Machine Learning in Statoil

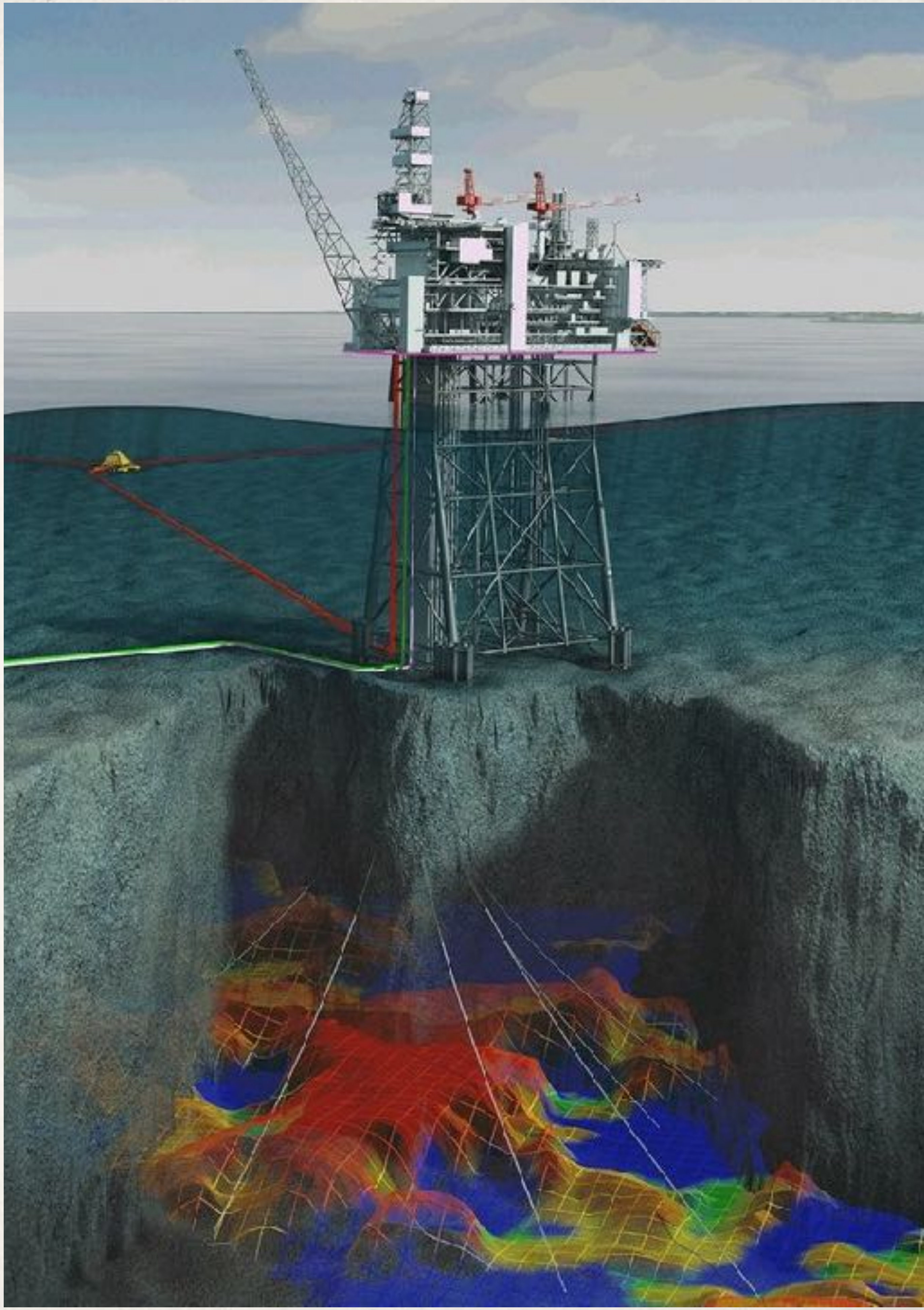
Kristian Flikka

14. May 2018

kflik@statoil.com / kristian@flikka.net

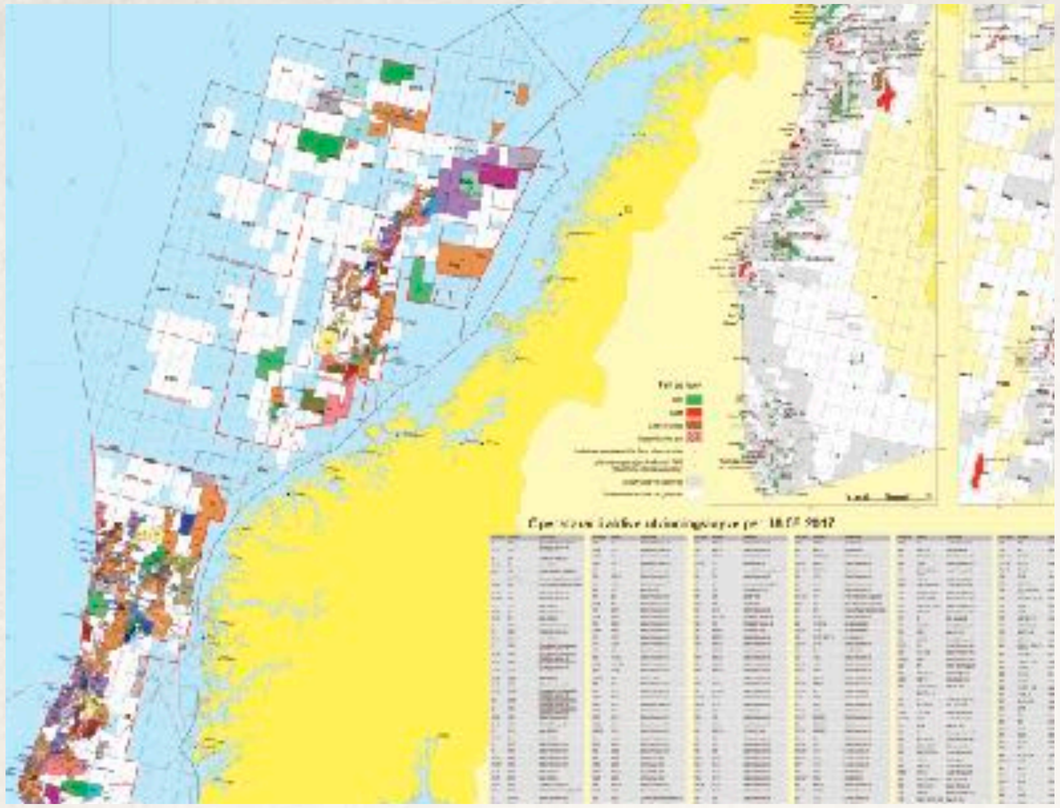
Topics

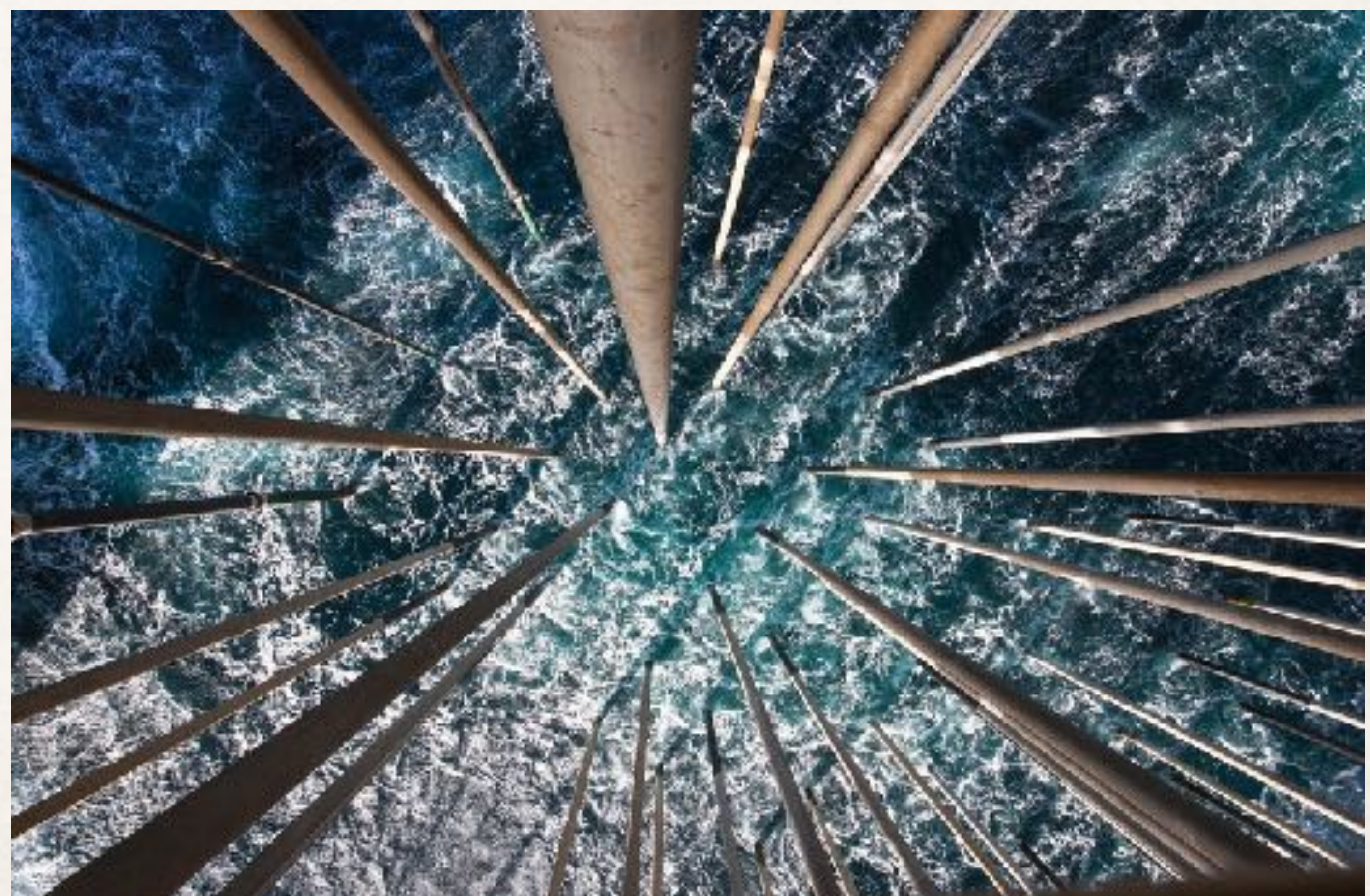
- What does Statoil do?
- Machine learning overall experiences in Statoil?
- Examples



Statoil's business

- Operations in more than 30 countries
- Explore, develop and produce oil and gas
- Deliver wind power to 650,000 British households
- Mounting solar panels in Brazil

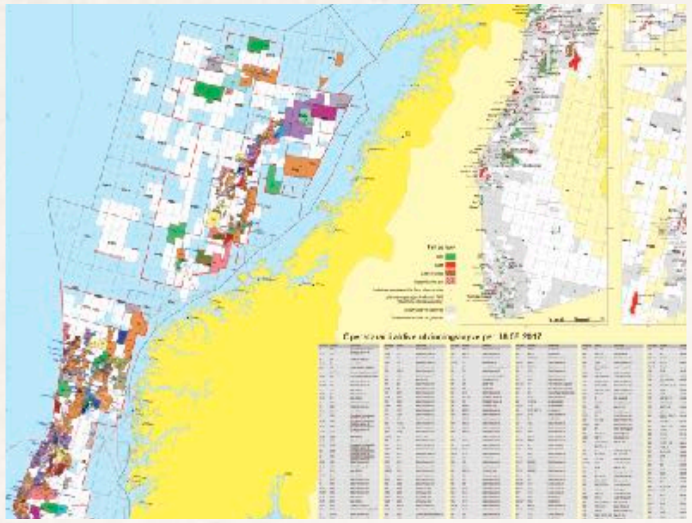


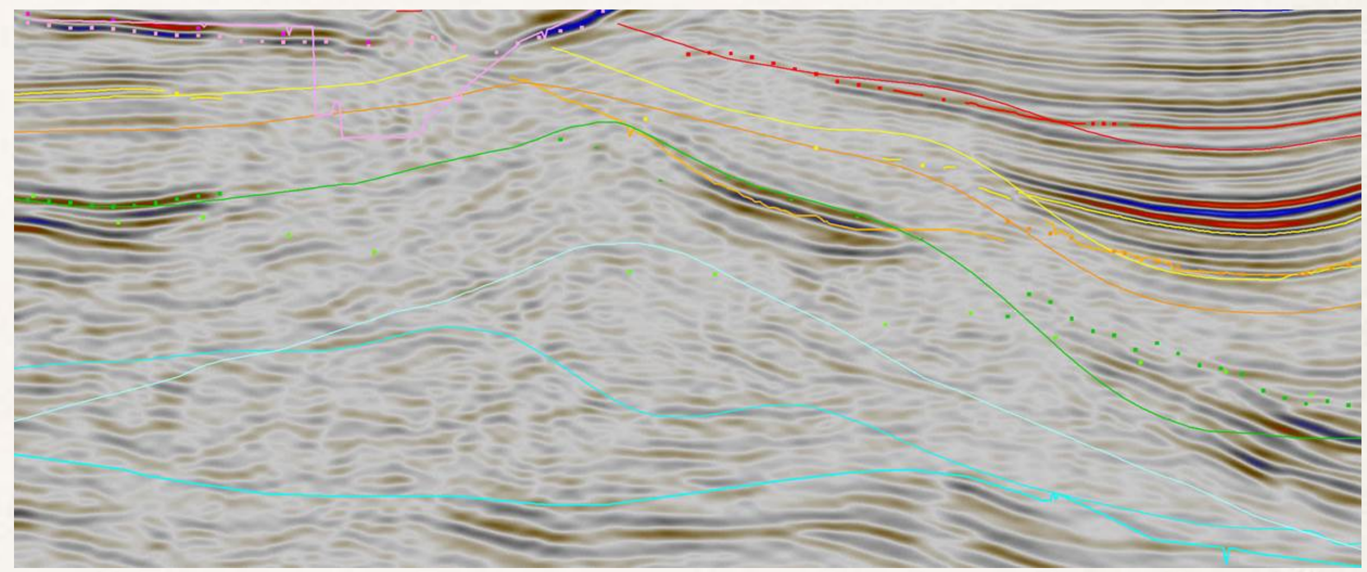



What is machine learning?

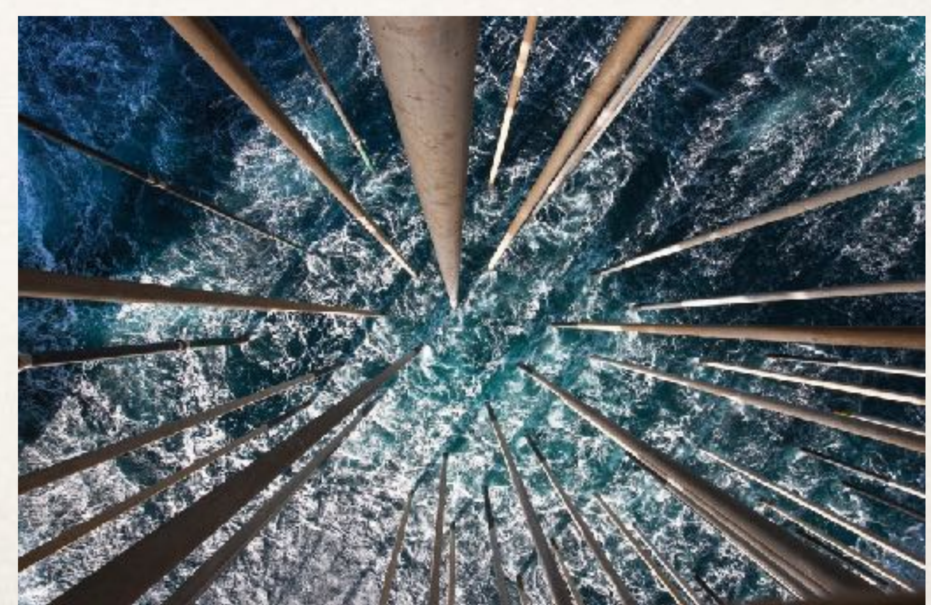
Machine learning is a field of computer science that give computer systems the ability to "learn" with data, without being explicitly programmed

How can machine learning help?





NPD map (freely available)
Photos: Statoil



Making ML work

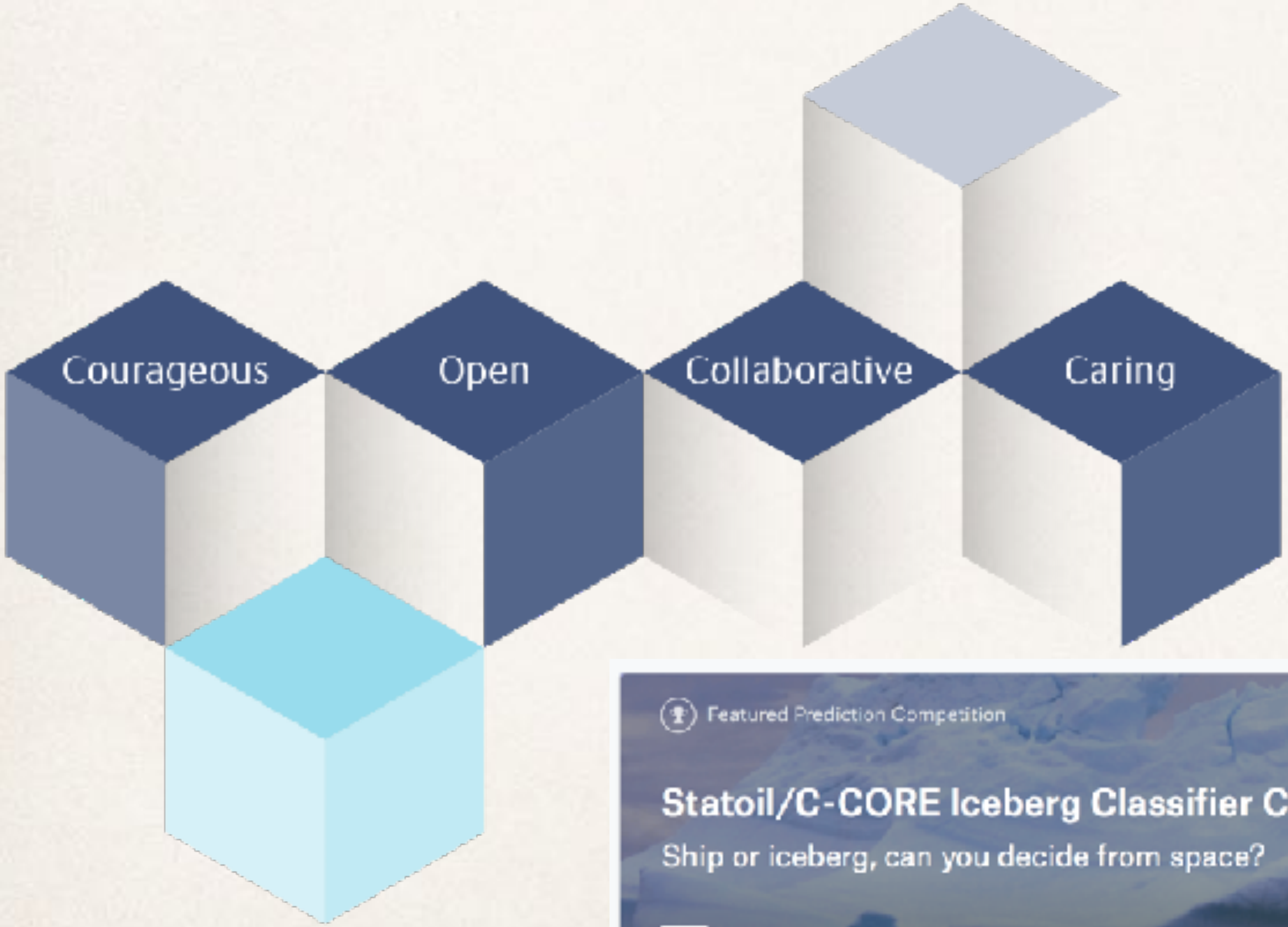
- **Focus on the problem**
- Do we have data?
- Does the features have a predictive potential?
- Can we put it into use? What is the product?

Focus:
The problem, not the solution

- We are solving problems and removing pain.
- Solving wrong problems, or choosing the wrong solution?
- We need to approach our problems and cases through some questions

By the way: Statoil values open

https://github.com/Statoil



Distribute lessons learned?

https://github.com/Statoil/data-science-template

1 Business Understanding and Problem Definition

1.1.1 Who is the sponsor?

1.1.3 What is the business goal/objective?

1.1.6 How is the business efficiency measured?

1.1.8 Who is the SME (Subject Matter Expert)?

1.1.9 Who is the end user?

1.2 Understanding business challenge

1.2.1 What is the business challenge/pain point/bottleneck?

1.2.2 What are the identified use cases?

1.2.4 What is the current solution?

1.2.5 What are the constraints?

1.5 Identifying DS opportunity

1.5.1 What is the problem type from DS perspective, e.g. regression, classification, clustering etc.?

1.5.3 What are the previous relevant experience/components that can be reused?

1.5.5 What is the business value the DS solution can bring?

1.5.6 What is the consequence of a potential DS solution error? How can we control it?

1.5.7 How can the DS solution fit into the business process?

1.5.8 What are the main risks to fail? How can we control them?

1.5.9 What is the feedback from SME on the proposed DS solution?

Making ML work

- Focus on the problem
- **Do we have data?**
- Does the features have a predictive potential?
- Can we put it into use?

Data, data, data

Data scientists spend a lot of time (80%?) on finding and aligning data

5.1	0.222222	3.5	0.625	1.4	0.067797	0.2	0.041667	setosa
4.9	0.166667	3	0.416667	1.4	0.067797	0.2	0.041667	setosa
4.7	0.111111	3.2	0.5	1.3	0.050847	0.2	0.041667	setosa
4.6	0.083333	3.1	0.458333	1.5	0.084746	0.2	0.041667	setosa
5	0.194444	3.6	0.666667	1.4	0.067797	0.2	0.041667	setosa
5.4	0.305556	3.9	0.791667	1.7	0.118644	0.4	0.125	setosa
4.6	0.083333	3.4	0.583333	1.4	0.067797	0.3	0.083333	setosa
5	0.194444	3.4	0.583333	1.5	0.084746	0.2	0.041667	setosa
4.4	0.027778	2.9	0.375	1.4	0.067797	0.2	0.041667	setosa
4.9	0.166667	3.1	0.458333	1.5	0.084746	0.1	0.01	setosa
5.4	0.305556	3.7	0.708333	1.5	0.084746	0.2	0.041667	setosa
4.8	0.138889	3.4	0.583333	1.6	0.101695	0.2	0.041667	setosa
4.8	0.138889	3	0.416667	1.4	0.067797	0.1	0.01	setosa
4.3	0.01	3	0.416667	1.1	0.016949	0.1	0.01	setosa
5.8	0.416667	4	0.833333	1.2	0.033898	0.2	0.041667	setosa
5.7	0.388889	4.4	0.9999	1.5	0.084746	0.4	0.125	setosa
5.4	0.305556	3.9	0.791667	1.3	0.050847	0.4	0.125	setosa
5.1	0.222222	3.5	0.625	1.4	0.067797	0.3	0.083333	setosa
5.7	0.388889	3.8	0.75	1.7	0.118644	0.3	0.083333	setosa
5.1	0.222222	3.8	0.75	1.5	0.084746	0.3	0.083333	setosa
5.4	0.305556	3.4	0.583333	1.7	0.118644	0.2	0.041667	setosa

https://www.nytimes.com/2014/08/18/technology/for-big-data-scientists-hurdle-to-insights-is-janitor-work.html

Gather data, leave the place more tidy

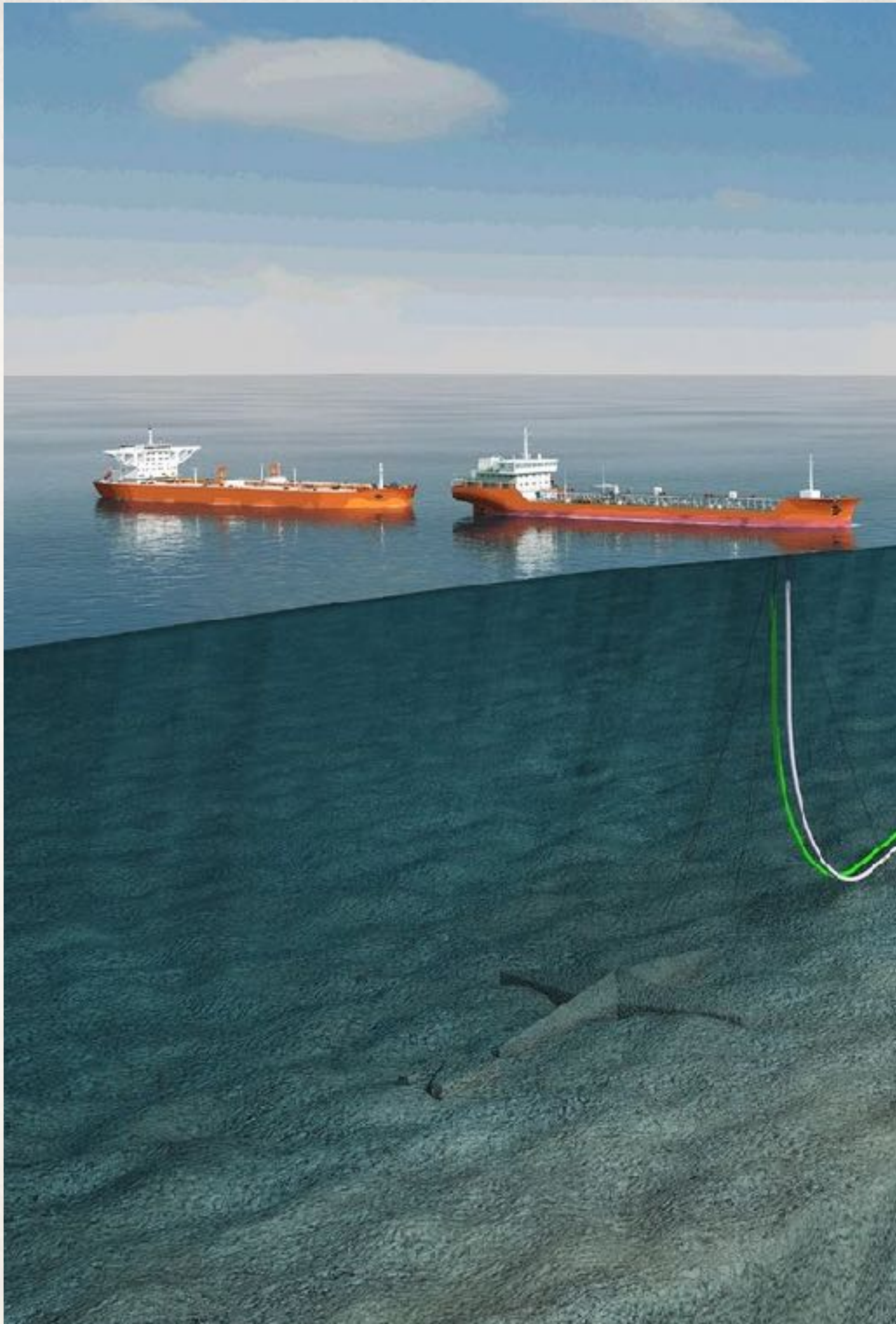
- Data engineers
- Data platform

Making ML work

- Focus on the problem
- Do we have data?
- **Does the features have a predictive potential?**
- Can we put it into use?

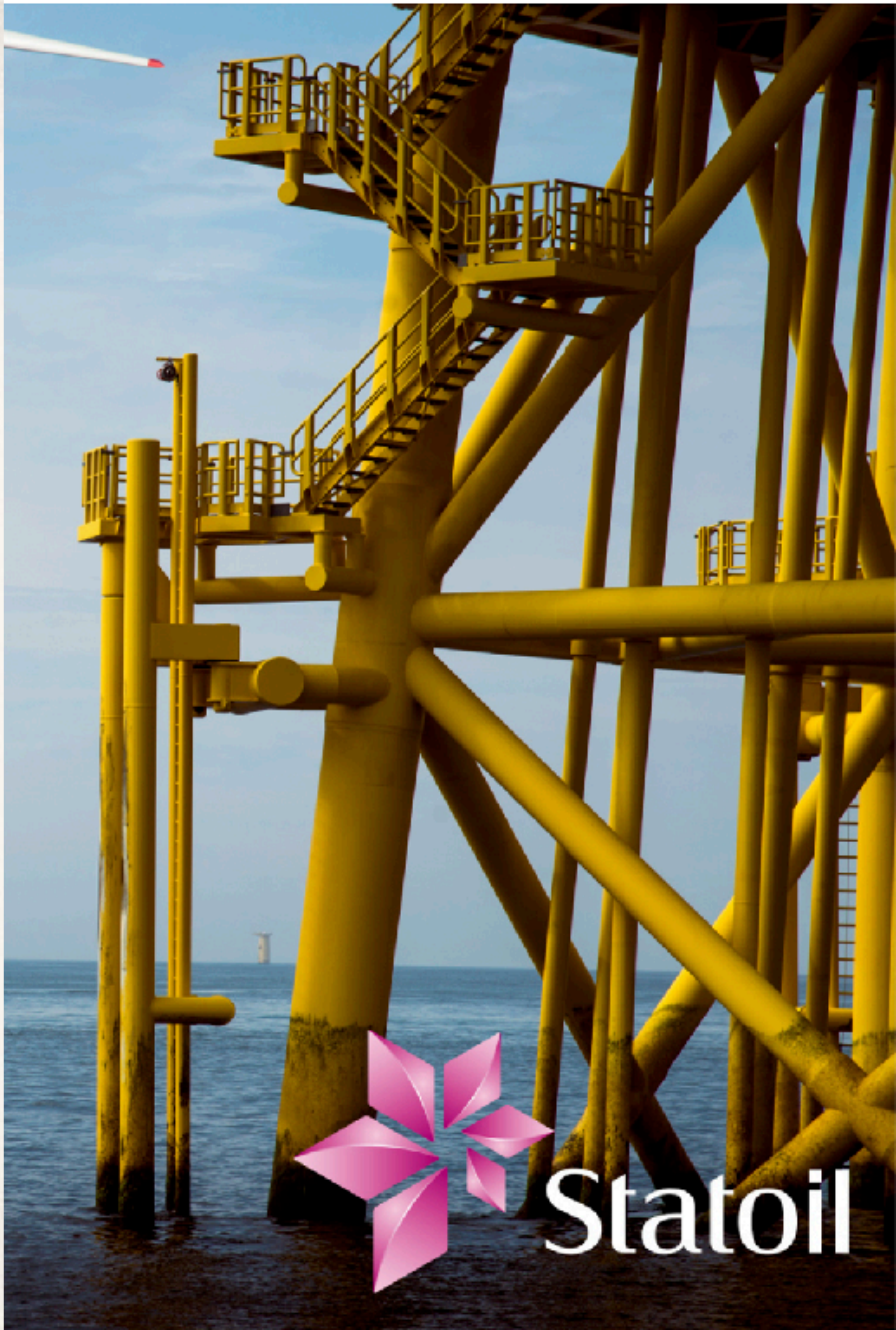
Talk to people

- Given the data, could YOU do it?
- Let experts own features



Making ML work

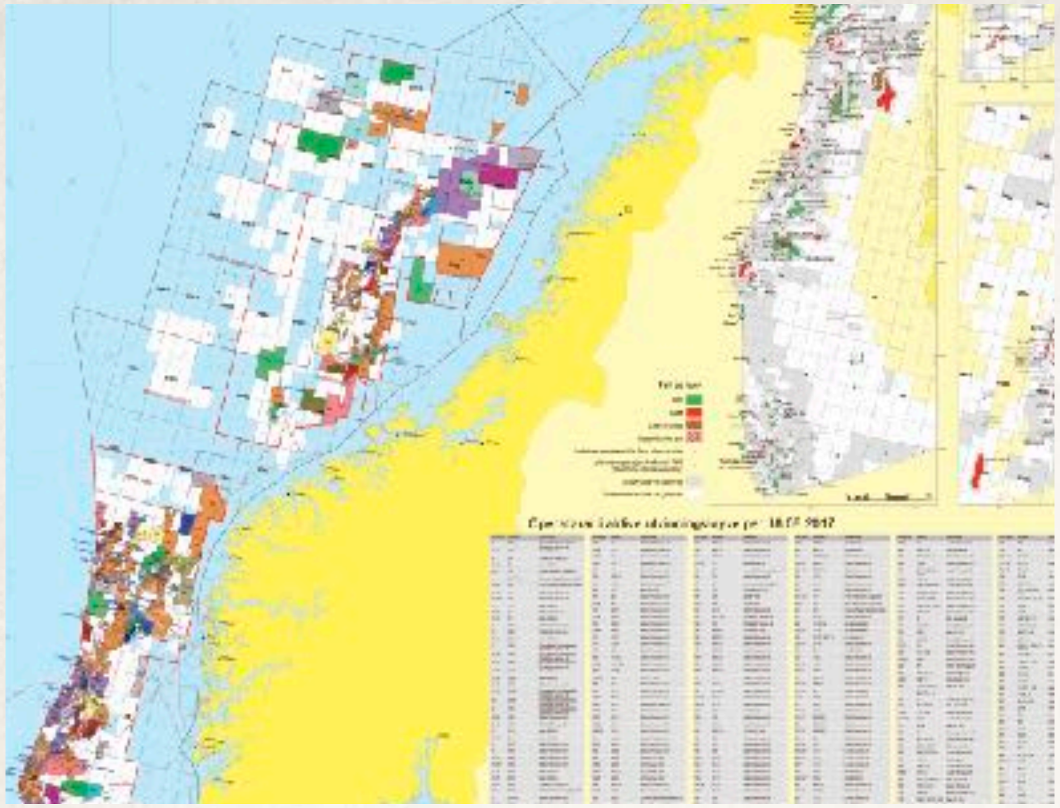
- Focus on the problem
- Do we have data?
- Does the features have a predictive potential?
- **Can we put it into use?**

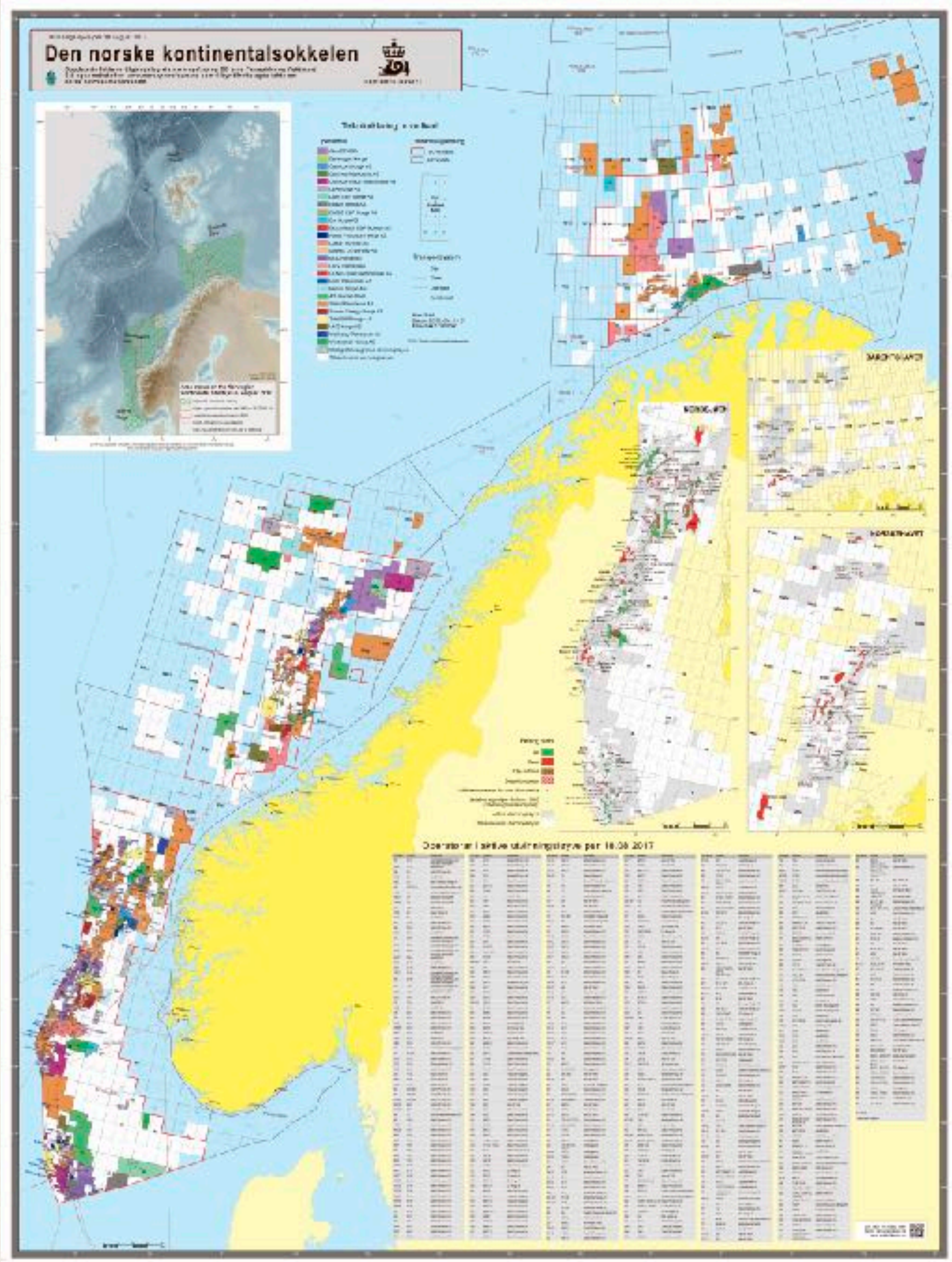


Machine Learning is not a product

- When are you done? When the problem is solved.
    - An ML experiment is probably not enough
    - A report (let's buy this field / company / …)?
    - Most often -> the result is software

Cases





Each prospect is evaluated based on certain criteria.

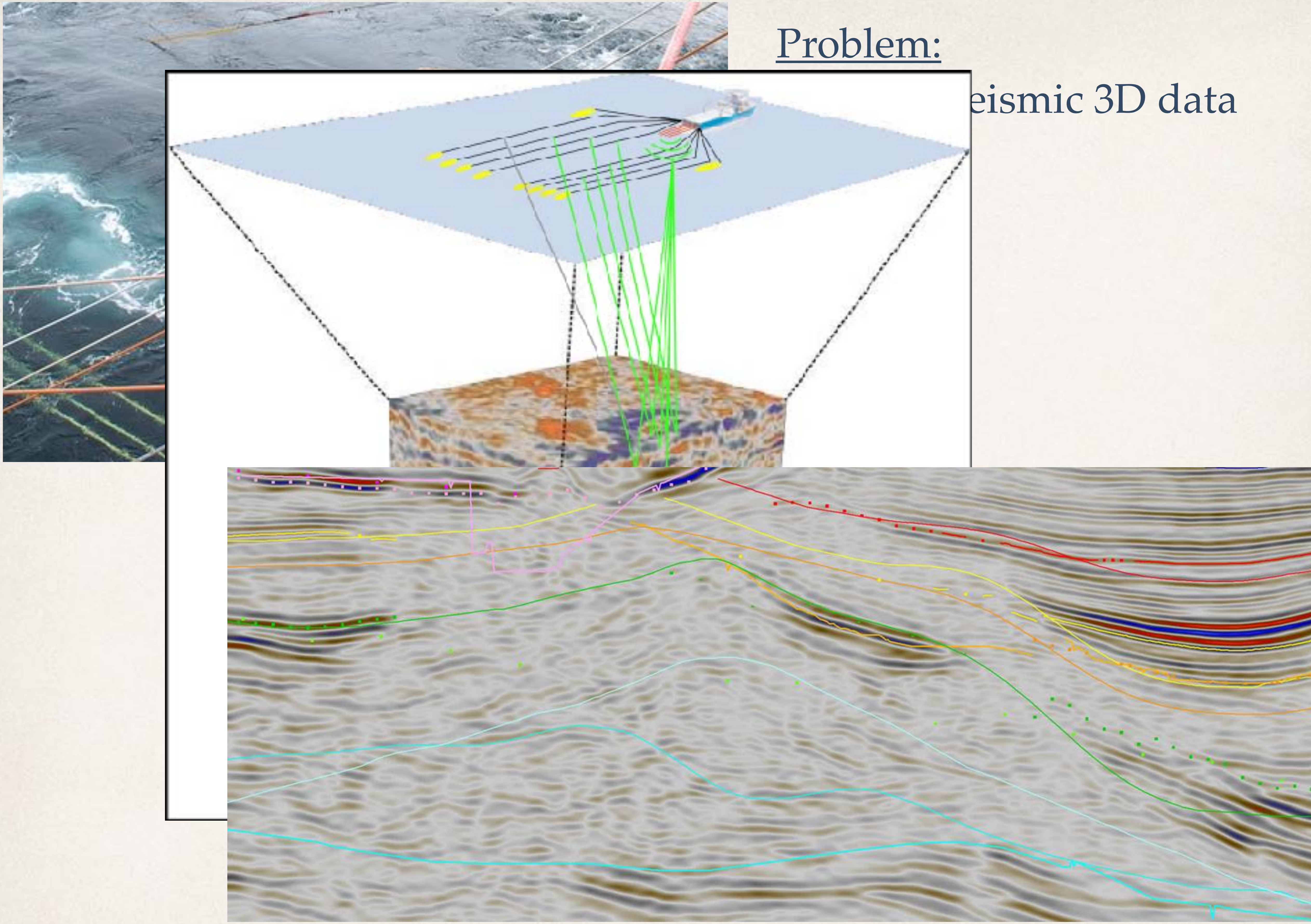
But, there is a lot of uncertainty. As time passes, one learns more.

<u>**Focus**</u>: Collect high quality data - then explore machine learning.

<u>**Solution**</u>**:** Some sort of anomaly detection (unsupervised learning)?

Problem:

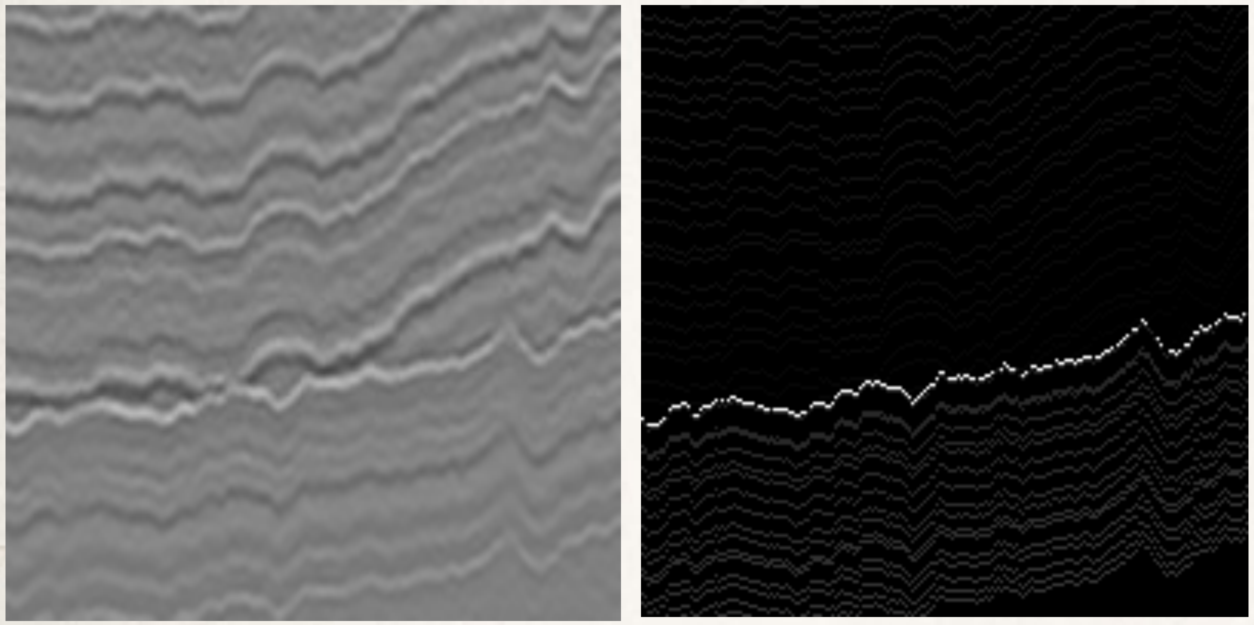
eismic 3D data



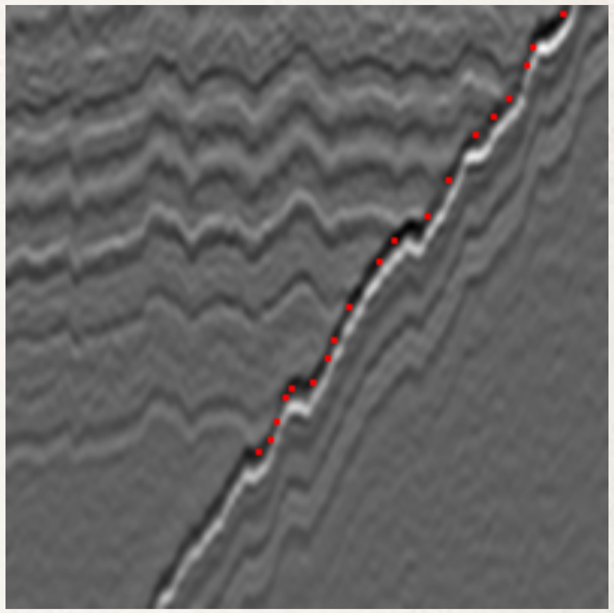
What is the correct interpretation?

(Figures: Statoil DELI Project and Alejandra L. Cameselle)

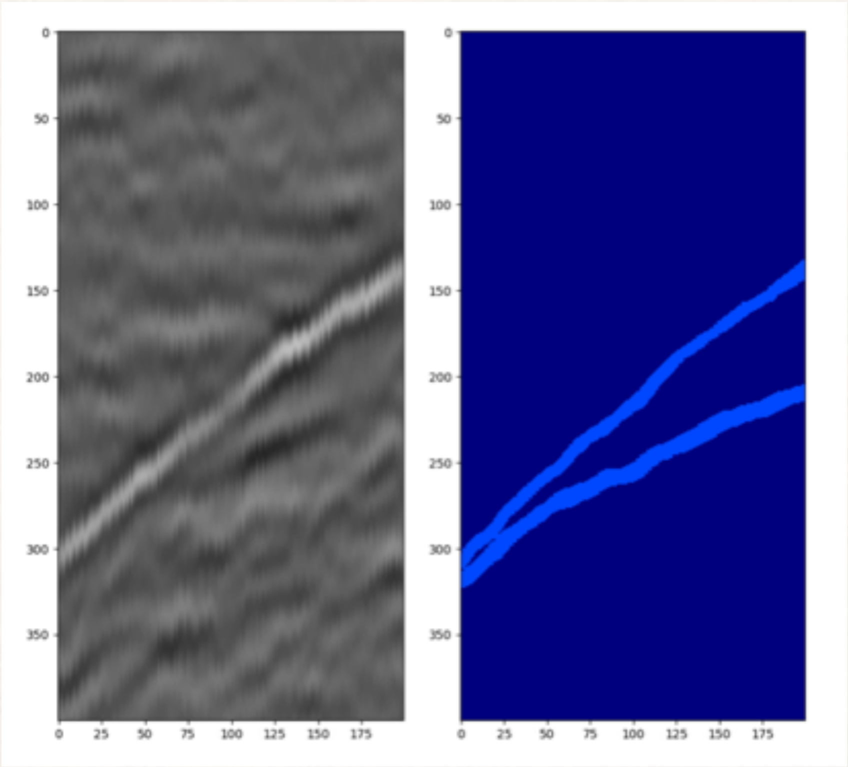
Key Horizon Segmentation



Geometric Detection



Horizon Segmentation 2D/3D



Figures: Statoil DELI Project

Trained on synthetic data, run on real data.
Still research, and not clear if it will be a success

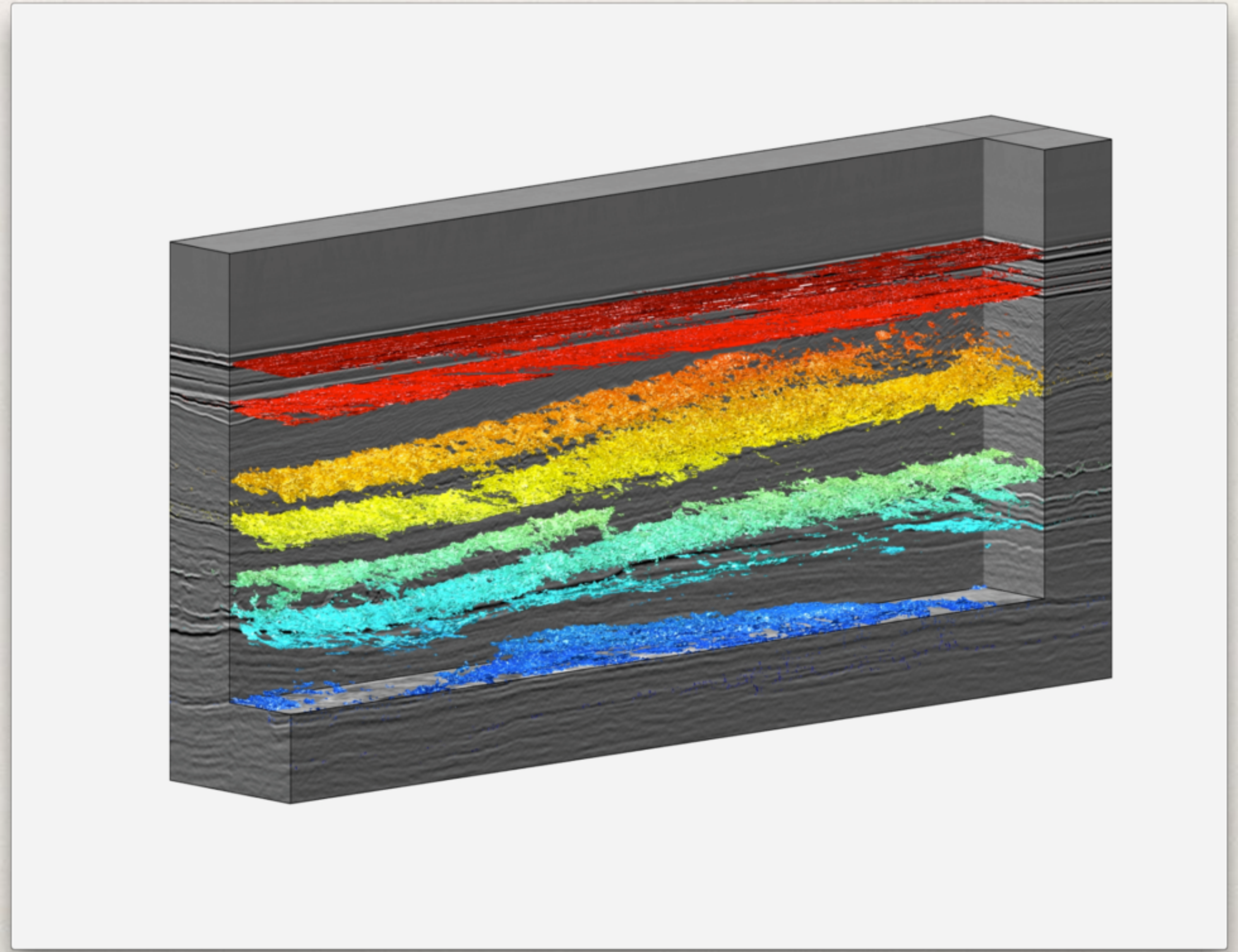
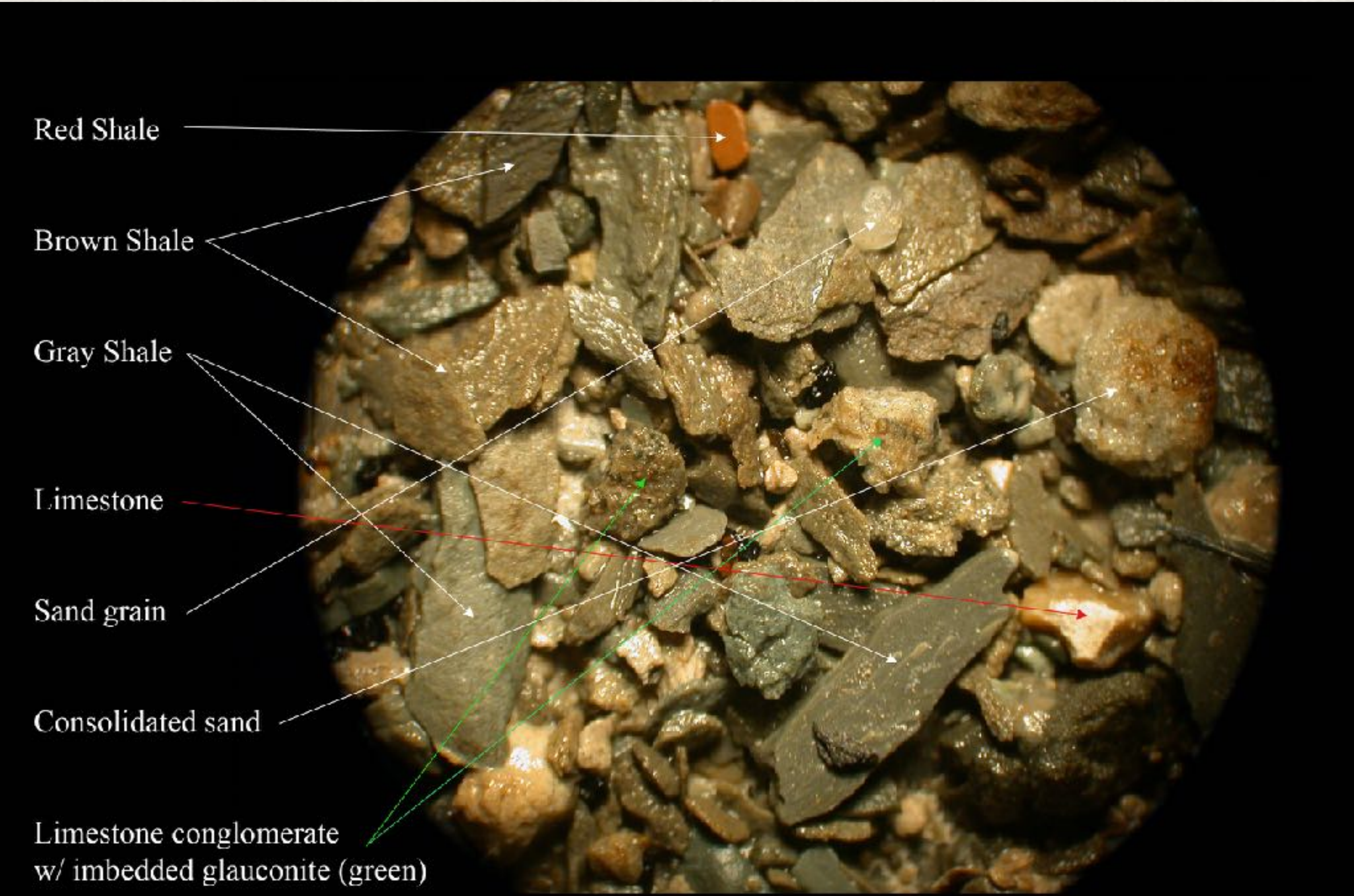


Figure: Statoil DELI Project

More details on the architecture

Task	Architecture	Training data	Train windows	Training Samples
Key horizon segmentation	Deep Unet 2D Segmentation (ResNet+UNet)	Synthetic key horizons Ignore on all other horizons.	200x1x200 or 1x200x200	**14,500** image slices for training and 1450 for validation. Data extracted from 136 different 200x200x200 synthetic cubes
3D horizon segmentation	Volumetric ConvNets with Mixed Residual Connections [Lequan et al, AAAI 2017]	Synthetic horizons with added noise, jitter and quasi-faults	16x128x64	**65,000** sub-cubes of 16x128x64 extracted from approximately 600 different 200x200x400 synthetic seismic cubes
Geometry detection	ResNet18 2D Detection	Synthetic onlaps	32x32	**820,000** sub-images of 32x32 extracted from approximately 136 different 200x200x200 synthetic seismic cubes


OCEANEERING
SW
263 W
NW
Tether Out 75
Deepsea Atlantic
DPT 106.9
ALT 0.0
Statoil
Johan Sverdrup D-11
00:20:21
01.03.16



Sample of drill cuttings under a 10x microscope

Problem:
Manual (or no) classification and logging of cuttings

- Machine learning
  - Automatically classify samples from the well
- Data
  - Take pictures of the rubble (stones)
  - Note down depths
  - Annotate samples
- A supervised machine learning problem

CUILLIN: CUttings Image LithoLogy INterpretation

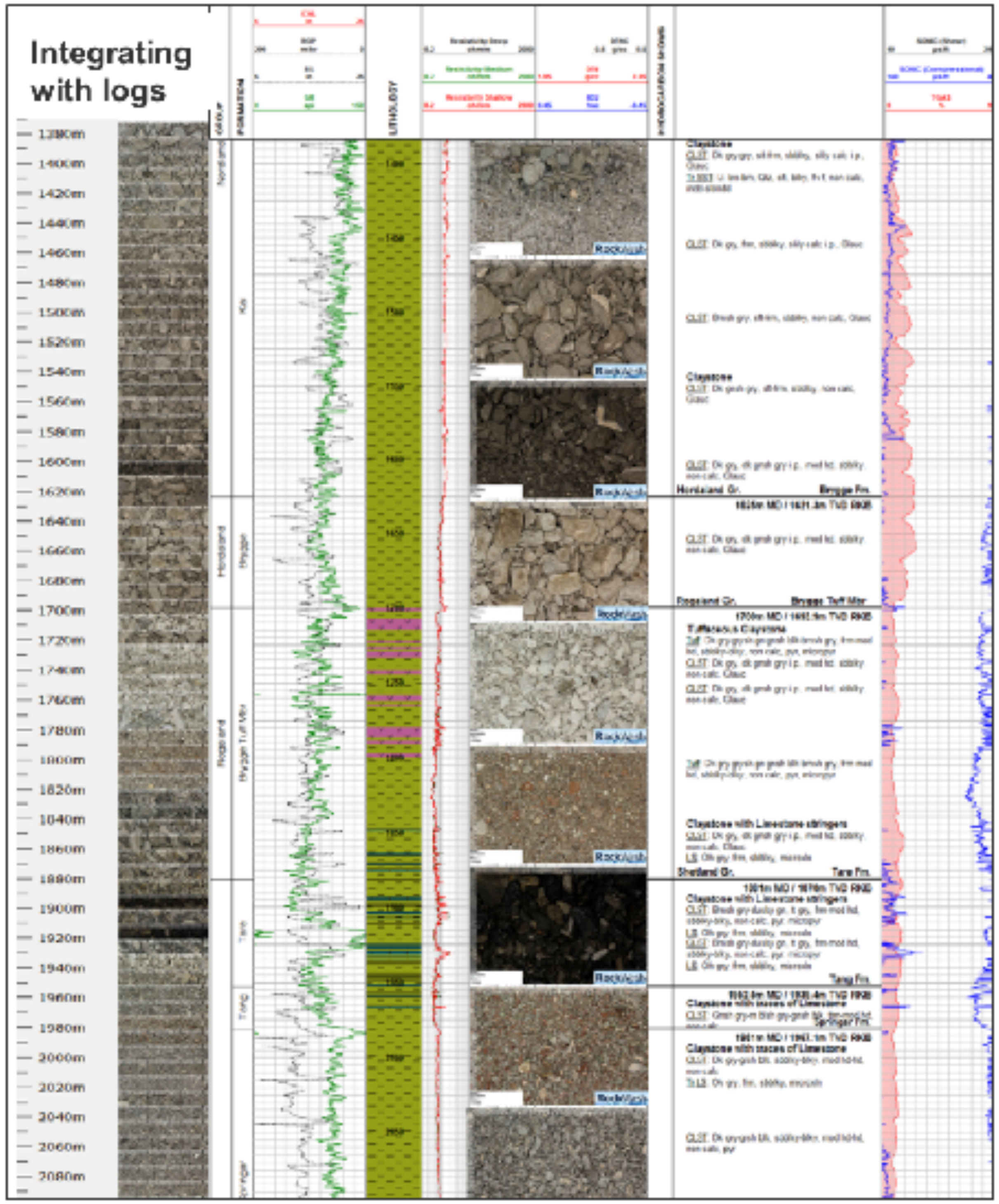
Digitizing cuttings

- Photographing
- Integrating with logs
- Predicting lithology

From storage to photos

30g washed and dried cuttings from Rockwash. Entire well

30g washed and dried cuttings from Rockwash. 1 sample



Predicting lithology

through machine learning

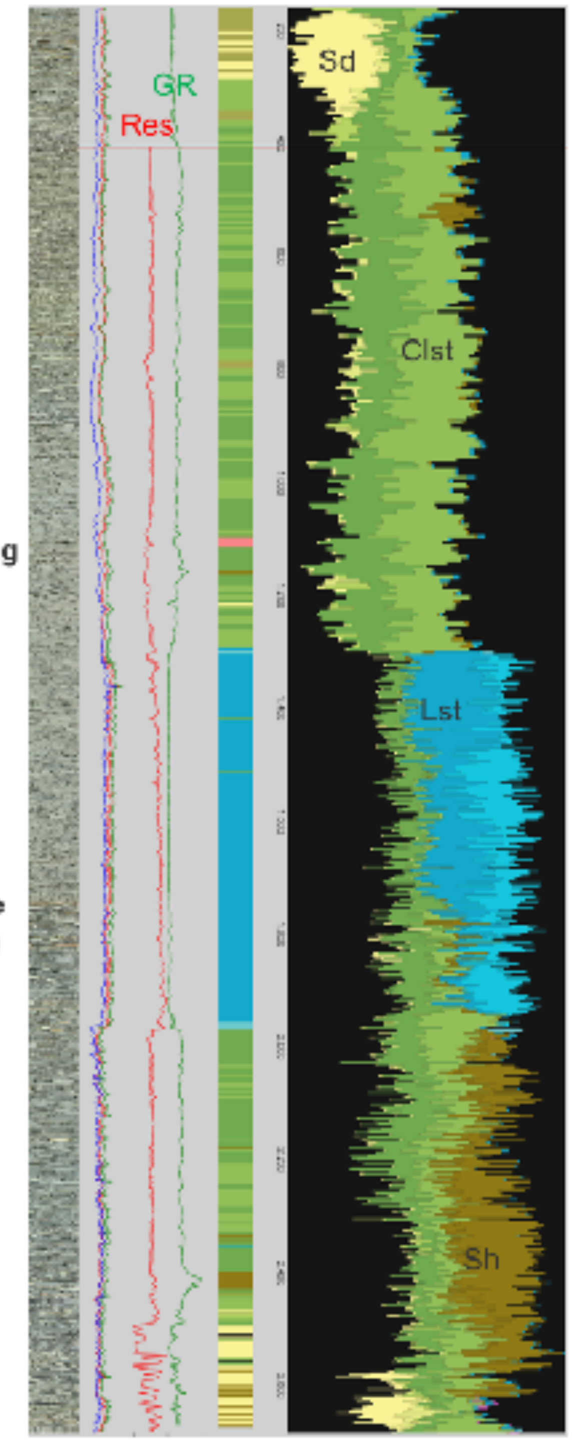


Photo RGB Logs Lith Prediction

Problem: How much cargo to move?

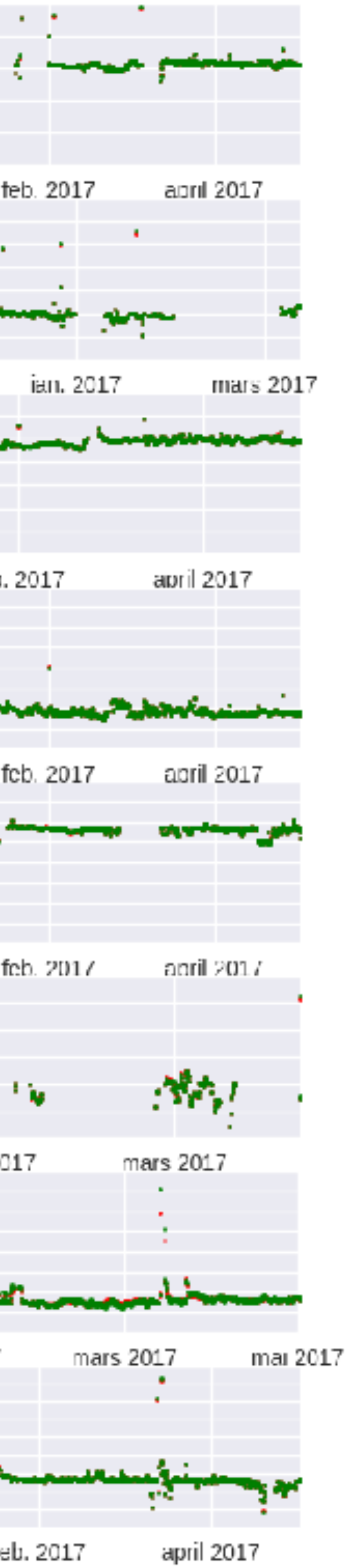
- How much "well building material" will we need to transport the next week, next month?
- Data from many different sources. A lot of pain to integrate.
- Challenge: Stakeholder and commitment

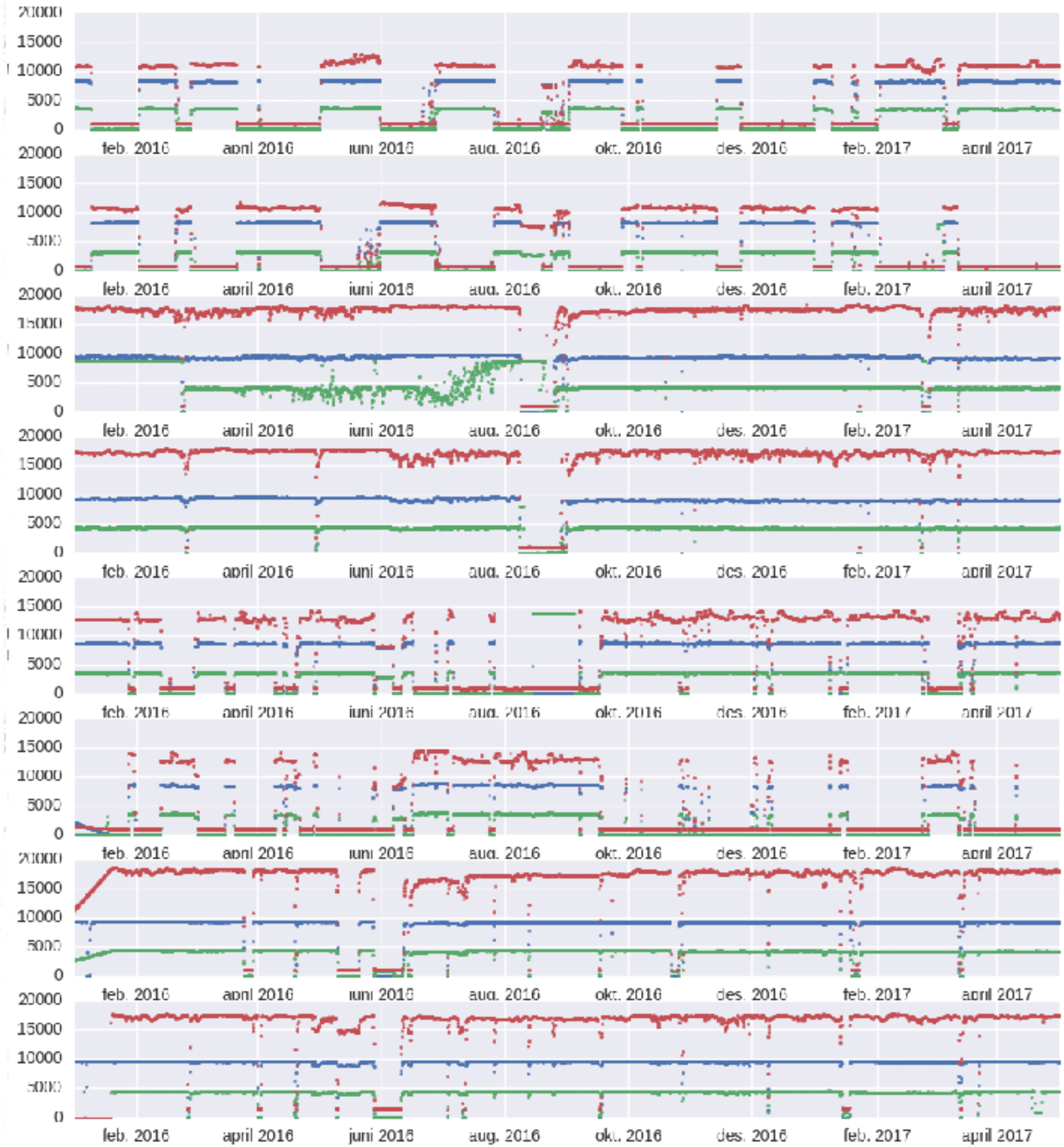
Turbine Example

- A turbine is used to on a plane, but also as the engine to drive a gas compressor (and other things)
    - **Problem**: When the turbine blades get dirty, its efficiency drops. But we can't measure that.
    - **Solution**: Provide efficiency with a simulator (physical model). Use that data to build an ML model.
- https://github.com/Statoil/turbinelearn

Turbine

- The efficiency drops after a period of operation.
- Cleaning helps. But when is the right time?


feb. 2017
april 2017
jan. 2017
mars 2017
april 2017
feb. 2017
april 2017
feb. 2017
april 2017
mars 2017
mars 2017
mai 2017
april 2017


20000
15000
10000
5000
0
feb. 2016
april 2016
juni 2016
aug. 2016
okt. 2016
des. 2016
feb. 2017
april 2017

Turbine, the data

- Based on 4 features, the efficiency can be predicted.

Efficiency (y)	Temperature 1	Temperature 2	Pressure 1	Pressure 2
88.1	22	270	0.6	11.79
80.0	10	268	0.59	10.90
90.2	3	290	0.62	11.22
88.0	5	276	0.63	11.78

Turbine, the solution

```
[pgdr@be-lx885510 ~/tmp/test/turbine_polynomial]$ bin/turbine -m=fcv data/TROLL_*
/prog/sdpsoft/python2.7.13/lib/python2.7/site-packages/IPython/html.py:14: ShimWarn
ing: The `IPython.html` package has been deprecated since IPython 4.0. You should i
mport from `notebook` instead. `IPython.html.widgets` has moved to `ipywidgets`.
  "`IPython.html.widgets` has moved to `ipywidgets`.", ShimWarning)

R^2 test score:
        - Average:      0.882416
        - Median:       0.950945
        - Std dev:      0.129998
        - Worst:        0.525366 (obtained when testing on: data/[redacted].csv, d
ata/[redacted].csv)
R^2 score: 0.992
203.2248 + 0.6255*T1 - 92.8032*P1 - 0.4337*T2 + 6.2251*P2 - 0.0017*T1^2 + 0.0116*T1
*P1 - 0.0005*T1*T2 - 0.0020*T1*P2 + 16.5365*P1^2 + 0.0428*P1*T2 + 0.5677*P1*P2 + 0.
0003*T2^2 + 0.0003*T2*P2 - 0.1515*P2^2
[pgdr@be-lx885510 ~/tmp/test/turbine_polynomial]$
```

Example in Python (if time permits):

Machine data, Python, plotting and some regression

Production Optimisation

- Oil in Water
- Slugging
- Gas Oil Rate

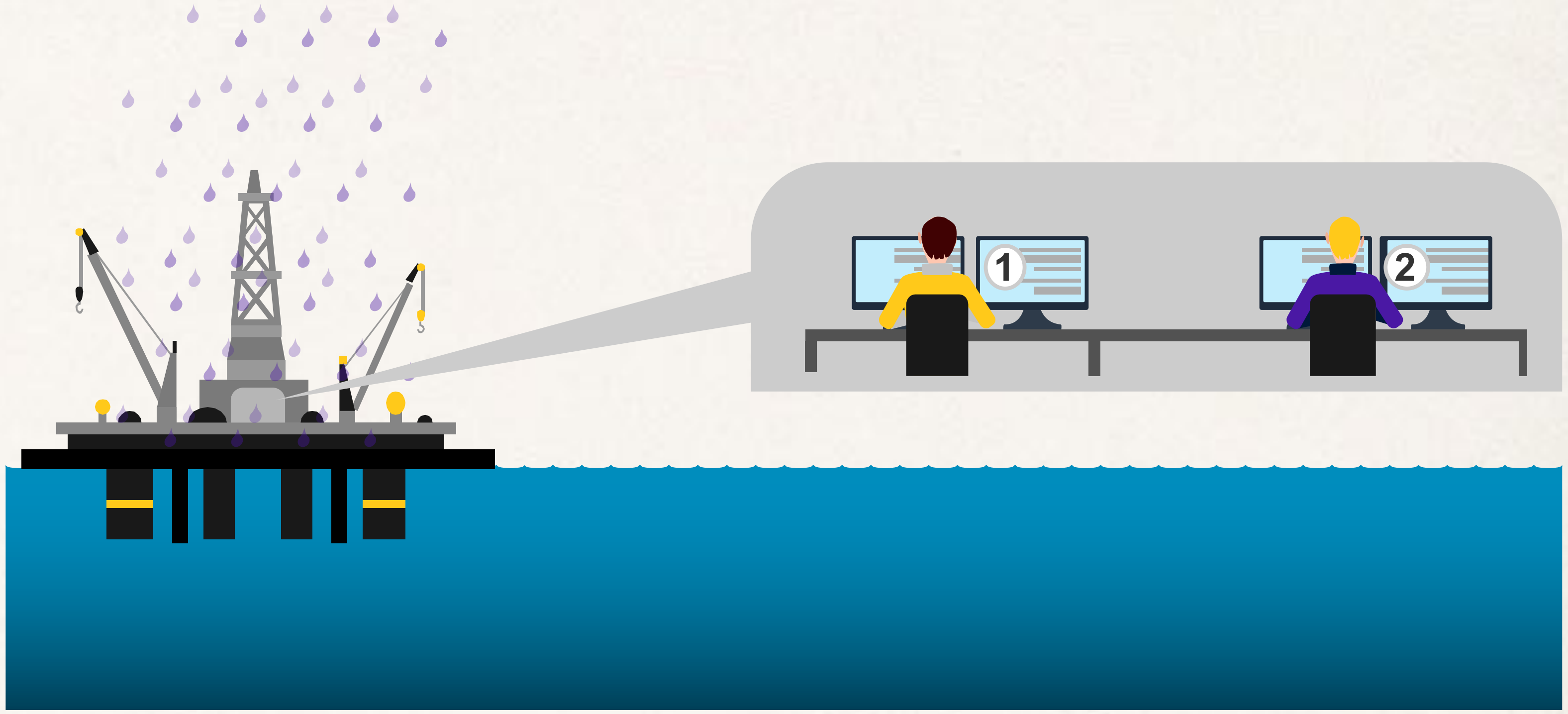
Oil in water

Trigger event

Weather conditions will change within 24 hours

Predicted impact

Risk of oil in water
+40p.p.



1

Mitigating action

Prepare by skimming separators

Impact

Risk of OiW
- 20p.p.

2

Mitigating action

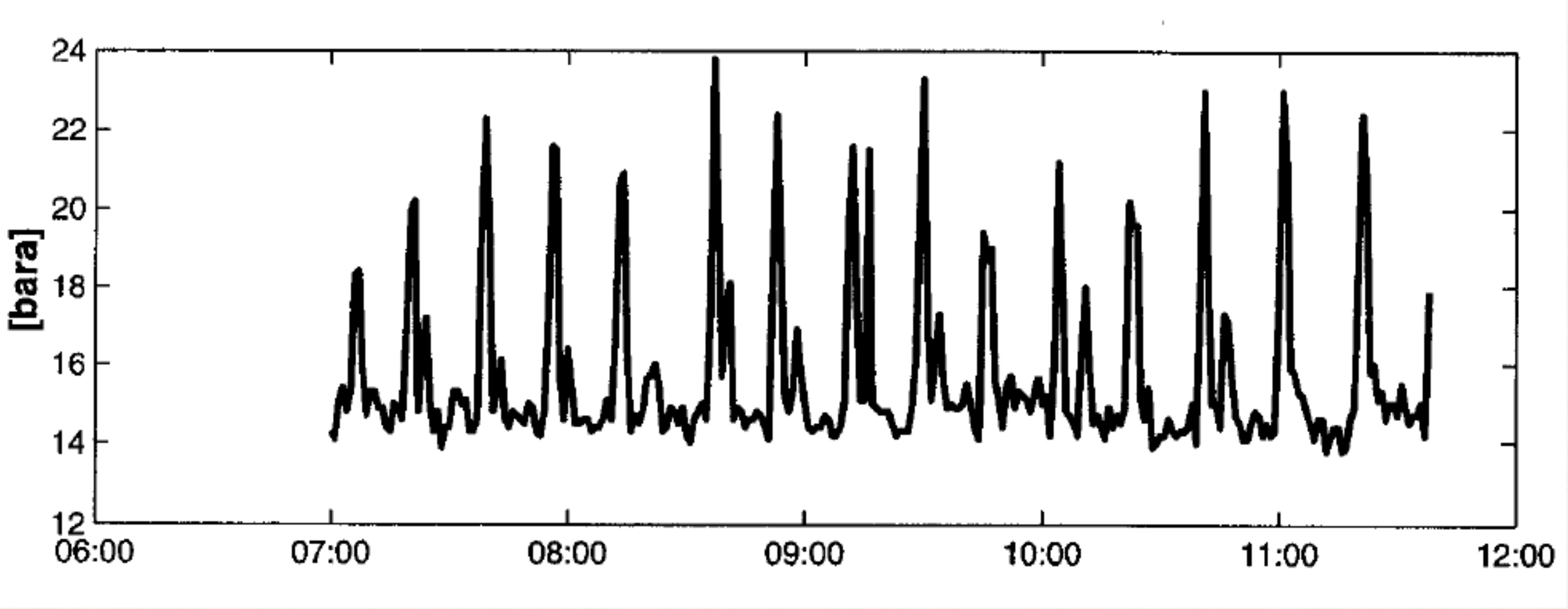
Close well A-11, open C-1 or T-12

Impact

Risk of OiW
- 10p.p.

Slugging

Problem: Inlet separator slugging forces a margin to avoid flaring

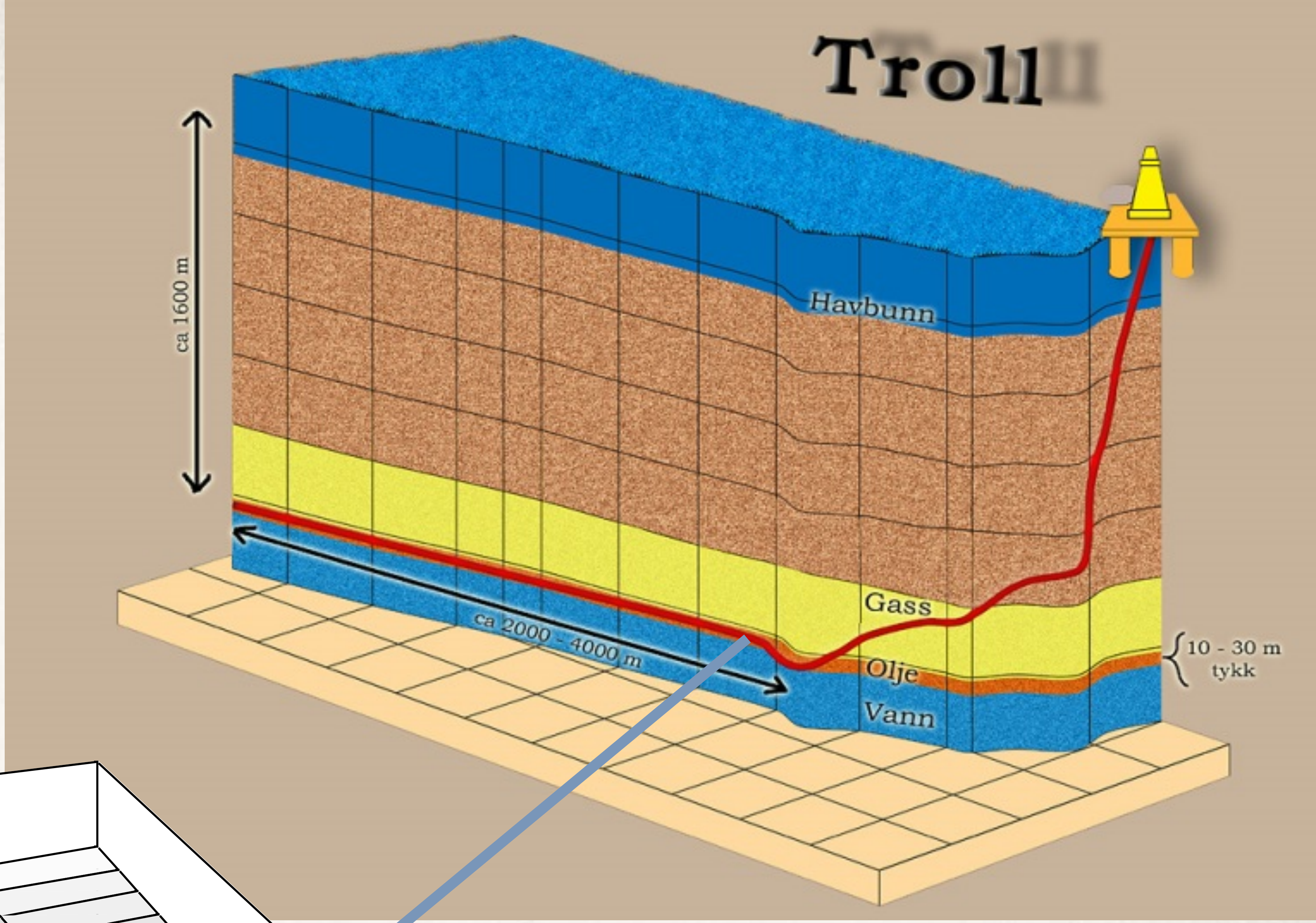


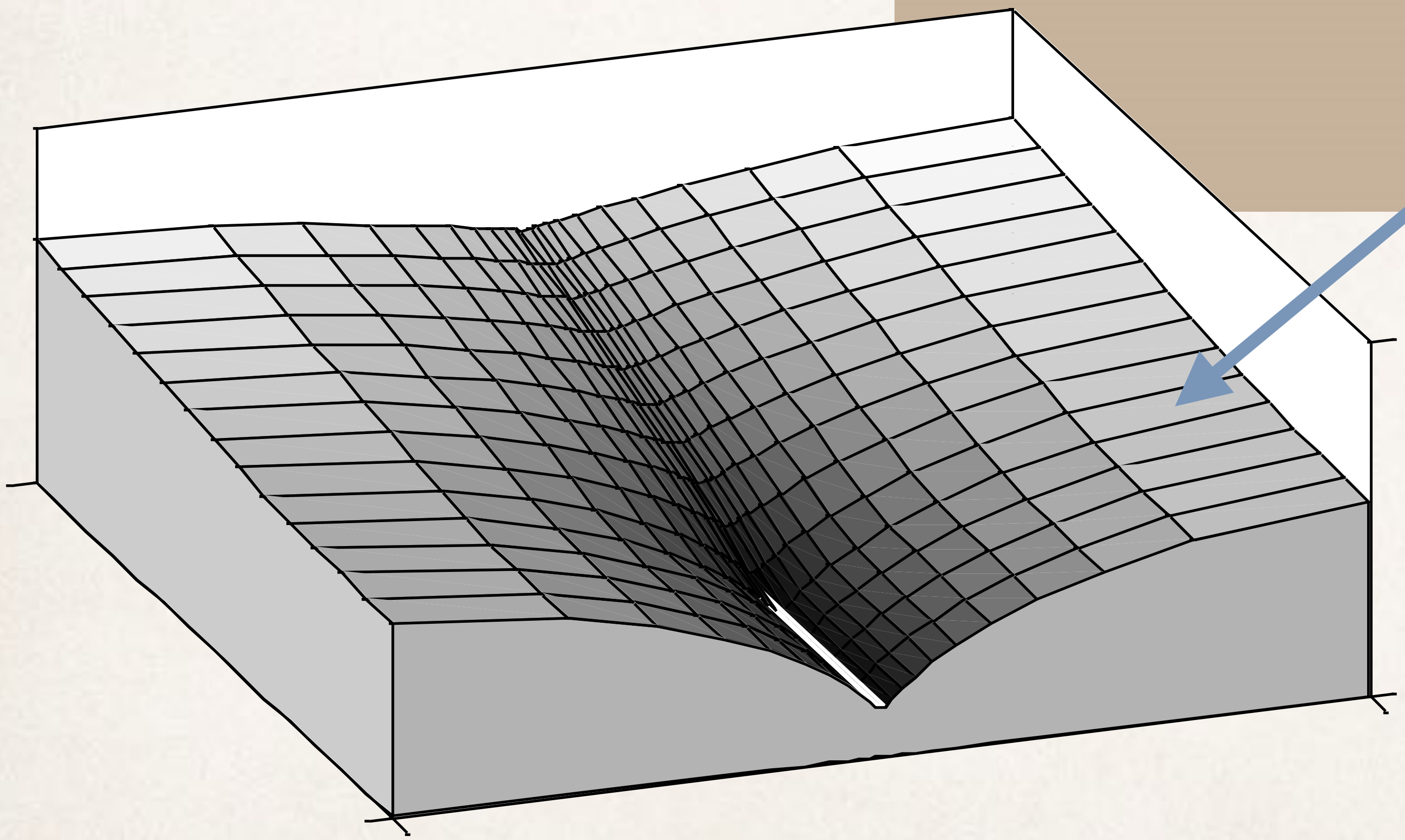
Opportunity: Increase pressure by identifying and/or predicting slugging behaviour (main contributors). Value potential: 100 – 200 Sm^3/d

Data and method: Predict pressure (spikes) based on flow and pressure from the various contributing wells and flow lines.

Gas / Oil ratio

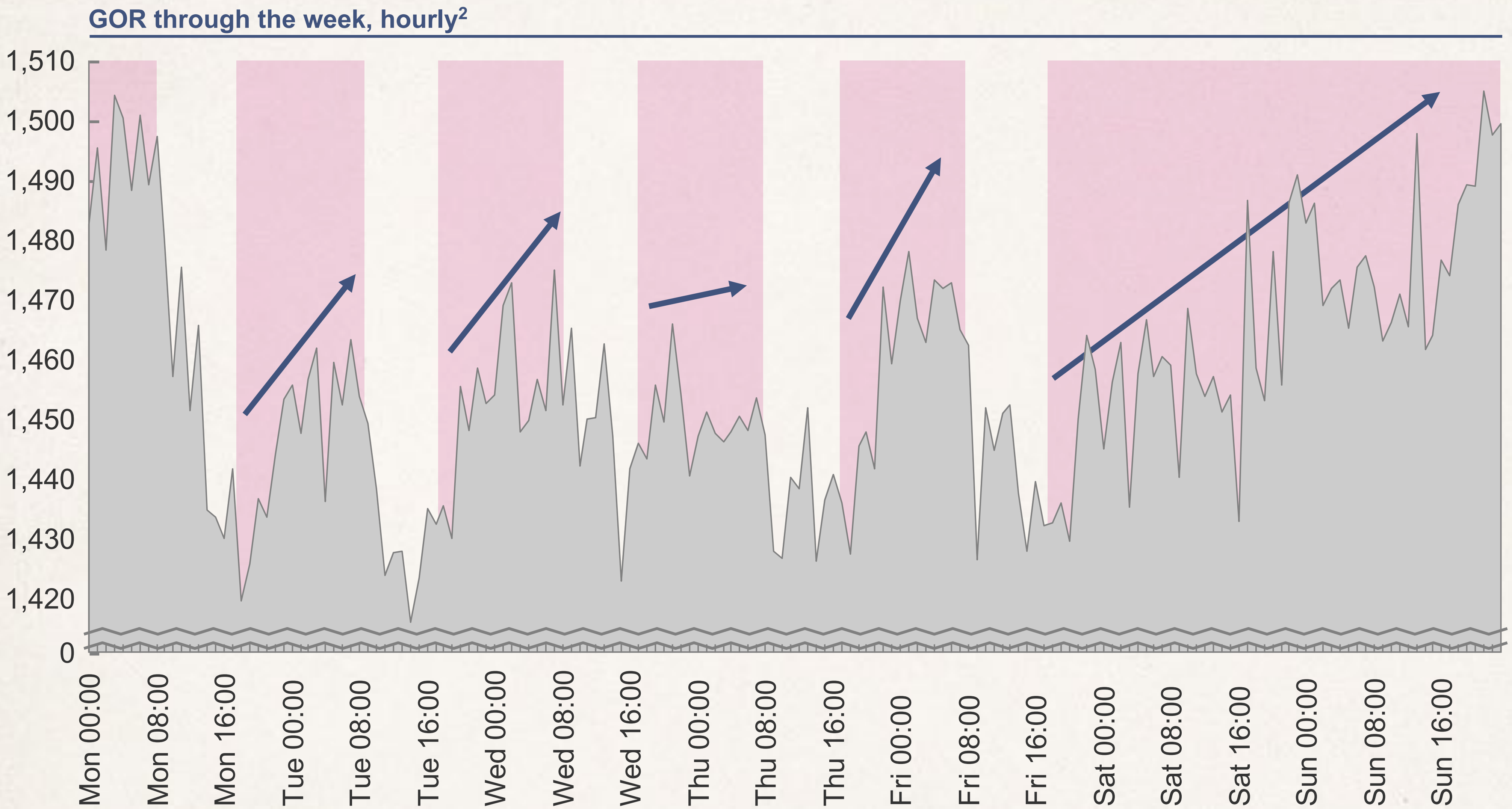
- What makes oil rims so special?
- Infinite" amounts of gas
- Thin oil column 1-5m





There is a coning effect, causing gas to be produced.
The ratio between gas and oil is called "GOR"

Gas / Oil ratio



Problem: We only know the Gas / Oil ratio for the whole field, not the well contributions

Each well contributes to the GOR (previous picture), but human interpretation during office hours is needed to adjust. Wells are turned off, and others are opened.

Solutions: Can we make a system that learns what the humans do? Or even better, can a system suggest better configurations that the human can?

Summary

We have many important problems that might be solved with machine learning
But hard to do it right, not only the ML

What is the product?
What is a sensible metric for your ML performance?
Can we generalise? We have 30 turbines, cannot have a project for each.