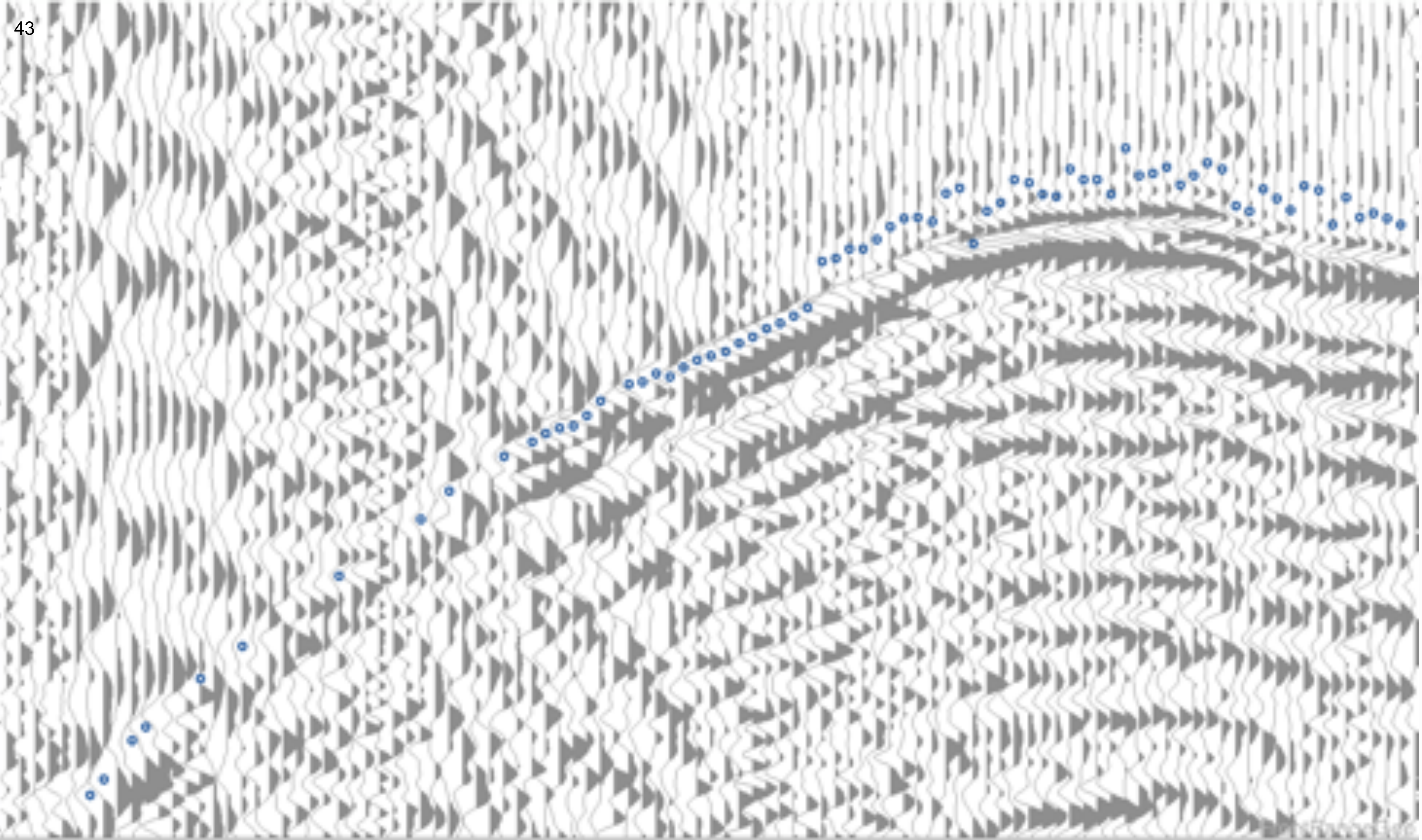


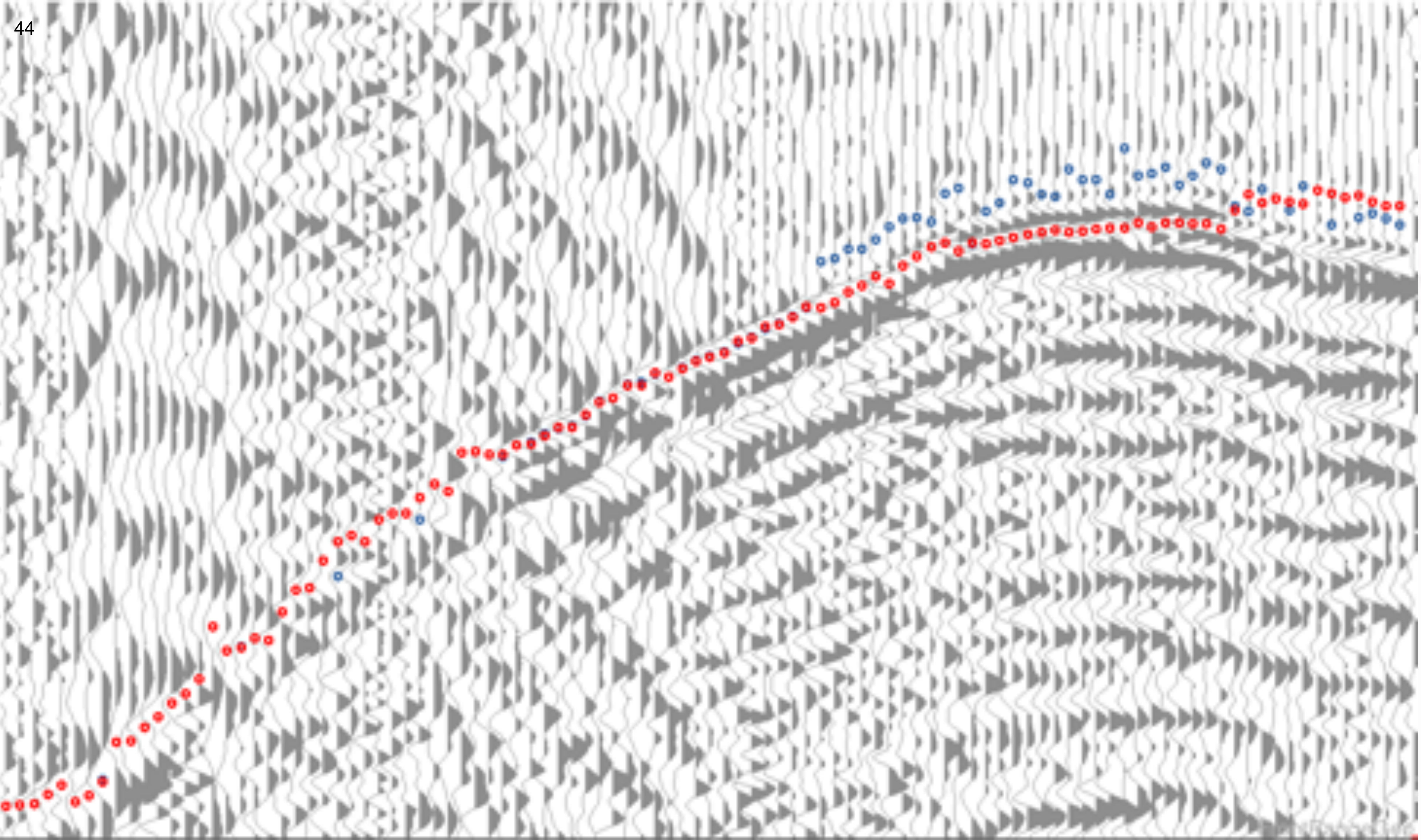


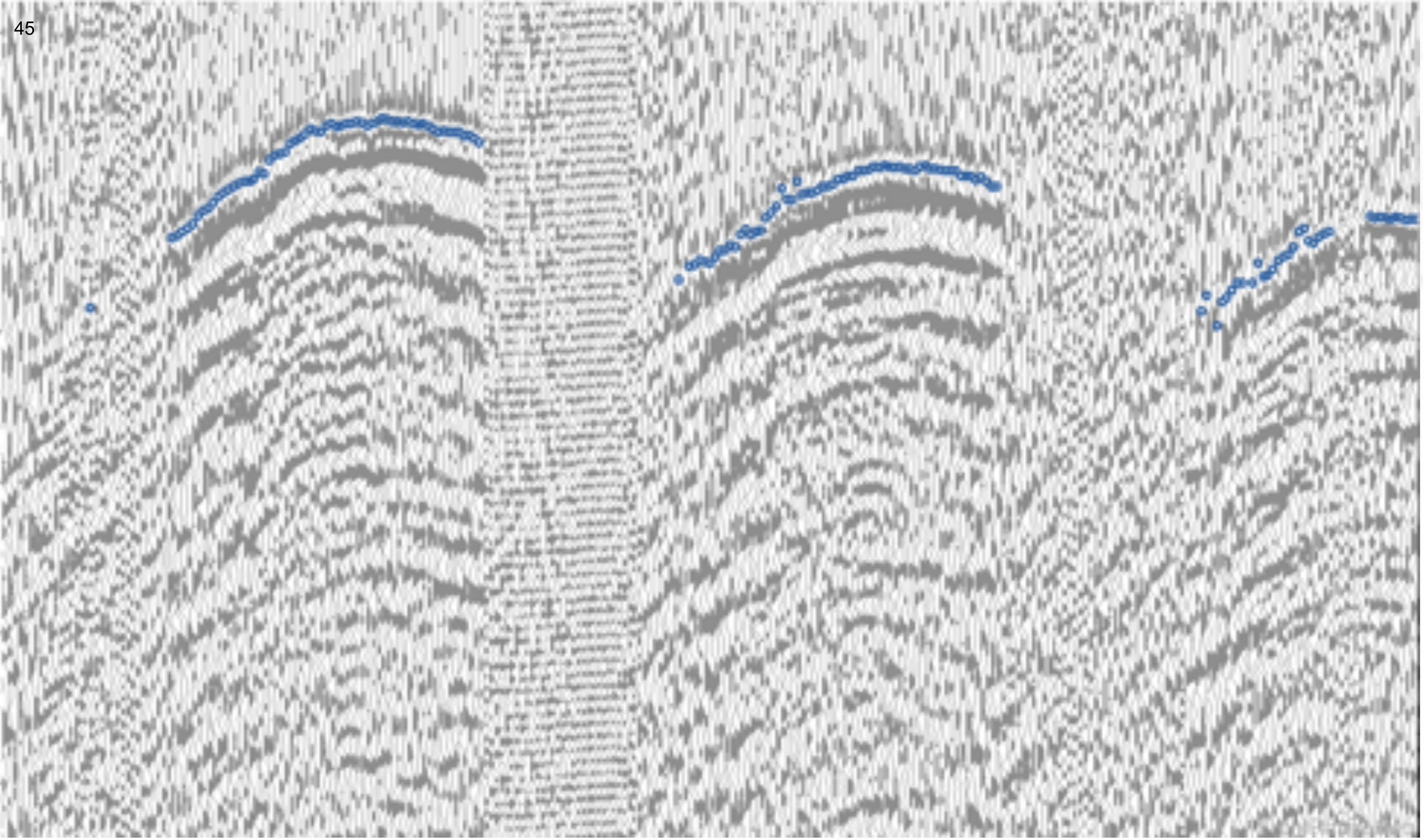


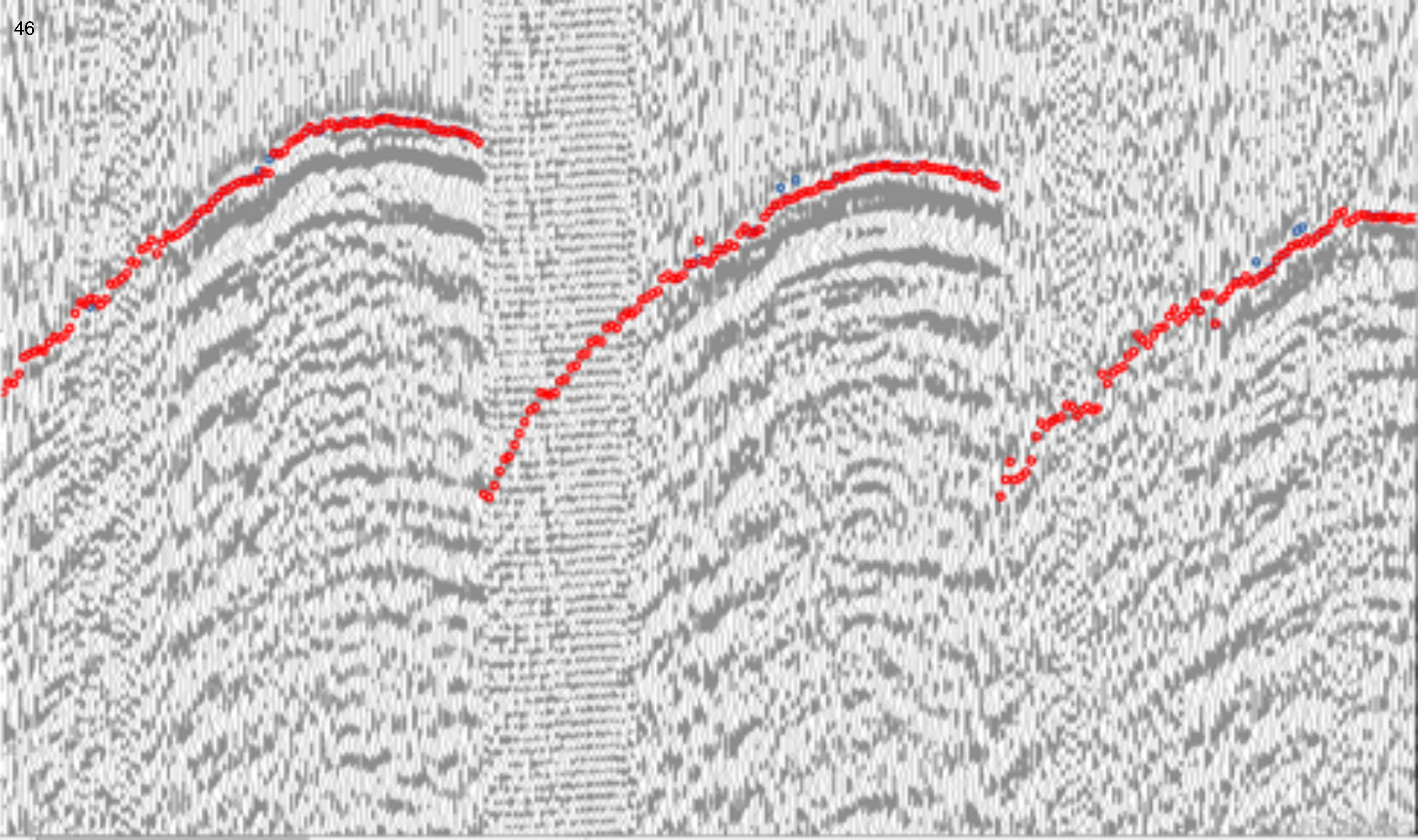


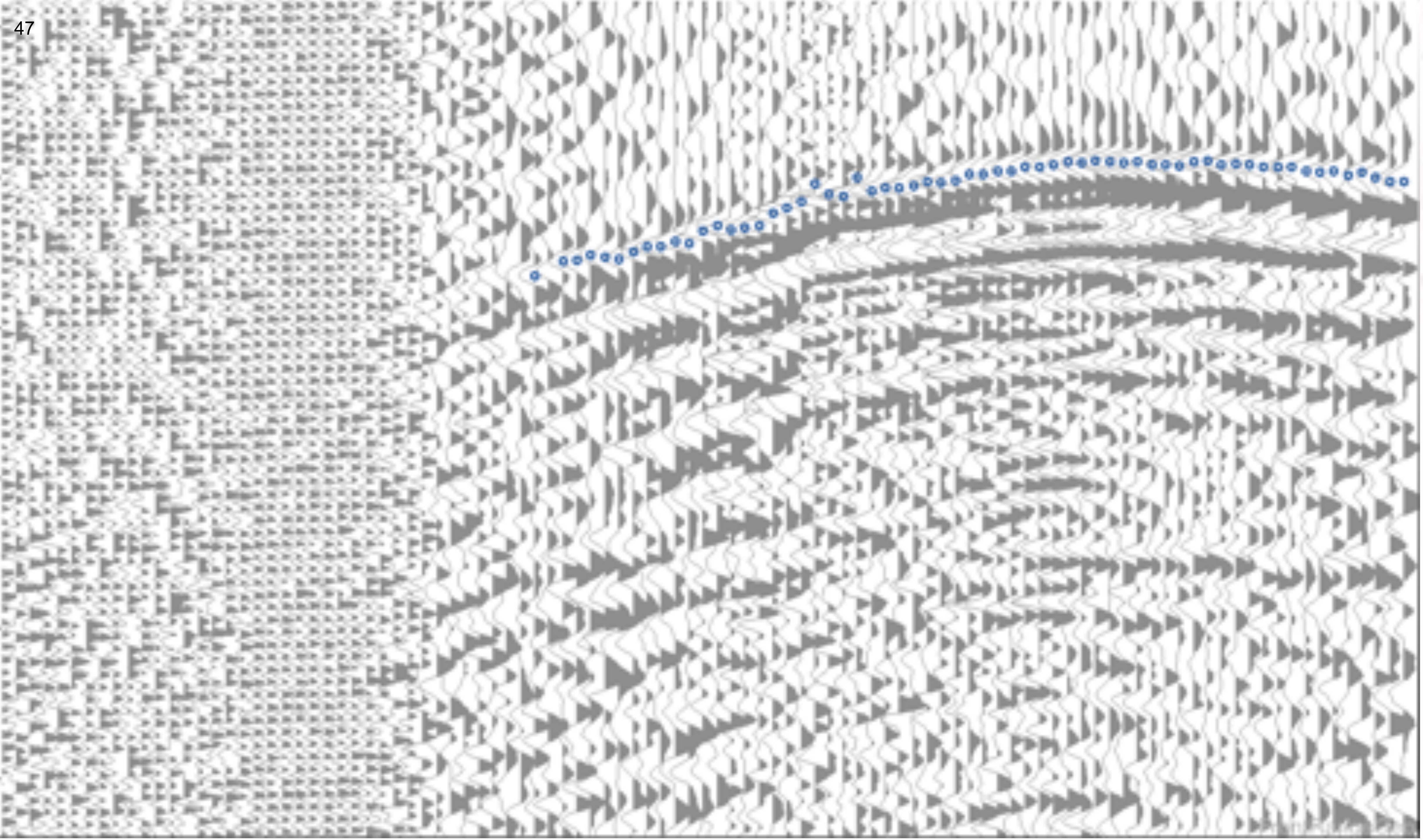


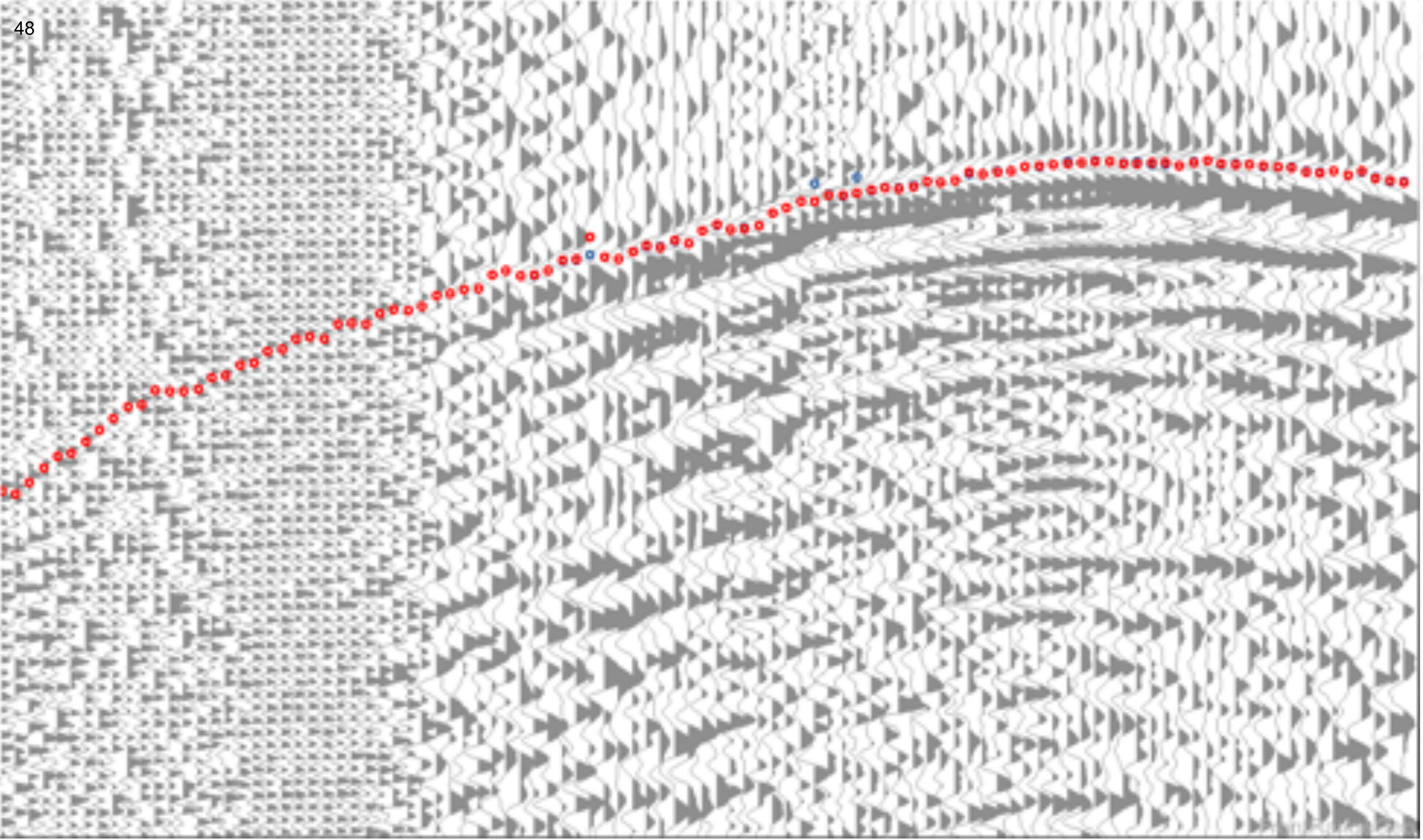


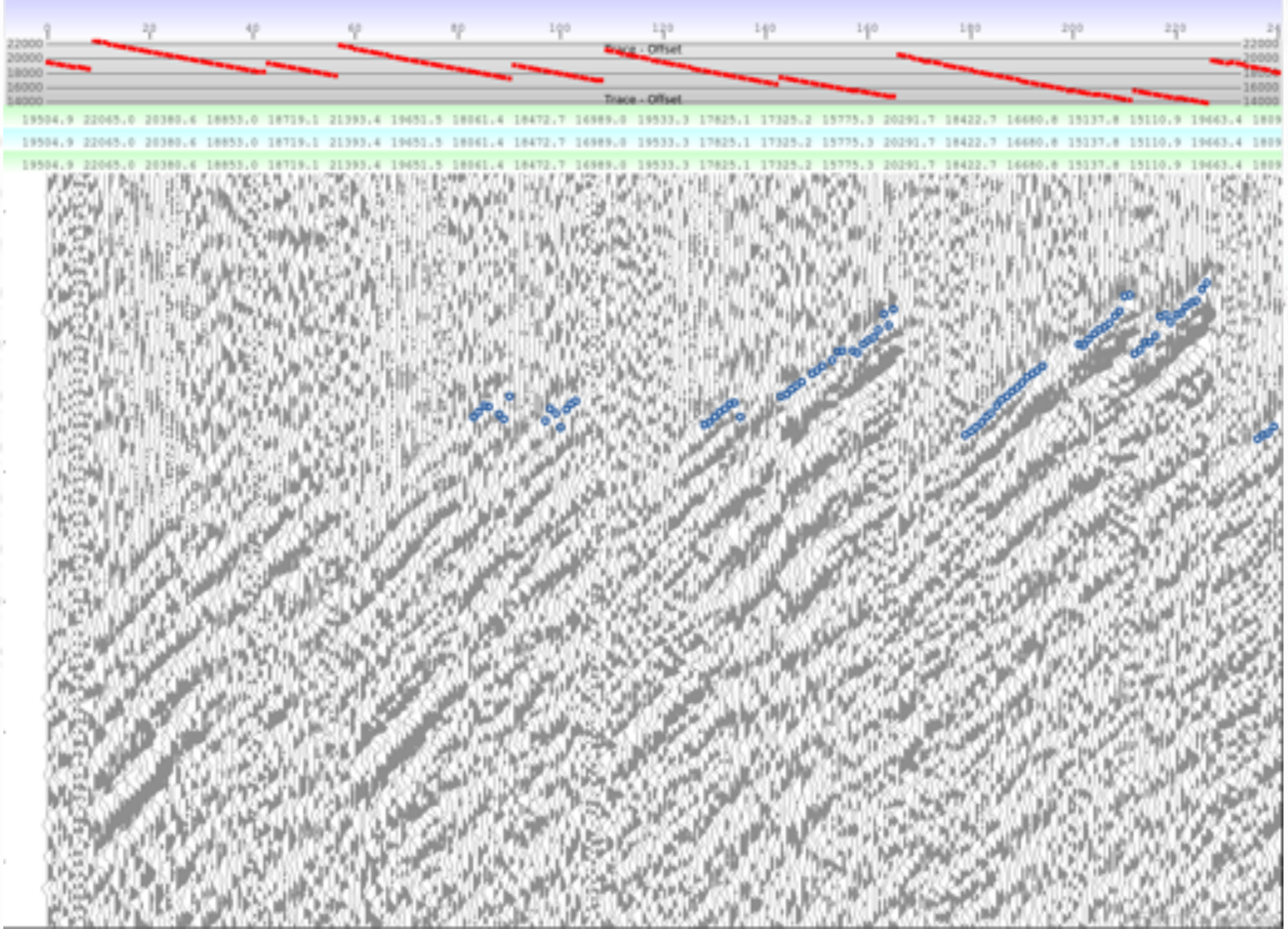


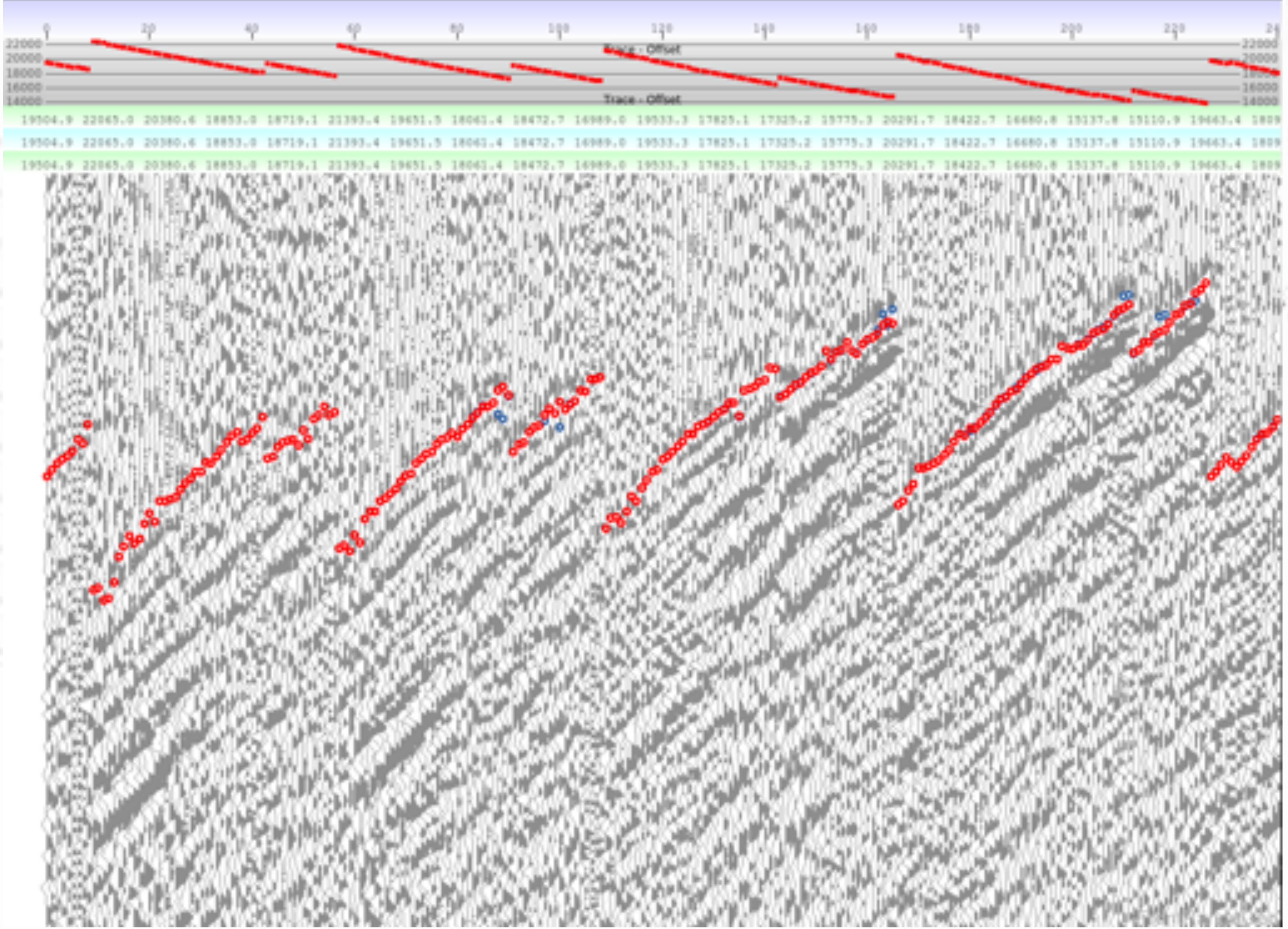


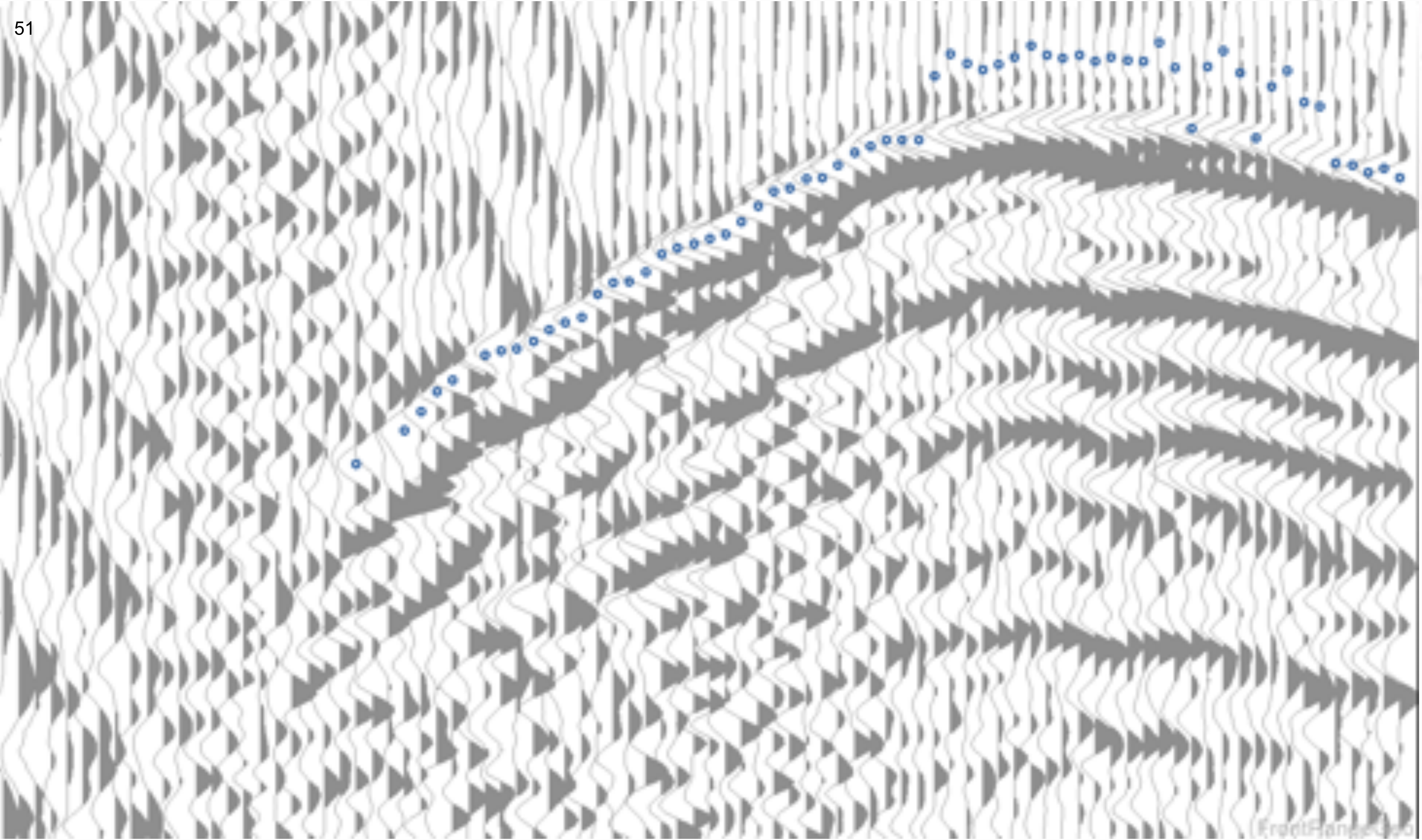


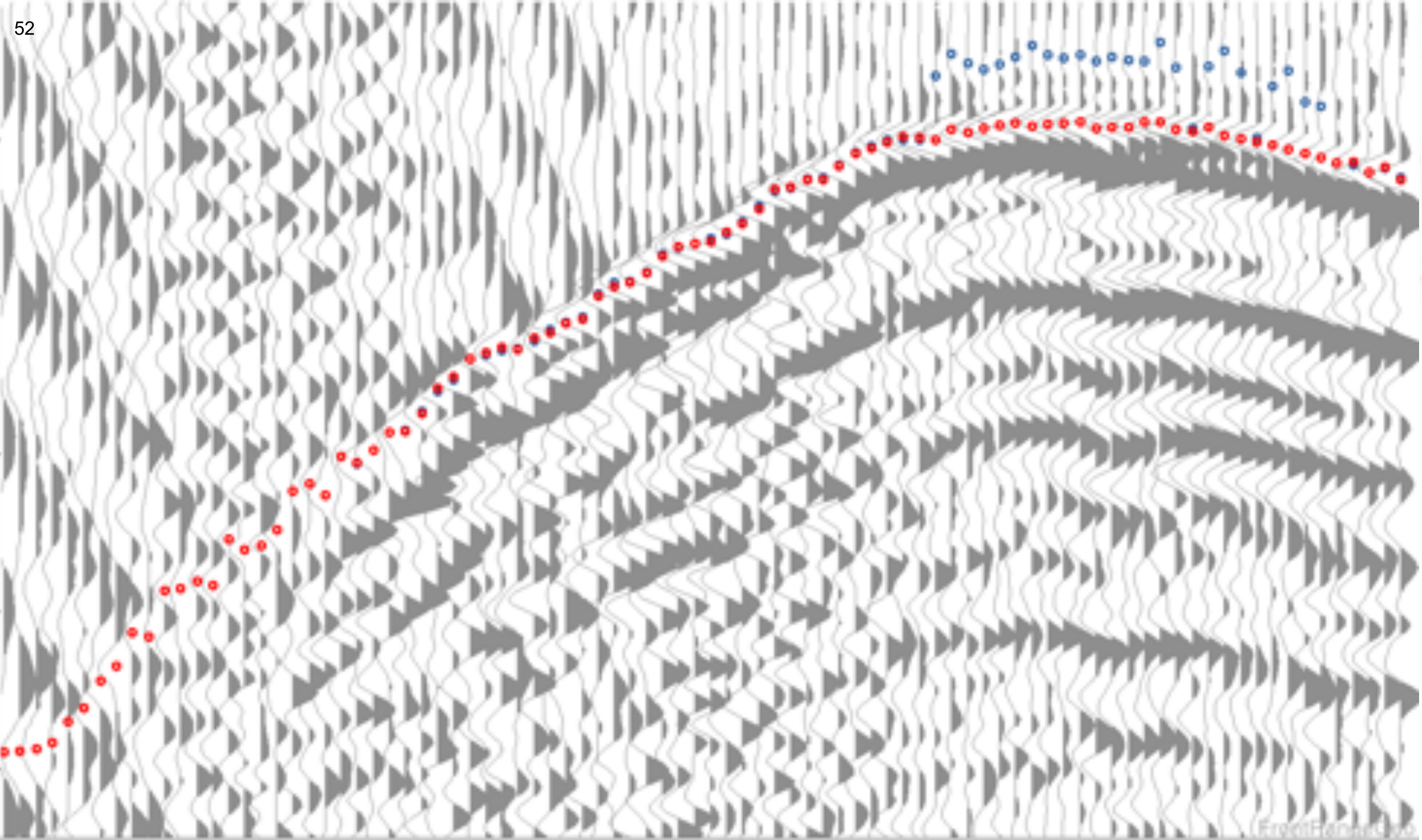


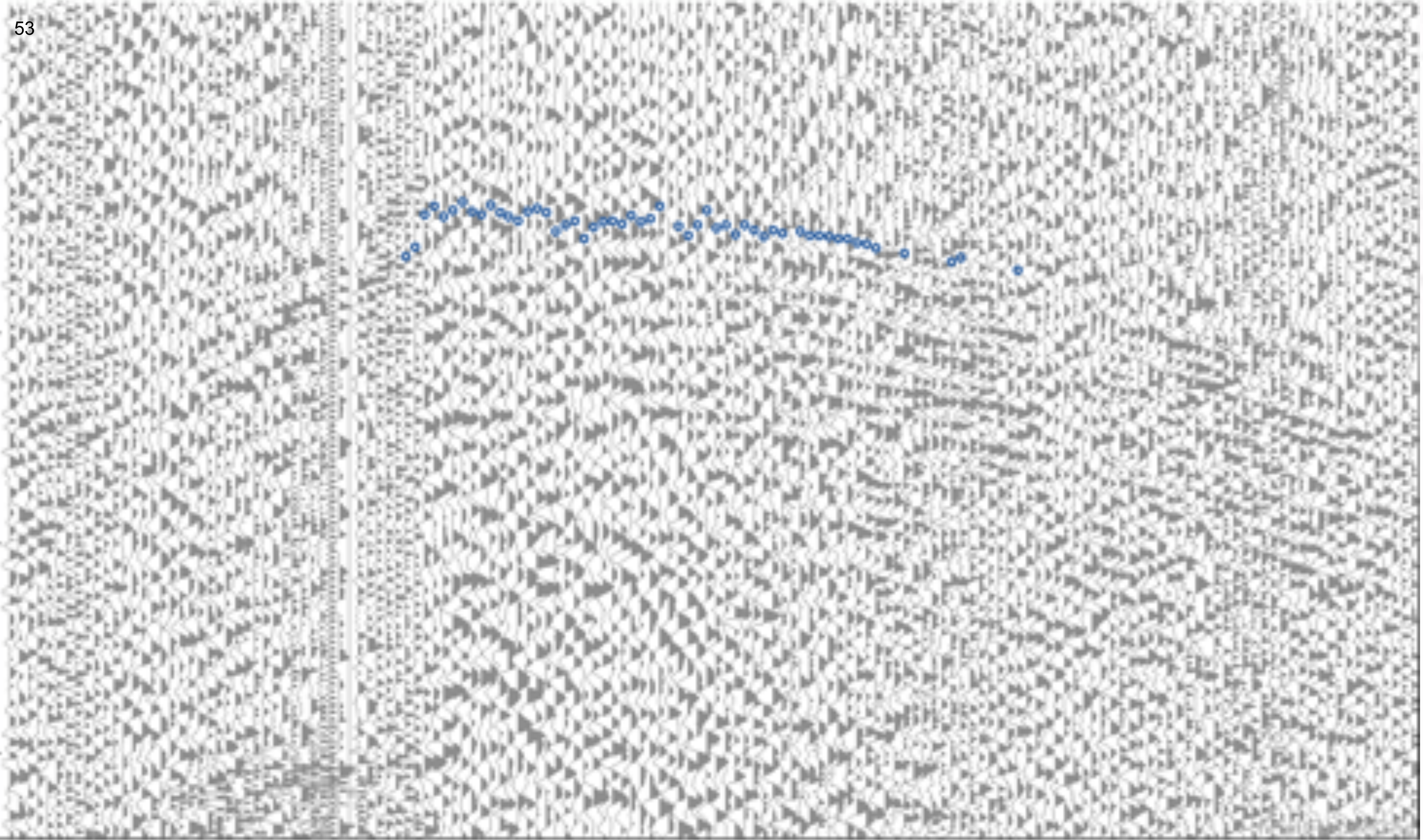


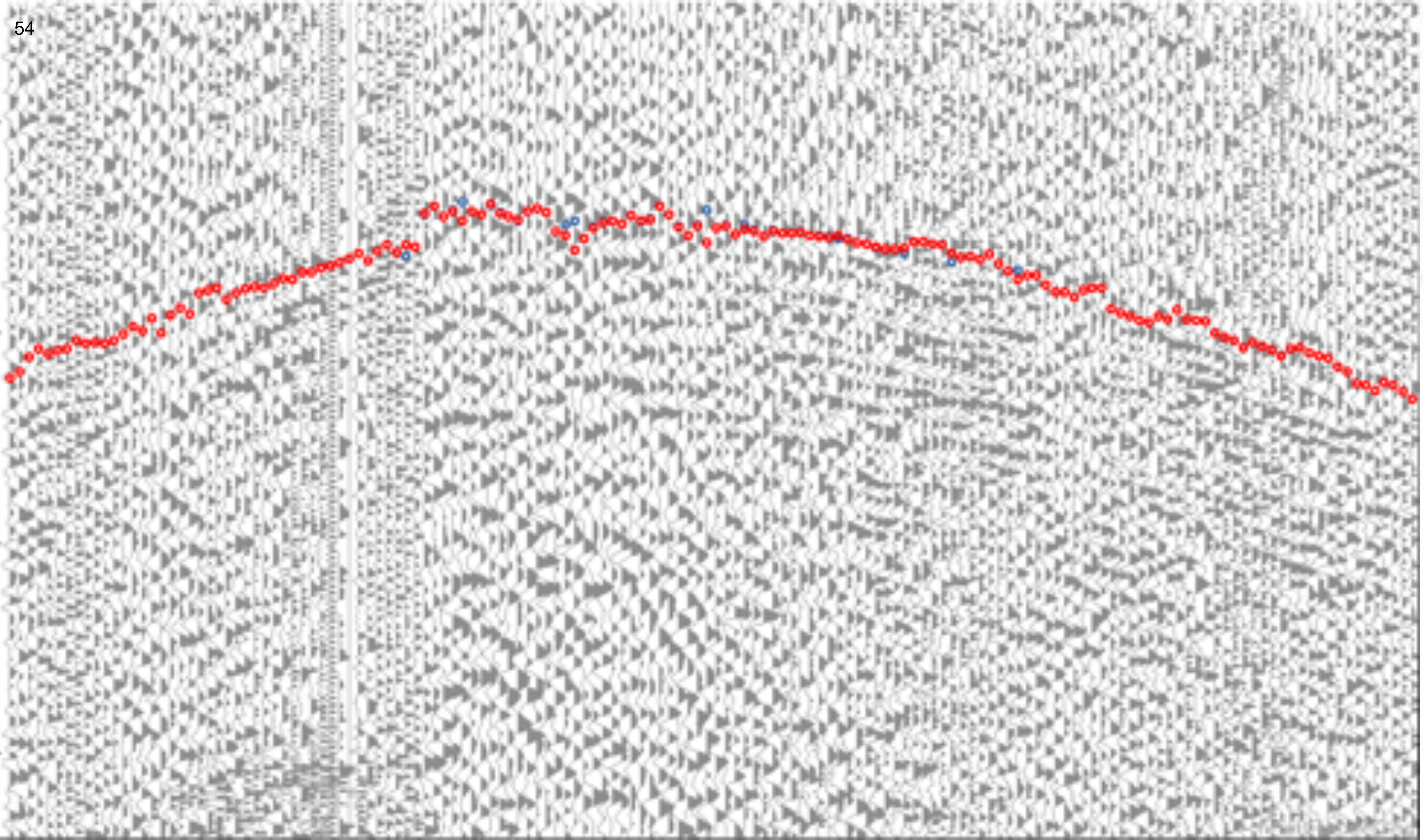

Trace - Offset
Trace - Offset


Trace - Offset
Trace - Offset









Phoenix

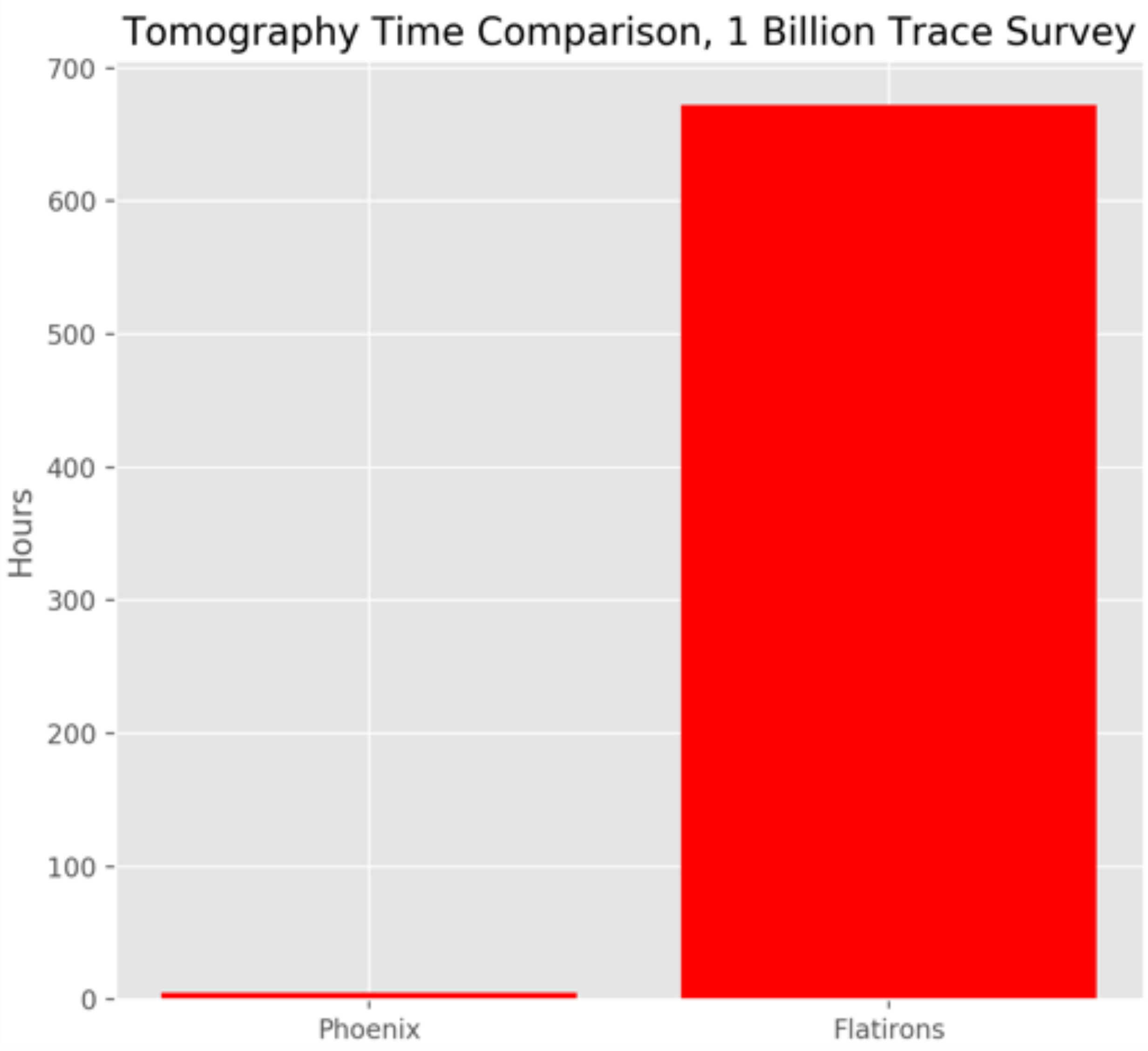
the next generation refraction statics solution

Phoenix

Phoenix is the next generation Refractions Statics solution, designed for the largest and most difficult surveys.

Phoenix achieves unheard-of speeds by massively parallelizing and threading most processes.

New physics and inversion constraint methods increase accuracy and quality of solution.

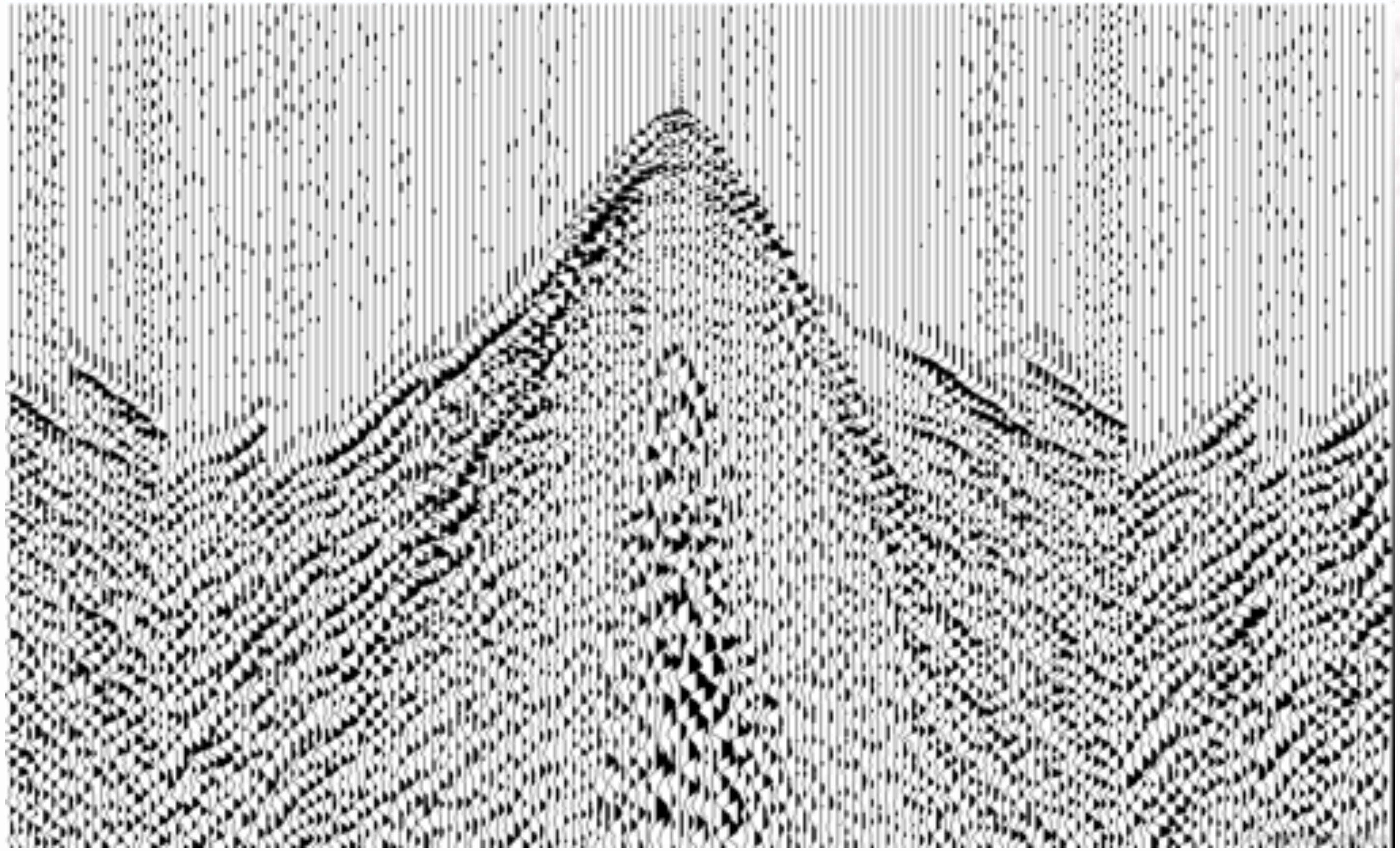


DeepTrace + Phoenix

By iteratively updating the physics model with new picks, and adjusting the network's view of the picks based on the model, we can quickly refine both the picks and model.

Phoenix and DeepTrace models *work together*, while avoiding circularity by taking different approaches (computer vision vs physics).

Phoenix and DeepTrace work together seamlessly to automate the entire near-surface processing workflow, requiring minimal human oversight.

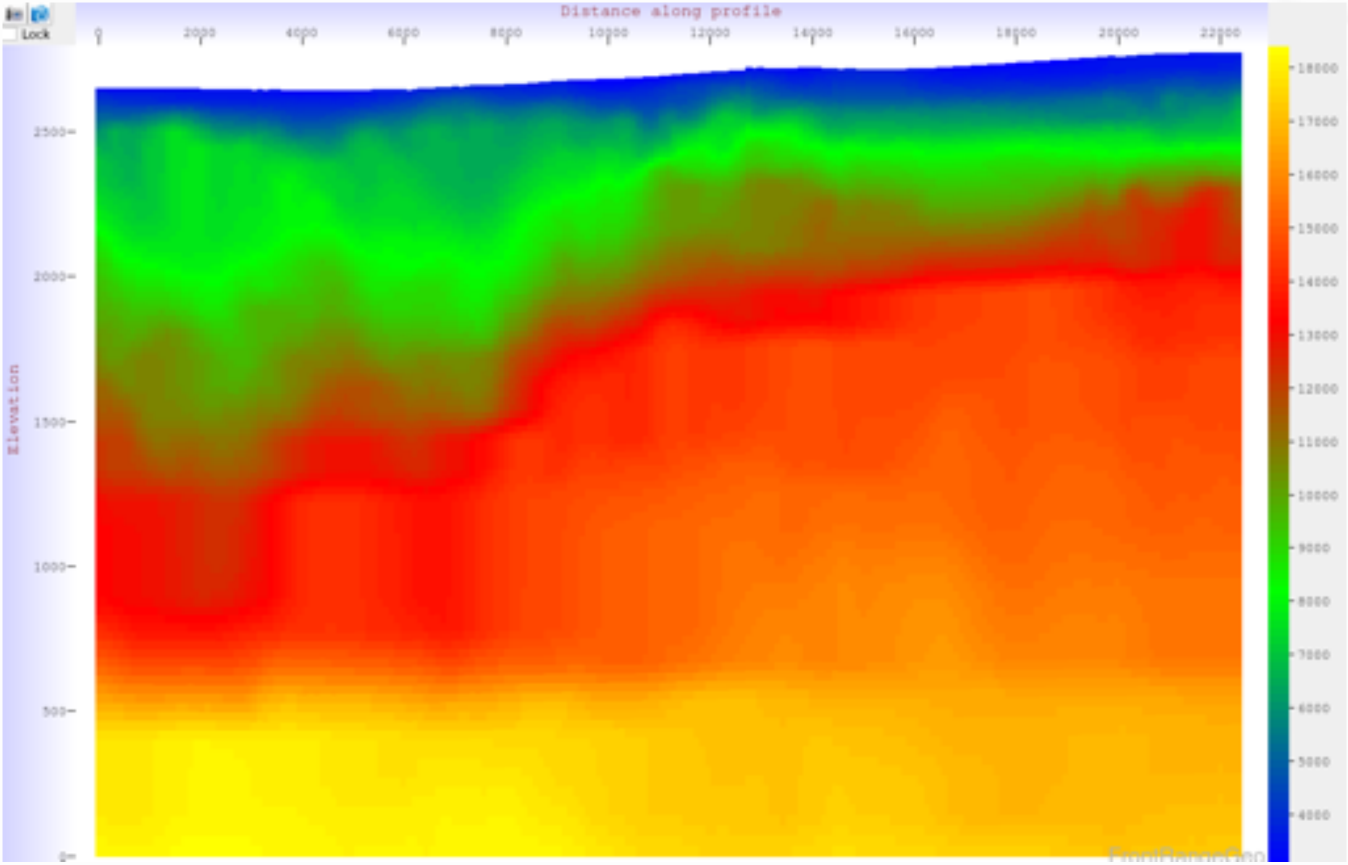


Animation: Shot line gather before and after applying tomographic solution.

Model Comparison

First: Velocity profile from human picks.

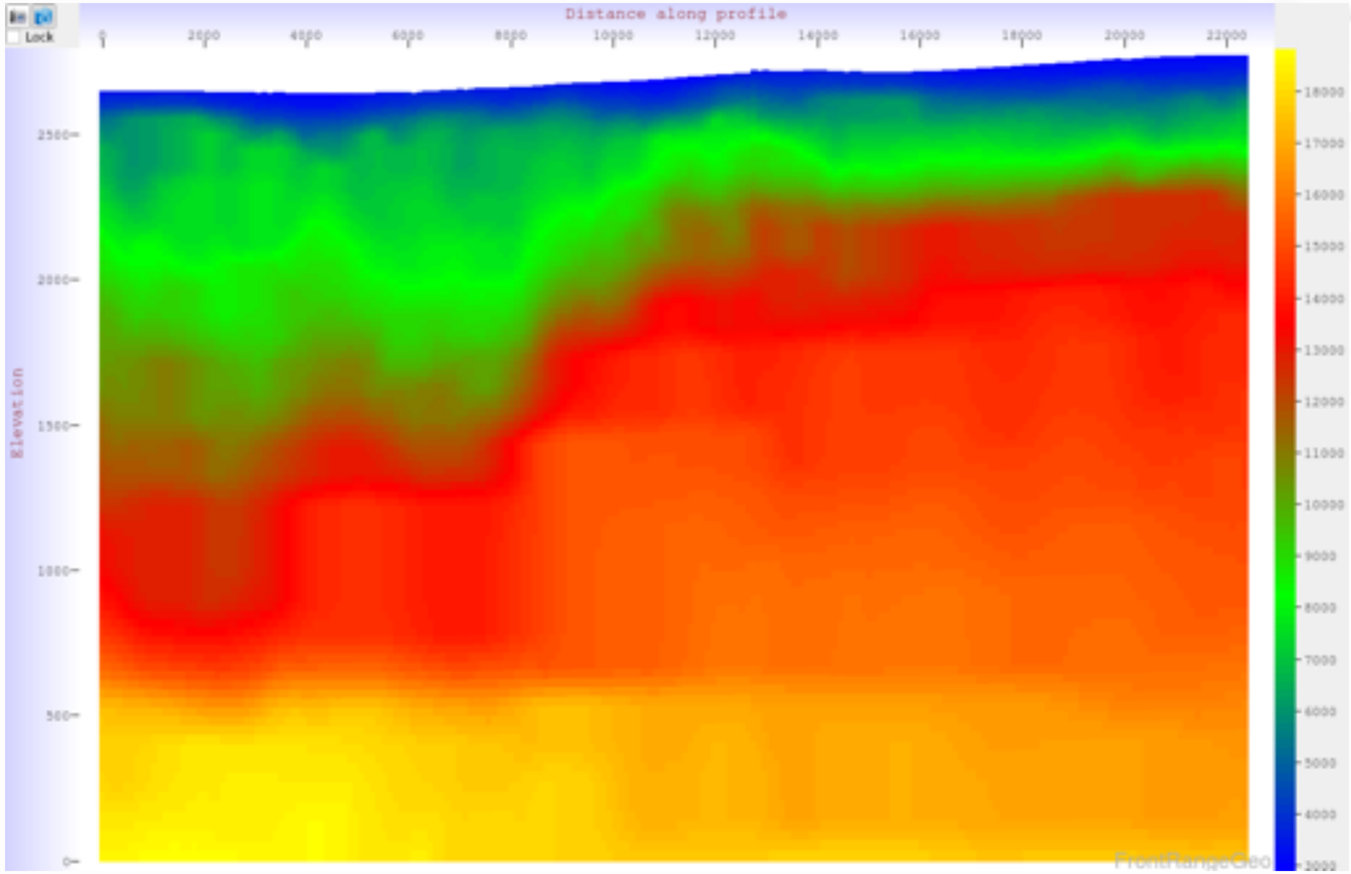
Second: Velocity profile from DeepTrace picks. No post-processing.



Model Comparison

First: Velocity profile from human picks.

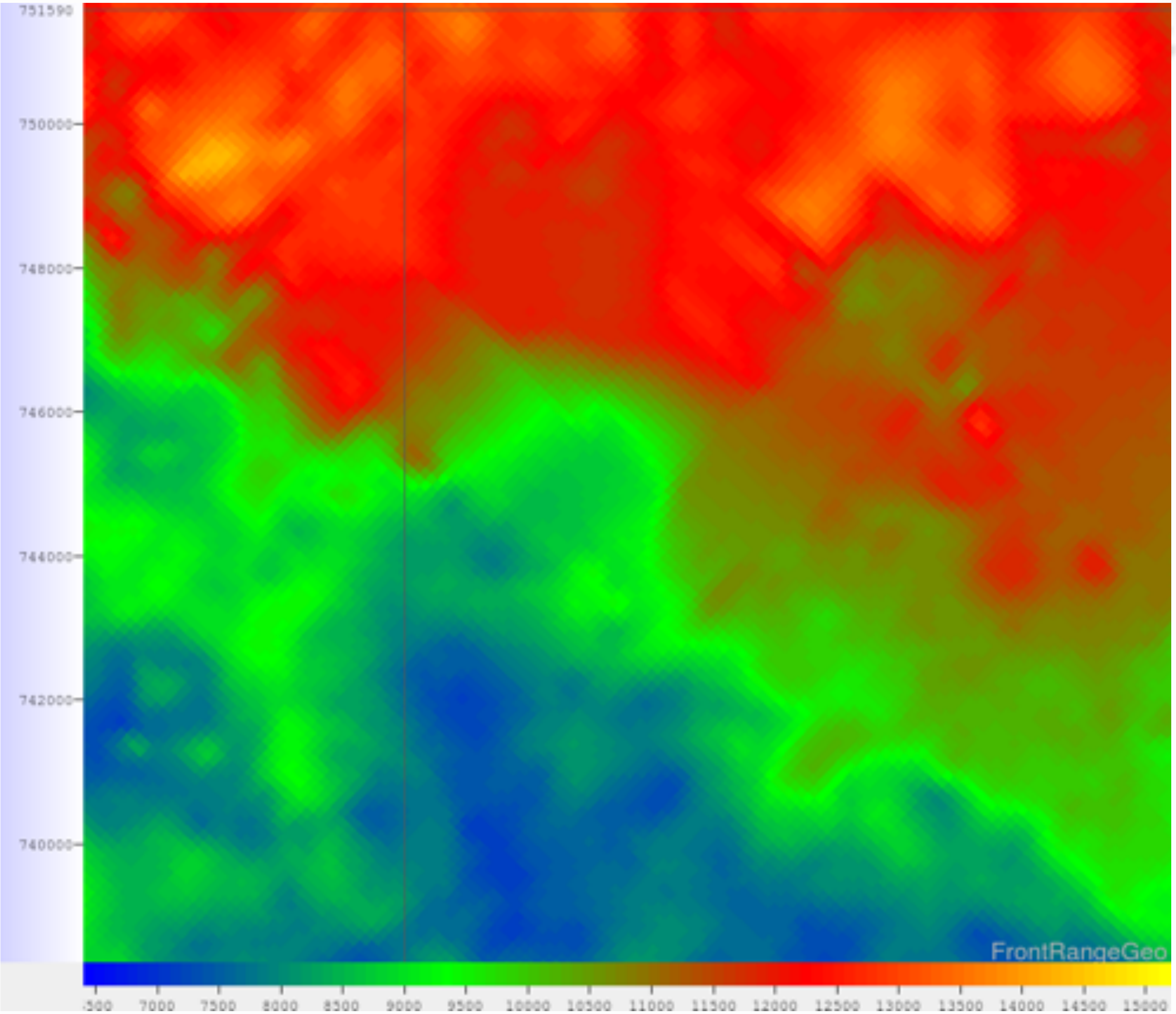
Second: Velocity profile from DeepTrace picks. No post-processing.



Model Comparison

First: Velocity depth slice from human picks.

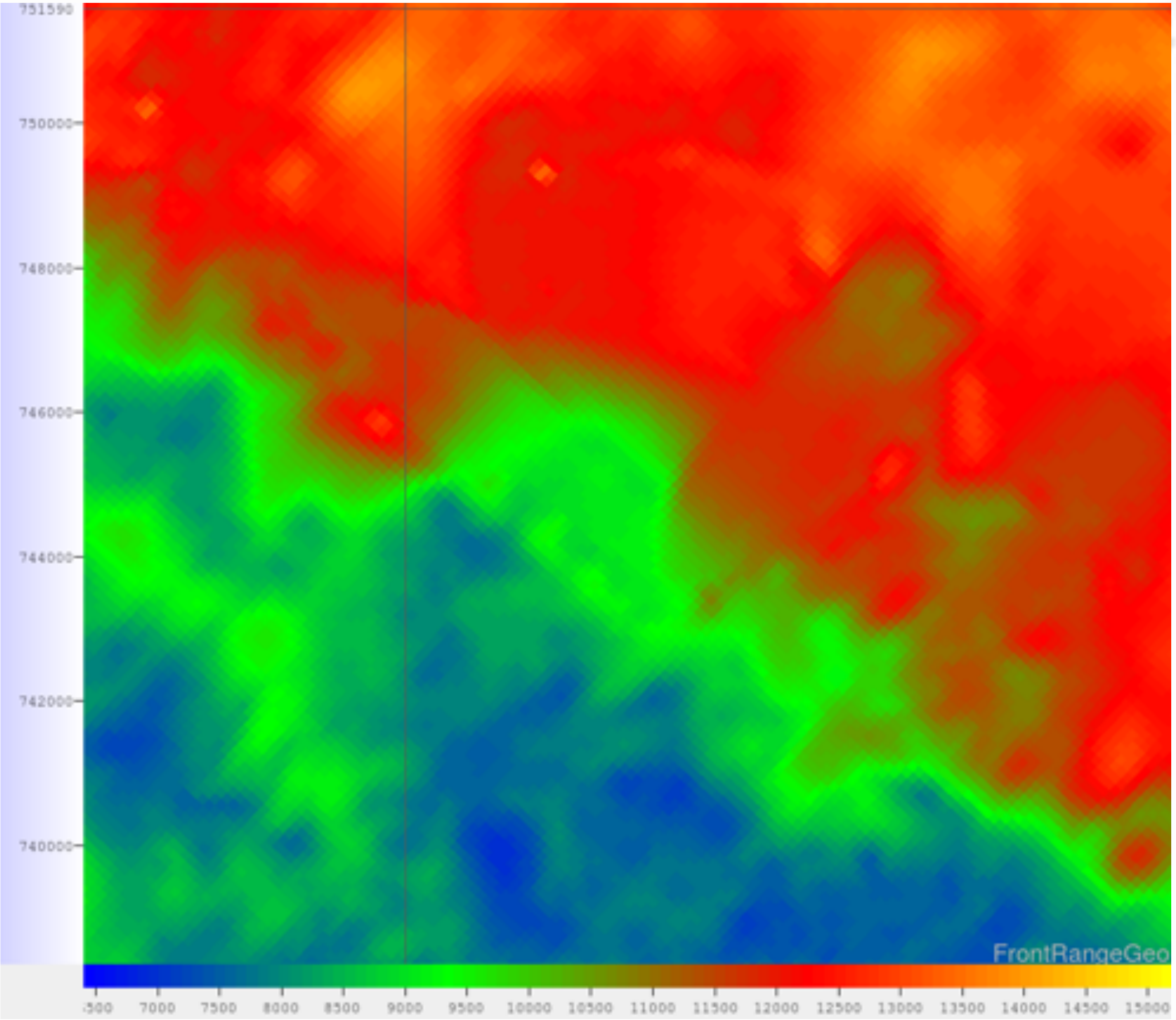
Second: Velocity depth slice from DeepTrace picks. No post-processing.



Model Comparison

First: Velocity depth slice from human picks.

Second: Velocity depth slice from DeepTrace picks. No post-processing.



Reliability

automating quality control

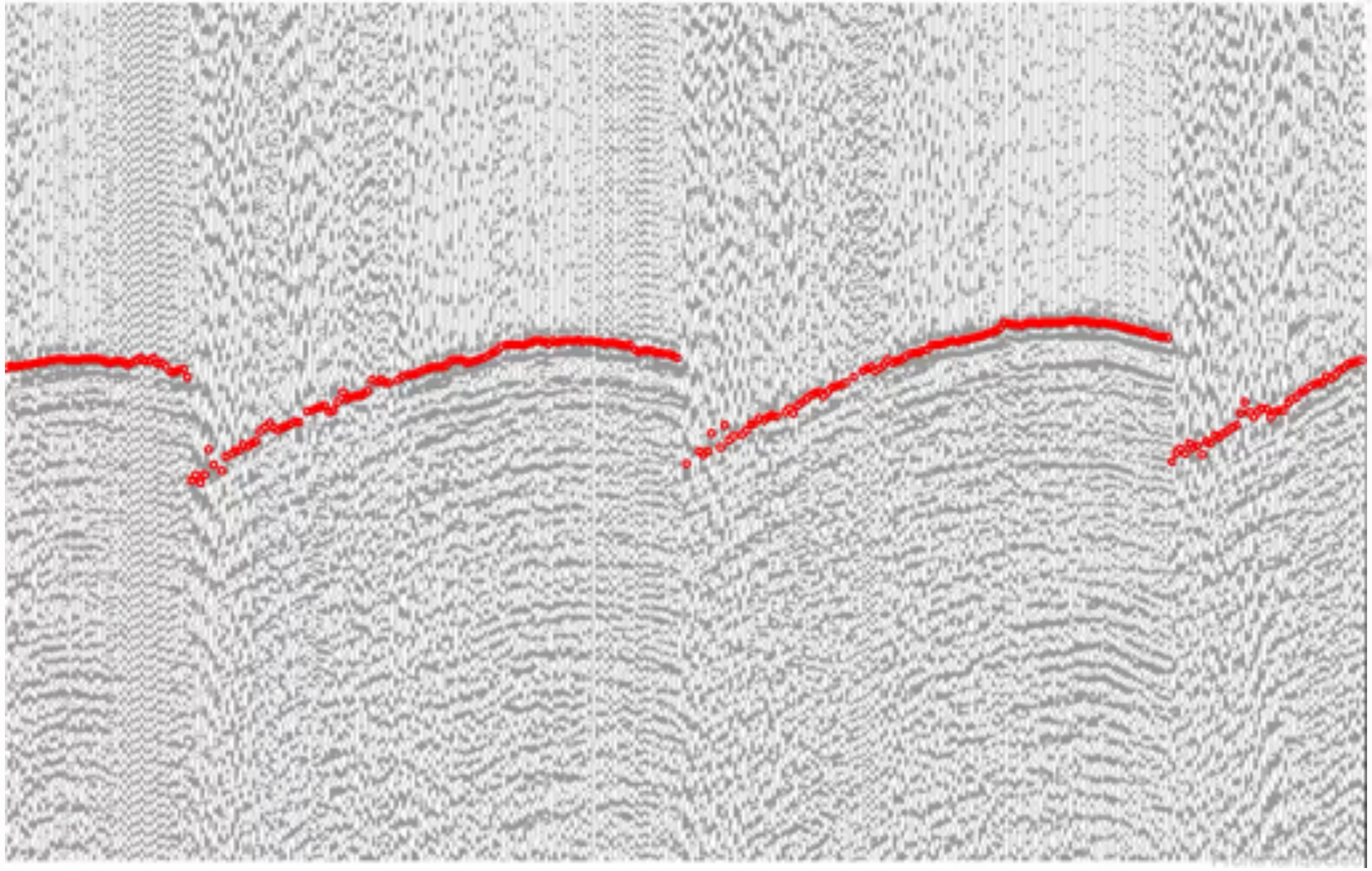
STA/LTA Thresholding

Wong et al. first proposed the STA/LTA method to find first breaks.

Their method involves calculating many STA/LTA windows per trace to find a first break.

We calculate the STA/LTA only once around the DeepTrace pick to use as a quantitative measure of confidence.

Users can specify an STA/LTA confidence threshold that the DeepTrace pick must satisfy to be considered valid.



Video: Killing picks based on varying STA/LTA threshold.

Based on the methods of:
Wong et al. *Automatic time-picking of first arrivals on noisy microseismic data*. 2009

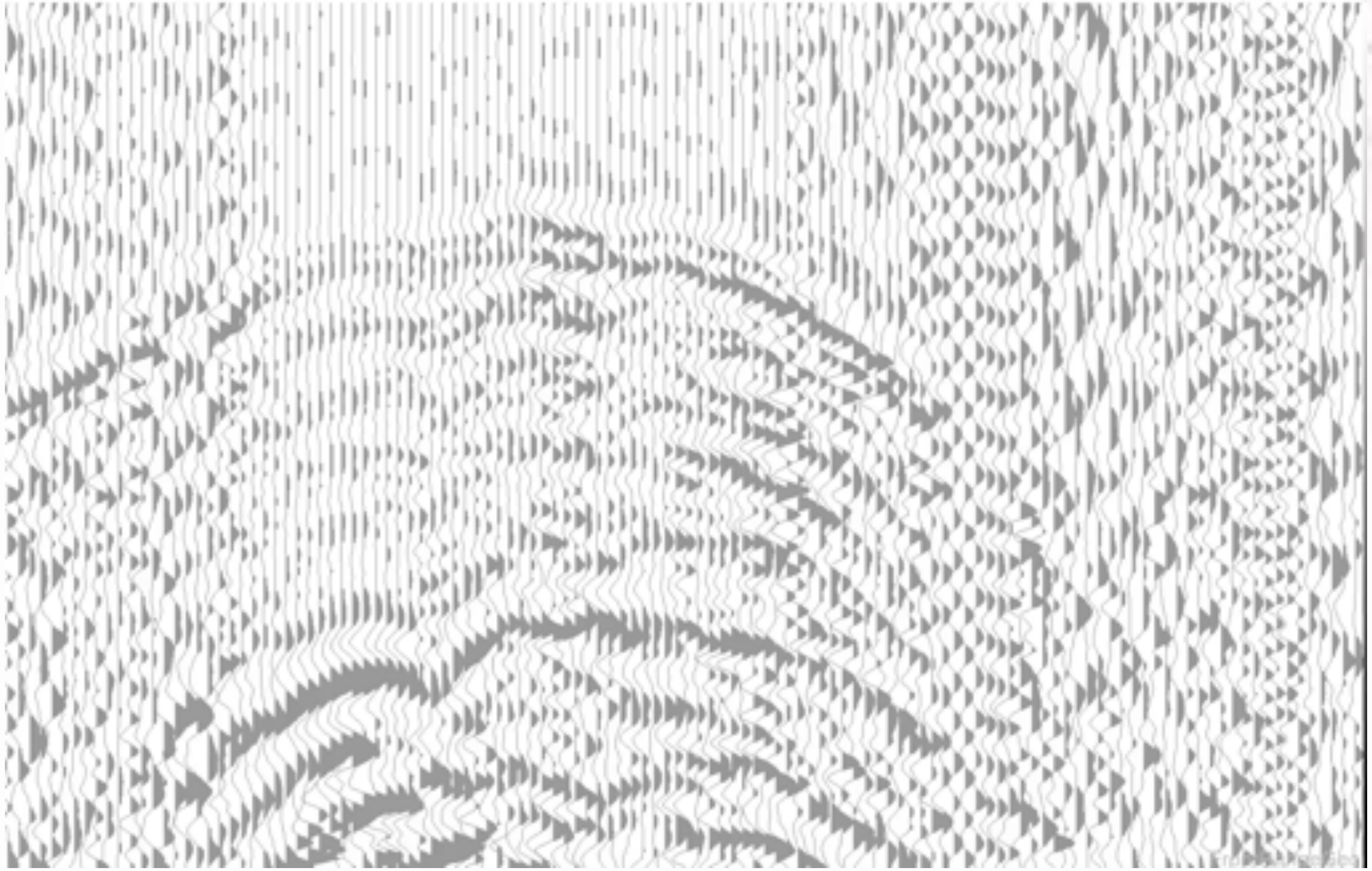
Model Ensembling

Accuracy increases by averaging over multiple models (generic deep learning result).

Variance of independent model picks gives another measure of confidence.

This distribution also conveniently solves our issue of assigning a single scalar value to a wavelet arrival.

Model ensembling is computationally expensive.



DeepTrace Model #1
DeepTrace Model #3
DeepTrace Model #2
DeepTrace Average

Summary

- Engineering a solution to first-break picking is nearly impossible.
- Modern compute enables training deep neural networks to learn the best approach to the problem.
- The machine learning paradigm of computing is here to stay. Large labeled data sets are the key to training A.I.
- Physics modeling and A.I. work together to give the best first-break picking results.

Thank you

Q&A?

We thank Lenovo for the opportunity to present at SEG 2019.

Contact: demoss@frontrangegeosciences.com