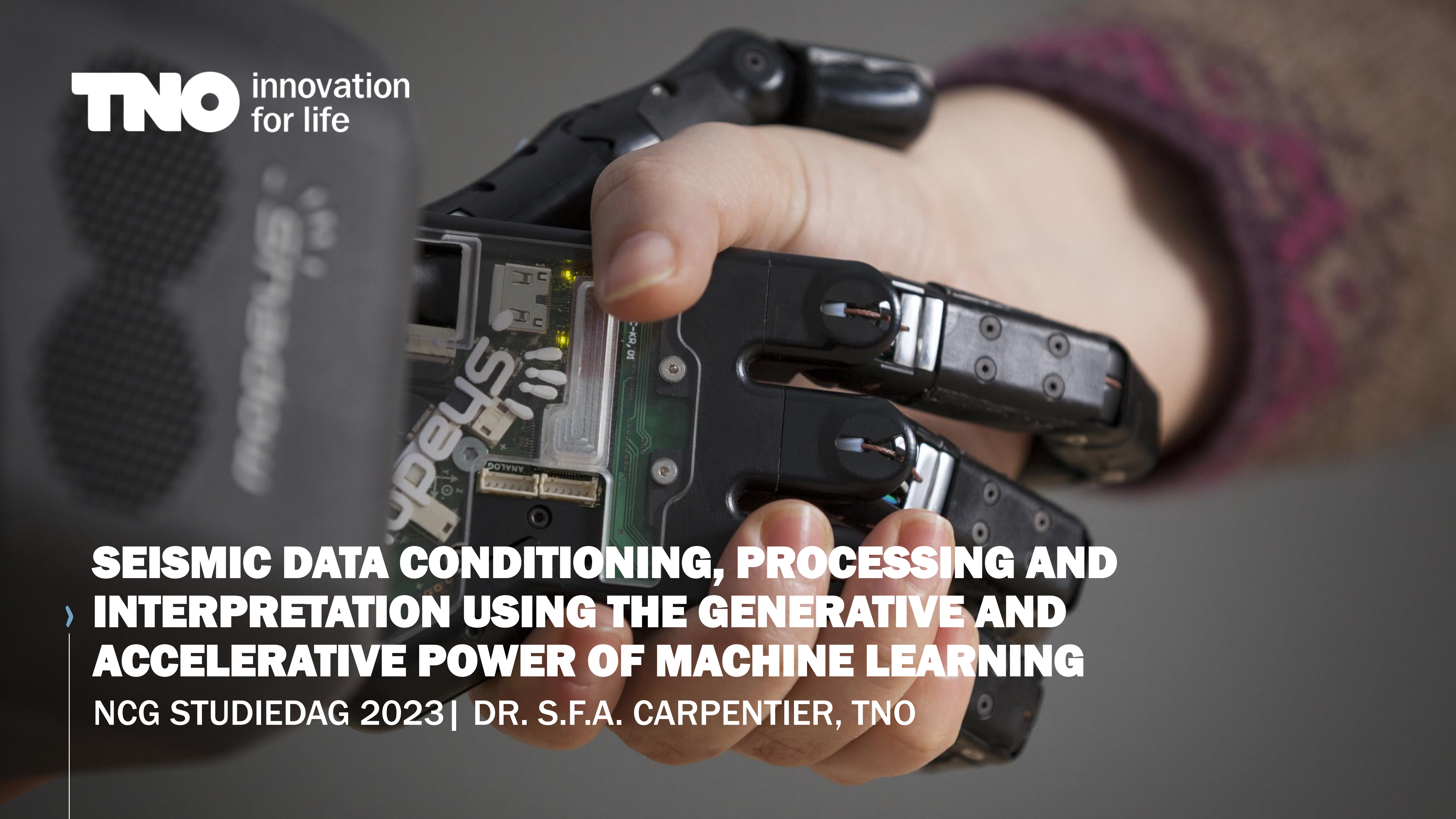


TNO innovation for life

SEISMIC DATA CONDITIONING, PROCESSING AND INTERPRETATION USING THE GENERATIVE AND ACCELERATIVE POWER OF MACHINE LEARNING

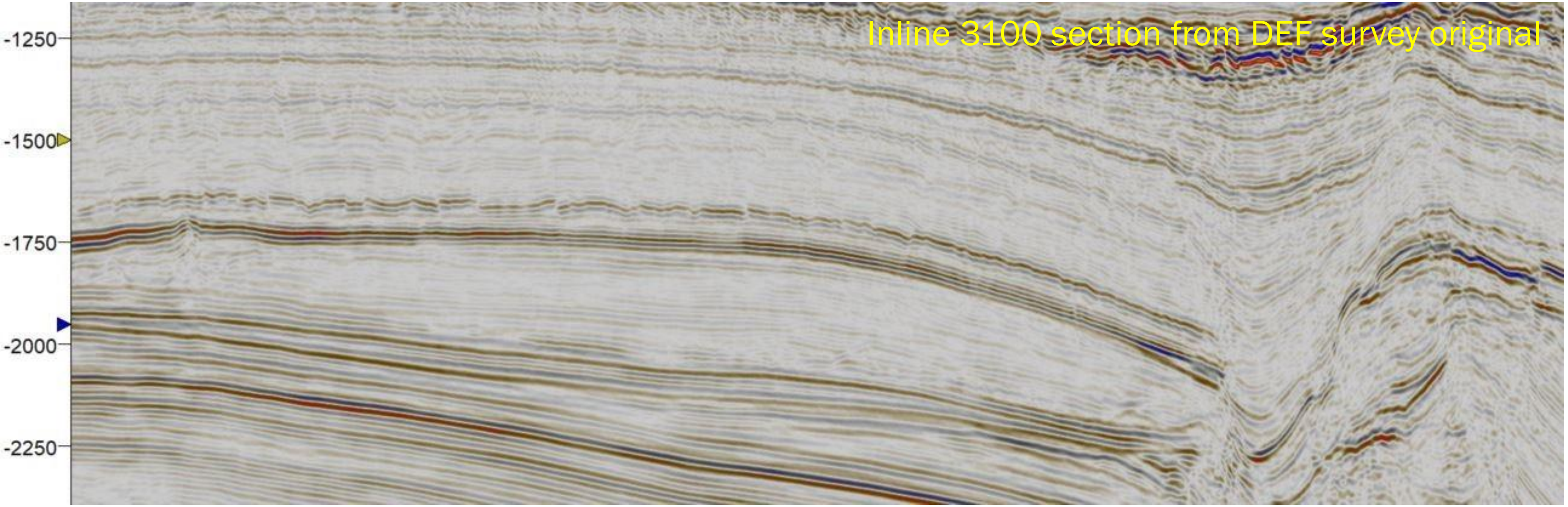
NCG STUDIEDAG 2023| DR. S.F.A. CARPENTIER, TNO

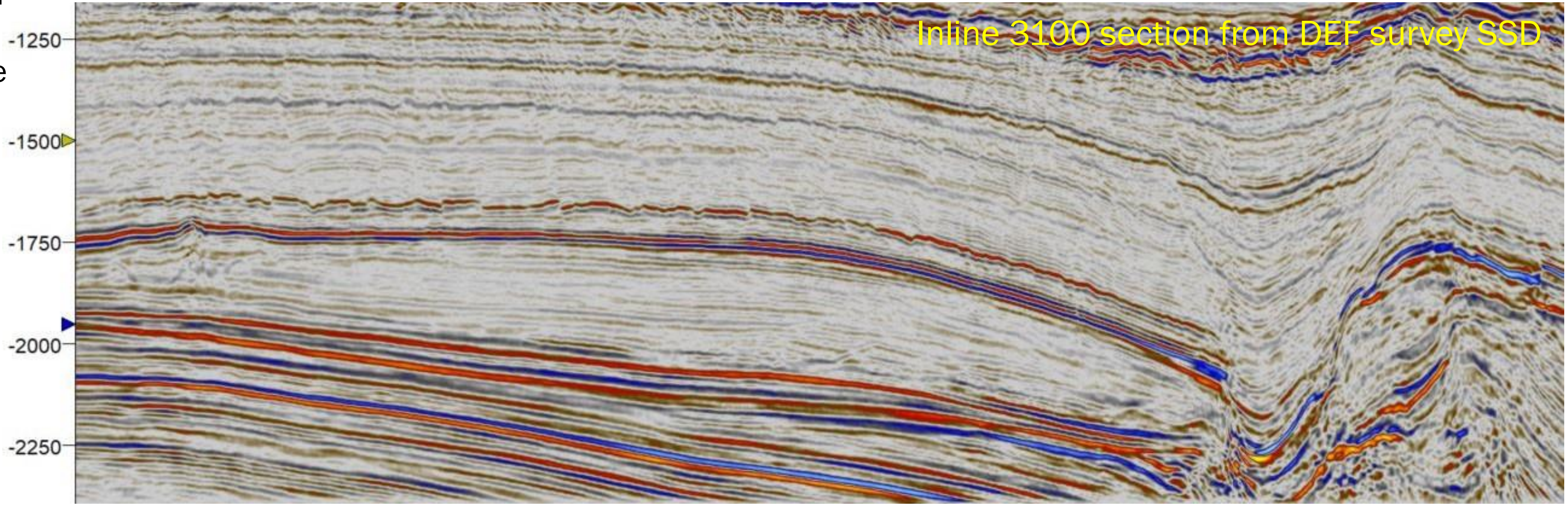
MIMIC – MACHINE INTELLIGENCE FOR MULTI-GEOPHYSICAL INTENSIVE COMPUTING

- A common problem these days in geological exploration, geophysical monitoring and derisking of sustainable energy applications like geothermal, CCS, hydrogen and storage is that it requires evermore intensive data surveying, data processing and data interpretation
- The computational power cannot keep up with the data volumes, until better software solutions arrive like improved parallelization, reservoir computing or better hardware like quantum computing. These solutions take long and the problem is now
- To tackle this problem, we can use for geophysical purposes cross-domain techniques from the Machine Learning (ML) realm as used in the multimedia and medical domain
- AI algorithms like GAN's (Generative Adversarial Networks) can mimic physics-based processing and simulation tools up to 99% accuracy, at a fraction of the computational power once trained
- We have set up a GAN tool which generates tremendously accelerated attributes and broadbanded and denoised seismic data for faster decision making in energy studies

PROBLEM

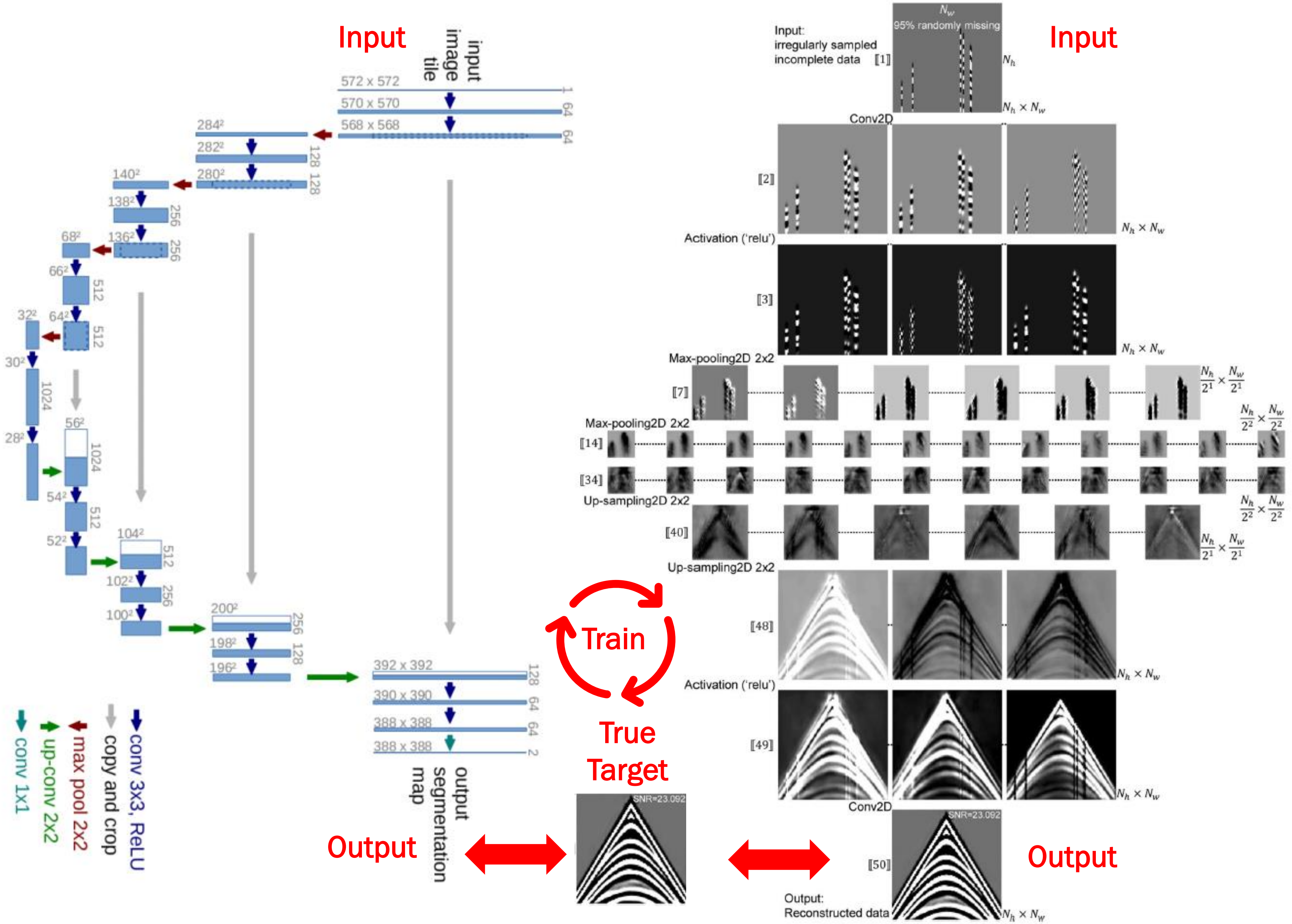
- PROBLEM: Diffraction imaging (DI), Sparse Spike Decon (SSD) broadbanding and Non Local Means (NLM) denoising take too long in computations





ACTION

- ACTION: Use Generative Adversarial Networks (GAN) as Machine Learning solution to approximate SSD and NLM computations to +- 95% accuracy at factor of +- 4000 faster



ACTION

- ACTION: Use Generative Adversarial Networks (GAN) as Machine Learning solution to approximate SSD and NLM computations to +- 95% accuracy at factor of +- 4000 faster

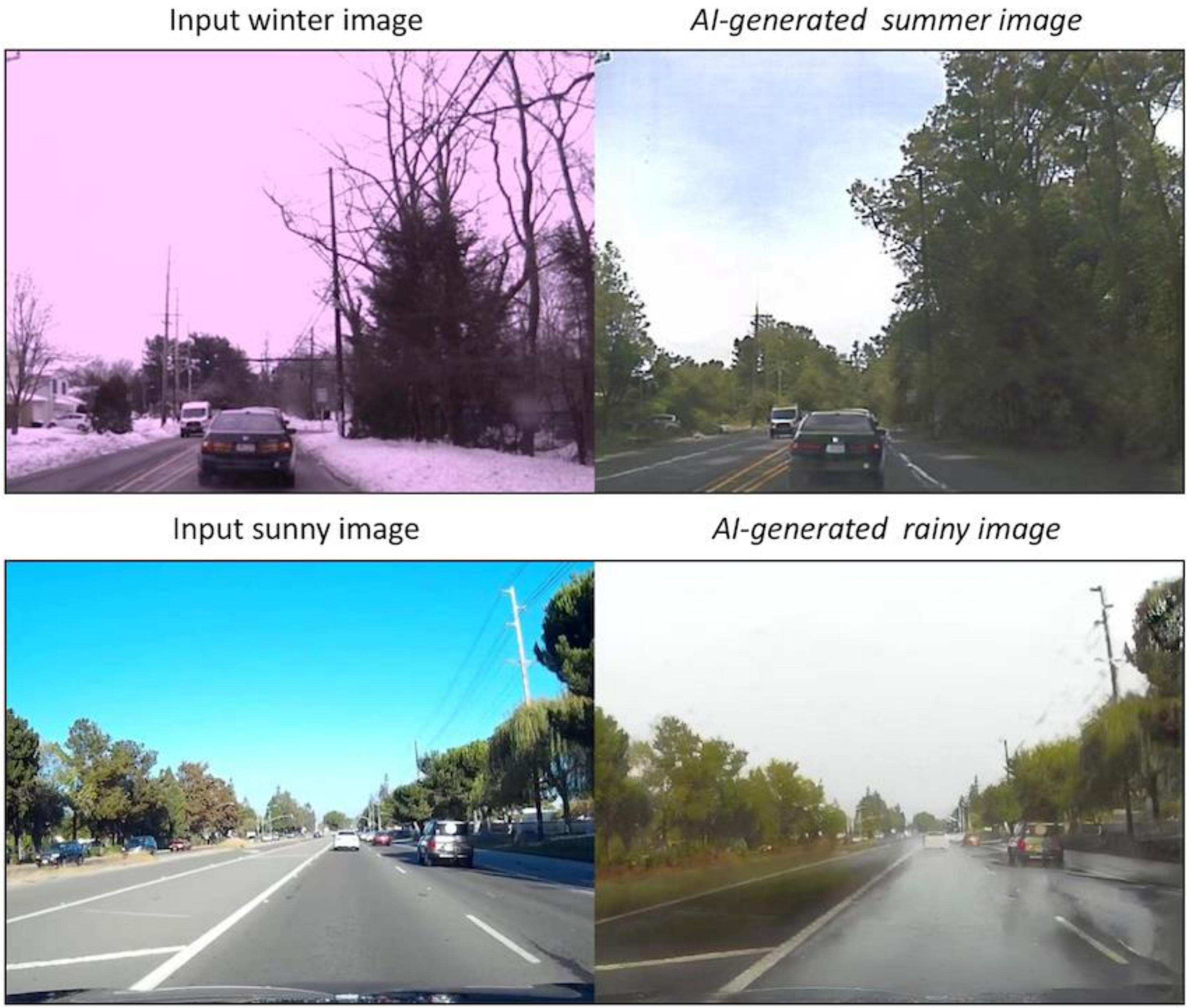
In simple words, the idea behind GANs can be summarized like this:

- Two Convolutional Neural Networks are involved.
- One of the networks, the Generator, starts off with a random data distribution and tries to replicate a particular type of distribution conditioned by a target.
- The other network, the Discriminator, through subsequent training, gets better at classifying a fake distribution from a real one.
- Both of these networks play a min-max game where one is trying to outsmart the other.
- GANs are generative models that learn a mapping from random noise vector z to output image y, $G : z \rightarrow y$. In contrast, conditional GANs learn a mapping from observed image x and random noise vector z, to y, $G : \{x, z\} \rightarrow y$.
- The generator G is trained to produce outputs that cannot be distinguished from "real" images by an adversarially trained discriminator, D, which is trained to do as well as possible at detecting the generator's "fakes".

ACTION

Non-Conditional GAN (https://github.com/junyanz/CycleGAN)

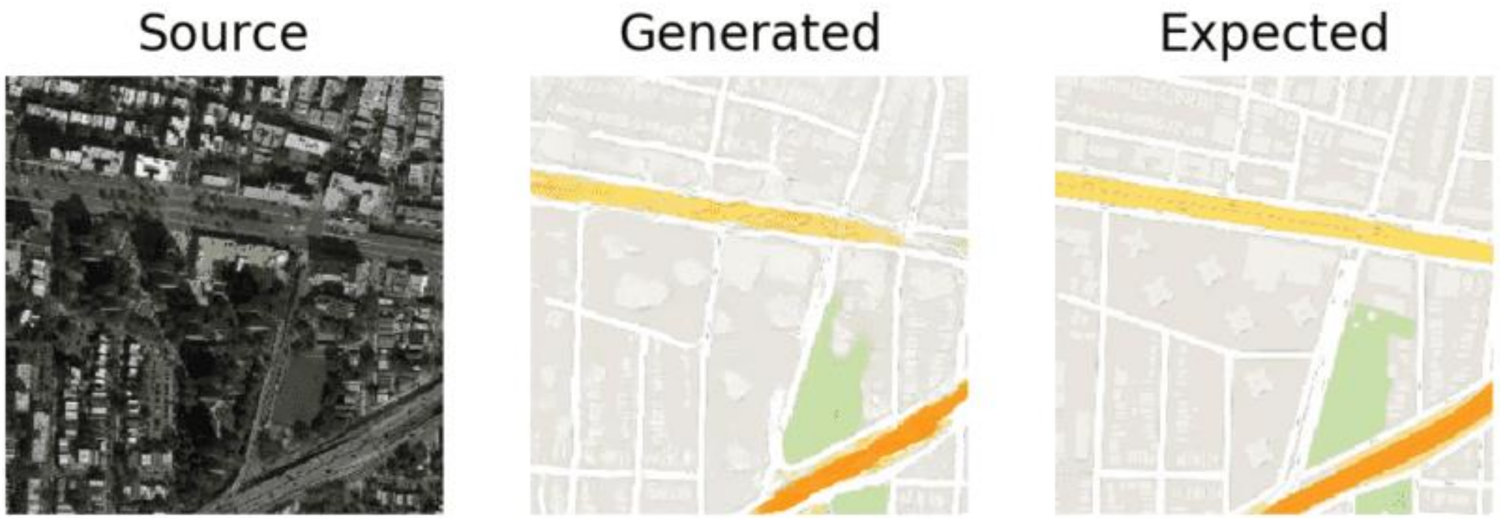
- ACTION: Use Generative Adversarial Networks (GAN) as Machine Learning solution to approximate SSD and NLM computations to +- 95% accuracy at factor of +- 4000 faster



ACTION

Conditional GAN

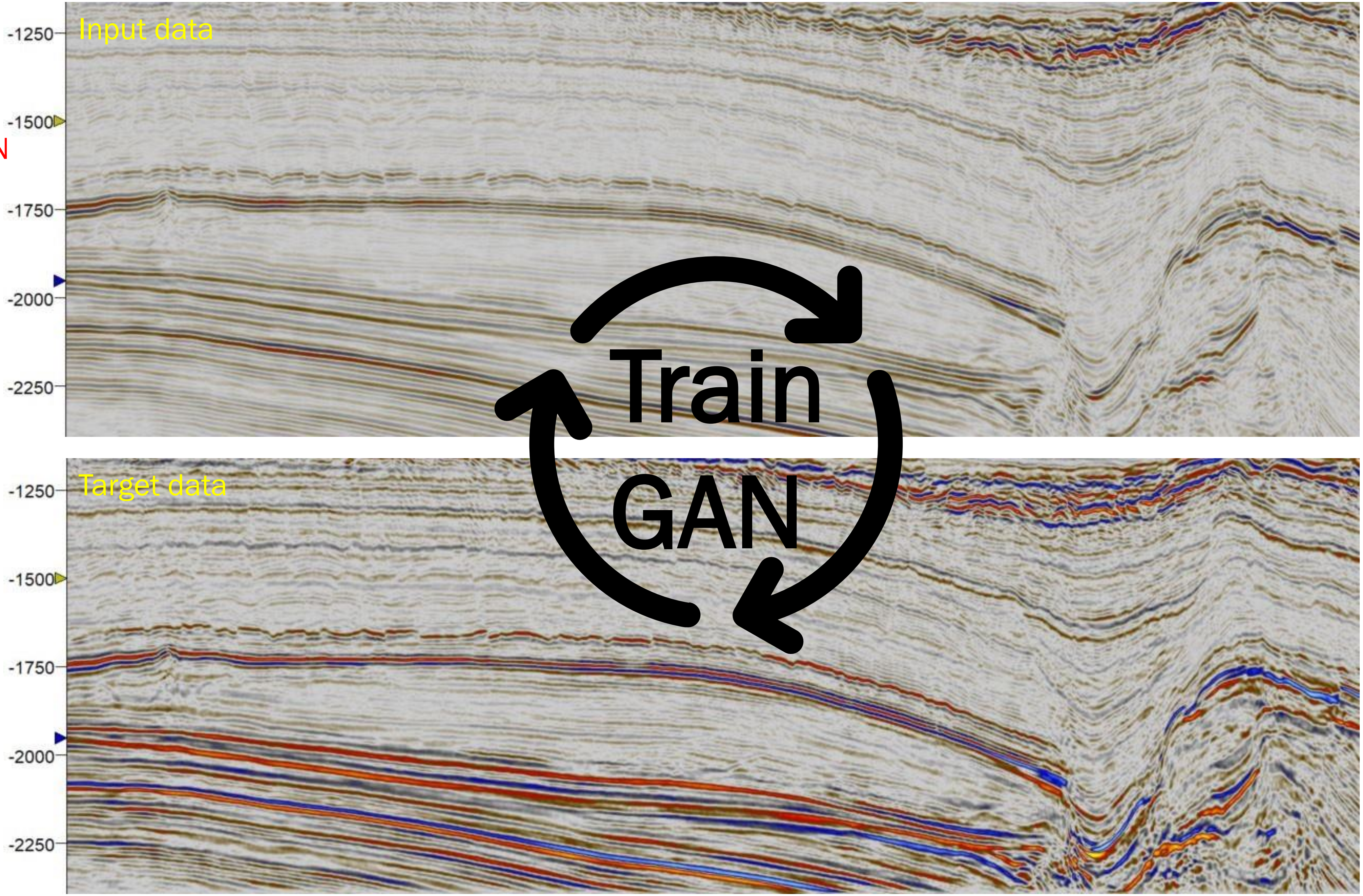
- ACTION: Use Generative Adversarial Networks (GAN) as Machine Learning solution to approximate SSD and NLM computations to +- 95% accuracy at factor of +- 4000 faster



ACTION

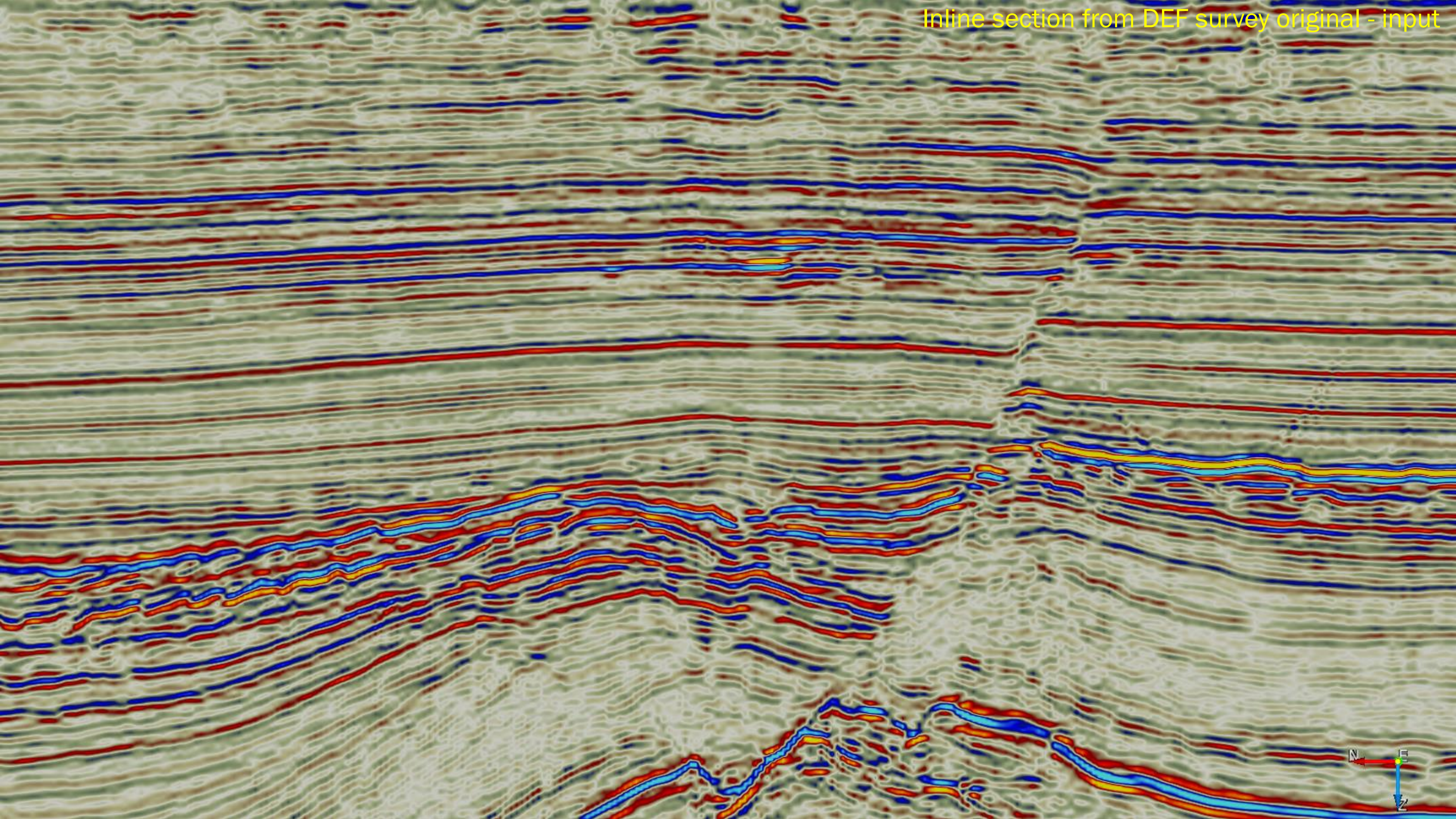
Conditional GAN

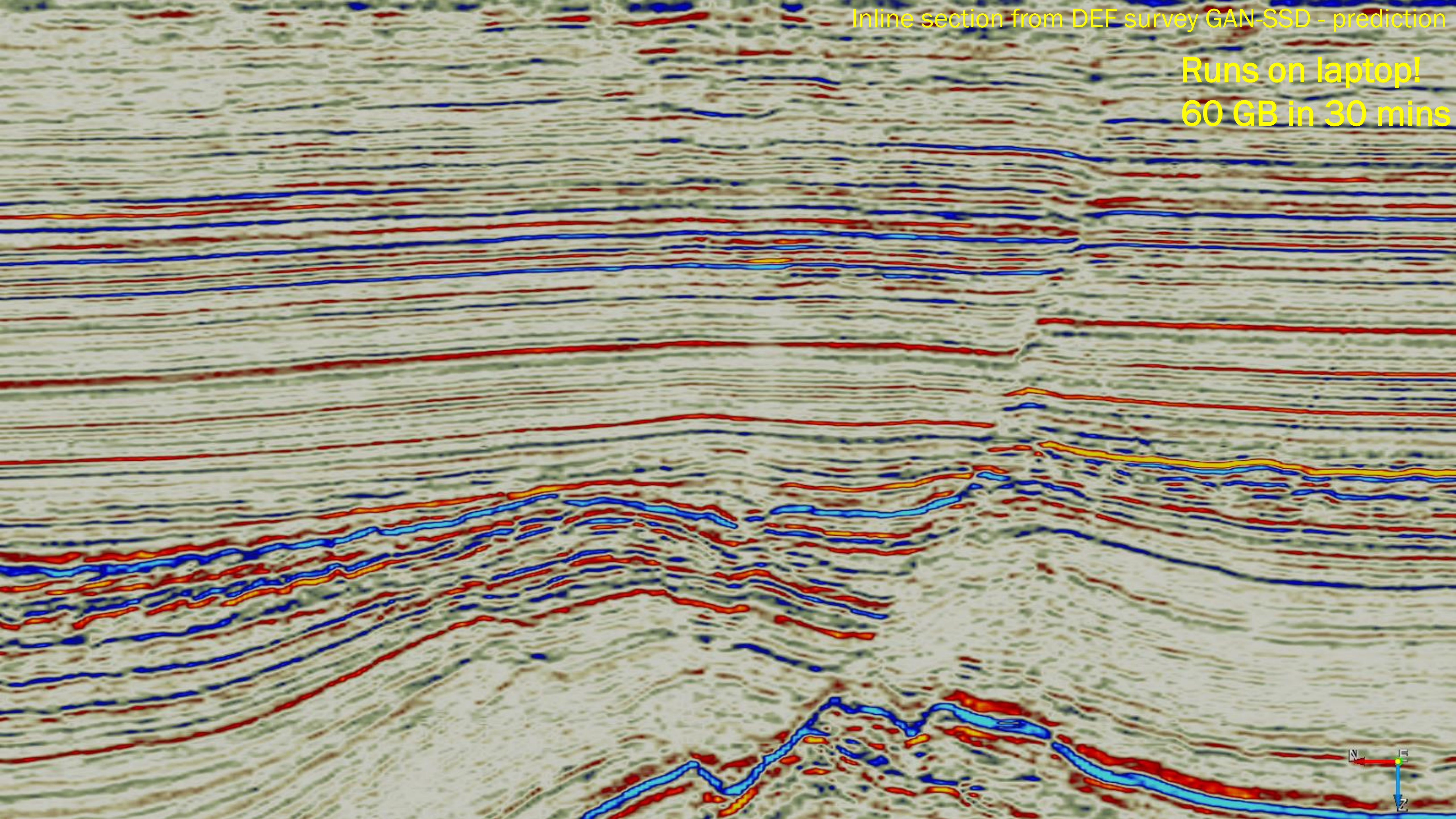
- ACTION: Use Generative Adversarial Networks (GAN) as Machine Learning solution to approximate SSD and NLM computations to +- 95% accuracy at factor of +- 4000 faster

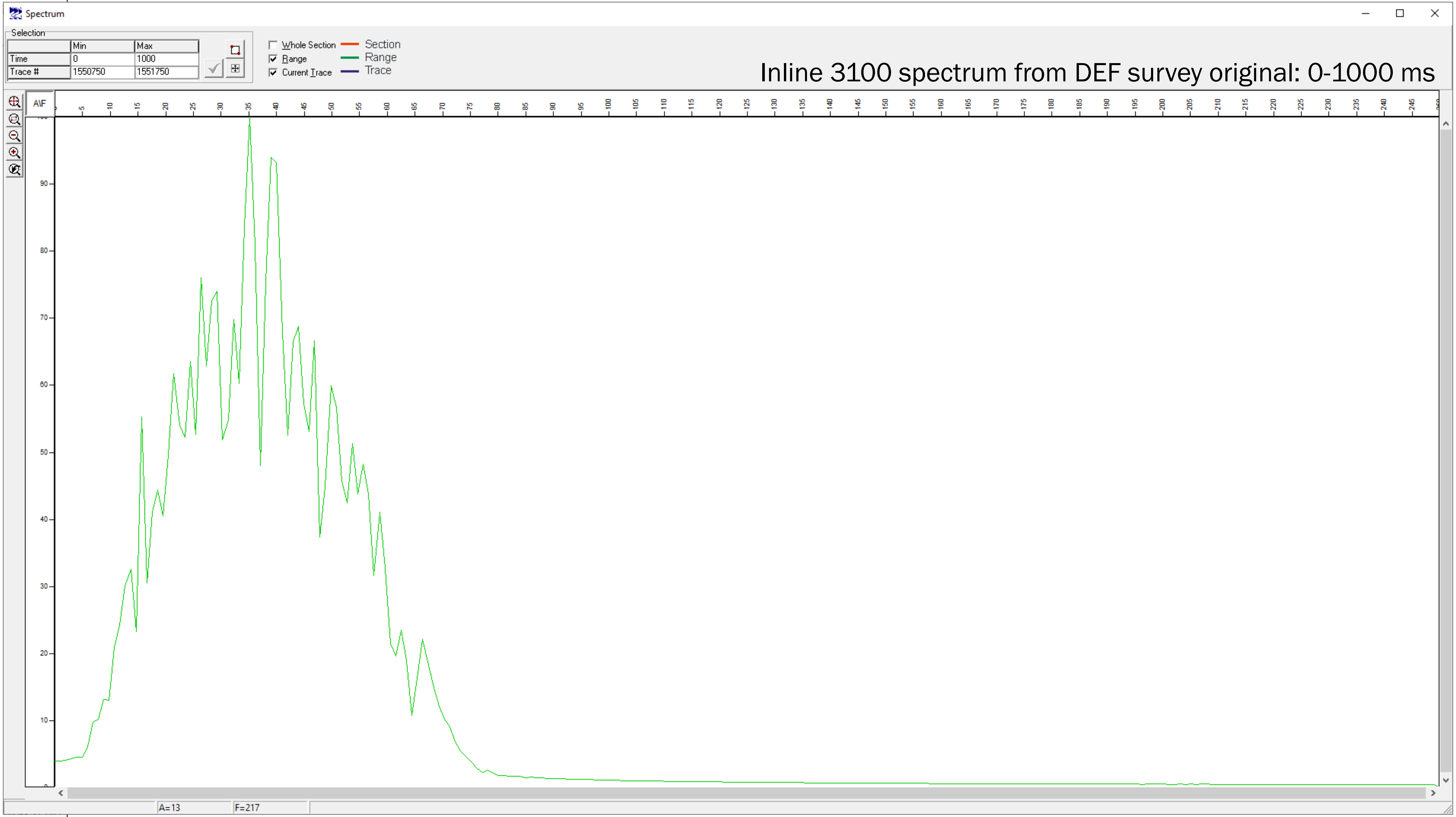


RESULT: GAN ON BROADBANDING – SPARSE SPIKE DECON (SSD)

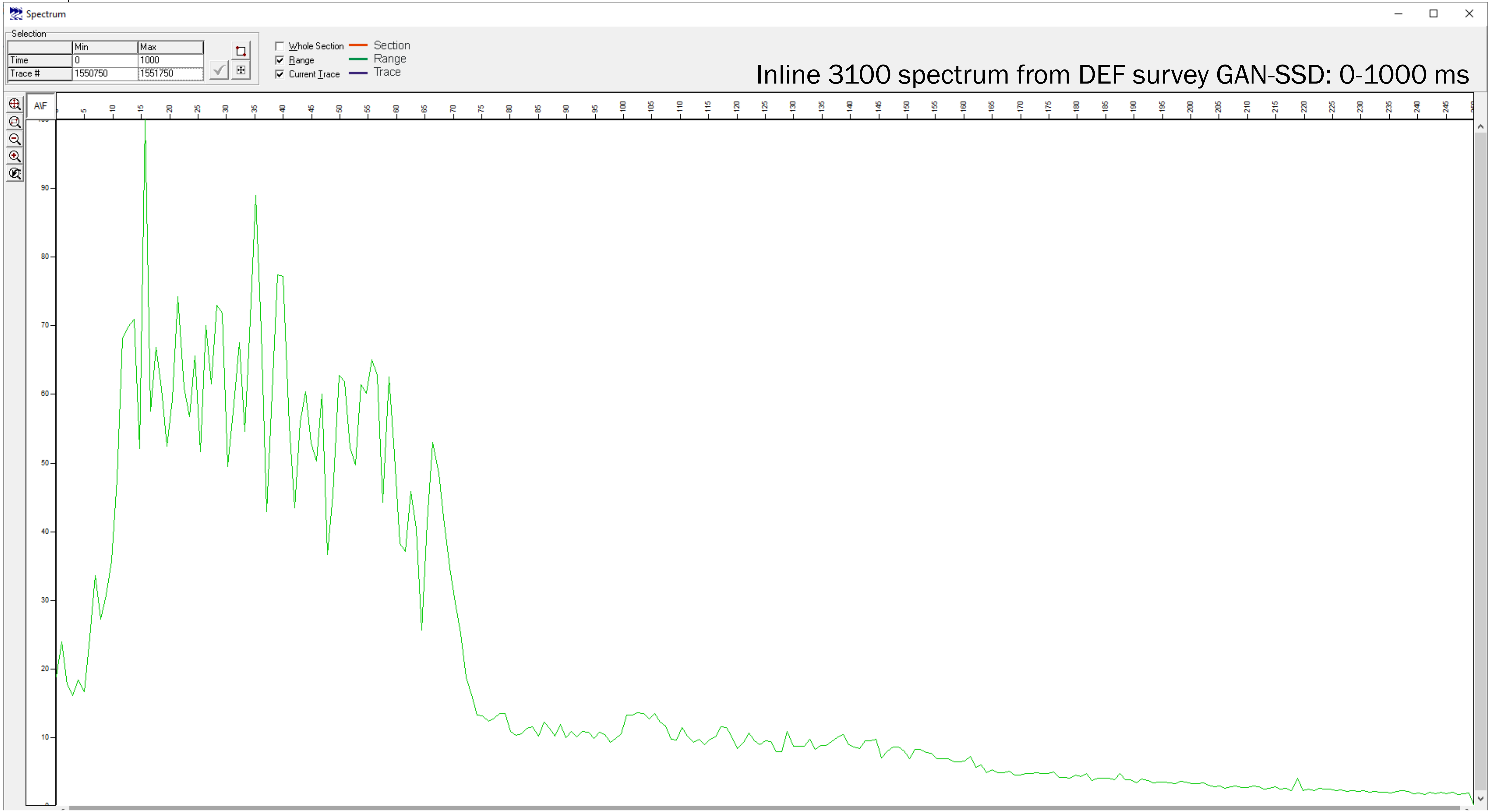
- RESULT: GAN application on broadbanding – sparse spike decon (SSD)
- Effectively: superresolution!


Inline section from DEF survey original - input


Inline section from DEF survey GAN-SSD - prediction
Runs on laptop!
60 GB in 30 mins



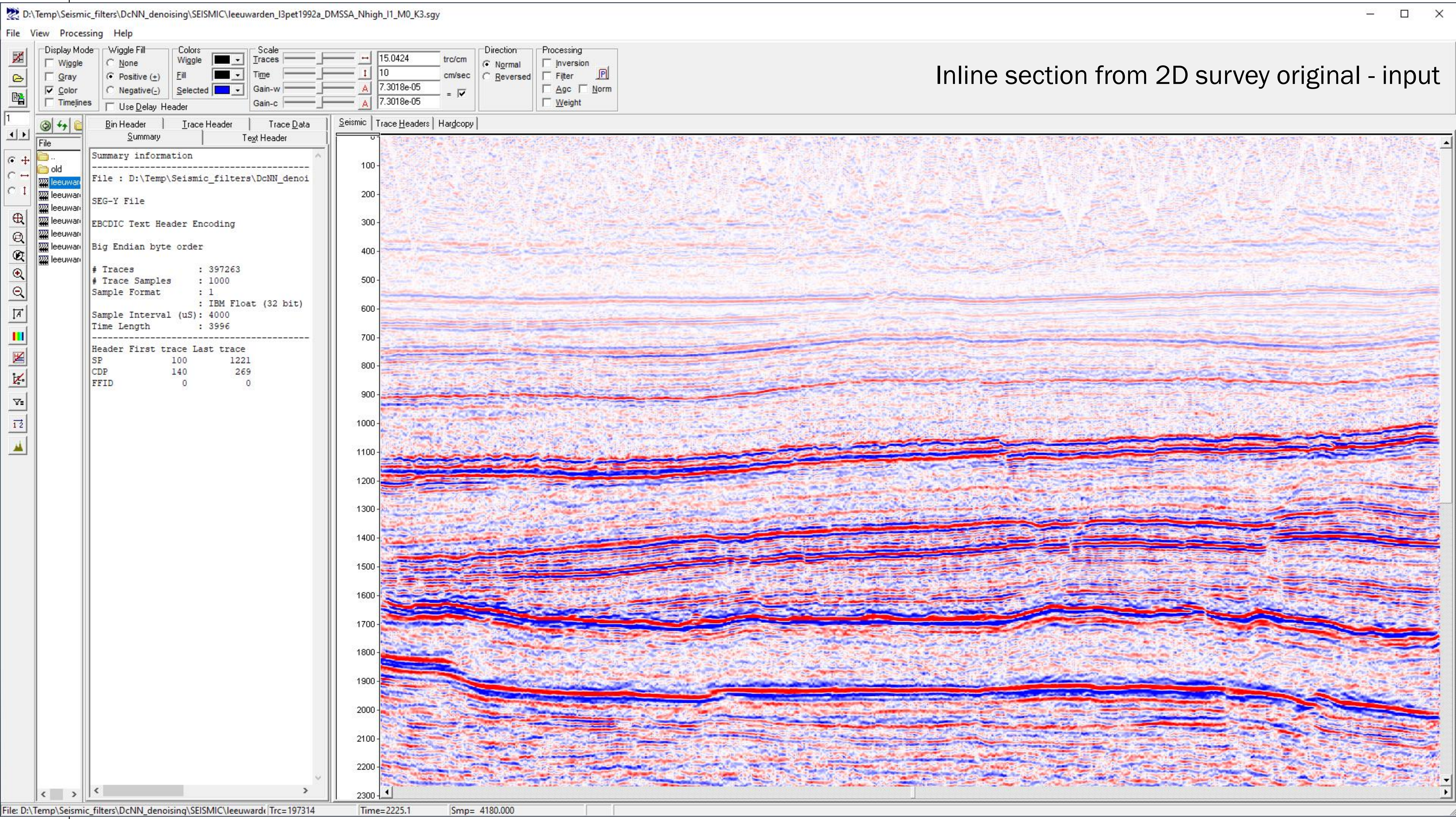
Spectrum
Selection
Min
Max
Time
0
1000
Trace #
1550750
1551750
Whole Section
Range
Current Trace
Section
Range
Trace
Inline 3100 spectrum from DEF survey original: 0-1000 ms
A\F
A=13
F=217

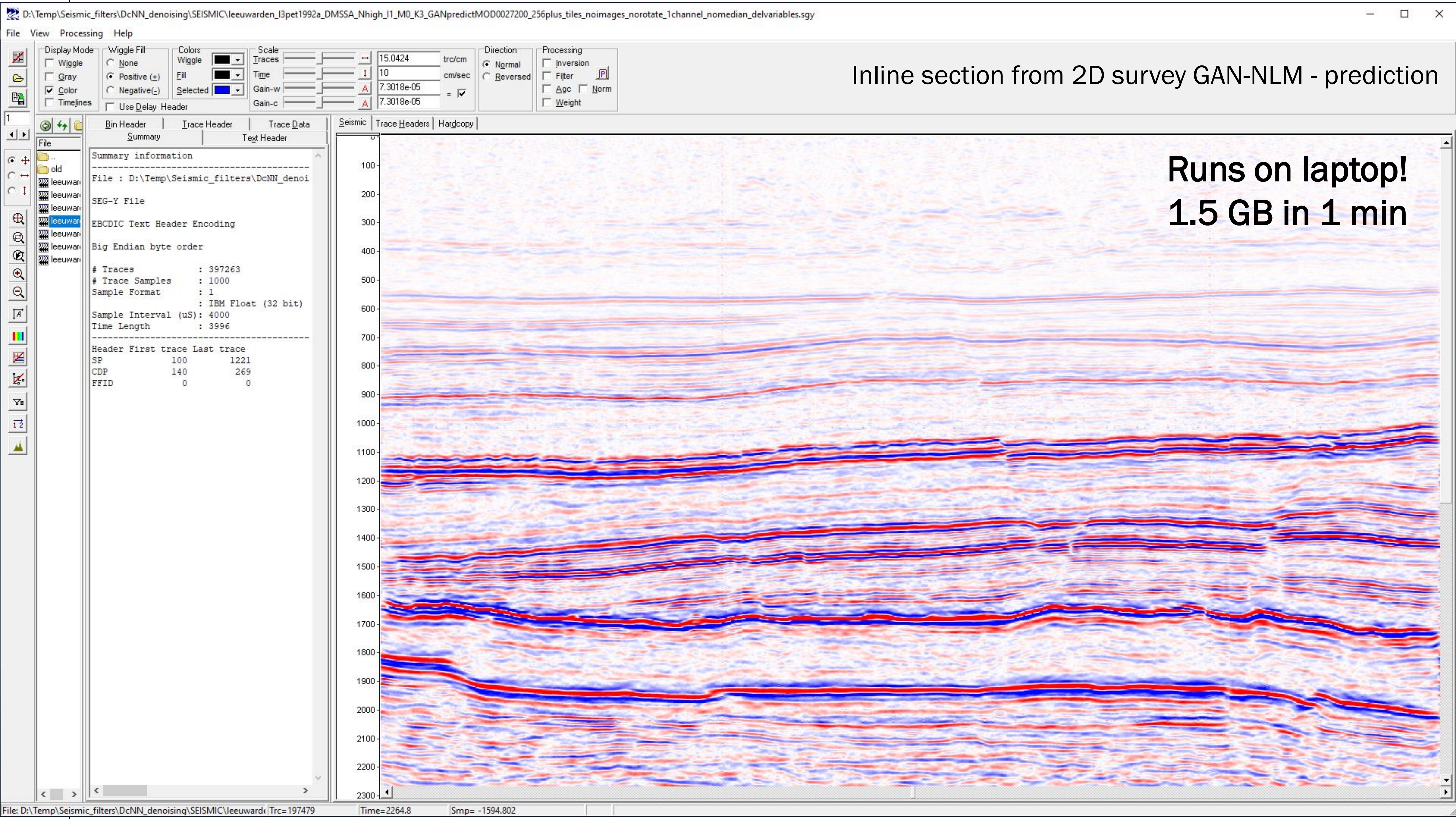

Spectrum
Selection
Min
Max
Time
0
1000
Trace #
1550750
1551750
Whole Section
Range
Current Trace
Section
Range
Trace
Inline 3100 spectrum from DEF survey GAN-SSD: 0-1000 ms
A\F
5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245
90
80
70
60
50
40
30
20
10

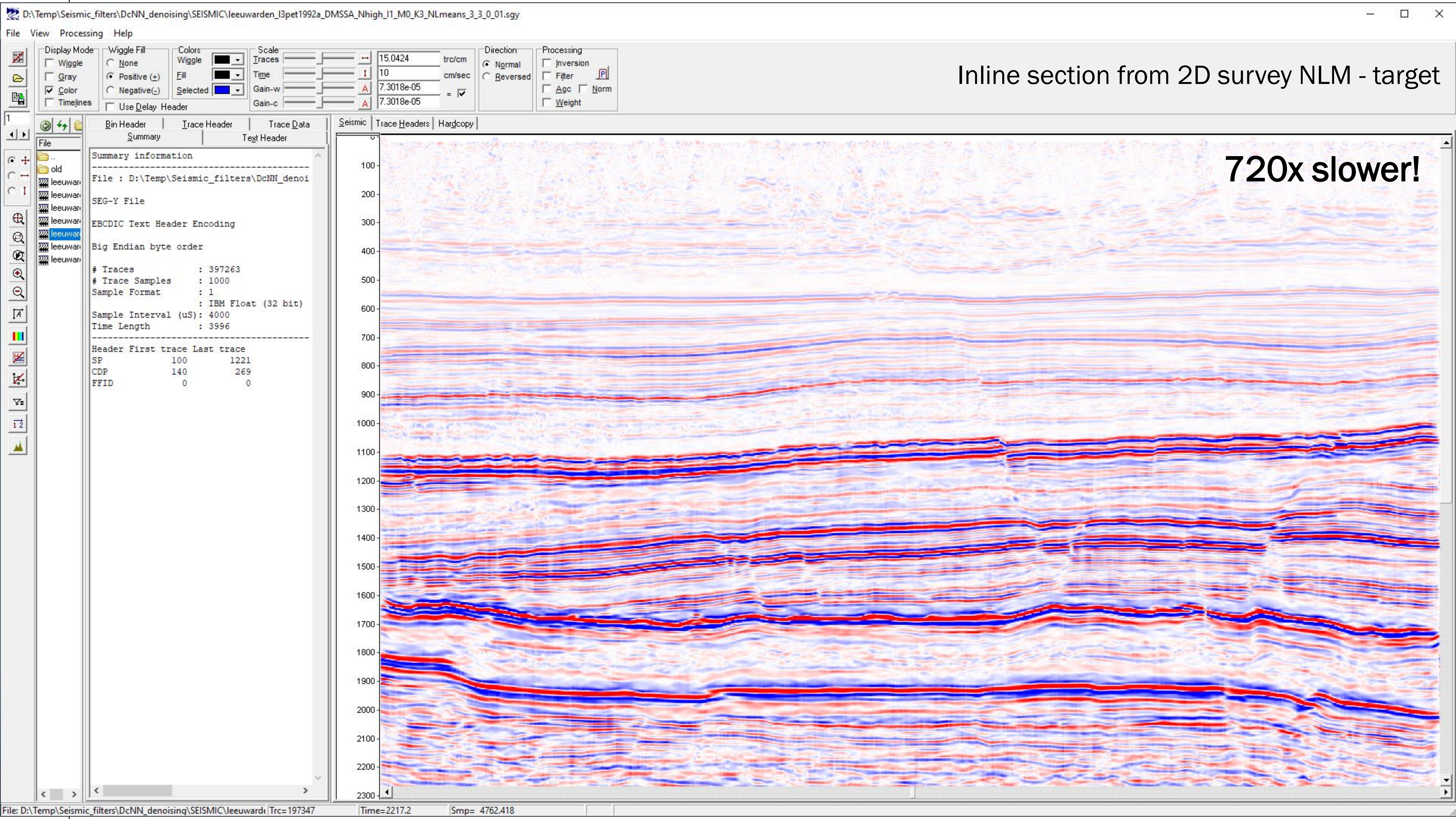
RESULT: GAN ON DENOISING NON-LOCAL MEANS (NLM)

- RESULT: GAN application on denoising non-local means (NLM)
- Effectively: cleaning data!

Inline section from 2D survey original - input




Inline section from 2D survey GAN-NLM - prediction
Runs on laptop!
1.5 GB in 1 min
D:\Temp\Seismic_filters\DcNN_denoising\SEISMIC\leeuwarden_l3pet1992a_DMSSA_Nhigh_l1_M0_K3_GANpredictMOD0027200_256plus_tiles_noimages_norotate_1channel_nomedian_delvariables.sgy
File View Processing Help
Display Mode
Wiggle
Gray
Color
Timelines
Wiggle Fill
None
Positive (+)
Negative(-)
Use Delay Header
Colors
Wiggle
Fill
Selected
Scale
Traces
Time
Gain-w
Gain-c
15.0424
10
7.3018e-05
7.3018e-05
trc/cm
cm/sec
Direction
Normal
Reversed
Processing
Inversion
Filter
Agc
Norm
Weight
Bin Header
Trace Header
Trace Data
Summary
Text Header
Seismic
Trace Headers
Hardcopy
Summary information
File : D:\Temp\Seismic_filters\DcNN_denoi
SEG-Y File
EBCDIC Text Header Encoding
Big Endian byte order
Traces : 397263
Trace Samples : 1000
Sample Format : 1
: IBM Float (32 bit)
Sample Interval (uS): 4000
Time Length : 3996
Header First trace Last trace
SP 100 1221
CDP 140 269
FFID 0 0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
File: D:\Temp\Seismic_filters\DcNN_denoising\SEISMIC\leeuward
Trc=197479
Time=2264.8
Smp= -1594.802



Inline section from 2D survey NLM - target
720x slower!
D:\Temp\Seismic_filters\DcNN_denoising\SEISMIC\leeuwarden_l3pet1992a_DMSSA_Nhigh_l1_M0_K3_NLmeans_3_3_0_01.sgy
File View Processing Help
Display Mode
Wiggle
Gray
Color
Timelines
Wiggle Fill
None
Positive (+)
Negative(-)
Use Delay Header
Colors
Wiggle
Fill
Selected
Scale
Traces
Time
Gain-w
Gain-c
15.0424
10
7.3018e-05
7.3018e-05
trc/cm
cm/sec
Direction
Normal
Reversed
Processing
Inversion
Filter
Agc
Norm
Weight
Bin Header
Trace Header
Trace Data
Summary
Text Header
Seismic
Trace Headers
Hardcopy
File
old
Summary information
File : D:\Temp\Seismic_filters\DcNN_denoi
SEG-Y File
EBCDIC Text Header Encoding
Big Endian byte order
Traces : 397263
Trace Samples : 1000
Sample Format : 1
: IBM Float (32 bit)
Sample Interval (uS): 4000
Time Length : 3996
Header First trace Last trace
SP 100 1221
CDP 140 269
FFID 0 0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
2000
2100
2200
2300
File: D:\Temp\Seismic_filters\DcNN_denoising\SEISMIC\leeuwarde
Trc=197347
Time=2217.2
Smp= 4762.418

ML INTERPOLATION OF SPARSE 3D SEISMIC DATA USING GAN

- Big Data in Offshore Windfarms:
  - Research question: is it possible to generate a 3D dip volume in a given survey area from 2D sparse seismic data?
  - Answer: yes, we think it is possible. We will use the Ten Noorden van Wadden Windfarm 2D HRS seismic dataset to attempt:
  - 1) gridding an arbitrary set of 2D lines onto a 3D grid based on coordinates
  - 2) bin and stack the seismic traces into a sparse 3D volume
  - 3) interpolate the sparse 3D data into a dense 3D volume using state-of-the-art Machine Learning: MDA GAN

ML DATA INTERPOLATION: MDA GAN (MULTI-DIMENSIONAL ADVERSARIAL)

innovation for life

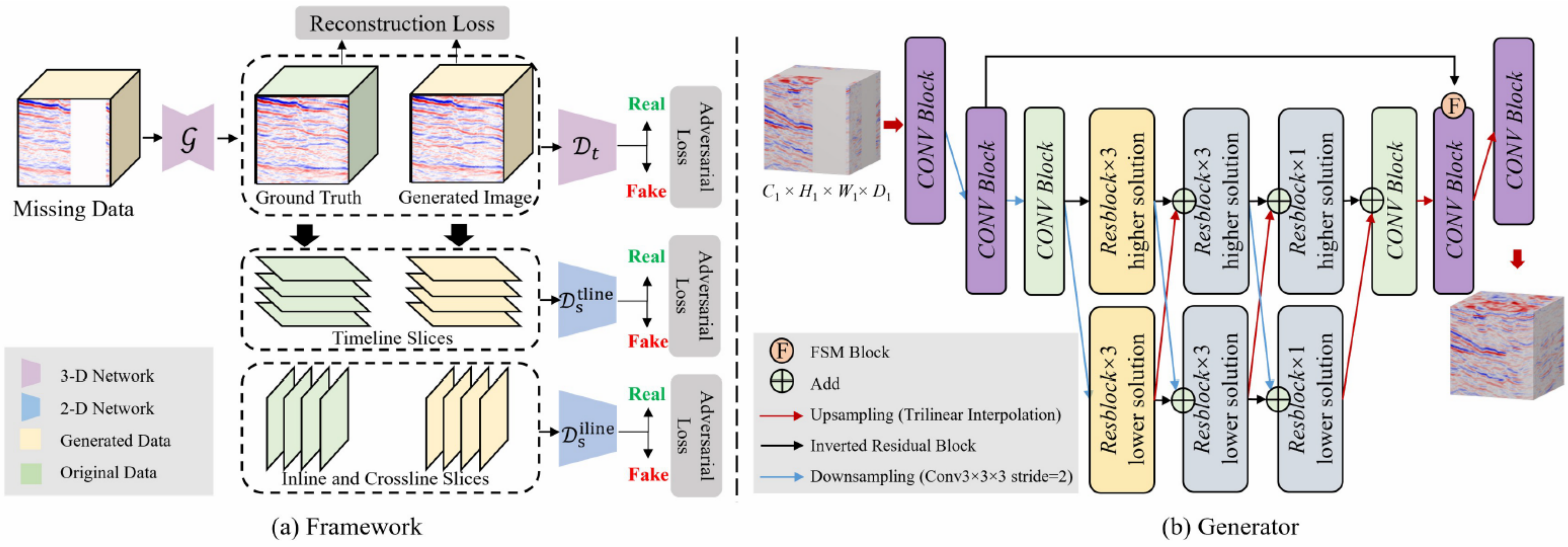


Fig. 1. In (a), the framework consists of one 3-D generator, one 3-D discriminator and two 2-D discriminators. For training, the input to the 3D network is data of size $128 \times 128 \times 128$, and the batch size is b. To conserve the RAM, the 2D discriminator randomly draws 8 slices of 128×128 in the 3D data along the corresponding direction as input, and the batch size is $8 \times b$. While for inference, the input to the generator can be any size as allowed by the hardware. (b) is the detailed structure of the generator in the framework, and the discriminator follows the standard encoder structure.The CONV block consists of a 3×3 convolution, a normalization layer and a LeakyReLU activation function, Resblock was proposed by He et al [40].

ML DATA INTERPOLATION: MDA GAN PRINCIPLE

TNO innovation for life

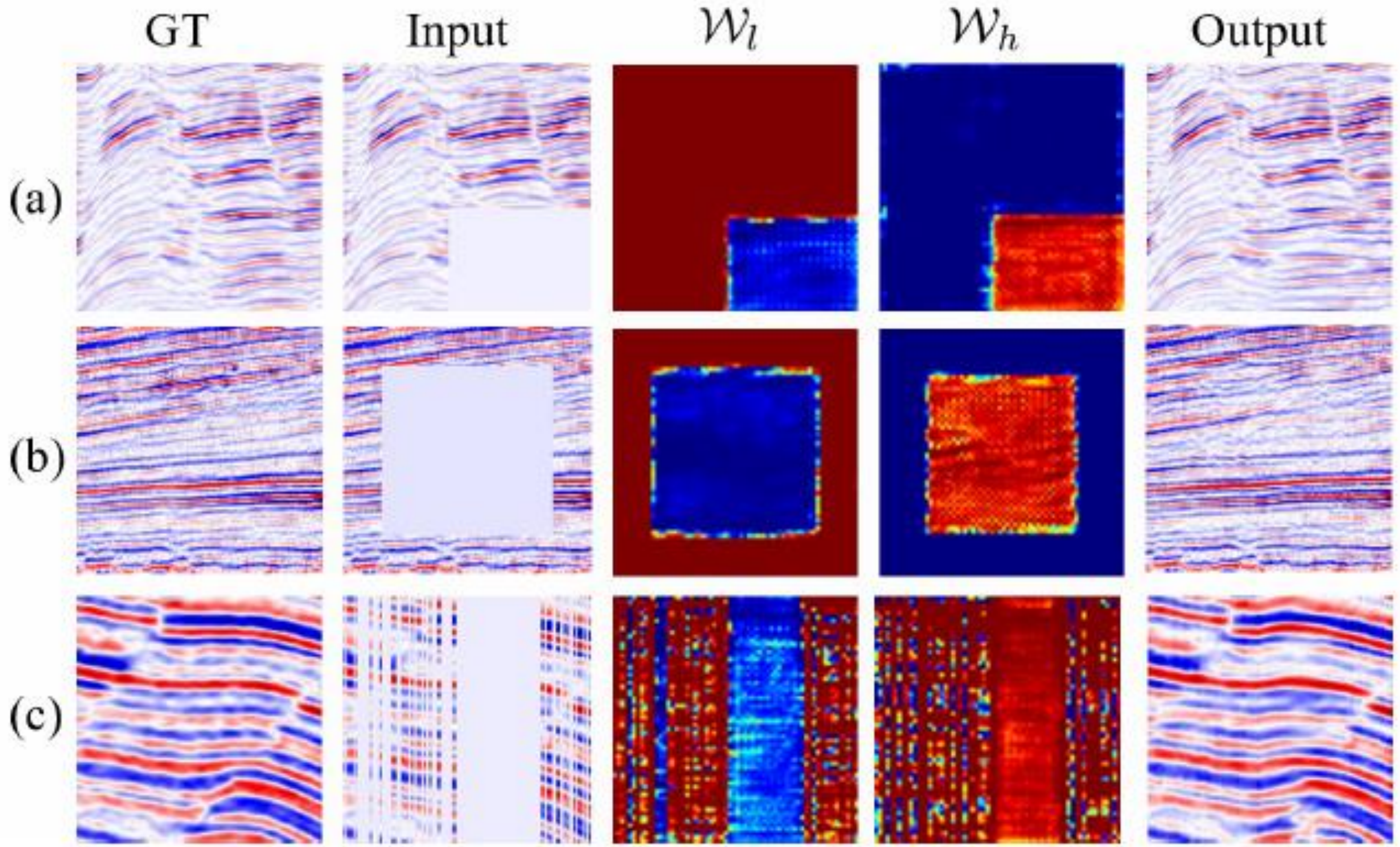


Fig. 3. The figure is shown as 2-D slices of 128^2 in 3-D volumes of 128^3, displaying the missing of the five modes. The FSM generates mask-like heatmaps without any mask supervision information.

ML DATA INTERPOLATION: MDA GAN APPLIED ON F3 3D DATA

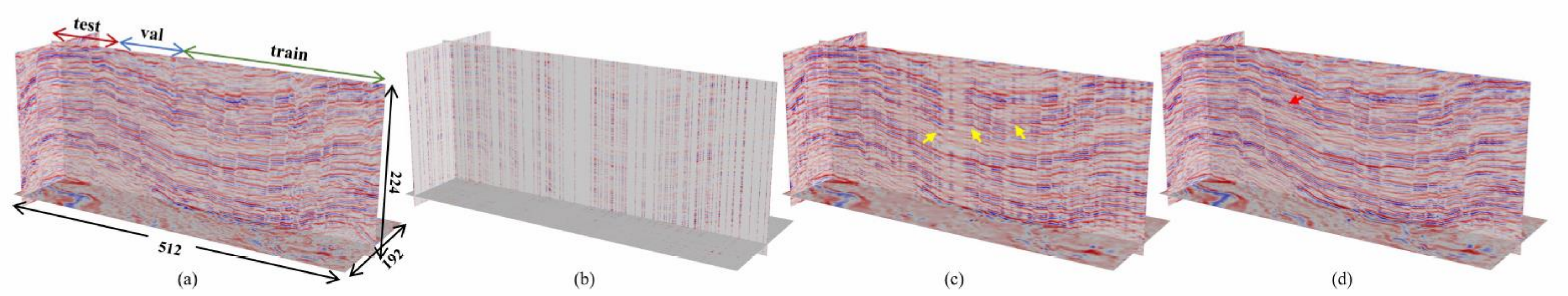


Fig. 12. (a) New Zealand Kerry original data, (b) 80% traces loss in both inline and crossline directions, (c) UNet interpolation results, (d) MDA GAN interpolation results.

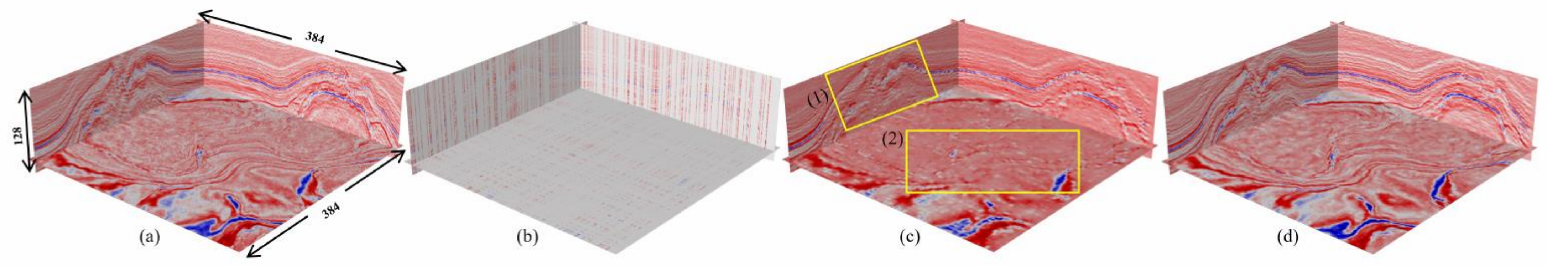


Fig. 13. (a) F3 Netherlands original data, (b) 80% traces loss in both inline and crossline directions, (c) UNet interpolation results, (d) MDA GAN interpolation results.

TNW 2D DATA EAST-WEST LINES

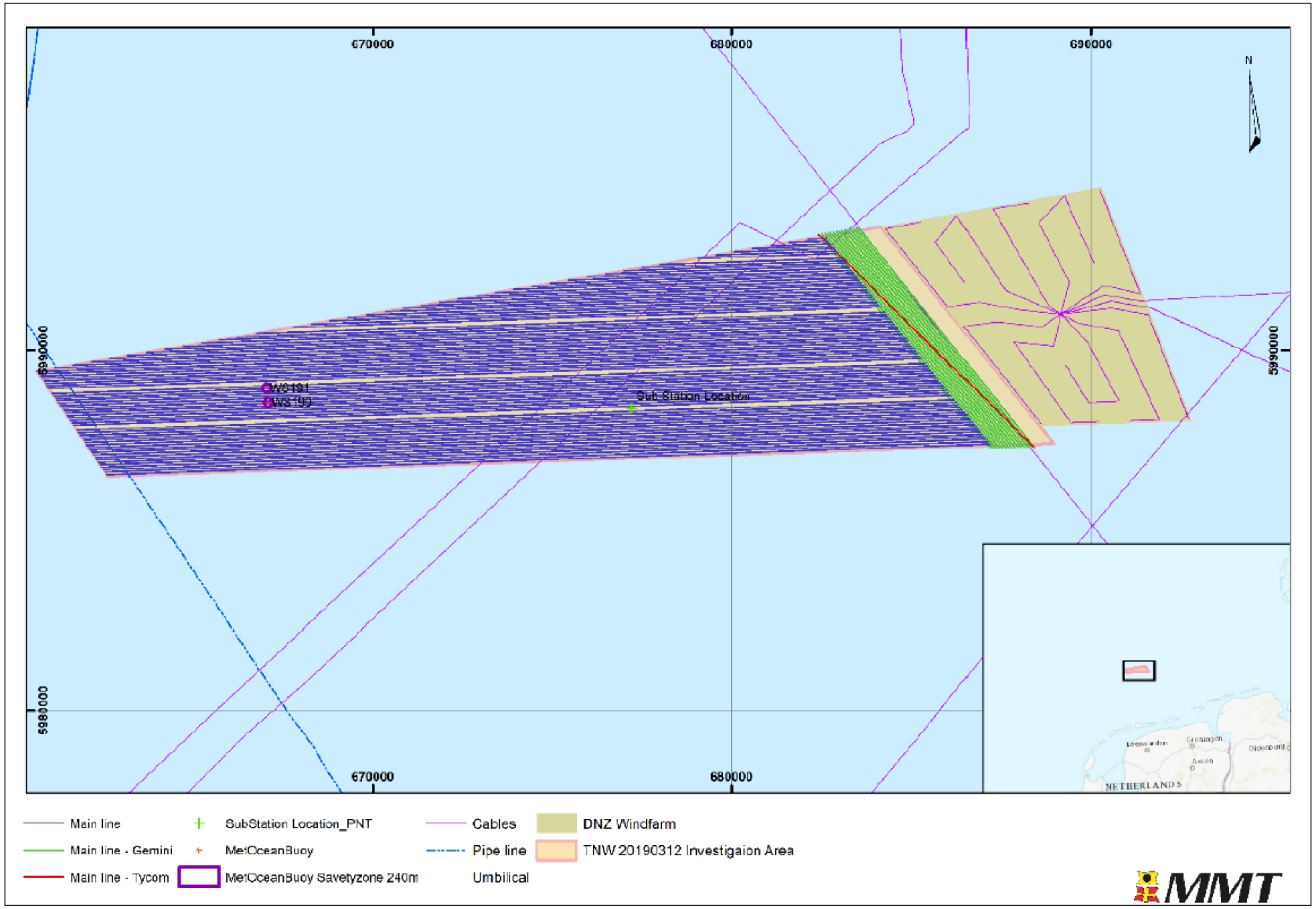


Figure 5 Line plan – main lines TNW-A (blue) and TNW-B (green)
Gaps in the line plan are where the reference lines were acquired

TNW 2D DATA NORTH-SOUTH LINES

Gaps in the line plan are where the reference lines were acquired

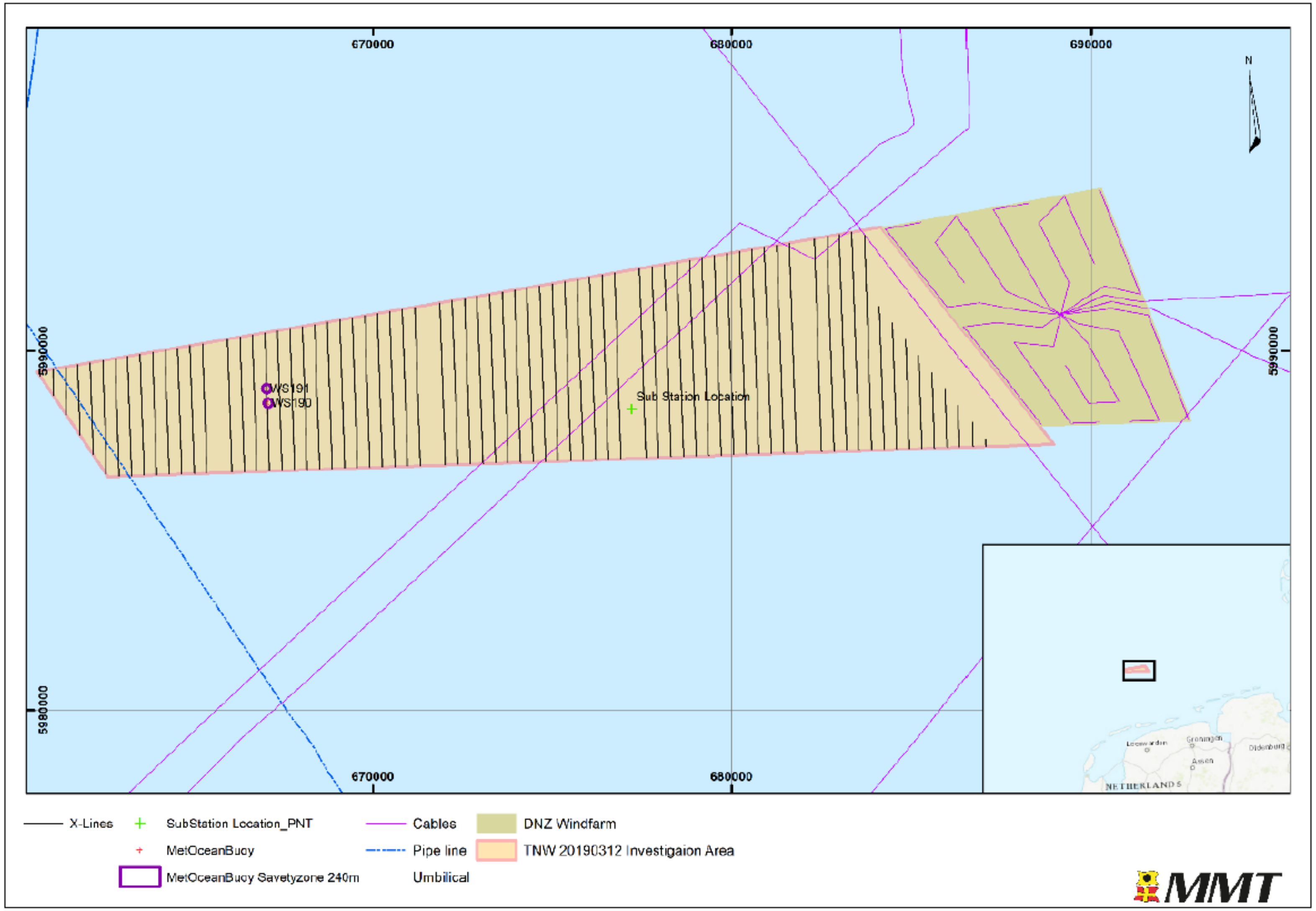
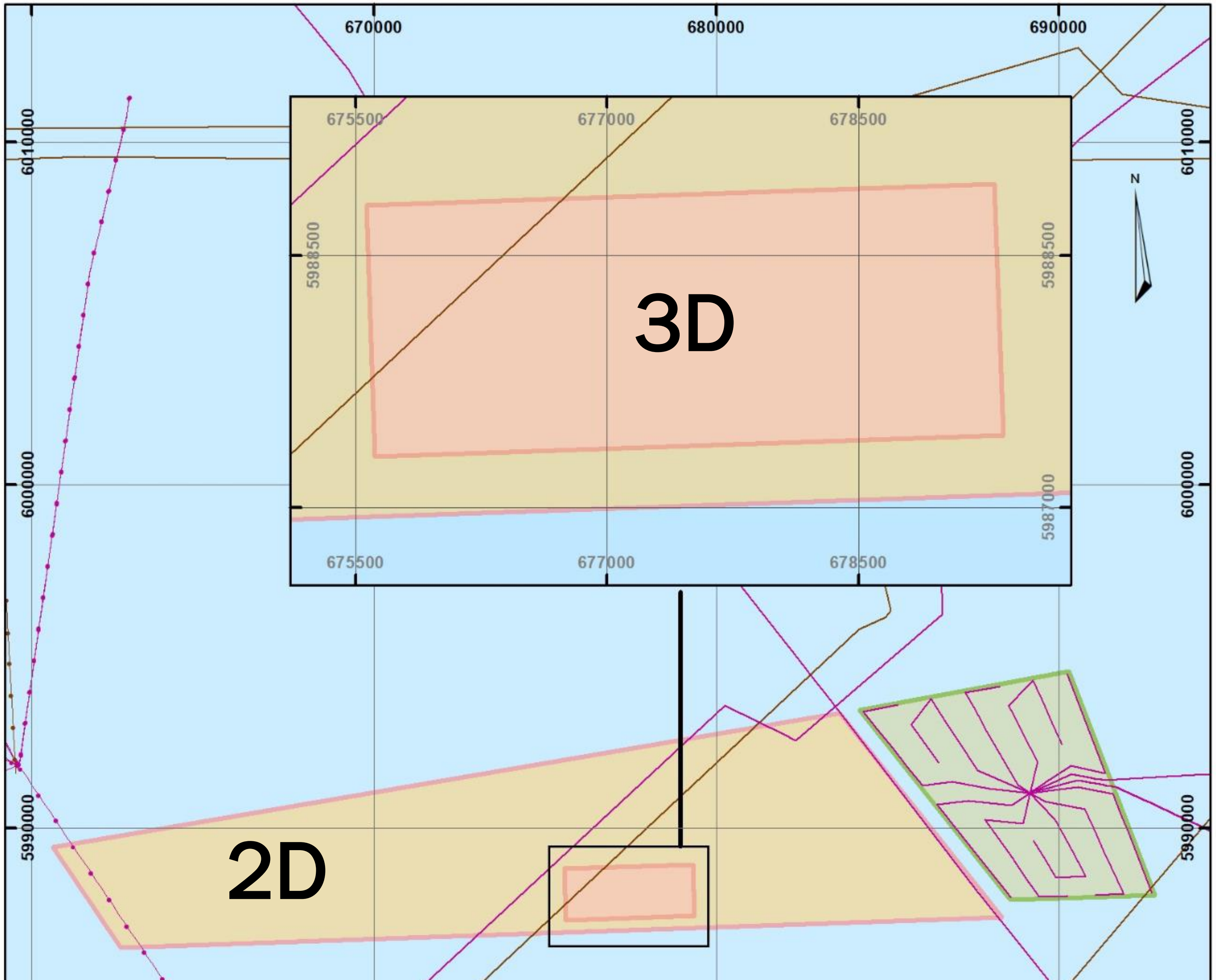


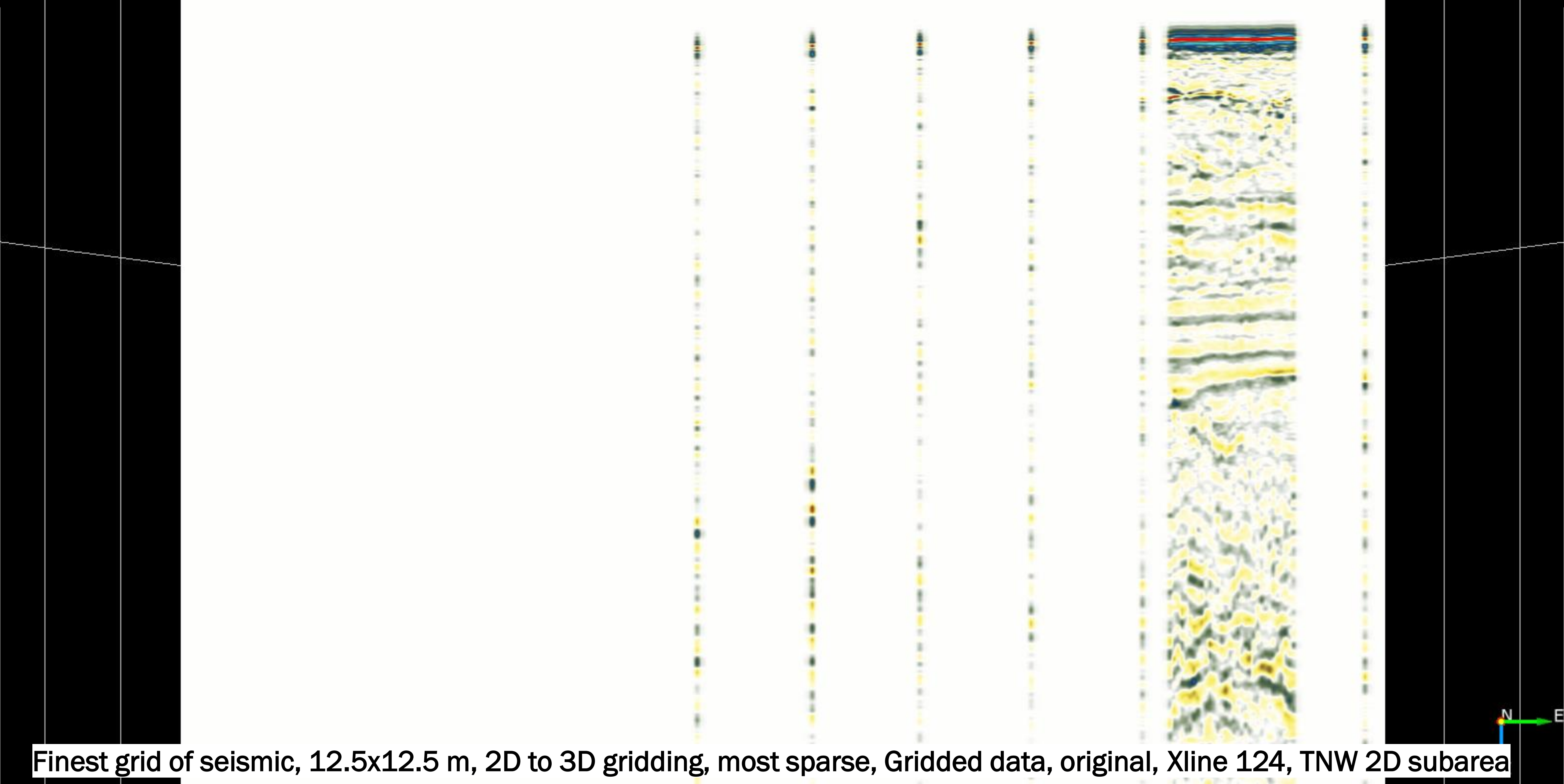
Figure 6 Line plan – cross lines TNW-A (black)
Gaps in the line plan are where the reference lines were acquired

TNW 3D HIGH-RES DATA



MDA GAN ON TNW 2D-3D DATA: ORG

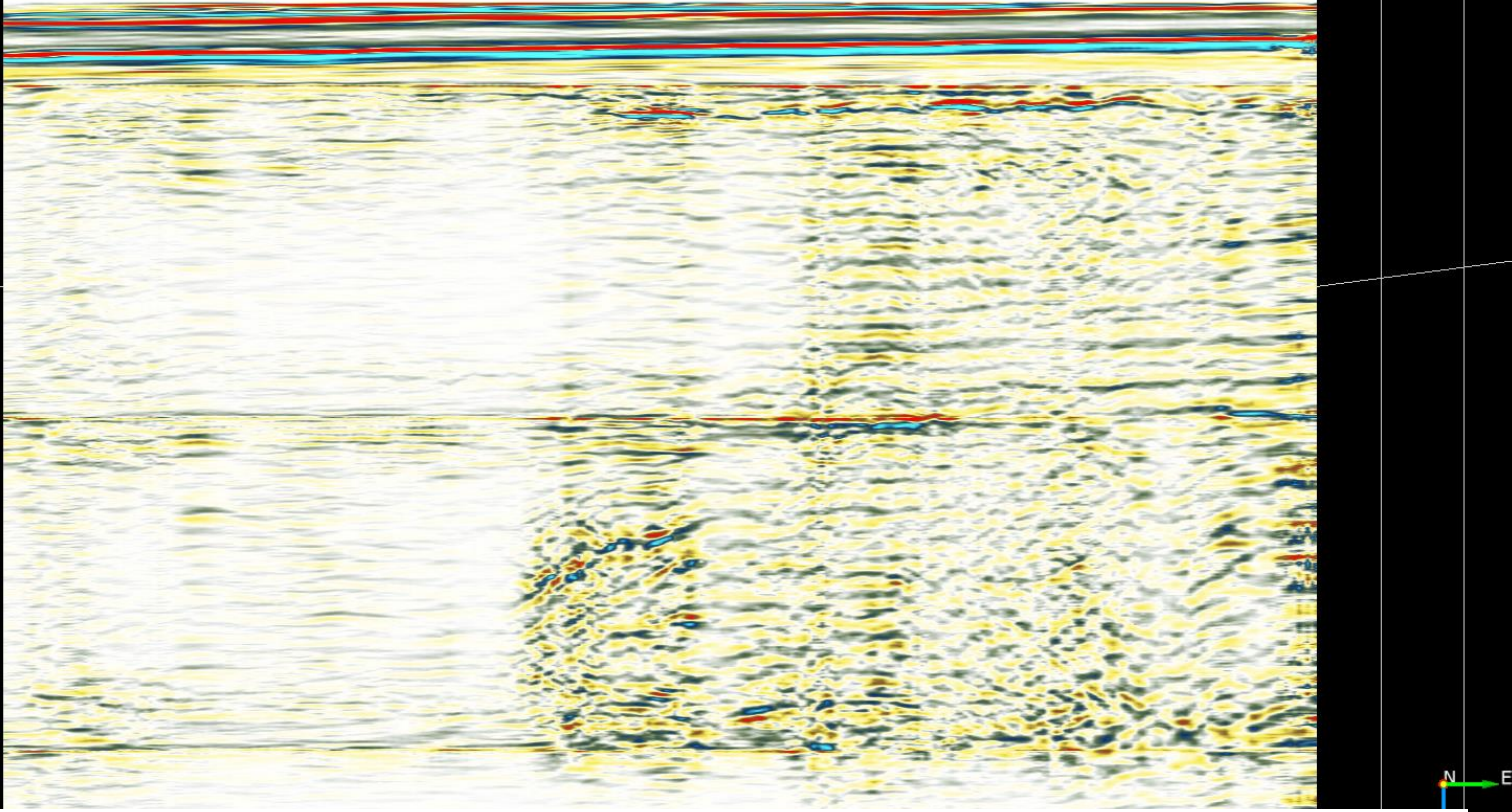
83% sparse data!



Finest grid of seismic, 12.5x12.5 m, 2D to 3D gridding, most sparse, Gridded data, original, Xline 124, TNW 2D subarea

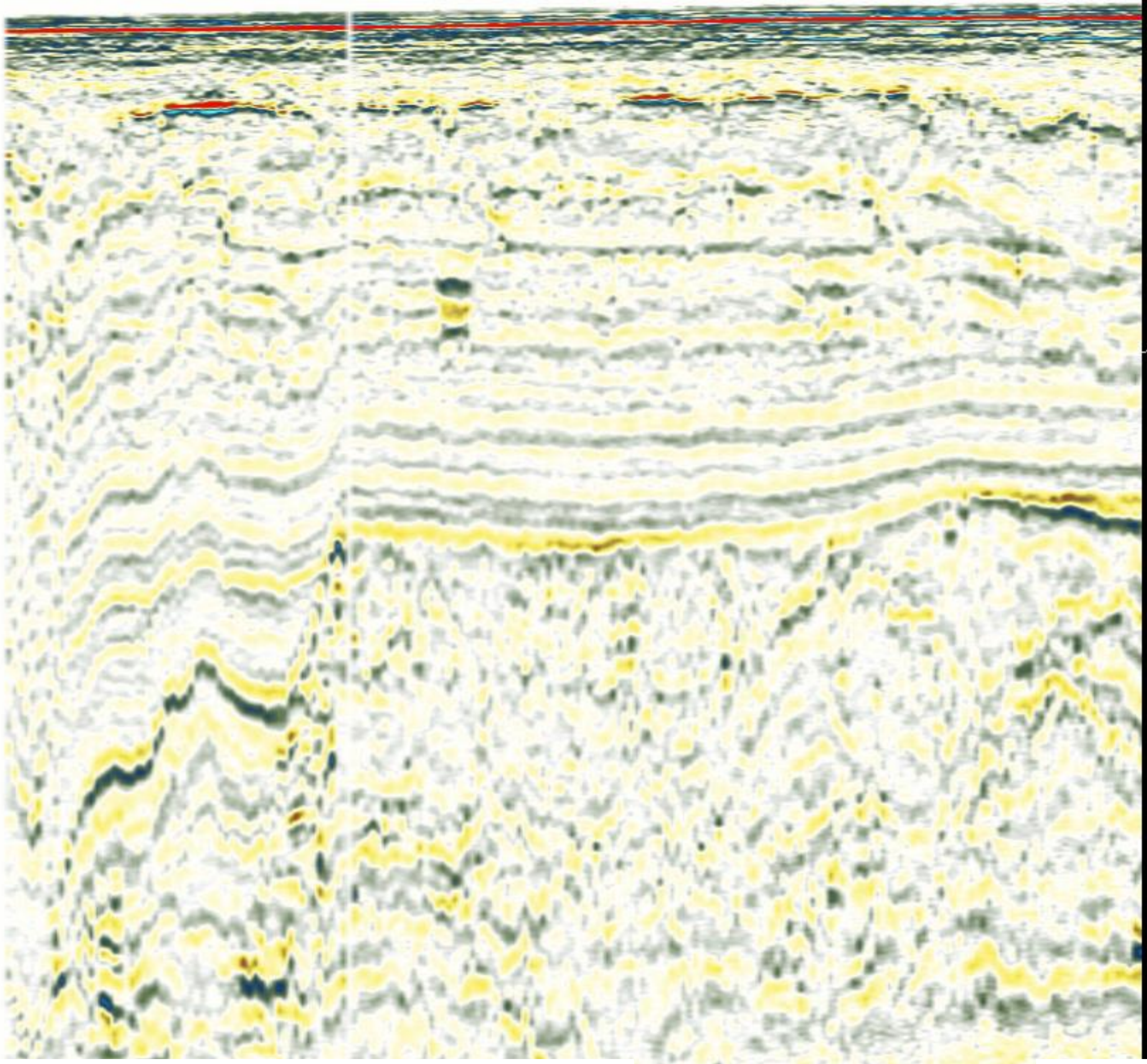
MDA GAN ON TNW 2D-3D DATA: MDA

Very dense data!



Finest grid of seismic, 12.5x12.5 m, 2D to 3D gridding, most sparse, Gridded data, ML interpolated steep, Xline 124, TNW

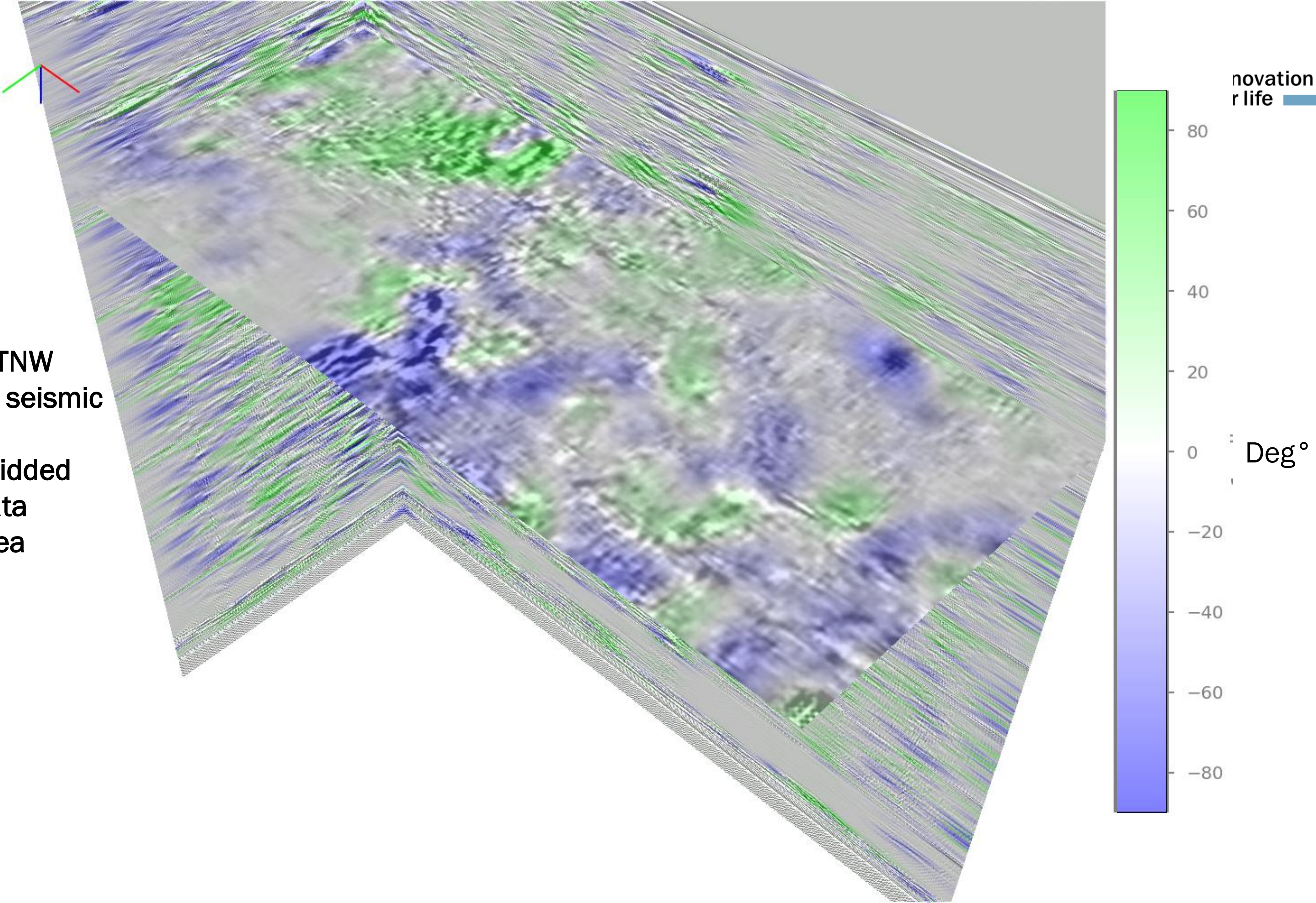
MDA GAN ON TNW 2D-3D DATA: 3D TRUTH



Finest grid of seismic, 12.5x12.5 m, 2D to 3D gridding, most sparse, Gridded data, 3D HRS ground truth, Xline 124, TNW

3D dipscan of TNW
superposed on seismic

2D migrated gridded
interpolated data
TNW 3D subarea



CONCLUSIONS

- The ‘MIMIC’ approach to approximate high-standard but expensive geophysical algorithms by Machine Learning routines appears to be promising
- +-95 % quality reproductions of geophysical algorithms at a speedup factor of +-1000 for diffractions, 720 for denoising and +- 30.000(!) for broadbanding. GAN interpolation on a typical 3D cube costs some minutes on a heavy laptop and GPU
- ML and GAN’s are promising departure point for seismic data conditioning, attributes, prestack processing and quantitative interpretation. The GAN’s already proved themselves in several TNO projects. Future is Diffusion Probabilistic Models
- Improvements: smarter subset training of GAN’s, transfer learning, active learning

THANK YOU FOR YOUR TIME

innovation
for life