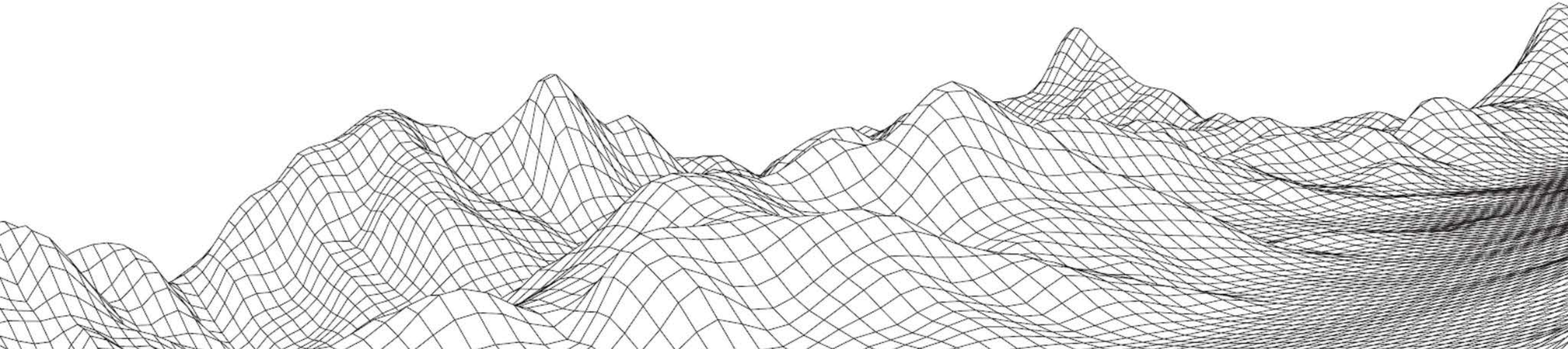
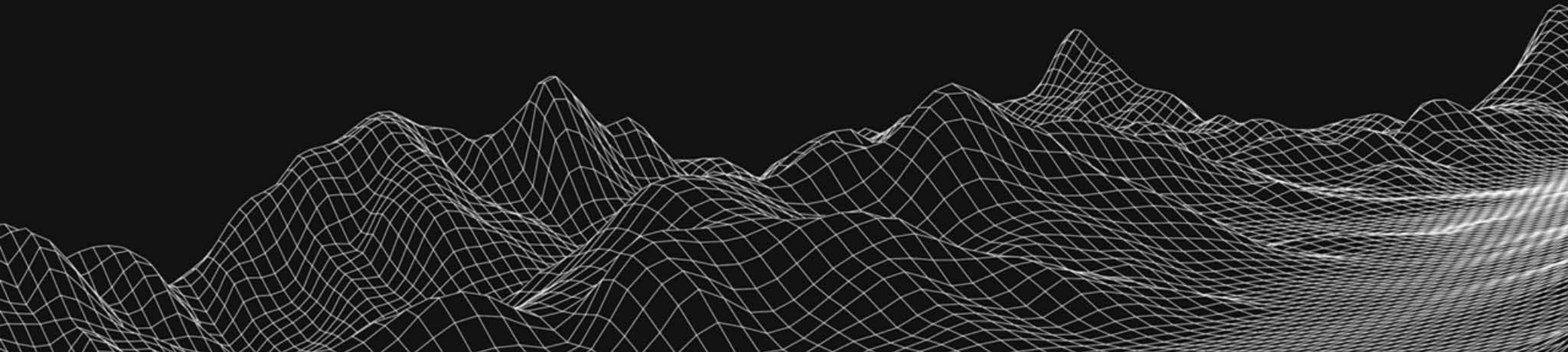
EARTH SCIENCE ANALYTICS



Machine Learning with Well Data

Porosity Shear Sonic Velocity Lithology

Coping with Imperfect Training Data



Can we use wireline logs as proxy for the properties we really care about?

Measured properties	Interesting properties
Neutron	Porosity
Density	Permeability
Sonic	Facies
Gamma Ray	Lithology
Resistivity	Saturation
...	Fluid type
	Hydrogen index
...	Total organic content
...	Vitrinite reflectance
...	Pay interval
...	Reservoir zones
...	Seal quality
...	Pressure

Traditional Methods

Shear sonic prediction

For some mudrocks: $V_p = 1.16V_s + 1.36,$

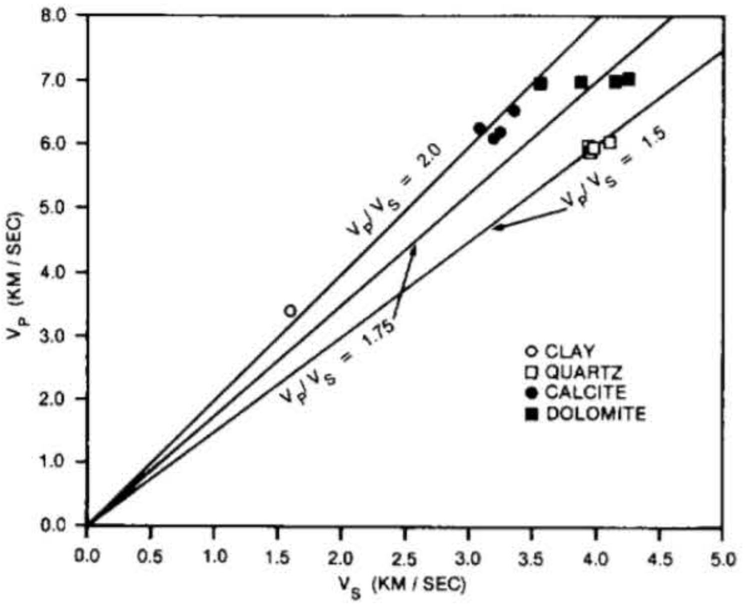


FIG. 2. Compressional and shear velocities for some minerals.

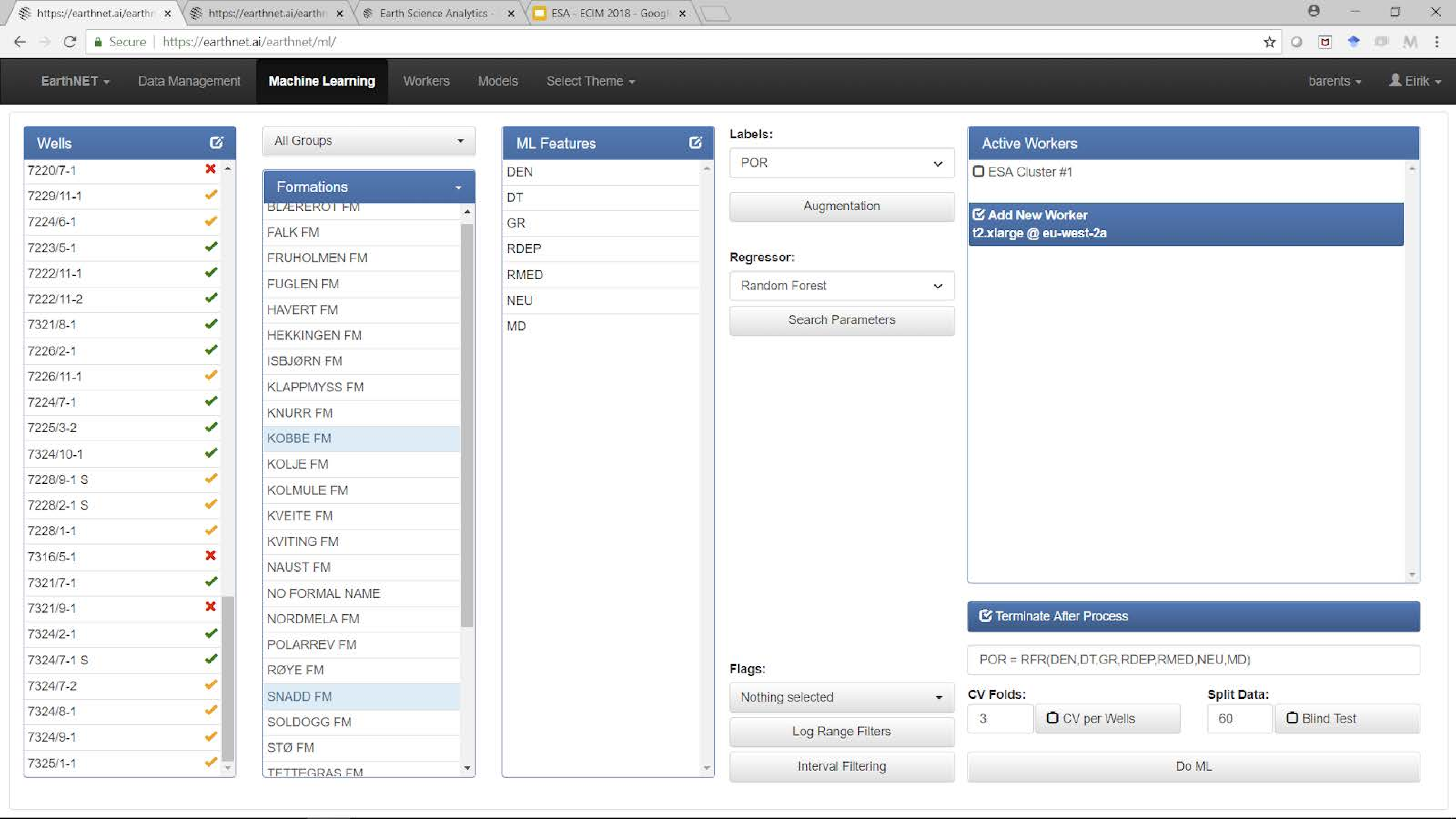
Plot from Castagna et al. 1985

$$\phi_D = \frac{\rho_{ma} - \rho_b}{\rho_{ma} - \rho_f}$$

$$\phi_S = \frac{\Delta t - \Delta t_{ma}}{\Delta t_f - \Delta t_{ma}}$$

Variables in equations needs tuning to:

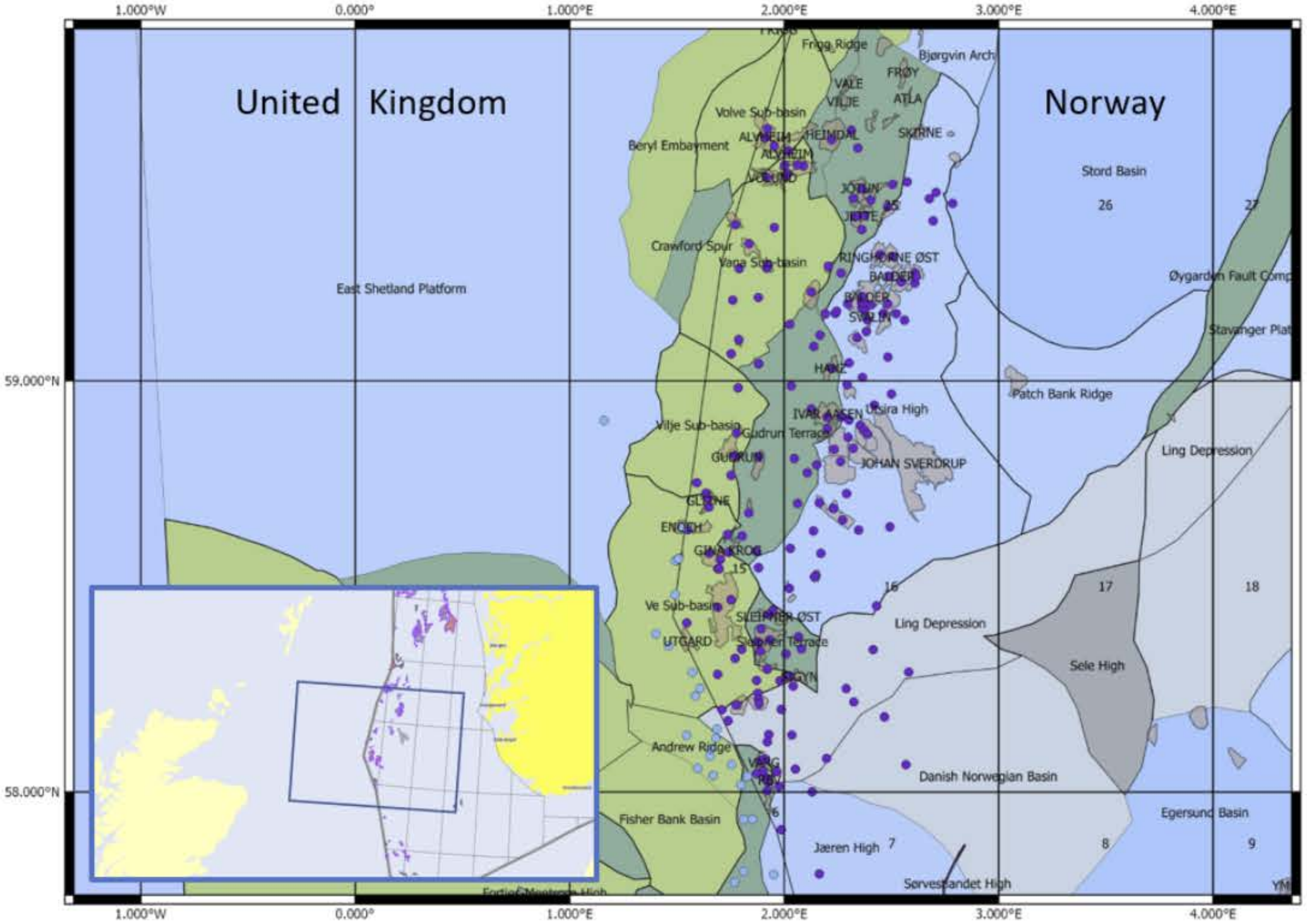
- lithology
- stratigraphic formations
- fluid fill


https://earthnet.ai/earthn
https://earthnet.ai/earthn
Earth Science Analytics -
ESA - ECIM 2018 - Googl
Secure https://earthnet.ai/earthnet/ml/
EarthNET
Data Management
Machine Learning
Workers
Models
Select Theme
barents
Eirik
Wells
7220/7-1
7229/11-1
7224/6-1
7223/5-1
7222/11-1
7222/11-2
7321/8-1
7226/2-1
7226/11-1
7224/7-1
7225/3-2
7324/10-1
7228/9-1 S
7228/2-1 S
7228/1-1
7316/5-1
7321/7-1
7321/9-1
7324/2-1
7324/7-1 S
7324/7-2
7324/8-1
7324/9-1
7325/1-1
All Groups
Formations
FALK FM
FRUHOLMEN FM
FUGLEN FM
HAVERT FM
HEKKINGEN FM
ISBJØRN FM
KLAPPMYSS FM
KNURR FM
KOBBE FM
KOLJE FM
KOLMULE FM
KVEITE FM
KVITING FM
NAUST FM
NO FORMAL NAME
NORDMELA FM
POLARREV FM
RØYE FM
SNADD FM
SOLDOGG FM
STØ FM
ML Features
DEN
DT
GR
RDEP
RMED
NEU
MD
Labels:
POR
Augmentation
Regressor:
Random Forest
Search Parameters
Flags:
Nothing selected
Log Range Filters
Interval Filtering
Active Workers
ESA Cluster #1
Add New Worker
t2.xlarge @ eu-west-2a
Terminate After Process
POR = RFR(DEN,DT,GR,RDEP,RMED,NEU,MD)
CV Folds:
3
CV per Wells
Split Data:
60
Blind Test
Do ML

A Case Study for ConocoPhillips

271 North Sea Wells

A Viking Graben Case Study - for ConocoPhillips



Goal

- Predict
  - porosity,
  - shear wave velocity
  - lithology

  in +250 wells

  in the entire stratigraphy

Feature set

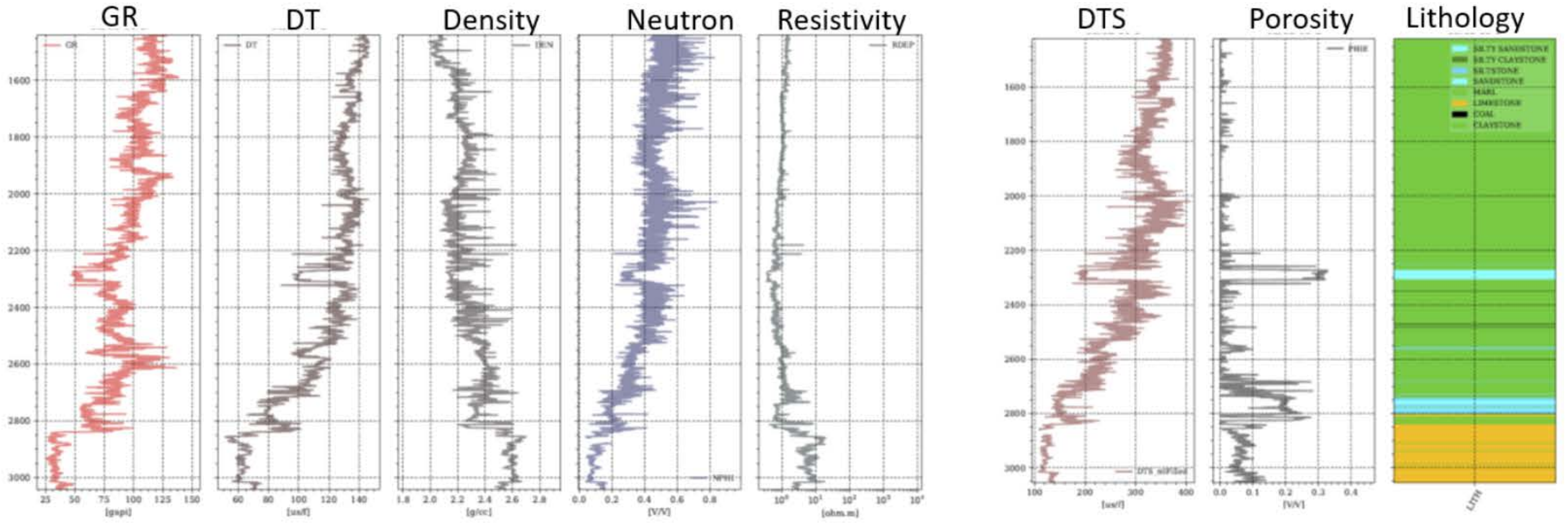
- Wireline logs

Label sets

- Human labelled lithology logs
- Porosity from CPI's
- Measured Vs logs

Project Database

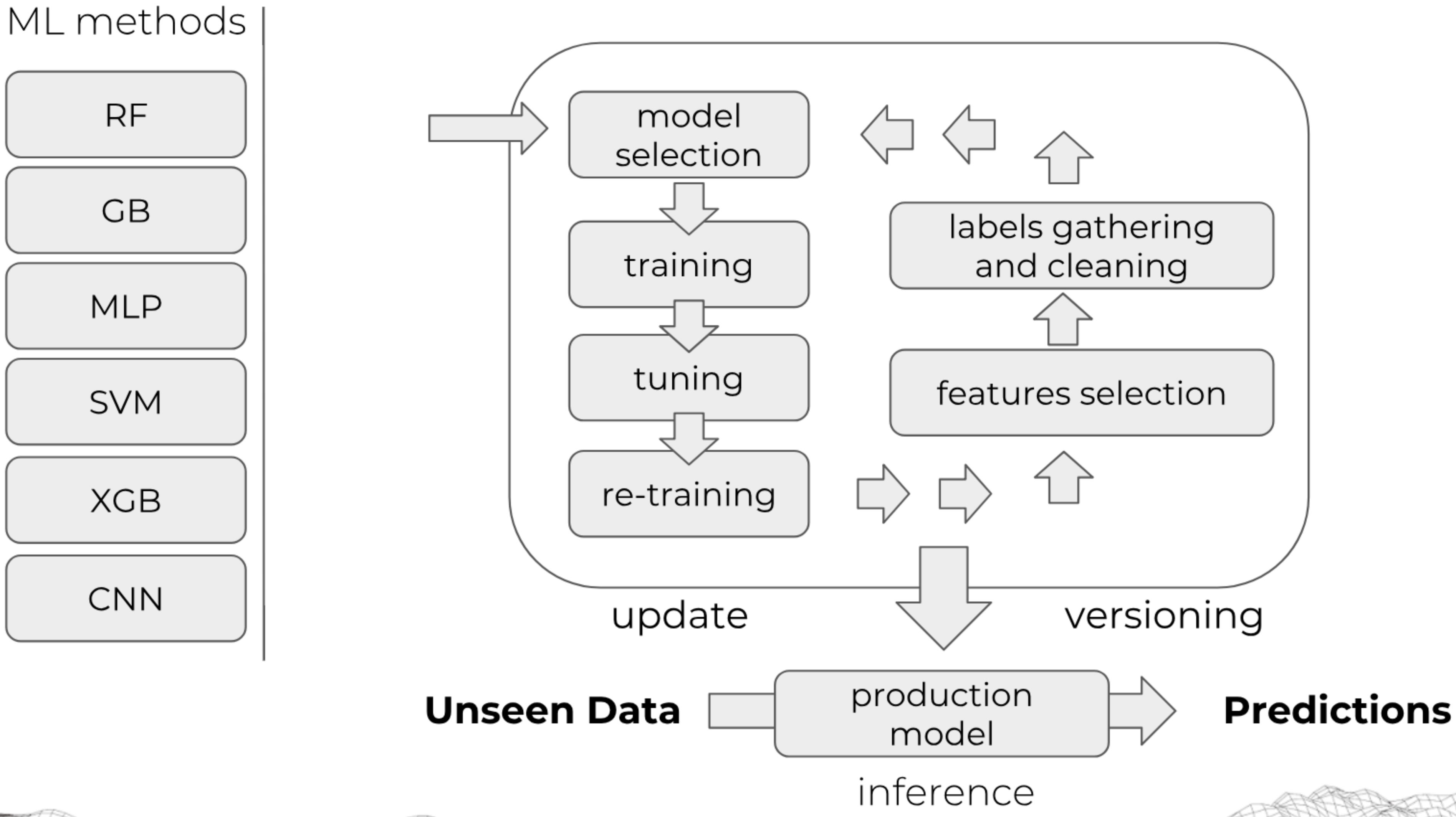
271 wells



197 wells | 83 wells | 26/58/105 wells | 205 wells

Features | Labels

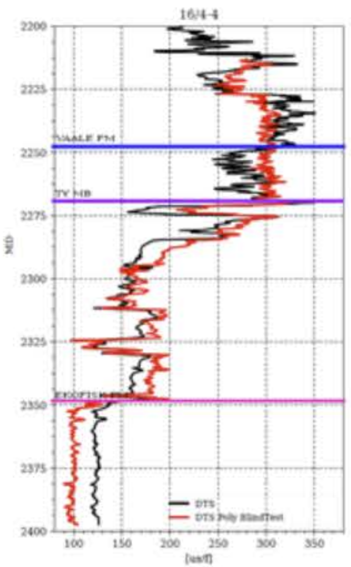
The Search for Better Models



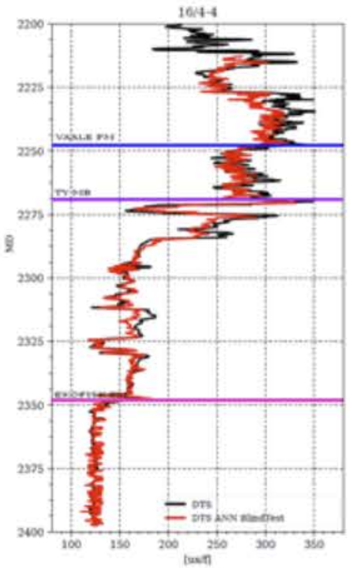
EARTH SCIENCE ANALYTICS

Shear-Sonic Prediction

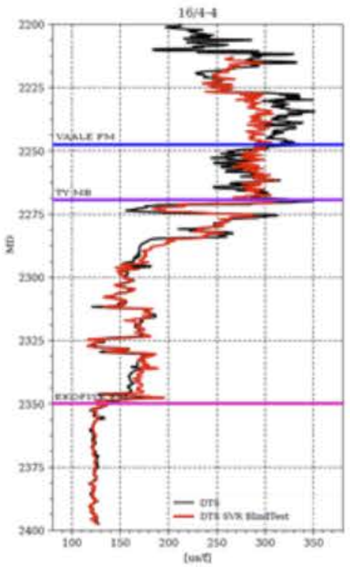
POLY



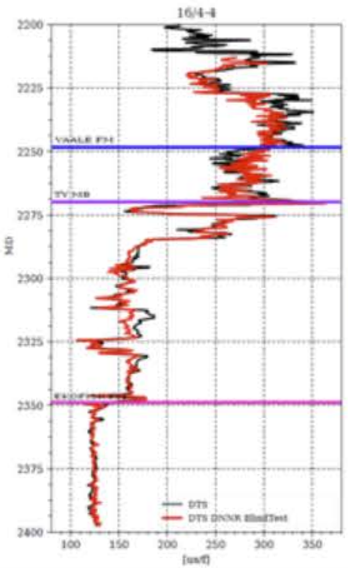
ANN



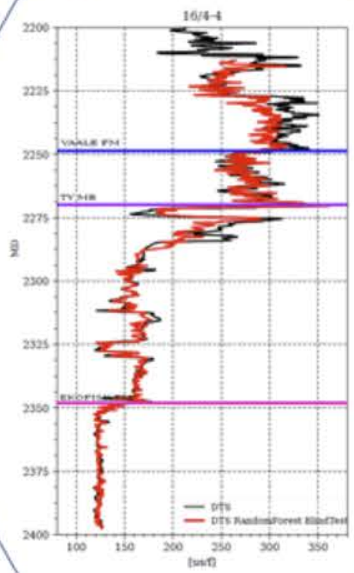
SVR



DNNR



RF

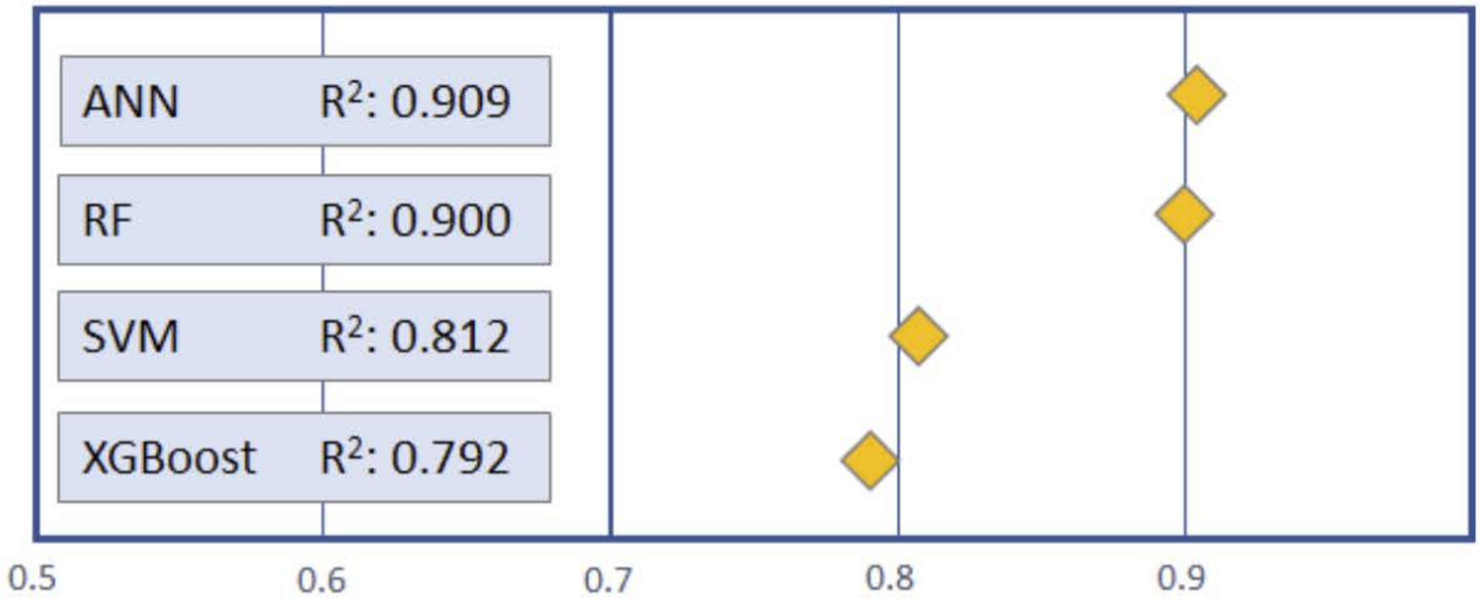


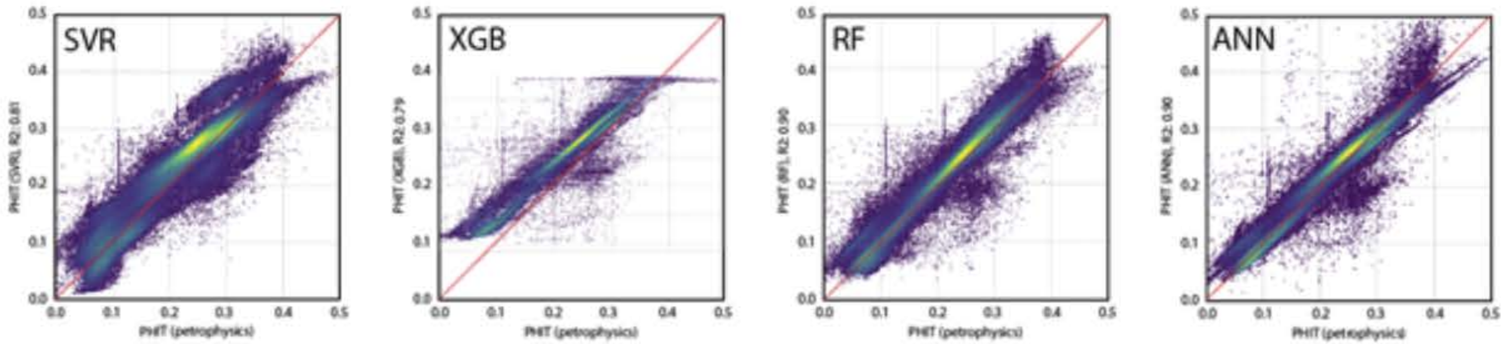
TEST OF MACHINE LEARNING ALGORITHMS FOR DTS PREDICTION

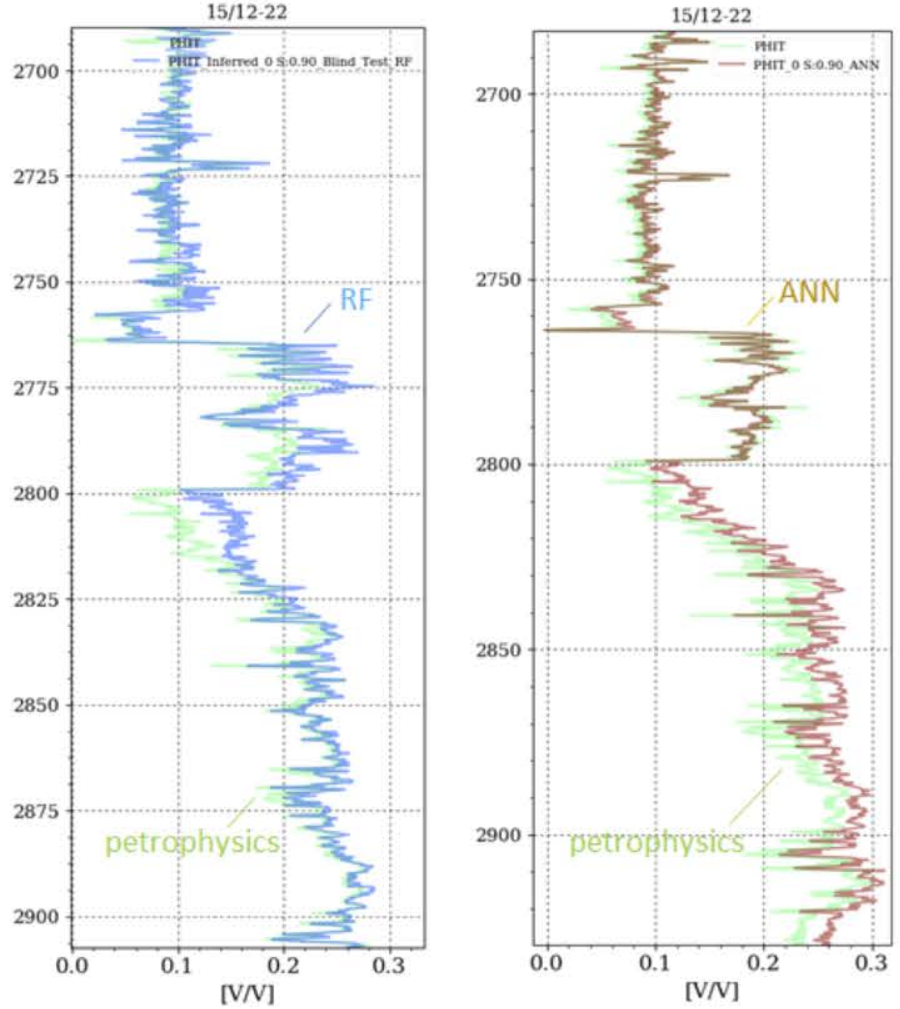
Algorithm	RS ALL	EVS ALL	MAE	MSE	R^2 Train	R^2 Test	Time [sec]
RF (Random Forest)	1	1	2.61	20.29	0.998	0.906	334.25
GB (Gradient Boost)	0.95	0.95	15.31	451.94	0.953	0.909	340.74
XGB (Extra Gradient Boost)	0.95	0.95	15.17	464.69	0.952	0.906	244.64
DNNR (Deep NN Regression)	0.95	0.95	15.73	497.1	0.948	0.911	5649.95
AB (ADA Boost)	0.93	0.93	20.21	717.14	0.925	0.898	426.06
SVR (Support Vector Regression)	0.92	0.93	19.07	720.93	0.925	0.907	273893.07
ANN (Artificial Neural Network)	0.92	0.92	20.58	776.41	0.919	0.904	3291.21
POLY (Polynomial Regression)	0.91	0.91	22.89	904.18	0.906	0.894	268.71
SGD (Stochastic Gradient Descent)	0.9	0.9	23.21	922.73	0.904	0.893	8.99

Porosity Prediction

Blind test R^2 score





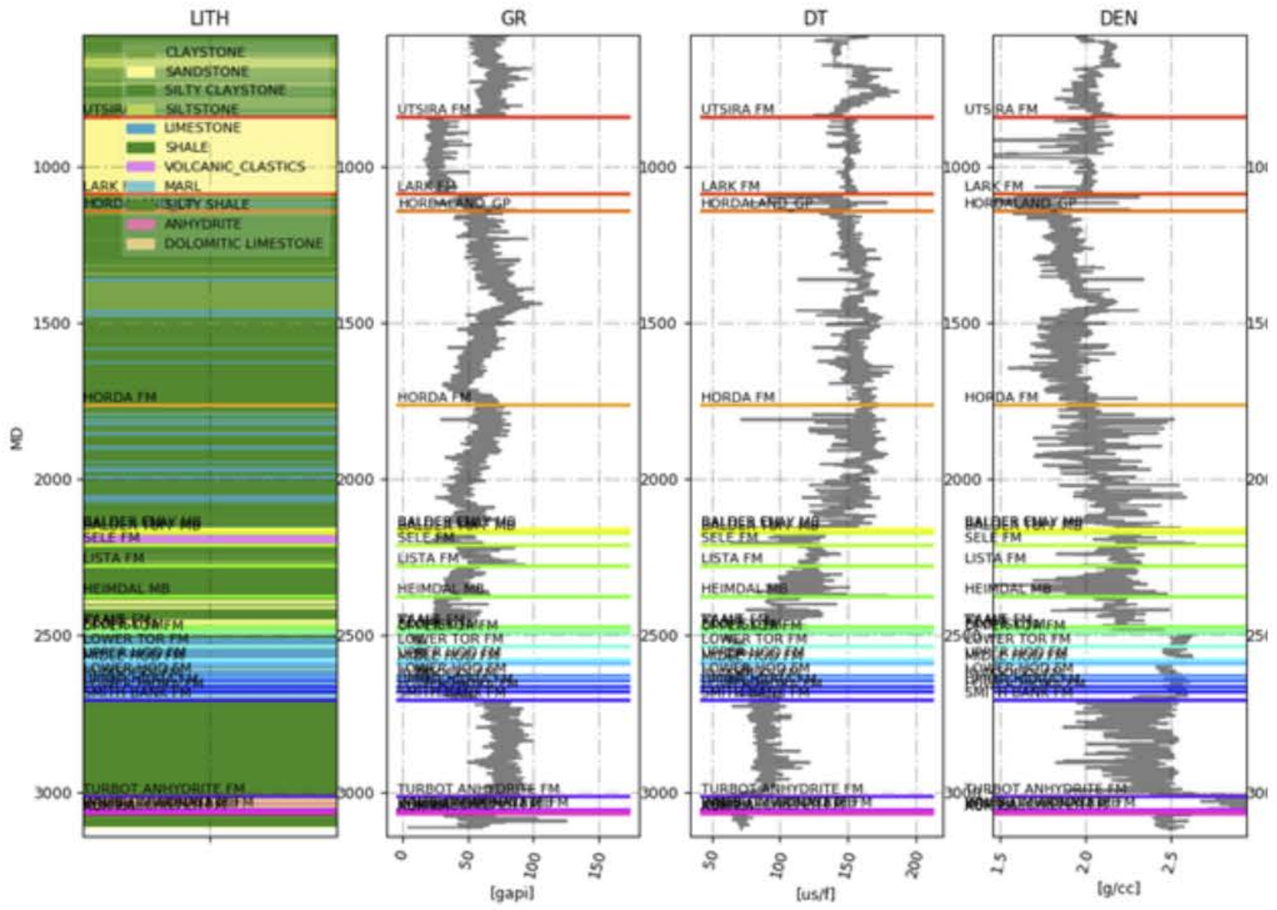


Lithology Classification

The original lithology data

- Lithology data from near seabed to TD in 205 wells

- Lithology assigned based on core, cuttings and log response



EARTH SCIENCE ANALYTICS

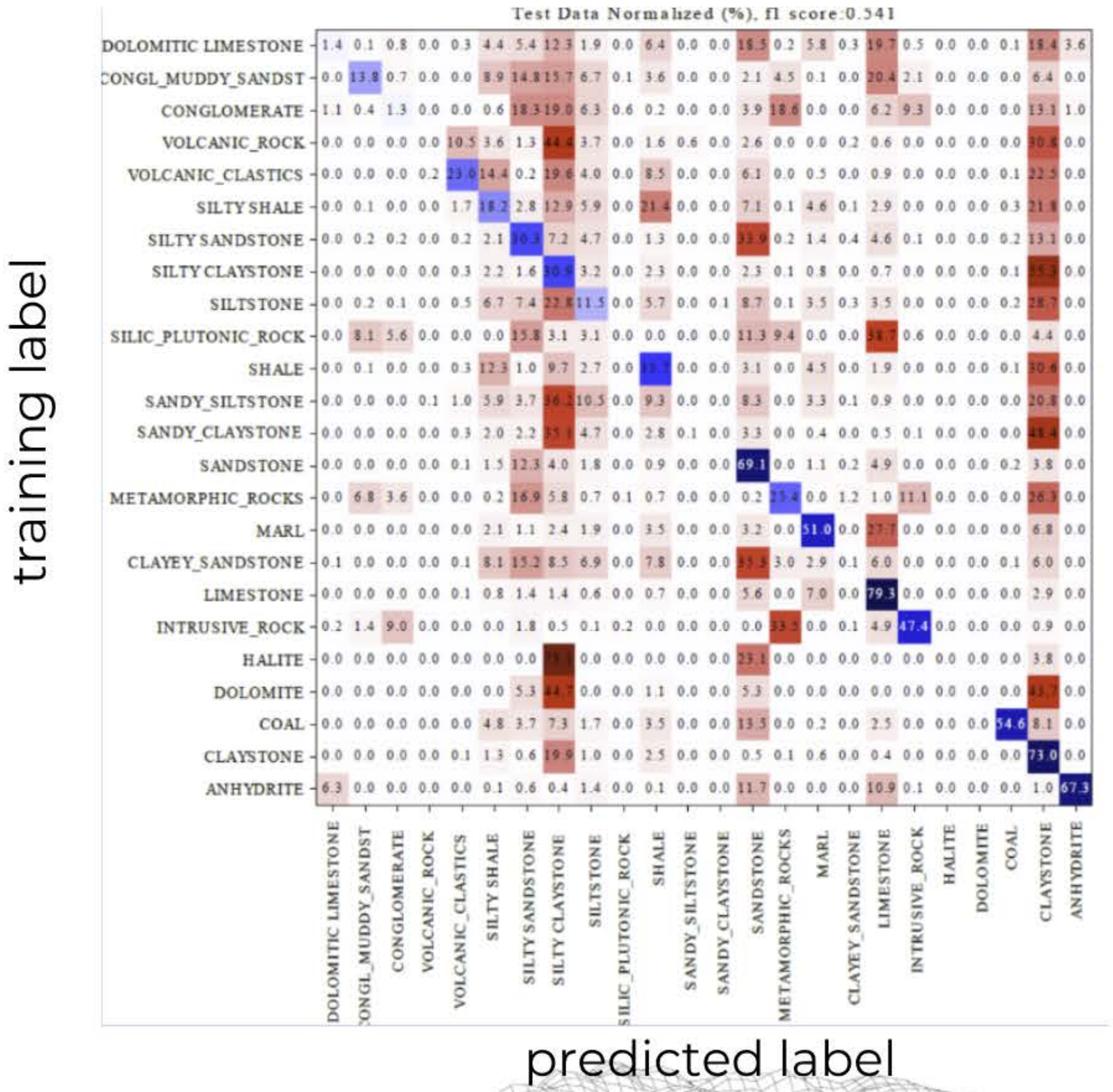
Lithology Classification

A first try

- Initial RF model trained on the raw original labels

- When blind testing for every well the aggregated F1 score is 0.53

- Lots of confusion:
    - Unable to classify several lithologies
    - Many confusions are irrelevant, and are practically the same rocks
    - Sandstone and claystone are predicted reasonably, as a start



Lithology Classification

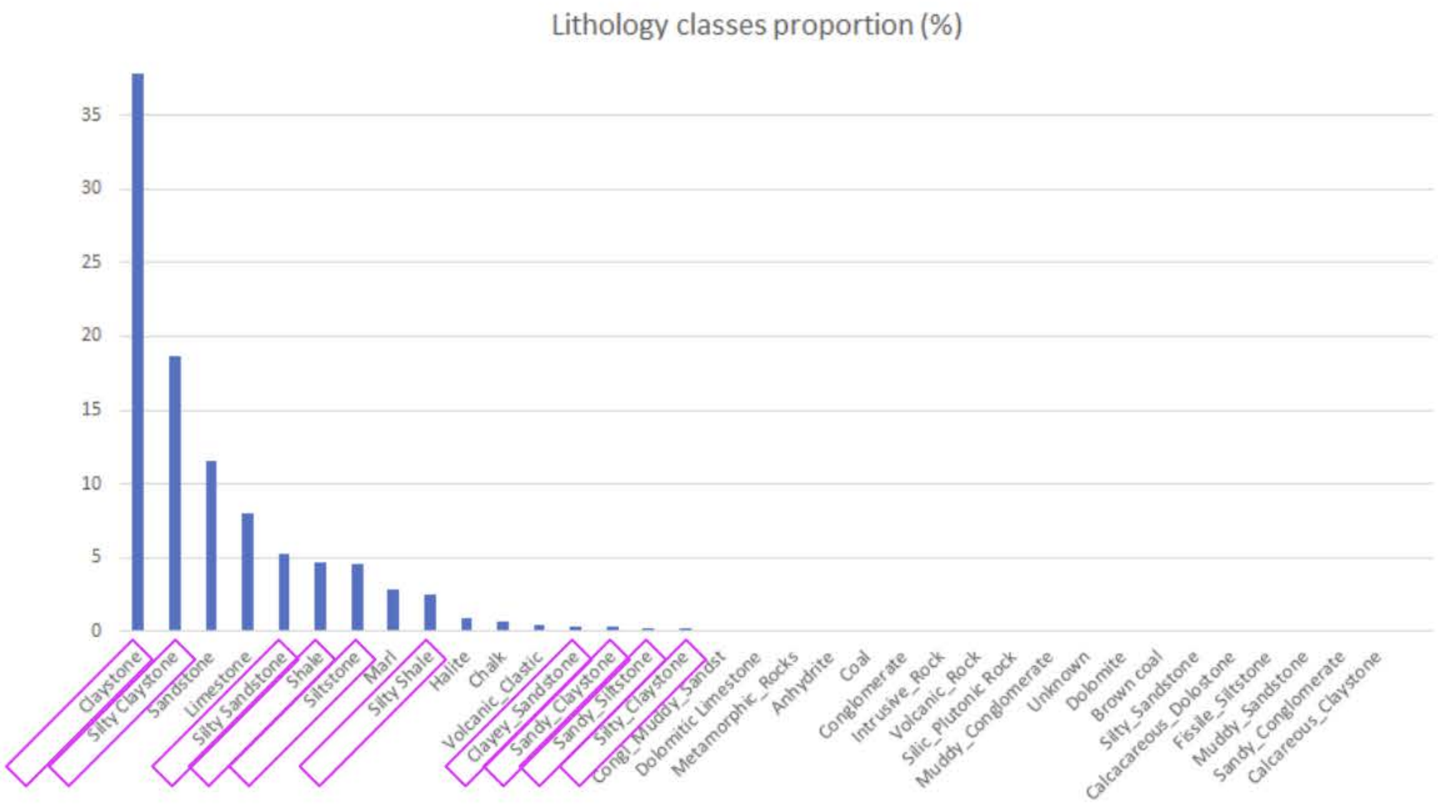
The original label set

- 22 lithology classes

- Some of the original classes practically are the same rocks
    - Let's combine some of them

- Some of the original classes have very few data points
    - Let's leave those out

- The data set is highly unbalanced
    - Let's collect more data



Lithology Classification

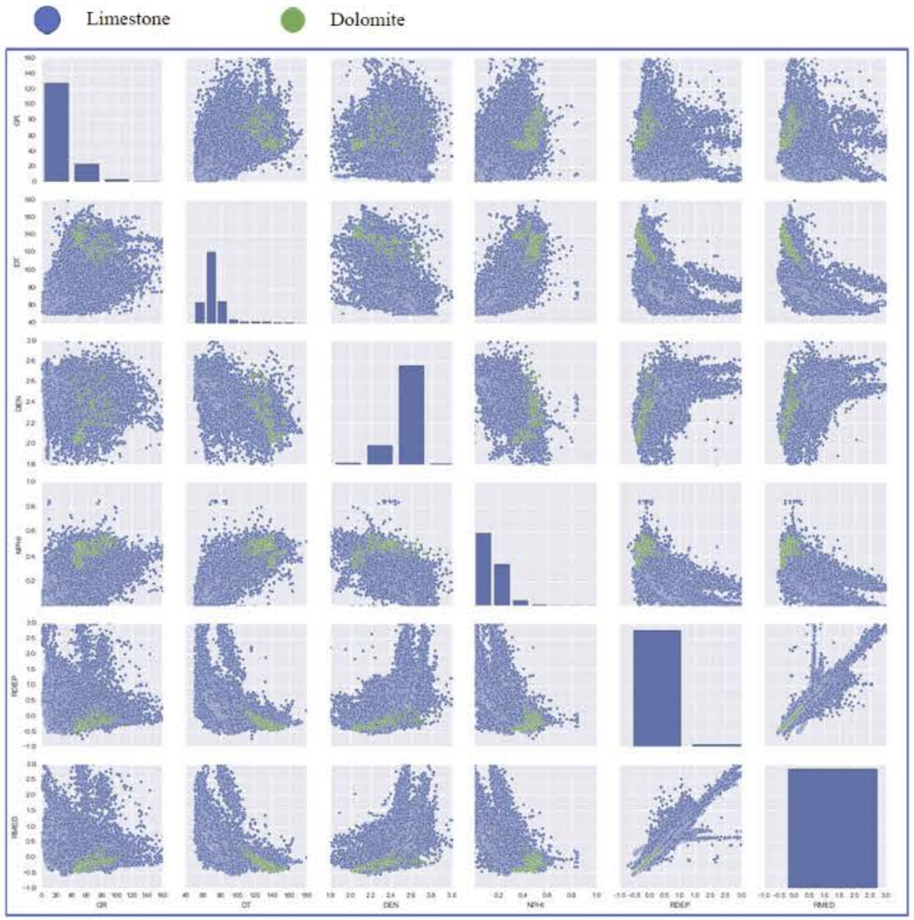
Data analysis

- Explore 22 lithology classes

- Some of the original classes practically are the same rocks
    - Let's combine some of them

- Some of the original classes have very few data points
    - Let's leave those out

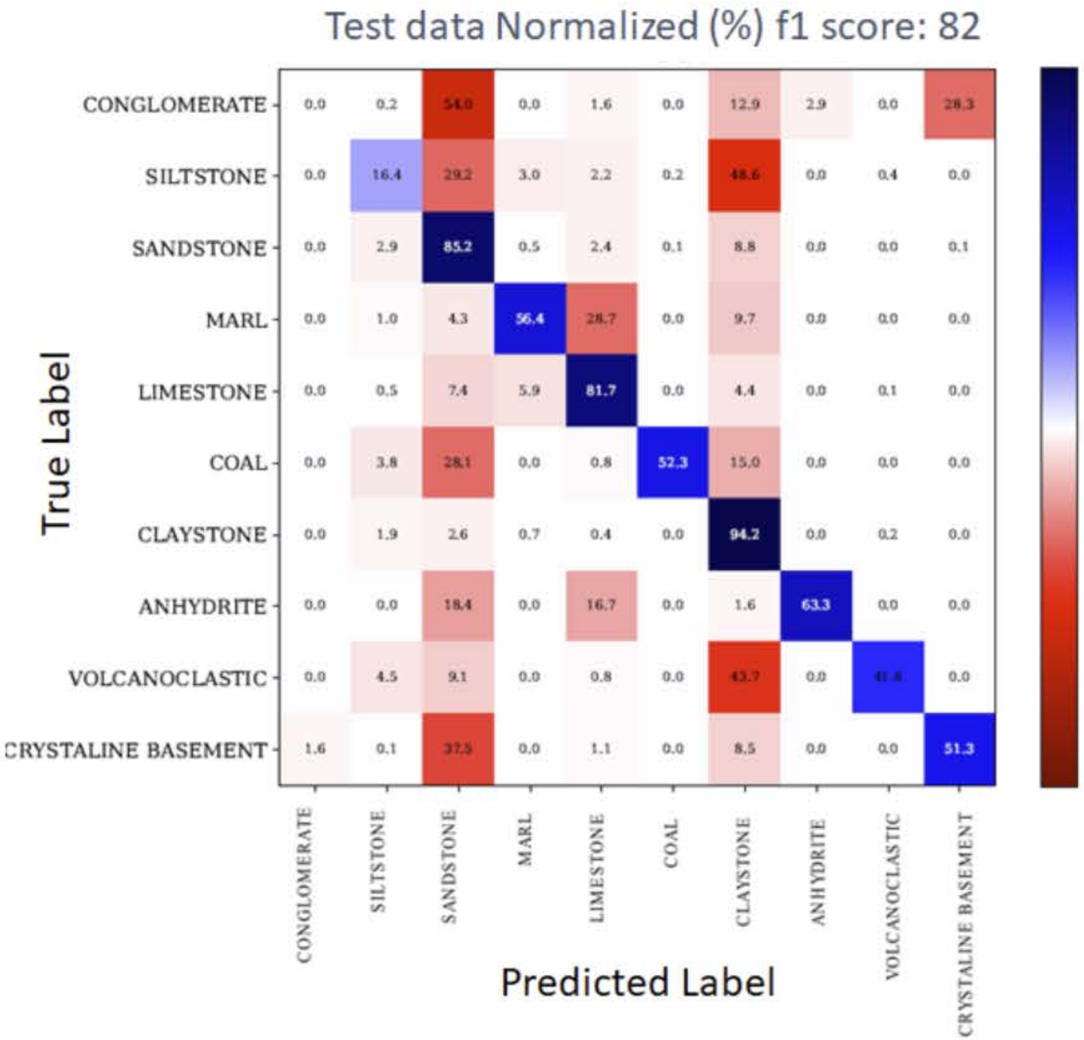
- The data set is highly unbalanced
    - Let's collect more data



Lithology Classification

A second try

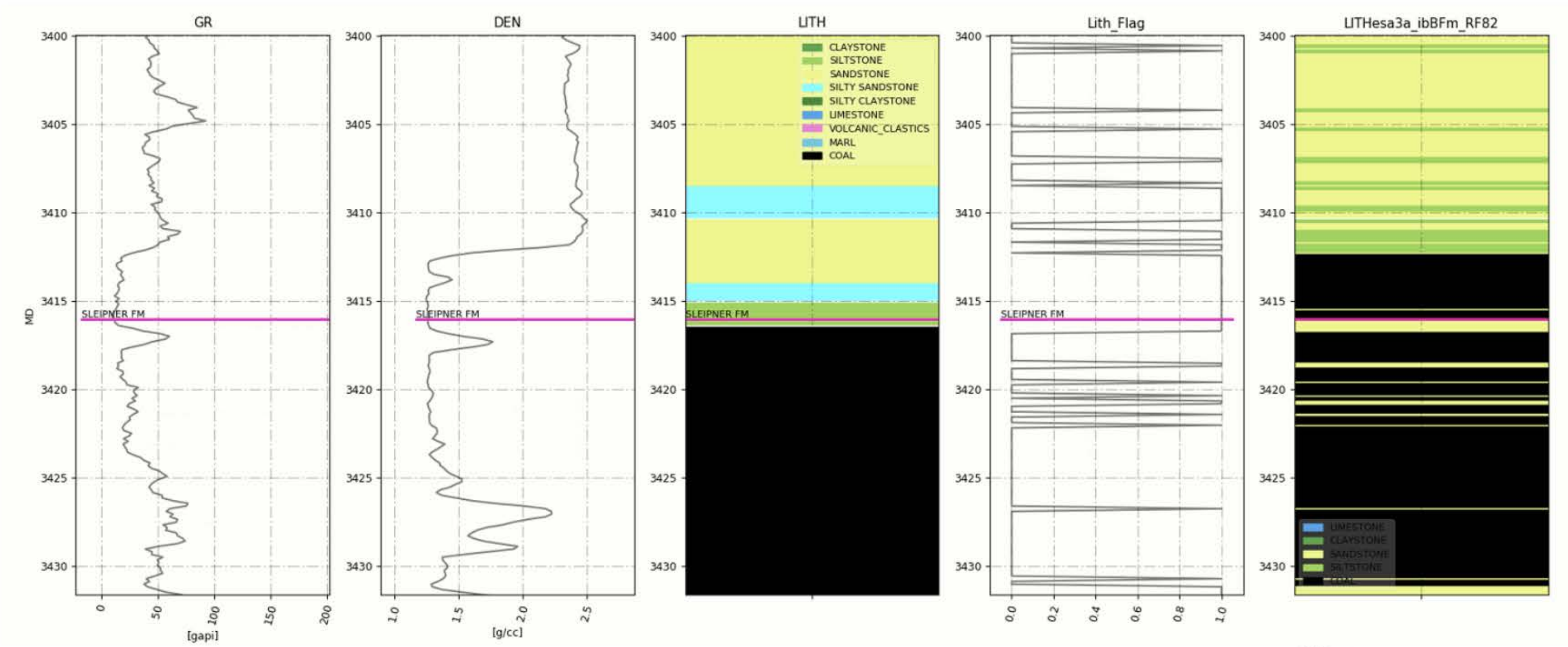
- Initial RF model on grouped labels
- Still lots of confusion:
    - Unable to classify conglomerates (to few training labels)
    - Silt confused with sandstone and claystone (silt is intermediate between the two)
    - Mediocre score for coal (to few training labels)



Lithology Classification

Identifying and correcting errors

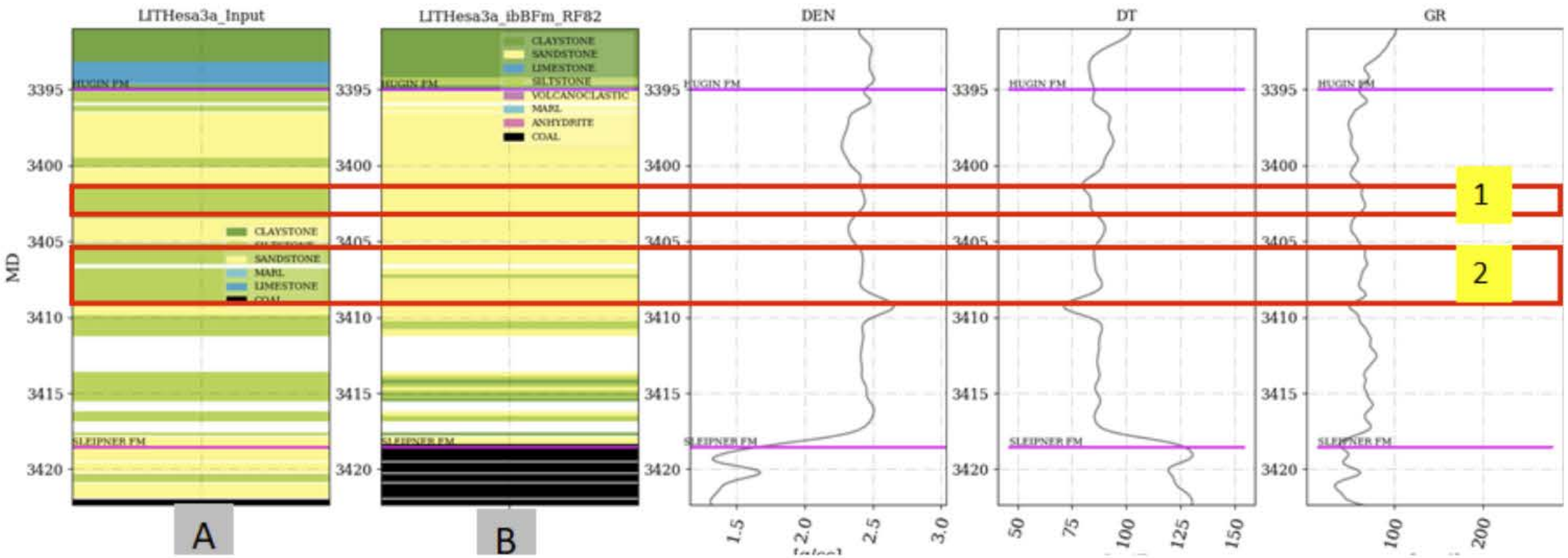
Coal label error
Flag shows the difference between original and predicted

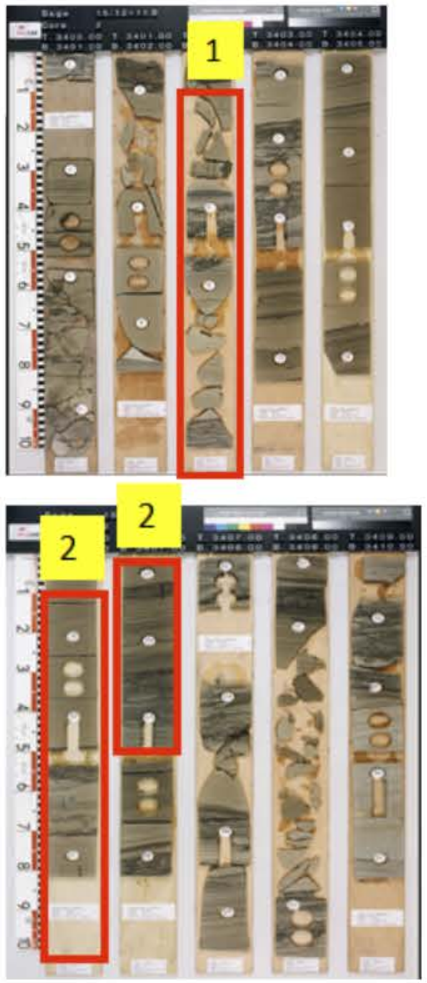


Lithology Classification

Identifying and correcting errors

Checking the potential errors using core images





Features

What is the optimal combination of logs?

- To many possible combinations to try them all

Features	Blind f1 Score (%)
GR	60
GR, RMED	72
GR, RDEP	70
GR,RMED,RDEP	74
GR,RMED,RDEP, DT	77
GR,RMED,RDEP, DT,DTS	76
GR,RMED,RDEP, DT,DTS, DEN	81
GR,RMED,RDEP, DT,DTS, DEN, NPHI	79
GR,RMED,RDEP, DT,DTS, DEN, NPHI,Fm	80
GR,RMED,RDEP, DT,DTS, DEN, NPHI,Fm, MD	78

Feature Engineering

We can engineer features from our logs that capture their vertical pattern

- Windows
- Statistics

Investigation window (m)	Statistical Measures	Regression Train Data Accuracy (R^2-Train)	Regression Blind Test Data Accuracy (R^2-Test)
no augmentation	-	99.8	90.9
1,2	Slope, R, Mean	99.7	91.7
5,10	Slope, R, Mean	99.8	90.9
1,2,4,10,20	Slope, R, Mean, Median	99.5	91.4

Results for DTS prediction with Random Forest algorithm

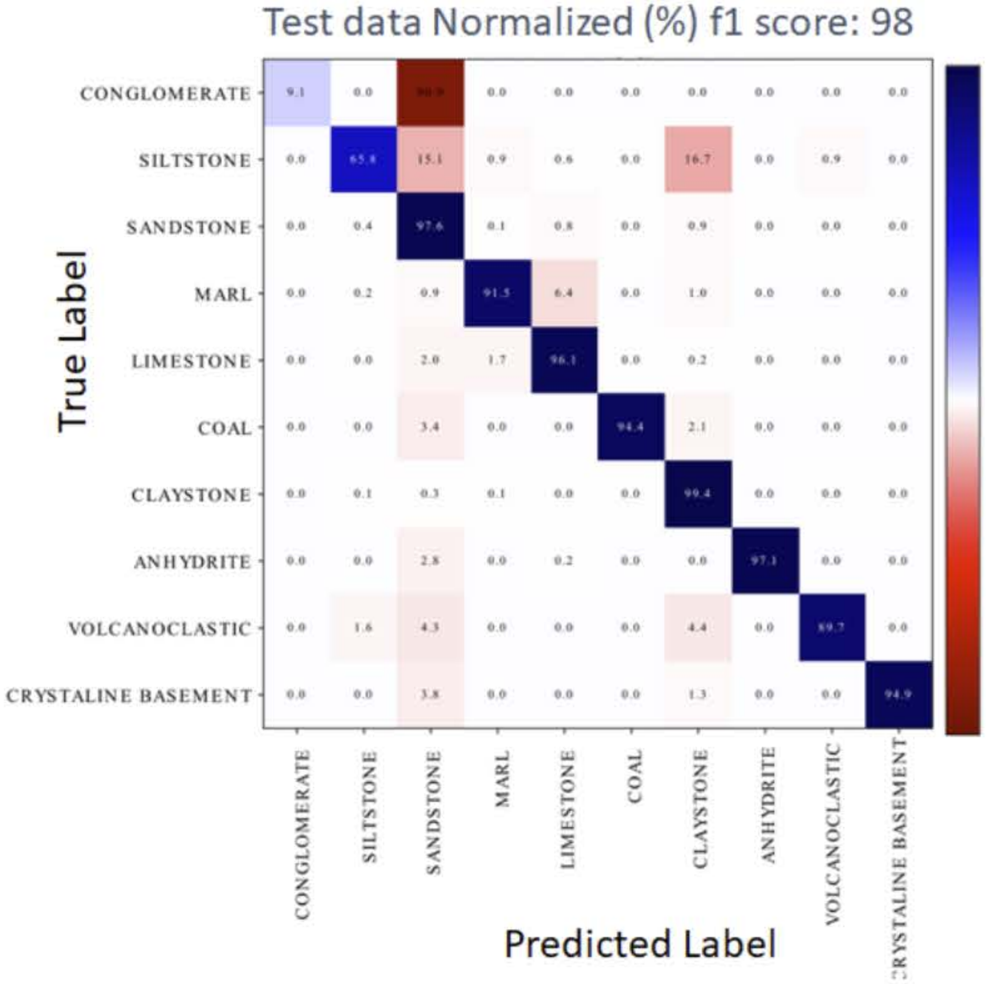
The Production Model

EARTH SCIENCE ANALYTICS

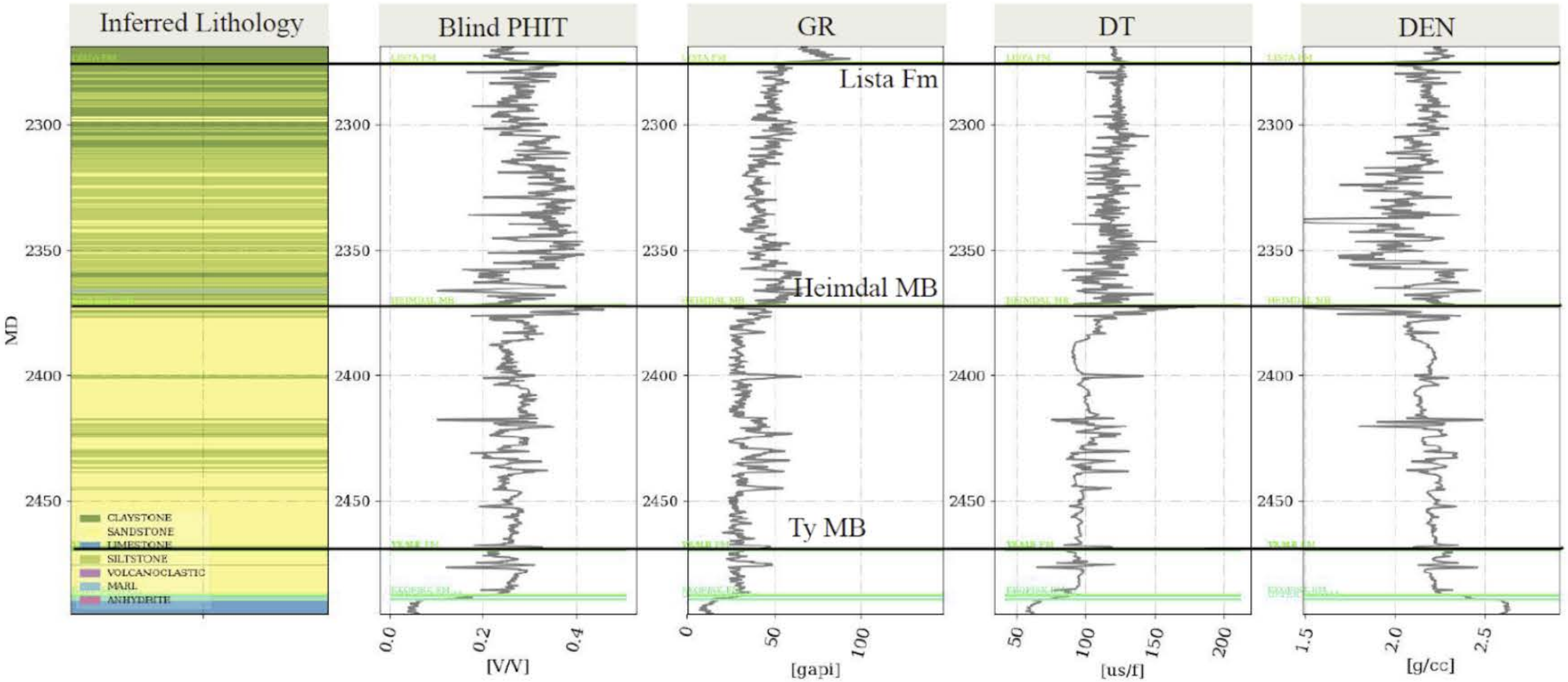
1.Manual cleaning of lithology labels with clear log signature/core description data

2.Label filtering
Blind test approach (total score 98%)
Basic features: Den, DT, DTS, GR, NPHI, RMED, RDEP

Random Forest algorithm
Number of estimators: 8
No augmentation



The Production Model



Integration

Data Platform

Data Analytics

https://earthnet.ai/earthnet/models/

EarthNET | Data Management | Machine Learning | Workers | **Models** | Select Theme | barents | Eirik

All Projects | All Labels | eirik.larsen | All Wells Selected | Show Hidden

Datetime	Project	Label	Label Type	Model Name	Method	Train Size	State	User	Train Score	Test Score	
9/7/2018, 11:17:52 PM	barents	POR	POR	POR = RFR(DEN,DT,GR,RDEP,RMED,NEU,MD)	RFR	60 %	processing	eirik.larsen			View Logs
9/7/2018, 10:31:50 PM	barents	POR	POR	POR = RFR(DEN,DT,GR,RDEP,RMED,NEU,MD)	RFR	60 %	created	eirik.larsen	97.7 %	91.4 %	Review
9/7/2018, 10:26:22 PM	barents	POR	POR	POR = RFR(DEN,DT,GR,RDEP,RMED,NEU,MD)	RFR	60 %	created	eirik.larsen	97.4 %	91.0 %	Review
9/7/2018, 10:23:32 PM	barents	POR	POR	POR = GB(DEN,DT,GR,RDEP,RMED,NEU,MD)	GB	60 %	processing	eirik.larsen			View Logs
9/7/2018, 9:44:39 PM	barents	LITHesa	LITH	LITHesa = RFC(DEN,DT,GR,RDEP,RMED,NEU,MD)	RFC	60 %	created	eirik.larsen	76.3 %	73.1 %	Review
9/7/2018, 9:39:05 PM	barents	LITHesa	LITH	LITHesa = SVC(DEN,DT,GR,RDEP,RMED,NEU,MD)	SVC	Blind	created	eirik.larsen	90.0 %	86.6 %	Review
9/7/2018, 9:15:07 PM	barents	LITHesa	LITH	LITHesa = EXTREES(DEN,DT,GR,RDEP,RMED,NEU,MD)	EXTREES	60 %	created	eirik.larsen	100.0 %	94.2 %	Review
9/7/2018, 9:14:44 PM	barents	LITHesa	LITH	LITHesa = SVC(DEN,DT,GR,RDEP,RMED,NEU,MD)	SVC	60 %	created	eirik.larsen	96.8 %	96.2 %	Review
9/7/2018, 8:48:50 PM	barents	POR	POR	POR = RFR(DEN,DT,GR,RDEP,RMED,NEU,MD)	RFR	60 %	created	eirik.larsen	97.4 %	90.9 %	Review
9/7/2018, 8:47:09 PM	barents	POR	POR	POR = SGD(DEN,DT,GR,RDEP,RMED,NEU,MD)	SGD	60 %	created	eirik.larsen	71.0 %	70.3 %	Review
9/7/2018, 8:41:24 PM	barents	POR	POR	POR = SVR(DEN,DT,GR,RDEP,RELPOS,RMED,NEU,MD,DTS_mlFilled)	SVR	60 %	created	eirik.larsen	76.5 %	76.0 %	Review
9/7/2018, 7:52:50 PM	barents	POR	POR	POR = GB(DEN,DT,GR,RDEP,RELPOS,RMED,NEU)	GB	60 %	created	eirik.larsen	83.5 %	82.8 %	Review
9/7/2018, 7:37:41 PM	barents	LITHesa	LITH	LITHesa = SVC(DEN,DT,GR,RDEP,RELPOS,RMED,NEU)	SVC	60 %	created	eirik.larsen	99.1 %	98.5 %	Review
9/7/2018, 7:35:01 PM	barents	LITHesa	LITH	LITHesa = DNNC(DEN,DT,GR,RDEP,RELPOS,RMED,NEU)	DNNC	Blind	created	eirik.larsen	92.7 %	71.9 %	Review
9/7/2018, 7:34:44 PM	barents	LITHesa	LITH	LITHesa = DNNC(DEN,DT,GR,RDEP,RELPOS,RMED,NEU)	DNNC	Blind	created	eirik.larsen	92.9 %	61.9 %	Review
9/7/2018, 7:28:49 PM	barents	HC	HC	HC = EXTREES(DEN,DT,GR,RDEP,RELPOS,RMED,NEU)	EXTREES	60 %	created	eirik.larsen	100.0 %	98.6 %	Review
9/7/2018, 7:24:01 PM	barents	LITHesa	LITH	LITHesa = RFC(DEN,DT,GR,RDEP,RELPOS,RMED,NEU)	RFC	60 %	created	eirik.larsen	95.0 %	93.8 %	Review
9/7/2018, 7:19:24 PM	barents	DTS	DTS	DTS = XGR(DEN,DT,GR,RDEP,RELPOS,RMED,NEU)	XGR	60 %	created	eirik.larsen	97.8 %	97.6 %	Review
8/28/2018, 9:39:04 PM	barents	DTS	DTS	DTS = GB(DEN,DT,GR,RDEP,RELPOS,RMED)	GB	Blind	created	eirik.larsen	95.2 %	93.1 %	Review

Ask the Data

Alexa.
Can you show me the porosity distribution;

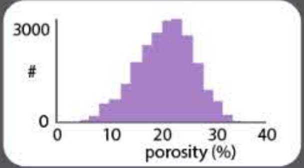
- for all the wells in Quad A
- for Formation B
- for the fluvial channels only

- using inferred data

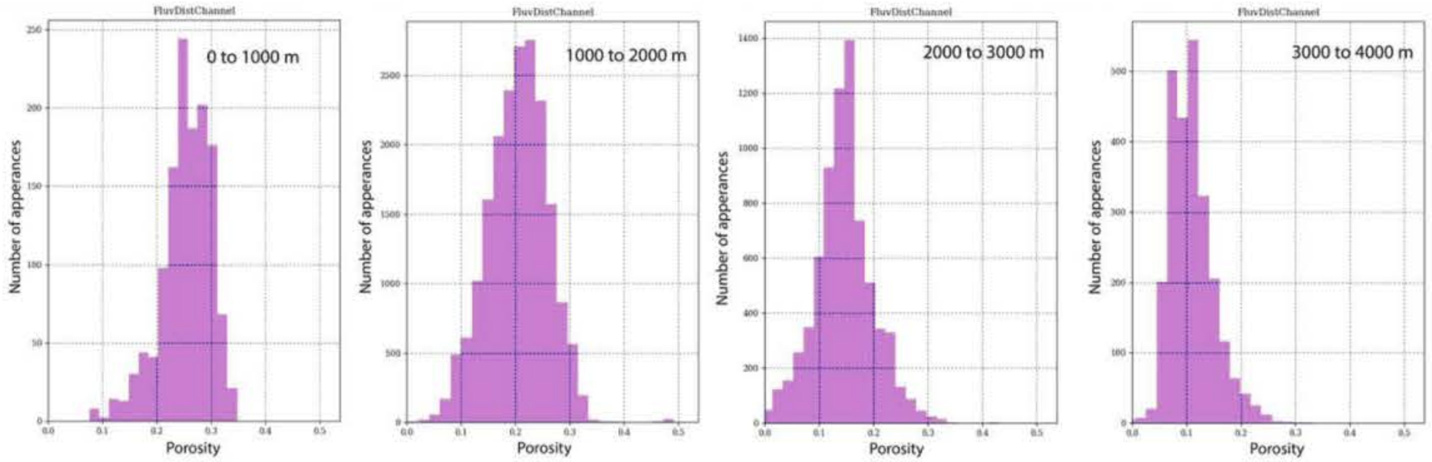
Do you want to specify a depth range?

Yes. Between 1000 and 2000 metres below sea level please

Here you go



Thanks! Let's use this for the volumetrics



earthanalytics.ai