Detecting flow anomalies from accelerometer data analytics

Wathsala Perera
Researcher
SINTEF
17-12-2024

Teknologi for et bedre samfunn

Introduction

- HARMONIC - HYDROPHONE & ACCELEROMETER REAL-TIME MONITORING AND INTERPRETATION IN COMPLEX FLUIDS.

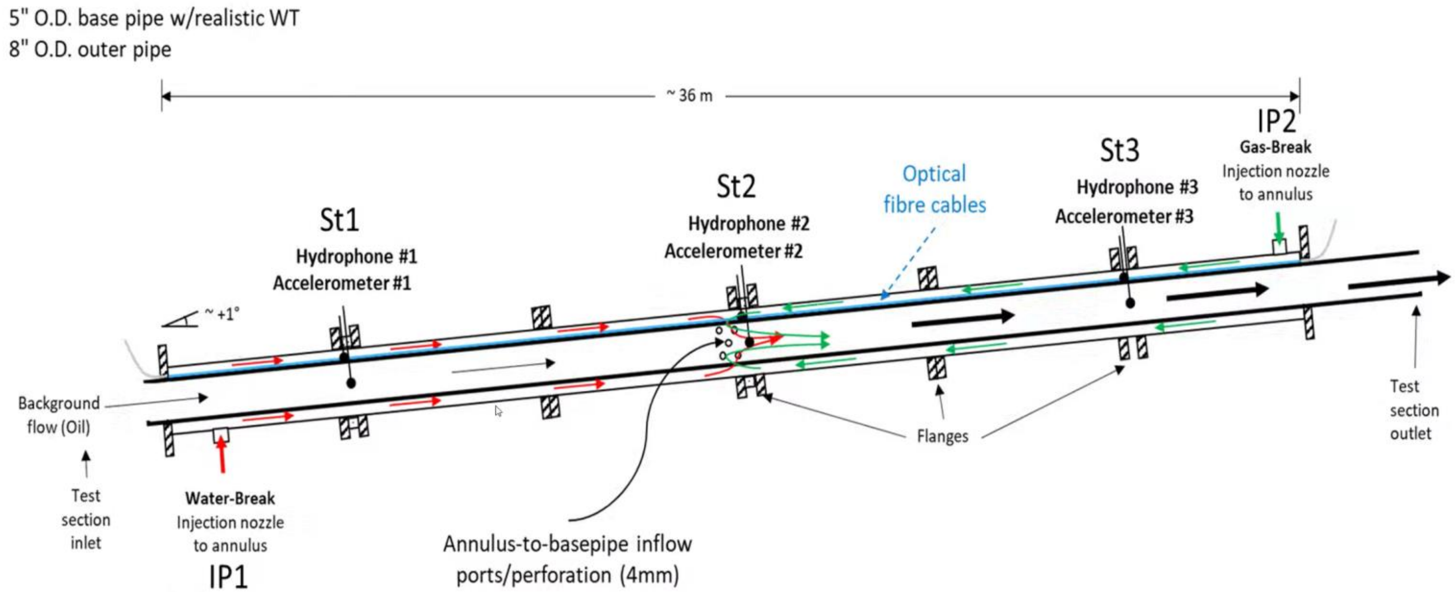
- HARMONIC
  - DIFIPRO (SINTEF, NORCE, UIB, UIS)
    - DAS (Fibre optics)
    - Tri-axial accelerometers
    - Hydrophones
  - Develop correlations between accelerations, hydrostatic pressure, DAS signals

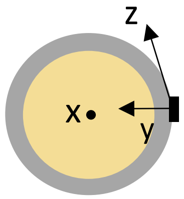
  - HARMONIC will see how to use large amounts of data in effectively interpreting multiphase flow characters with the use of only accelerometers.

Goals

Problem	Goal 1	Goal 2
Complexity of using accelerometer data in predictive models	Improve the efficiency of feature extraction from accelerometer data	Use extracted features in classification algorithms to classify water/gas breaks
		Use extracted features in predictive algorithms to obtain good performance in prediction multiphase flow anomalies

Experimental set-up






Multiphase Flow Rig


Pipe-in-pipe w/ FO cables


Hydrophone reference probes


Accelerometer reference probes

Data

DIFIPRO Data

data.sintef.no

Data Acquisition

Only accelerometer data (62.5 kHz)

Only Steady-state experiments (202)

Each experiment 60 seconds long

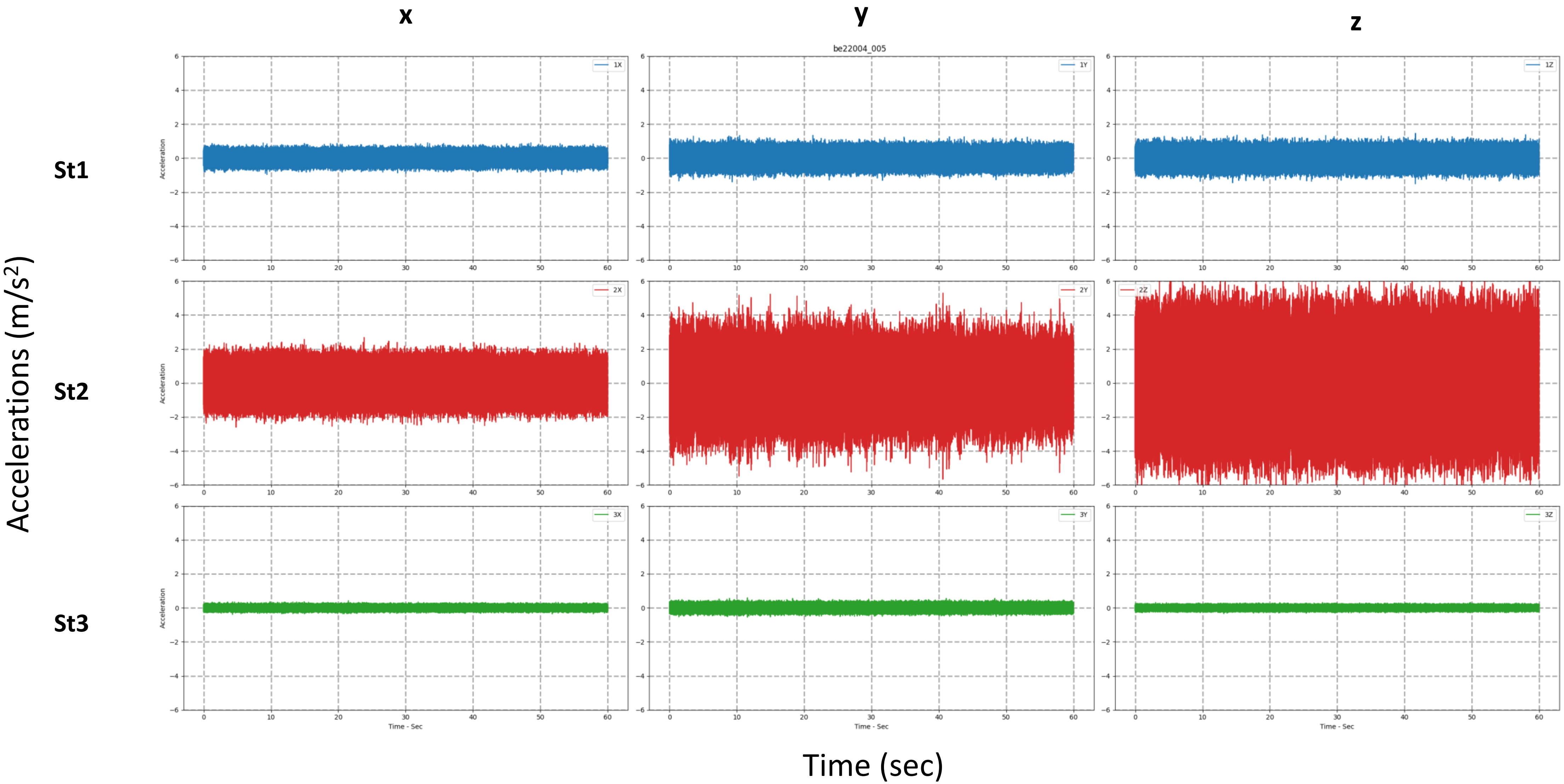
Reference measurements

Available for each experiment

1	Background oil mass flow rate	mfroil0
2	Background water mass flow rate	mfrwat0
3	Background gas mass flow rate	mfrgas0
4	Skid 1 oil mass flow rate	mfroil1
5	Skid 1 water mass flow rate	mfrwat1
6	Skid 1 gas mass flow rate	mfrgas1
7	Skid 2 oil mass flow rate	mfroil2
8	Skid 2 water mass flow rate	mfrwat2
9	Skid 2 gas mass flow rate	mfrgas2
10	Pressure #2 test section	pi72102

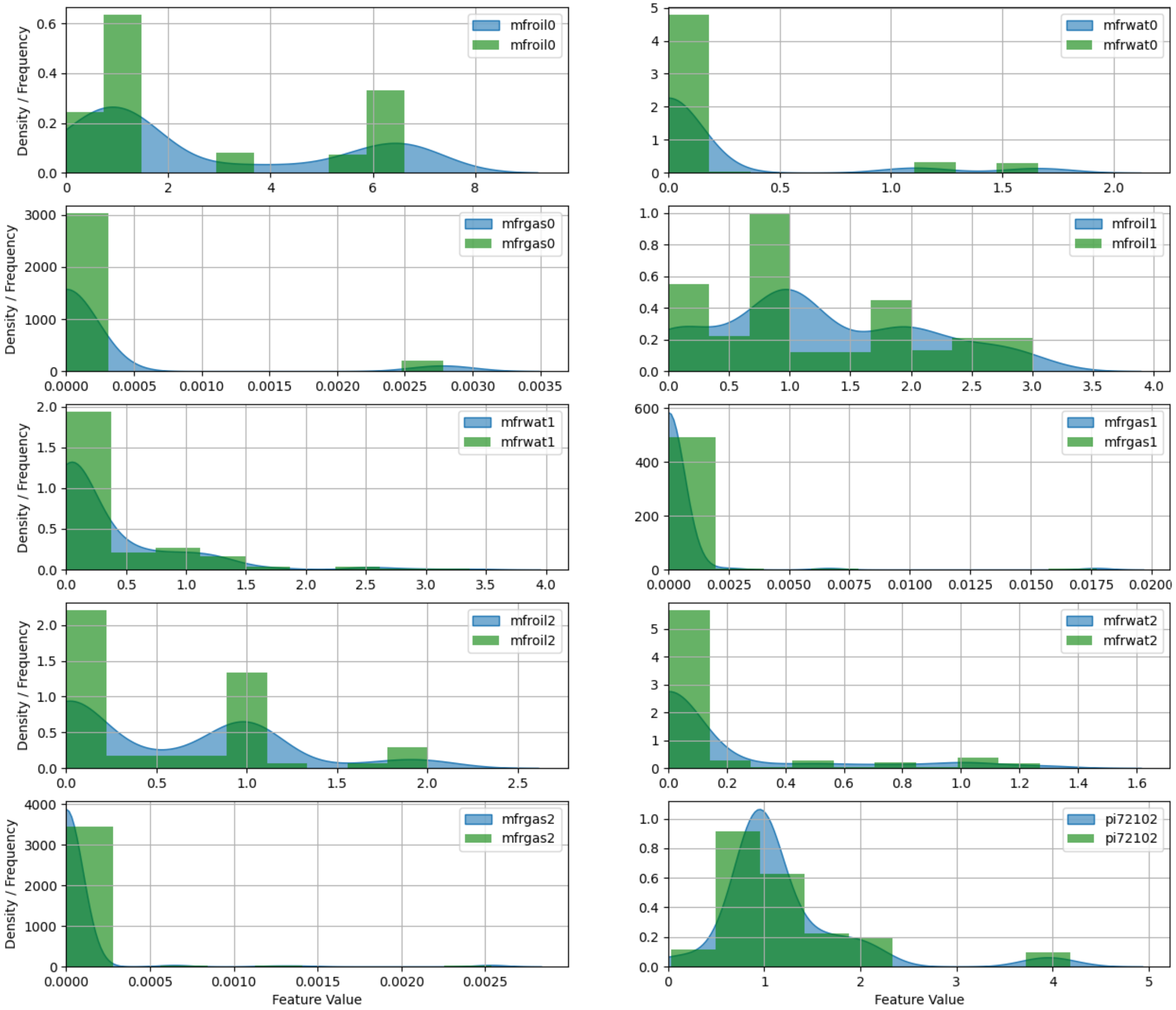
Accelerometer data example (exp_id: be22004_005)

Time Stamp	station1-Accel 1x	station1-Accel 1y	station1-Accel 1z	station2-Accel 2x	station2-Accel 1y	station2-Accel 1z	station3-Accel 3x	station3-Accel 3y	station3-Accel 3z



Reference measurements for steady state experiments

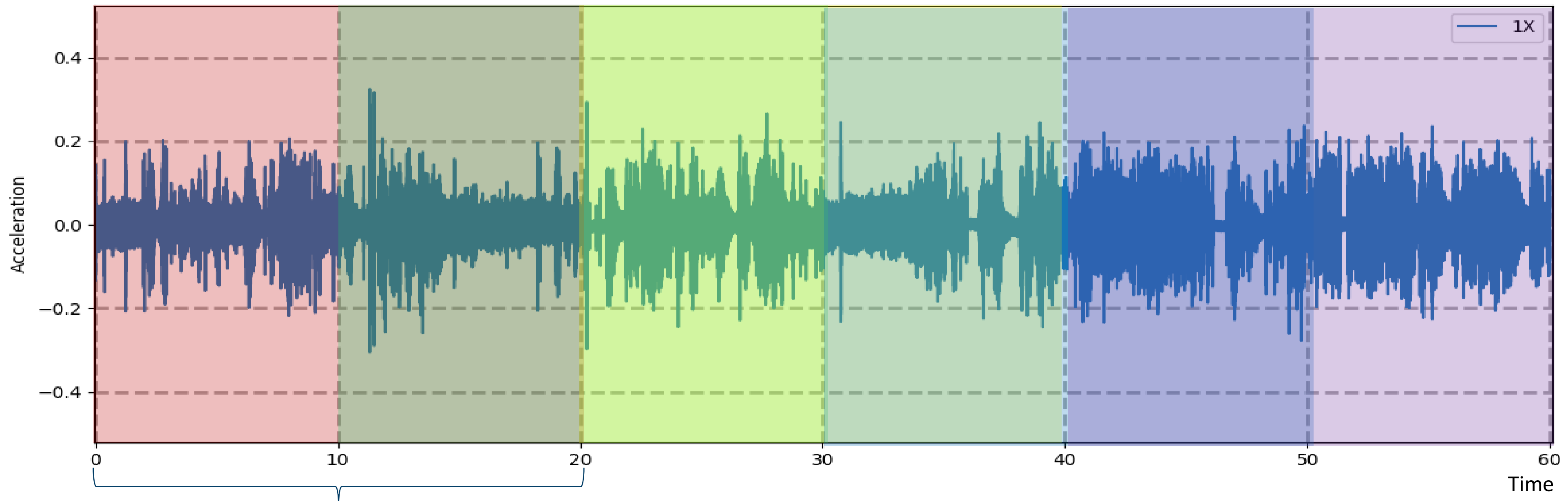
1	Background oil mass flow rate	mfroil0
2	Background water mass flow rate	mfrwat0
3	Background gas mass flow rate	mfrgas0
4	Skid 1 oil mass flow rate	mfroil1
5	Skid 1 water mass flow rate	mfrwat1
6	Skid 1 gas mass flow rate	mfrgas1
7	Skid 2 oil mass flow rate	mfroil2
8	Skid 2 water mass flow rate	mfrwat2
9	Skid 2 gas mass flow rate	mfrgas2
10	Pressure #2 test section	pi72102



Accelerometer data dimentionality reduction

- How to reduce the dimentionality of the acceleration signals so that they can be used in building the correlation between acceleration data and reference measurements?
  - 62500*60 (3 750 000) data /rows in each file
  - 9 acceleration columns

	TimeStamp Rel	Station 1-Accel 1X	Station 1-Accel 1Y	Station 1-Accel 1Z	Station 2-Accel 2X	Station 2-Accel 2Y	Station 2-Accel 2Z	Station 3-Accel 3X	Station 3-Accel 3Y	Station 3-Accel 3Z
0	0.00000	-0.02114	-0.02954	0.02594	0.09352	0.03161	-0.09297	-0.00643	0.01112	-0.01831
1	0.00002	0.05308	-0.05591	-0.03030	-0.07204	0.22964	0.34343	-0.00447	0.00360	-0.03422
2	0.00003	0.00643	-0.01602	0.00142	0.07924	-0.10485	-0.02354	0.00033	0.01668	-0.02278
3	0.00005	-0.01439	-0.02856	0.03782	0.05014	0.04981	-0.05133	0.01439	0.01068	0.00589
4	0.00006	0.00229	0.02376	0.01766	0.05439	-0.06834	0.07281	0.00349	0.00981	0.02670
5	0.00008	-0.00414	0.07771	-0.02627	-0.19520	0.14747	-0.00272	0.00022	0.01406	0.02790
6	0.00010	-0.01362	0.03030	-0.04054	-0.05275	-0.09667	-0.13602	0.00229	0.00948	-0.00610
7	0.00011	-0.02496	-0.00087	0.00490	-0.01460	-0.09417	-0.20185	-0.00730	0.01068	-0.03063
8	0.00013	-0.00120	0.02071	0.05221	0.03673	-0.06812	0.09068	-0.00861	0.01395	-0.02670
9	0.00014	0.07161	0.05820	-0.00185	0.13896	-0.10910	0.25755	-0.00905	0.00926	-0.00676
10	0.00016	-0.00425	0.09580	-0.04894	0.05155	0.19978	-0.02104	-0.01123	0.00905	0.02202



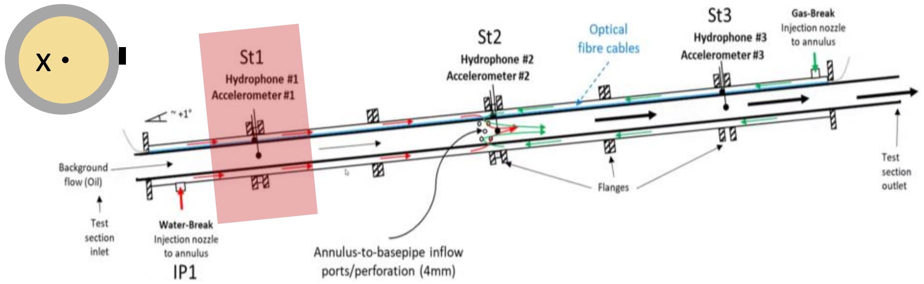
Time domain
RMS
Mean
STD
Skewness
Kurtosis

FFT
Dominant frequency
Spectral centroid
Spectral bandwidth

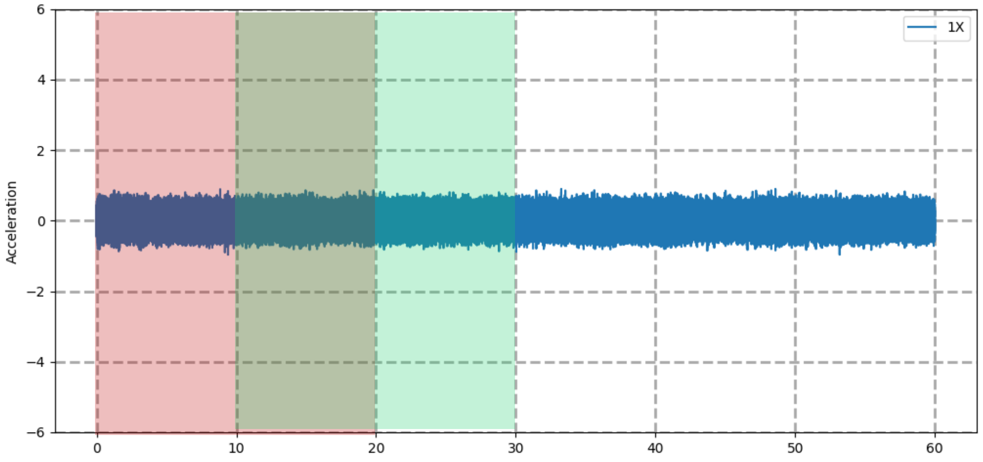
PSD
Peak frequency
Total power

SINTEF

Time domain features extracted from moving windows (ax @st1)



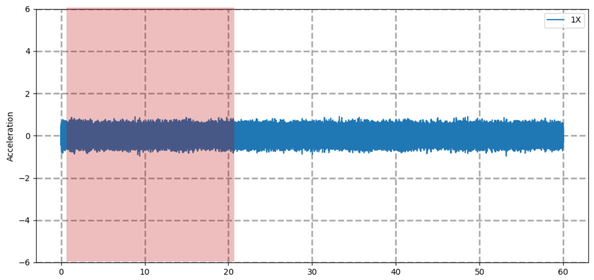
RMS	Average power of the signal over a specified time period.
Mean	The mean represents the average value of the signal over time. For a stationary signal, the mean is typically close to zero.
STD	Standard deviation measures the spread or variability of the signal around its mean.
Skewness	Skewness measures the asymmetry of the signal’s distribution around its mean. It indicates whether the acceleration signal has more extreme values (outliers) on one side of the mean than the other.
Kurtosis	Kurtosis measures the extent to which data resides in the tails of the distribution. A kurtosis value of approximately 3 suggests a normal, Gaussian-like distribution.

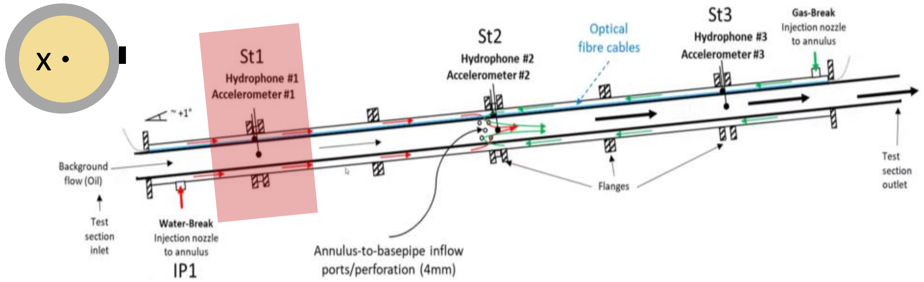


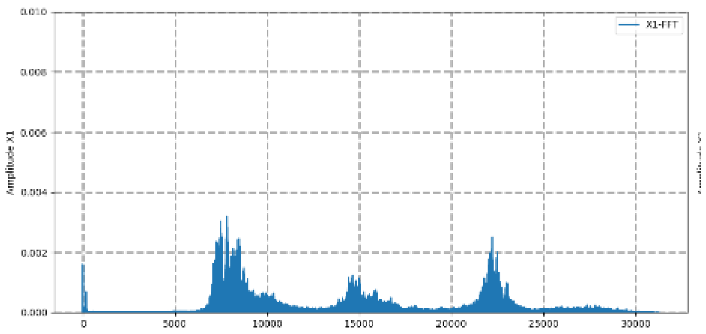
exp	1X_rms_w1	1X_mean_w1	1X_std_w1	1X_skewness_w1	1X_kurtosis_w1	1X_rms_w2	1X_mean_w2	1X_std_w2	1X_skewness_w2	1X_kurtosis_w2
be22003_028	0.021865443	0.000293829	0.021863468	-0.019325466	4.826746081	0.021910035	-0.000236193	0.021908762	-0.020812483	5.382503546
be22004_000	0.029516444	-0.000330157	0.029514597	-0.0702682	2.628766354	0.029329773	-0.001146609	0.029307352	-0.05243757	2.538553227
be22004_002	0.013996849	-0.001112041	0.013952603	0.154147128	1.8253907	0.01267164	0.000553775	0.012659534	0.26188925	1.957627798
be22004_003	0.017766911	-0.00115155	0.017729553	-0.012378195	0.172062264	0.017453987	0.001739375	0.017367102	-0.060142337	0.225907842
be22004_005	0.185038178	-9.35E-05	0.185038154	-0.006466214	0.01508194	0.184560565	-0.000123531	0.184560524	-0.005364633	0.006700935

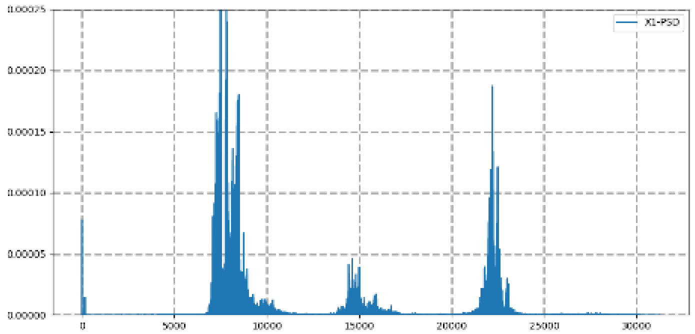
Frequency domain features extracted from moving windows (ax @st1)

Dominant Frequency	This is the frequency that best represents the primary periodic behaviour of the signal. It can indicate a characteristic flow pattern.
Spectral Centroid	The weighted average of all the frequencies in the signal.
Spectral Bandwidth	This describes how spread out the frequencies are around the dominant frequency.
Peak Frequency	The frequency with the highest power occurs in the power spectrum of a signal. It shows the strongest frequency component but may not reflect the fundamental oscillation.
Total Power	The sum of all spectral powers.





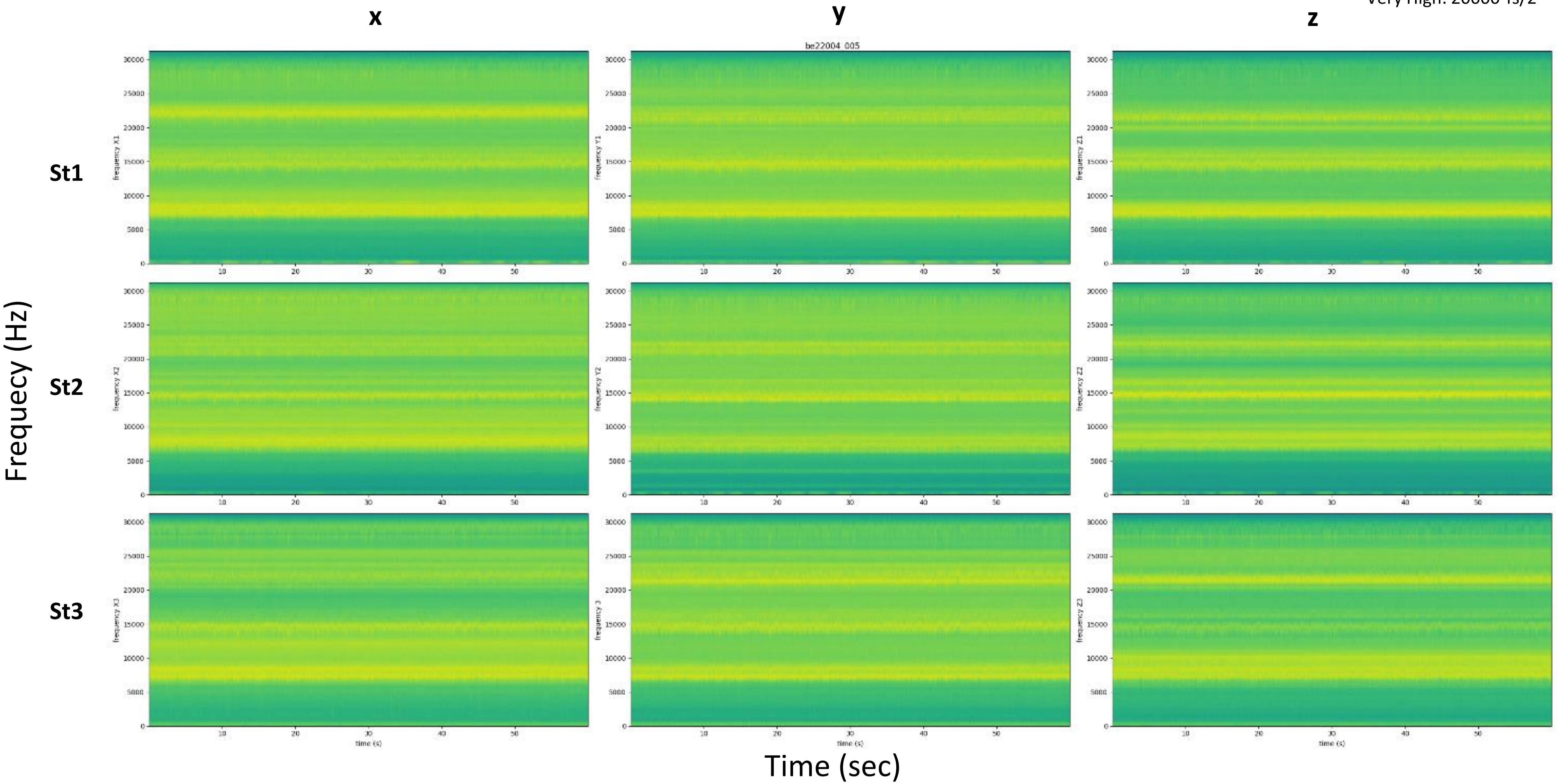




exp	1X_dominant_frequency_w1	1X_spectral_centroid_w1	1X_bandwidth_w1	1X_peak_frequency_w1	1X_total_power_w1
be22003_028	0.1	12910.00669	7953.302029	0.1	0.009560225
be22004_000	0.05	14790.51281	7412.906873	0.05	0.017422229
be22004_002	0.25	10146.06407	7352.956101	0.25	0.003893503
be22004_003	0.15	11282.34237	7857.106932	0.15	0.006286741
be22004_005	7797.9	14760.16314	6661.756712	7797.9	0.684782368

Acceleration data in frequency domain - Spectogram (exp_id: be22004_005)

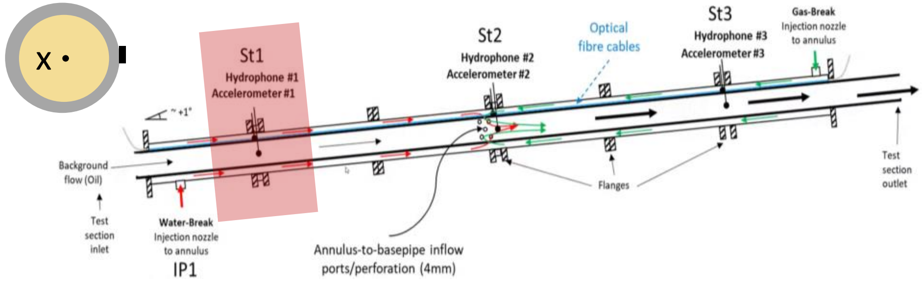
Low: 0-1000 Hz
Mid: 1000-10000 Hz
High: 10000-20000Hz
Very High: 20000-fs/2



SINTEF

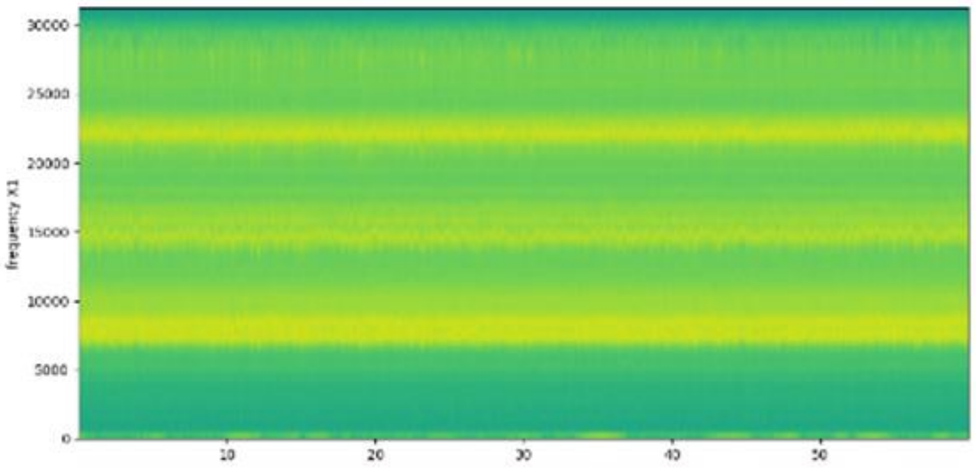
Frequency domain Features extracted from full acceleration signal

(ax @st1)



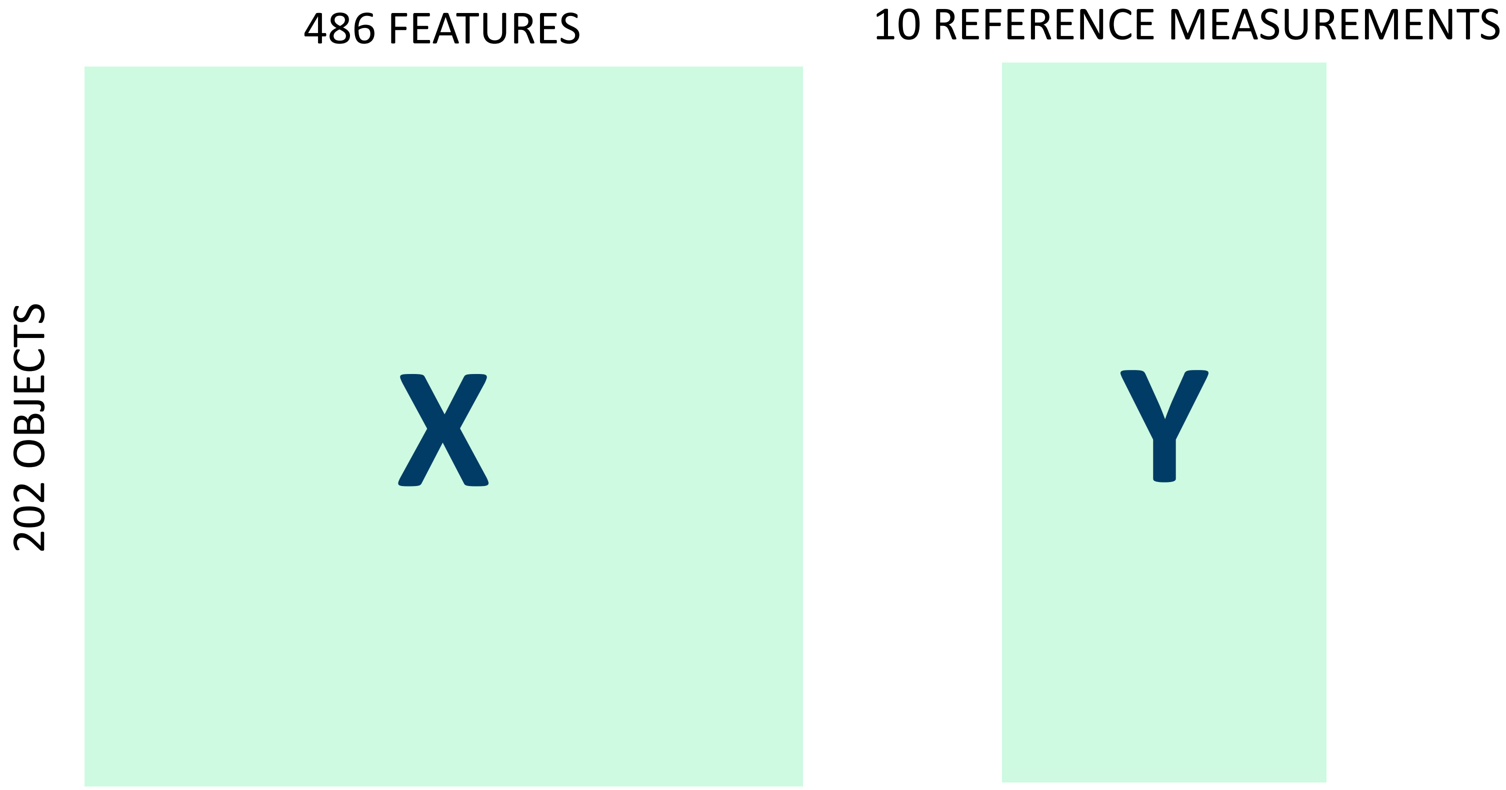
Low: 0-1000 Hz
Mid: 1000-10000 Hz
High: 10000-20000Hz
Very High: 20000-fs/2

Band power	The total power of a signal within a specific frequency range or "band." In signal processing and analysis, it measures how much energy or power is contained within a particular frequency band of a signal. PSD is often used to compute the band power.



exp	1X_band_low	1X_band_mid	1X_band_high	1X_band_very_high
be22003_028	0.001609032	0.014557812	0.008736004	0.008507113
be22004_000	0.001888022	0.014006371	0.014292896	0.020983241
be22004_002	0.002192458	0.00638898	0.000382533	0.000649816
be22004_003	0.005518025	0.009083846	0.001267716	0.002100833
be22004_005	0.000310533	1.21088707	0.271910936	0.538554306

Final data matrices

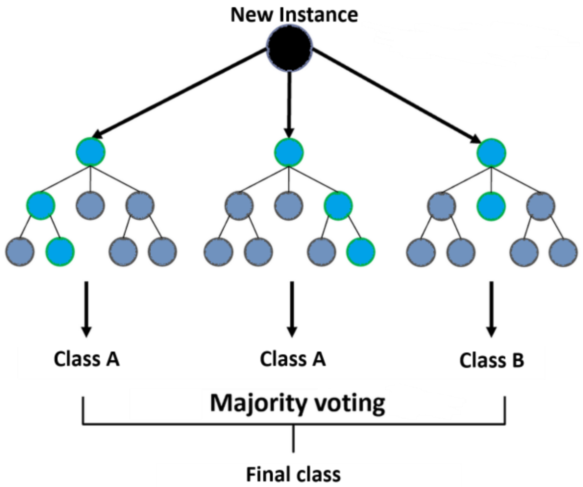


Binary Classification of water and gas breaks

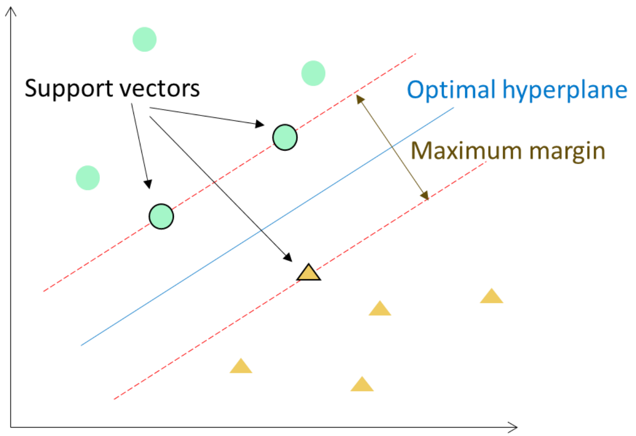
$$\text{Water break} = \frac{\text{Water mass flow rate}}{\text{Water + Oil + Gas mass flow rates}}$$

$$\text{Gas break} = \frac{\text{Gas mass flow rate}}{\text{Water + Oil + Gas mass flow rates}}$$

- Random Forest Classifier
  - Random Forest (RF) is a supervised machine learning algorithm that is constructed using decision tree algorithms.
  - Used to solve regression and classification problems
  - Utilizes ensemble learning techniques.



- Support Vector Machines
  - Support Vector Machine (SVM) is a supervised machine learning algorithm that is primarily used for classification tasks.
  - Primary objective of SVM is to identify the optimal decision boundary that maximizes the margin between these classes.



Random Forest Classification

mfrwat1

Actual \ Predicted	No Break	Break
No Break	82	14
Break	6	100

mfrwat2

Actual \ Predicted	No Break	Break
No Break	158	2
Break	15	27

mfrgas1

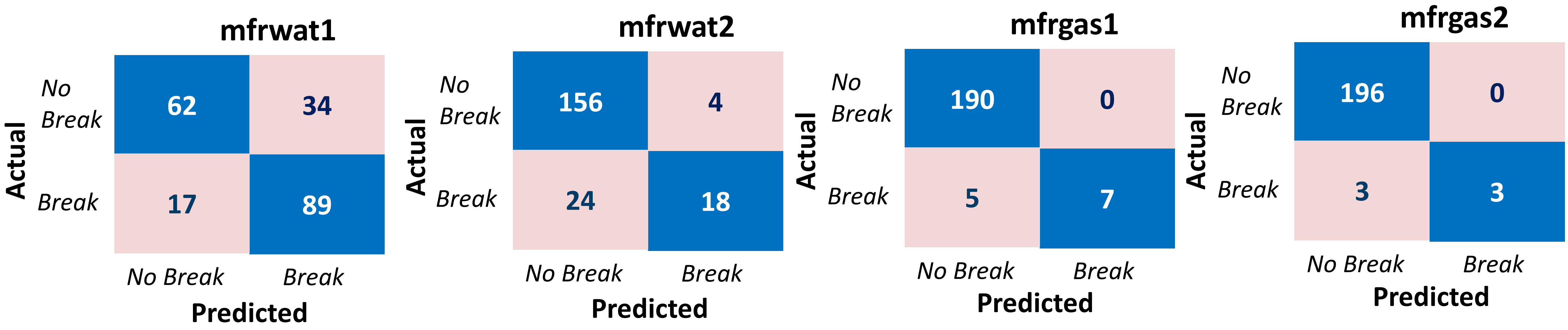
Actual \ Predicted	No Break	Break
No Break	190	0
Break	3	9

mfrgas2

Actual \ Predicted	No Break	Break
No Break	196	0
Break	4	2

Random Forest parameters	mfrwat1	mfrwat2	mfrgas1	mfrgas2
Max_depth	49	39	45	32
min_samples_split	4	5	6	5
n_estimators	107	86	80	143
Accuracy	90%	92%	98.5%	98%

Support Vector Classification

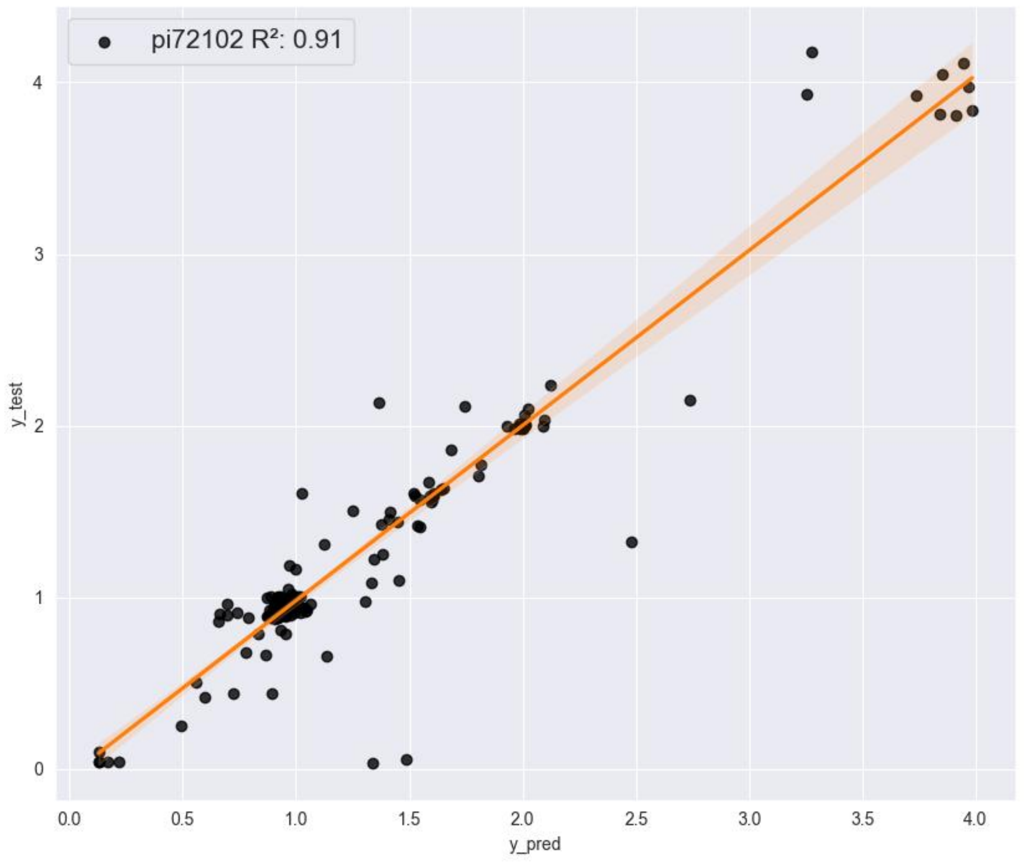


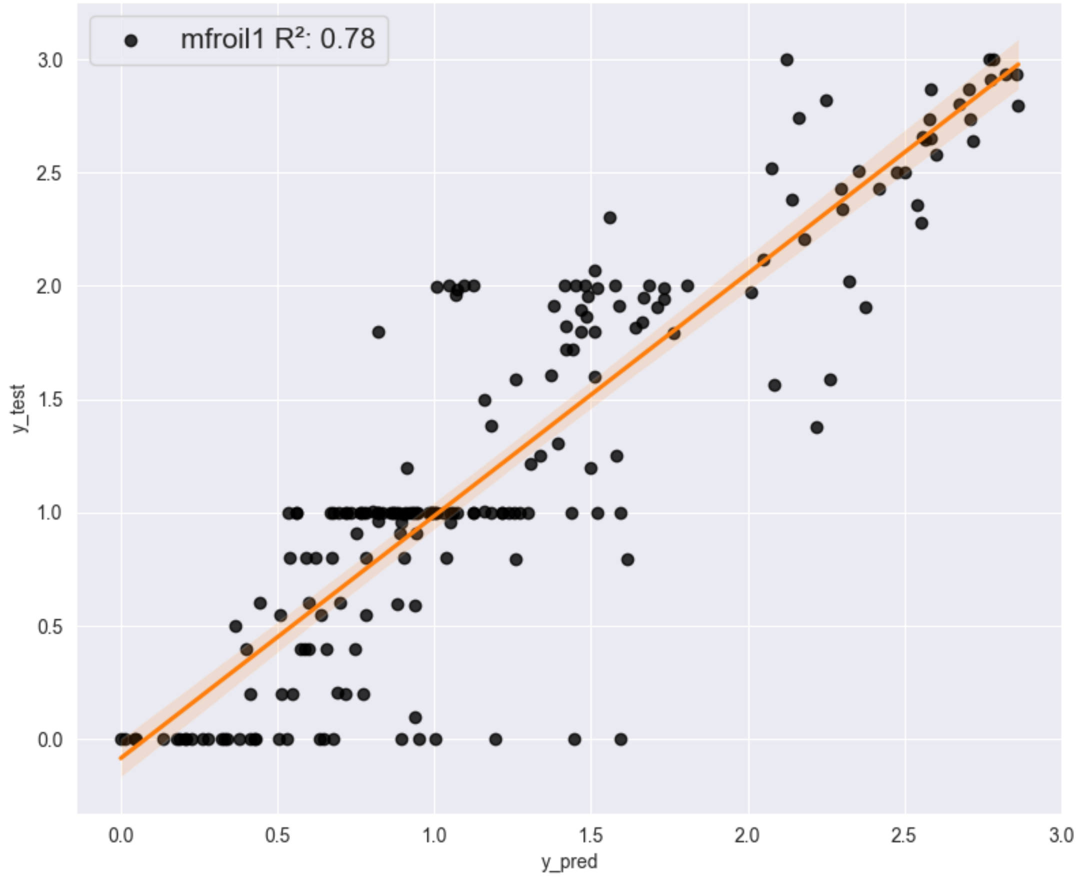
SVM parameters	mfrwat1	mfrwat2	mfrgas1	mfrgas2
C	79.506	401.43	11.974	115.34
Kernal	RBF	RBF	RBF	RBF
gamma	0.0078	0.0213	0.0022	1.6e-05
Accuracy	75%	86%	97.5%	98.5%

Prediction of reference measurements

- Gradient Boosting techniques - XGBoost
  - Gradient boosting is an ensemble technique that combines the predictions of several simpler models, usually decision trees, to make a stronger predictive model.
  - XGBoost is an advanced implementation of gradient boosting algorithm
    - optimized for speed and performance
    - efficient and scalable machine learning algorithm
    - can use in classification, regression, ranking
    - can handle large/complex datasets
    - can handle missing values

Prediction of reference data using extracted features using XGBoost





Summary

- **Efficient Data Handling:** The proposed method effectively processes large volumes of high-frequency accelerometer data.
- **Feature Extraction and Modeling:** Extracted features enable accurate classification and predictive modeling for time-series data.
- **Real-Time Applicability:** The approach is scalable for real-time operational applications.
- **Performance:** Demonstrates robust results with efficient execution and high accuracy.

Teknologi for et bedre samfunn