Supported by:

Federal Ministry for Economic Affairs and Climate Action

on the basis of a decision by the German Bundestag

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

OpenRAVE
-
Application of machine learning for error detection

RAVE Workshop 2022
03. February 2022

Nick Hansen[1], Muhammad Omer Khan[1]

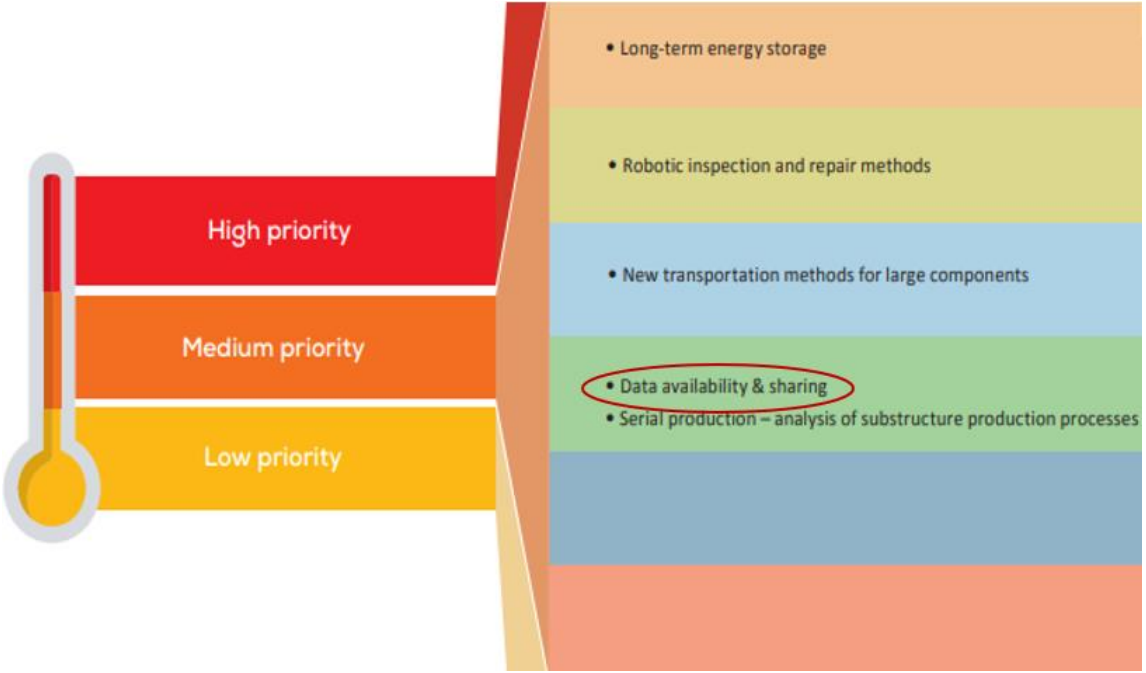
Richard Foreman[1], Thomas Neumann[1]

[1] Renewables, UL International GmbH (DE)

In collaboration with: DNV RAVE RESEARCH AT ALPHA VENTUS

Contents

Requirements for Innovation



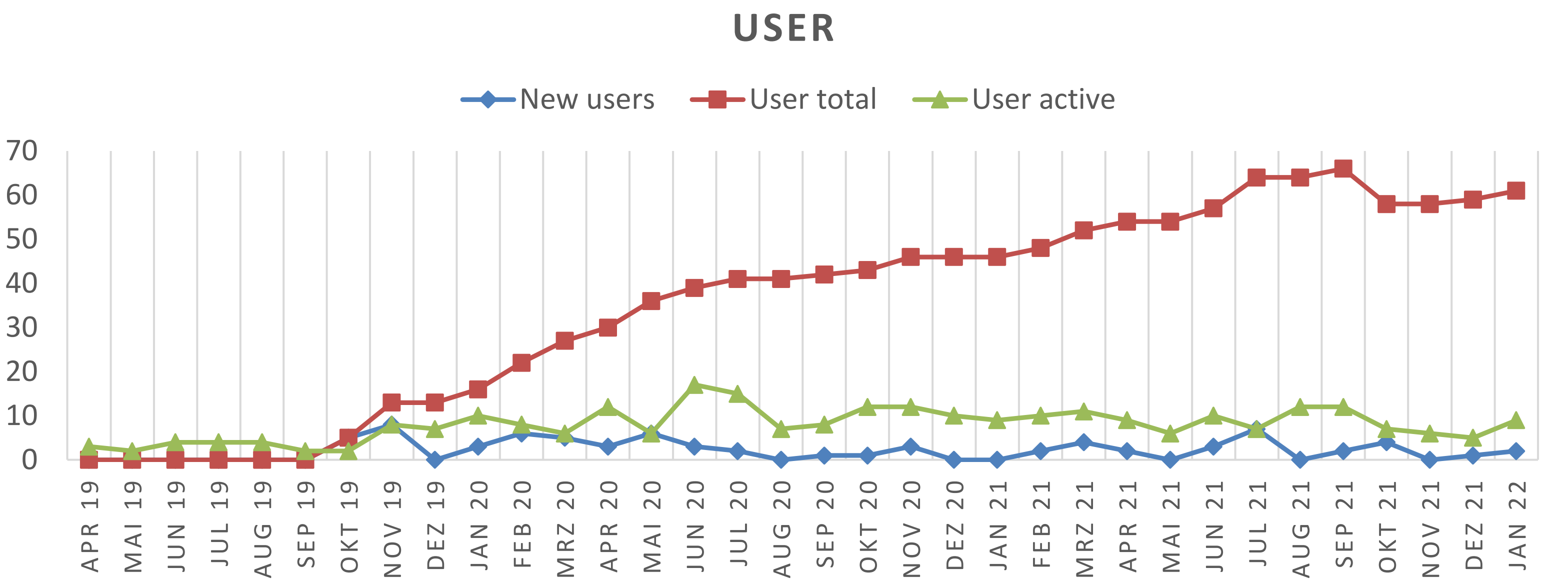
Research & Innovation priorities 2020-2027 [ETIPWind Roadmap]

Data availability
- Continuous time series since 2010

Data Quality
- Data Quality Control (DQC) implemented
- Historical data in progress

Data sharing
- Open access since 2019
- Meta data available since 2021

RAVE Database at BSH since 2019

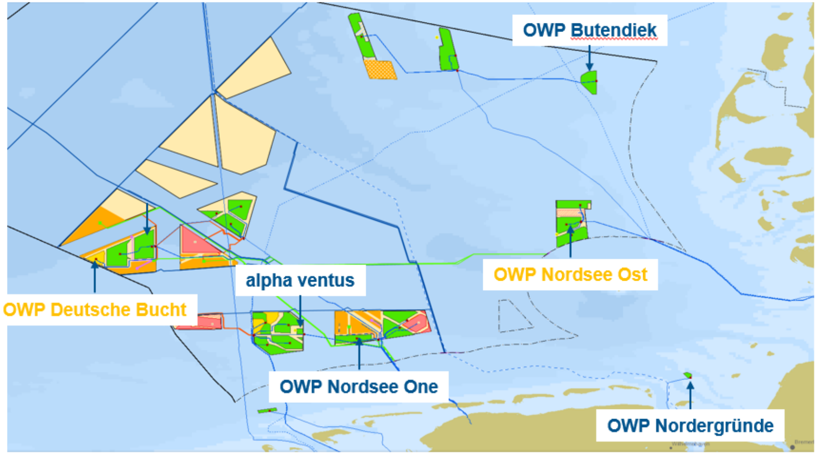


- Login via: [BSH-Login]
- Contact and Support: rave-forschungsarchiv@bsh.de

Meteorological and oceanographic measurements

Oceanographic measurements at **alpha ventus** and the **German Bight**

- Waves (Buoy, Radar and ADCP)
- Currents
- Water Level



Meteorological mast FINO 1 (separate project by BSH)

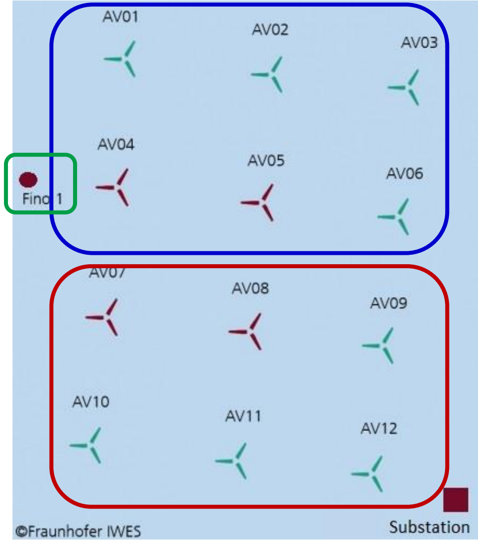
- Wind, temperature, humidity
- Lidar
- Waves
- Currents
- CTD
- …..

RAVE RESEARCH AT ALPHA VENTUS

Layout of alpha ventus

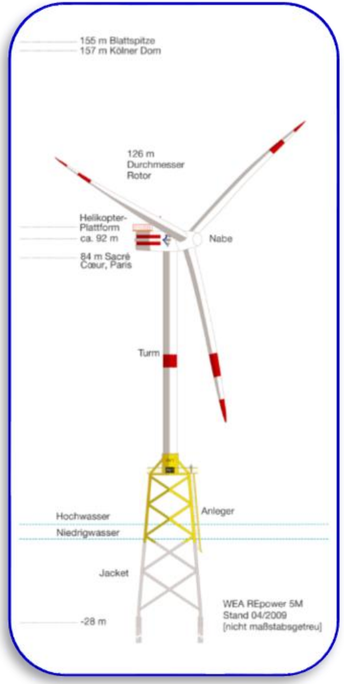
BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

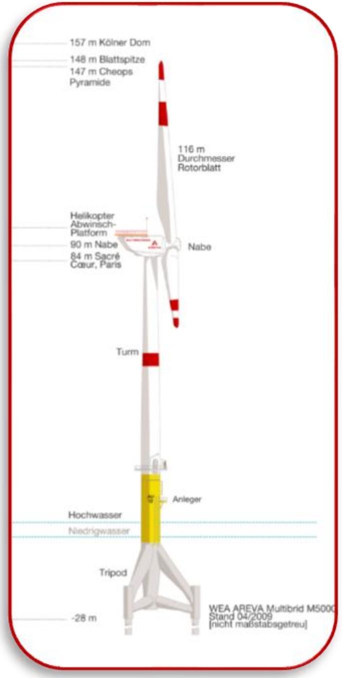
Senvion 5M



AREVA Wind M5000

©Fraunhofer IWES



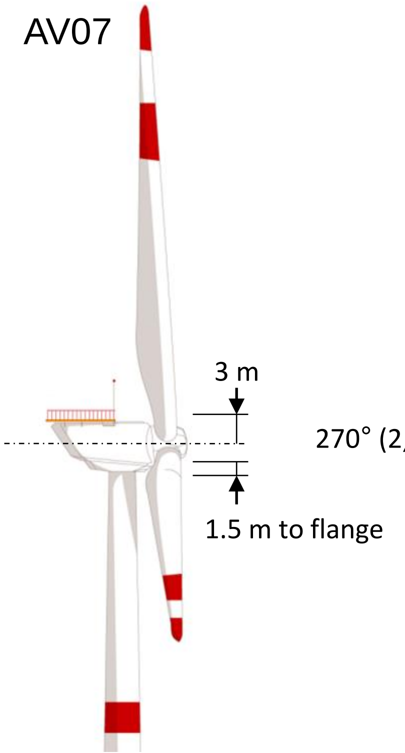


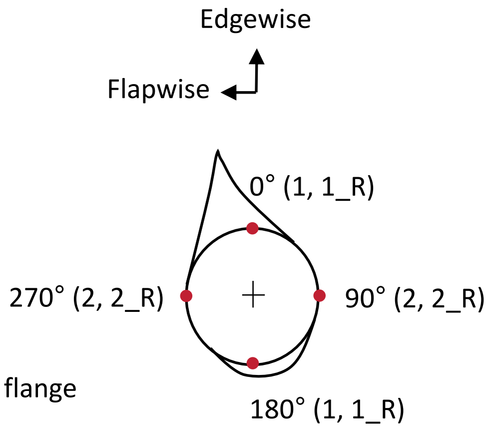
© https://www.alpha-ventus.de/technik

- strain gauges
- accelerometers
- acoustic sensors
- hydrographic sensors
- met data (sonic, lidar)
- sonars
- water pressure sensors
- SCADA
- corrosion
- video cam, radar

DNV
RAVE RESEARCH AT ALPHA VENTUS

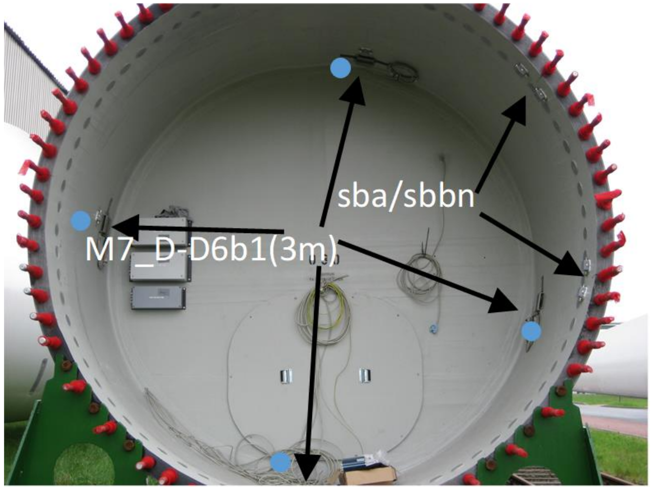
Example Case: M7_D-D6b1u(3m)1,2



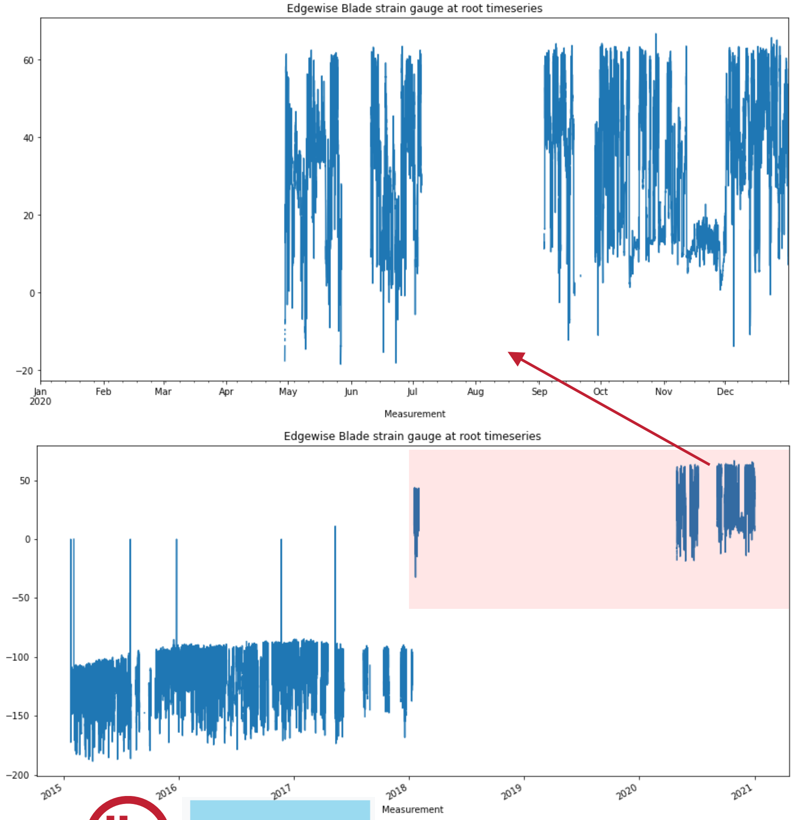


M7_D-D6b1u(3m)1,2
M7_D-D6b1u(3m)1,2_R

M7 Blade 1 Root



What is error in measured data?



In RAVE database, systematic errors, errors resulting from measuring devices [1], can occur, specifically offset and amplification and drift errors on long-term series.

→ RAVE Data Quality Control (DQC) [2] is unable to detect those errors

→ No assessment of long-term consistency

Motivation

Increasing the integrity of long-term data

Objective

→ **Flag systematic errors to support DQC**

Various approaches

Detection of errors or faults

Change point detection [3]

- Change points
- States
- Features of system states

Principal component analysis

- Fault or anomaly detection technique
- Loss of signal information [4]

Regression residual analysis

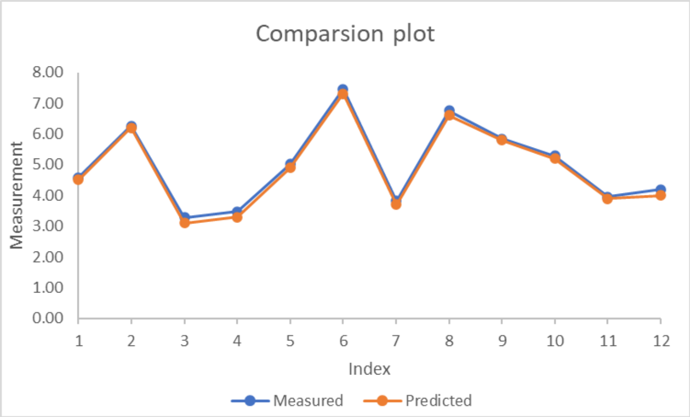
- Comparison of measurement and prediction
- Fault detection
- Generator bearing fault detection [5]

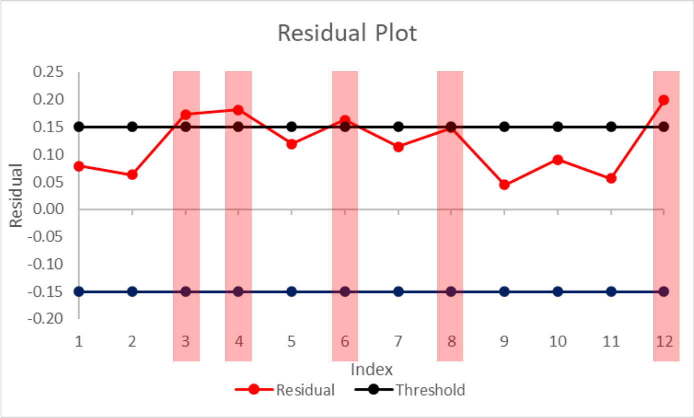
Methodology - Baseline approach

Considering the occurrence of systematic errors in data
→ Fault in system
→ Change in system

Following the machine learning approach by Orozco [5]
→ Develop a machine learning (ML) model
→ Predict sensor behavior
→ Comparing prediction and measurement
→ Defining appropriate error threshold and window
→ Flag systematic errors

Advantages, why ML is better
→ Flagging can be based on smaller window
→ Predicted behavior from ML model
Possibly can be used for error correction
Possibly error can be quantified to a certain accuracy





Methodology

→ **Data received from RAVE database (BSH)**
→ 10-min aggregated

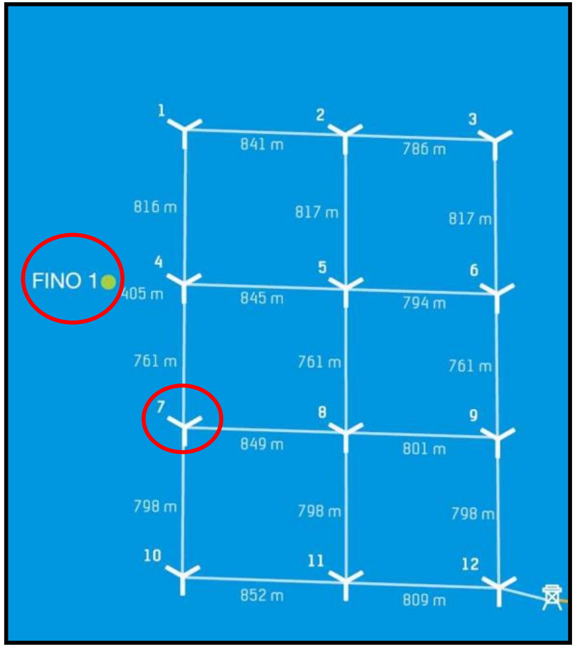
→ **Sensors selection and data cleaning***

→ **Data sets preparation***

→ **Model input space***
→ Feature selection and feature engineering

→ **Model selection and development**
→ Neural Network and sensitivity analysis

→ **Flagging criteria**
→ Absolute error thresholds in 1 hour window
→ >= 5 thresholds, Flag = 1 [7]
→ >= 3 thresholds <5, Uncertain flag = -1
→ < 3 thresholds, No flag = 0



Alpha Ventus wind farm layout [6]

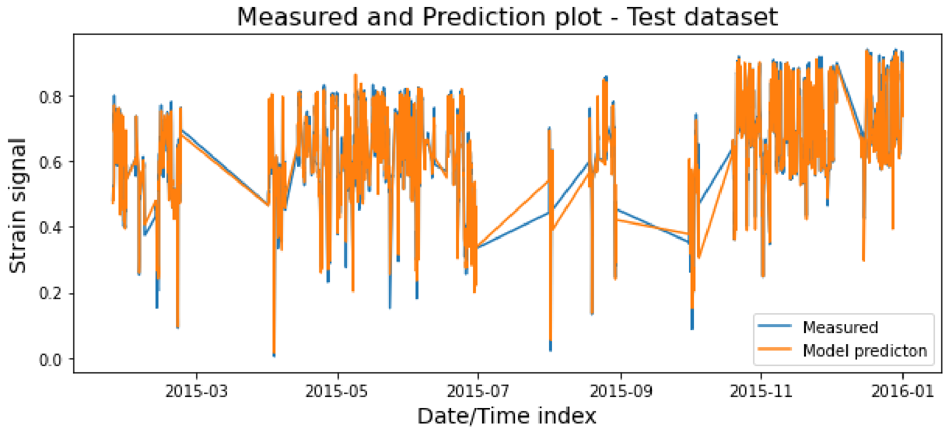
*Further details in Appendix A

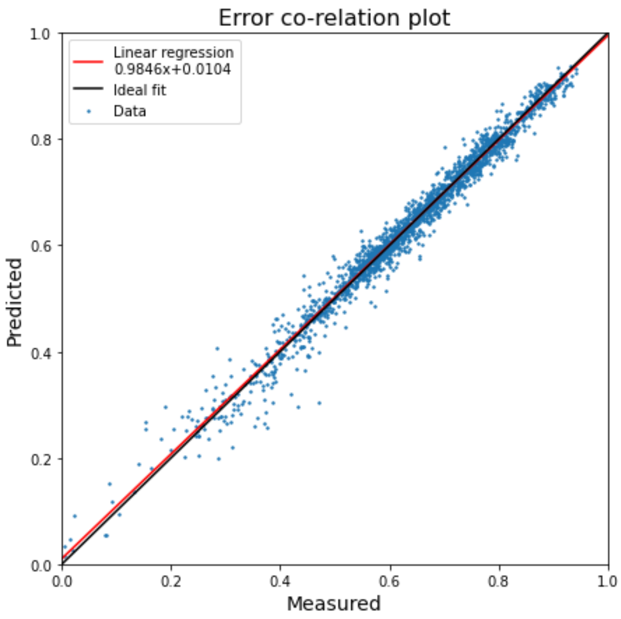
Results

Root edge-wise strain gauge

Neural network (NN) **V1-E**

Test dataset (Normalized)





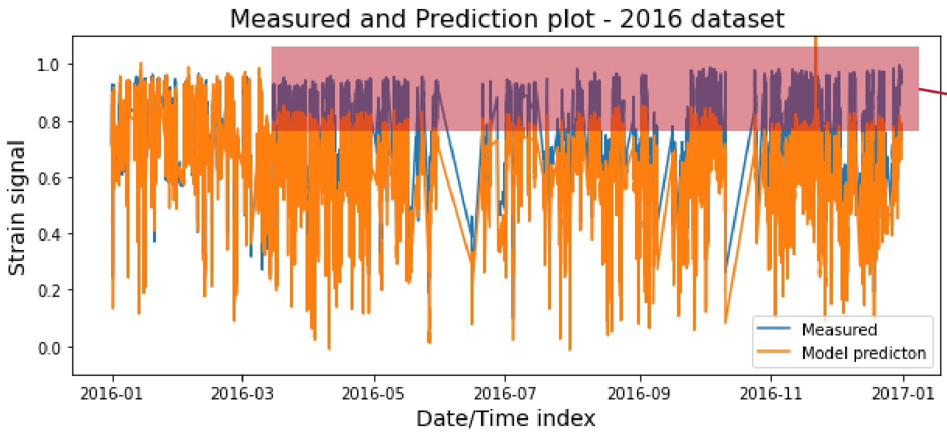
NN	MSE	RMSE	MAE	MAPE	R^2
V1-E	~0.0004	~0.0215	~0.0148	~3.46%	~0.9783

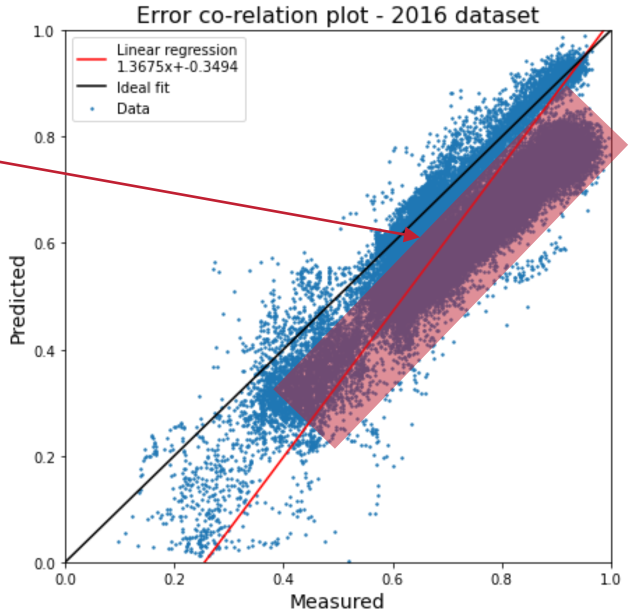
RAVE RESEARCH AT ALPHA VENTUS

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

Results

Prediction for year 2016 data with NN **V1-E**

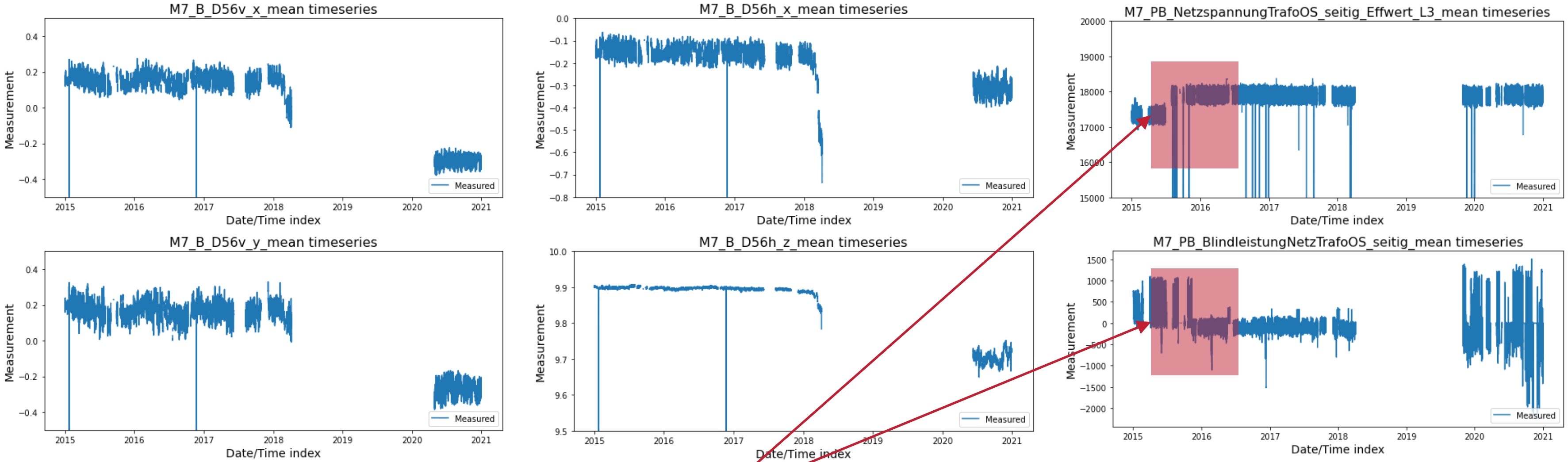


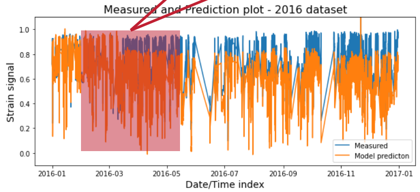


NN	Data set	MSE	RMSE	MAE	MAPE	R^2
V1-E	2016	~3.0844	~1.7562	~0.1123	~15.49%	~-134.53
	Test	~0.0004	~0.0215	~0.0148	~3.46%	~0.9783

Inconsistent features

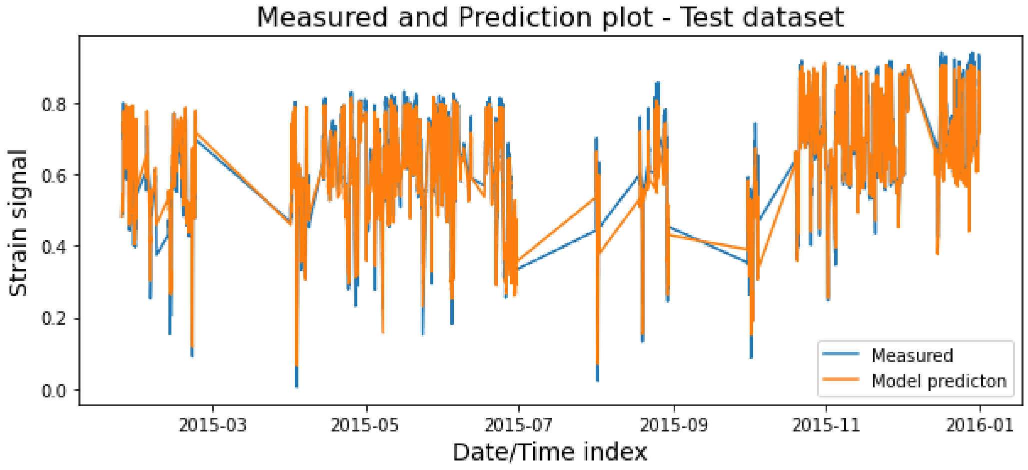
Manually checked input features and removed inconsistent measurement sensors

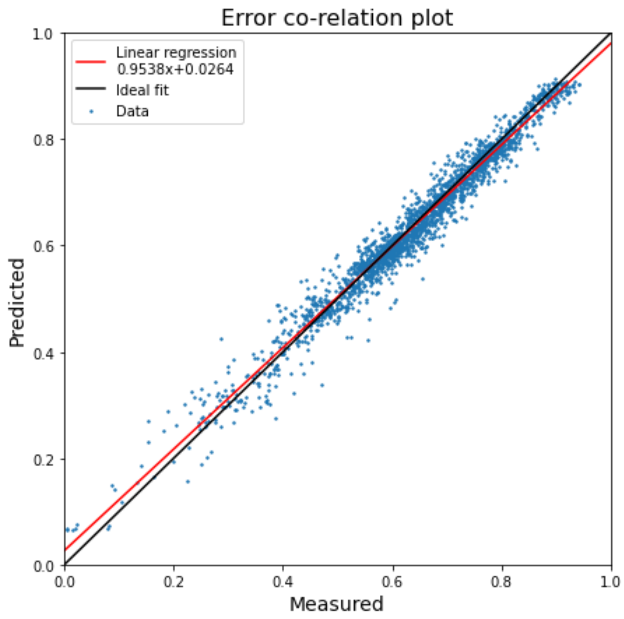




Results

Retrained NN V1-E → NN **V1.1-E**

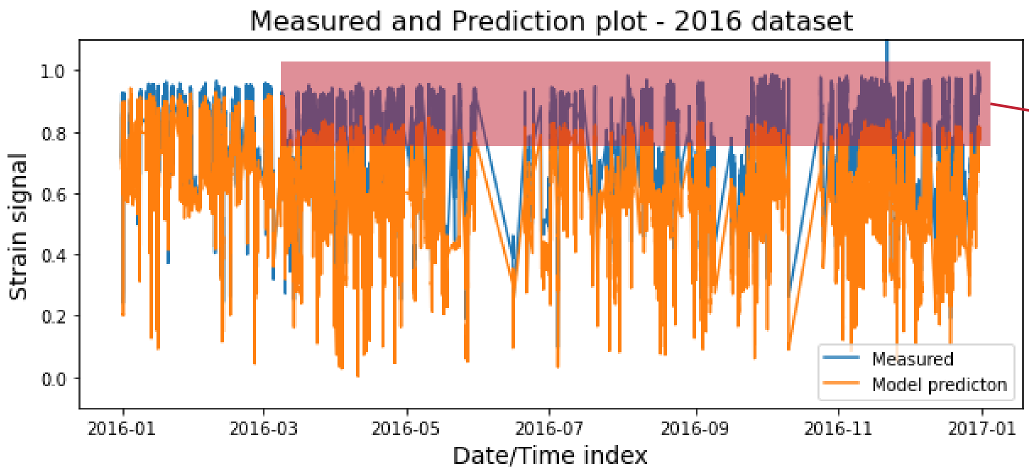




NN	Data set	MSE	RMSE	MAE	MAPE	R^2
V1.1-E	Test	~0.0007	~0.0270	~0.0198	~5.23%	~0.9659
V1-E	Test	~0.0004	~0.0215	~0.0148	~3.46%	~0.9783

Results

Prediction for year 2016 data with NN **V1.1-E**

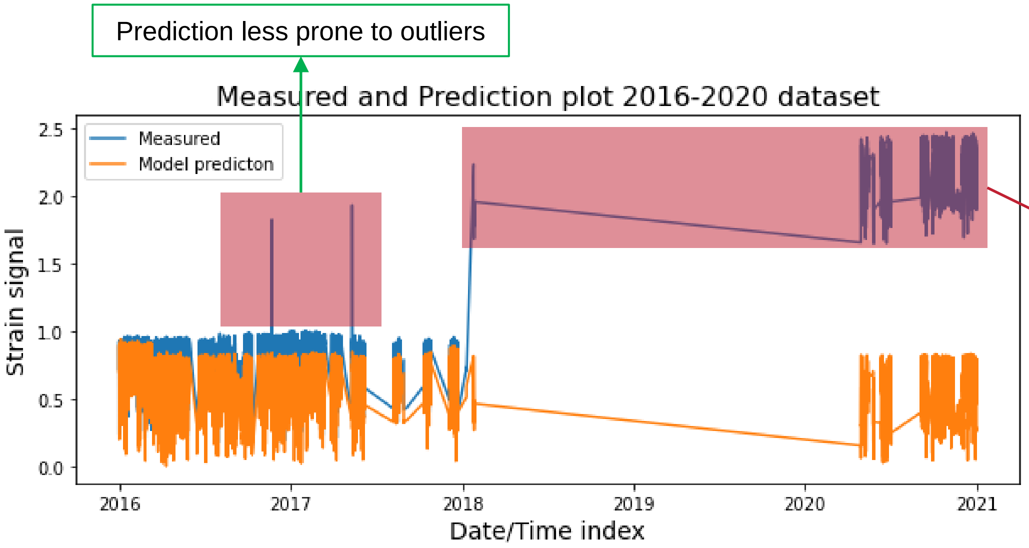


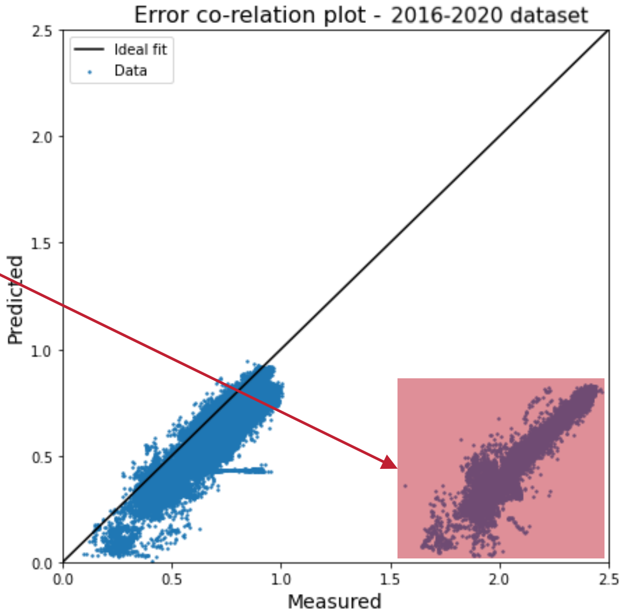
NN	Data set	MSE	RMSE	MAE	MAPE	R^2
V1.1-E	2016	~0.0159	~0.1262	~0.1111	~16.12%	~0.2992
V1-E	2016	~3.0844	~1.7562	~0.1123	~15.49%	~-134.53

RAVE
RESEARCH AT ALPHA VENTUS

Results

Prediction for year 2016 to 2020 data with NN **V1.1-E**

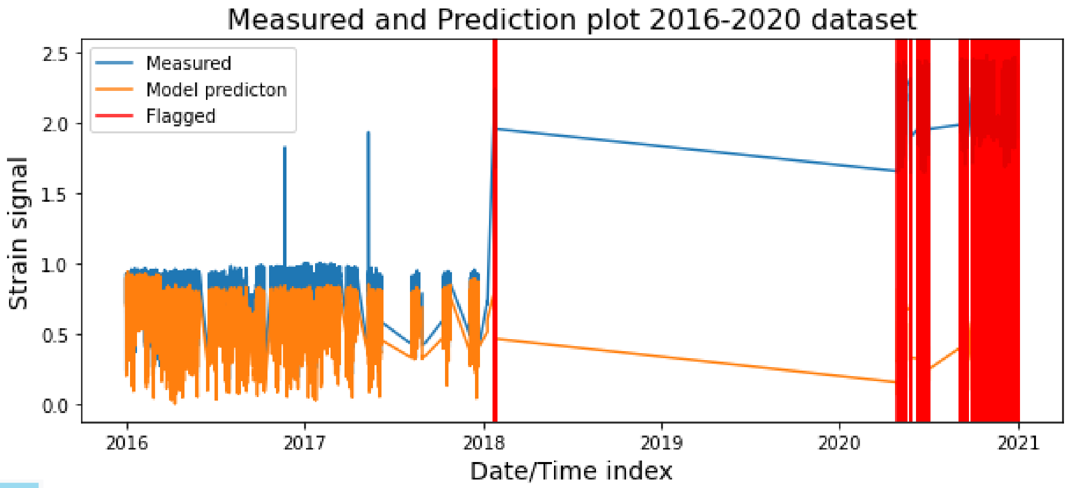




Error flagging - Blade root edge-wise

- → Either there are still inconsistent input features, or the target sensor is drifting (2016)
- → Later observed offset in measurement (2018 – 2020) is an error because of fault or change in system
  - → Possibility to flag with NN **V1.1-E** predictions

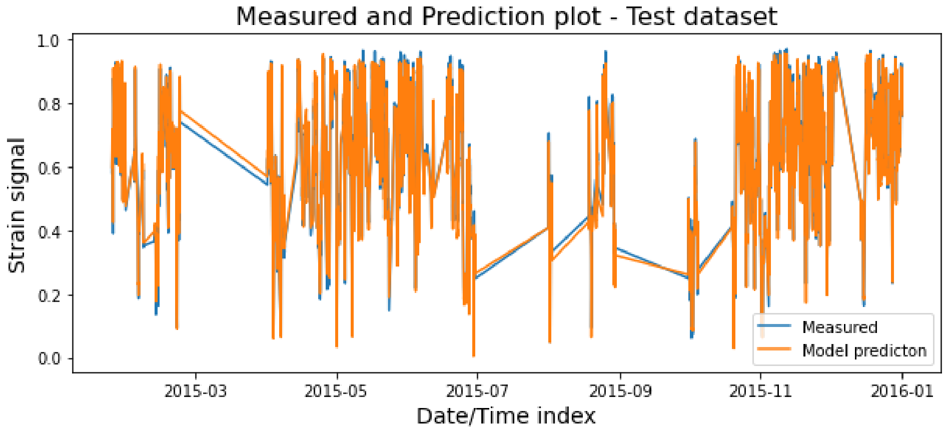
Flagging criteria → Absolute error threshold >= +-0.5

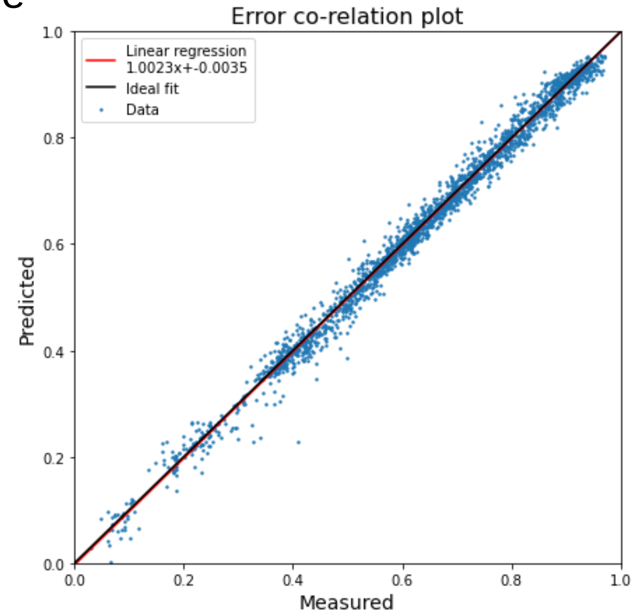


Results

Transferring: Blade root edge-wise → Blade root flap-wise

NN **V1.1-E** → NN **V1.1-F**

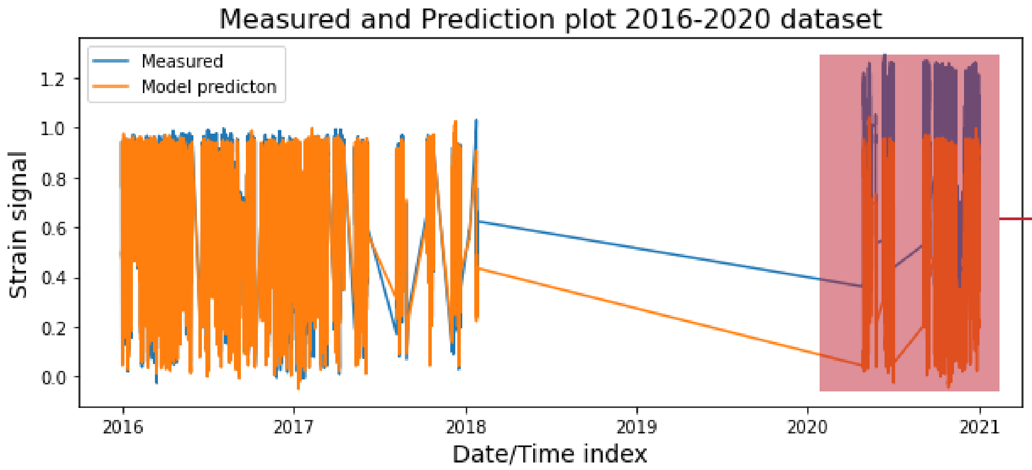




NN	Data set	MSE	RMSE	MAE	MAPE	R^2
V1.1-F	Test	~0.0004	~0.0204	~0.0152	~3.07%	~0.9898
V1.1-E	Test	~0.0007	~0.0270	~0.0198	~5.23%	~0.9659

Results

Prediction for year 2016 to 2020 data with NN **V1.1-F**

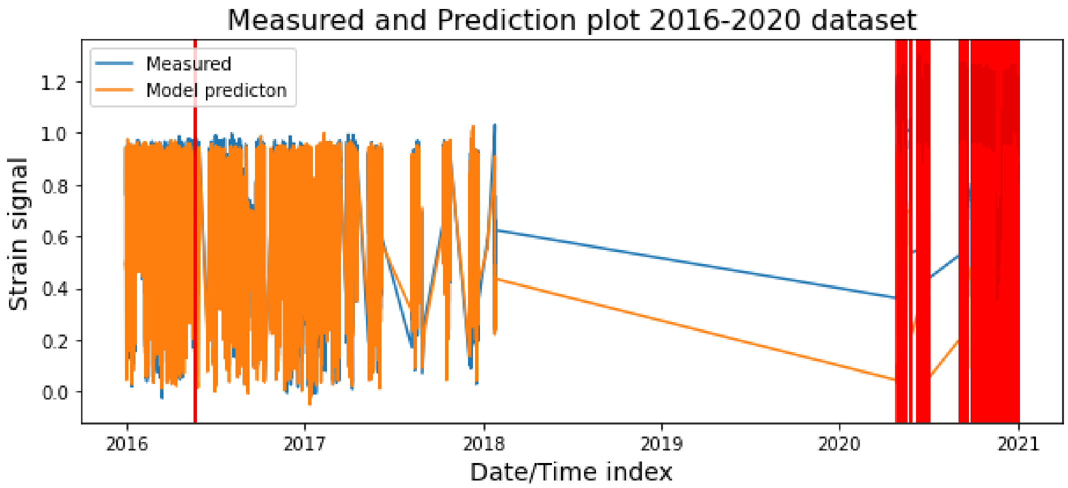


Error flagging - Blade root flap-wise

→ Observed offset in measurement (2020) is also an error because of fault or change in system

→ Possibility to flag with with NN **V1.1-F** predictions

Flagging criteria → Absolute error threshold >= +-0.3



Conclusion

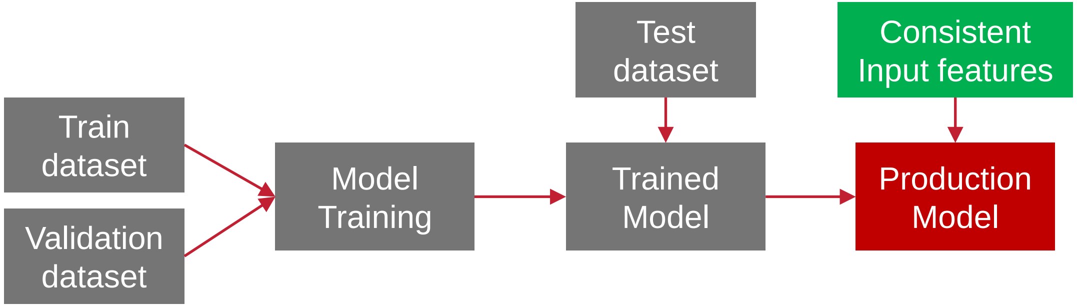
- Machine learning can be used to predict structural sensors behavior
- Transferability of neural network architecture is possible from one sensor to the other
- Preliminary results show the possibility of detection of systematic or measurement error
- Machine Learning shows the potential of increasing the integrity of long-term measurement series
- Absolute necessity for a good self-consistency test for neural network inputs

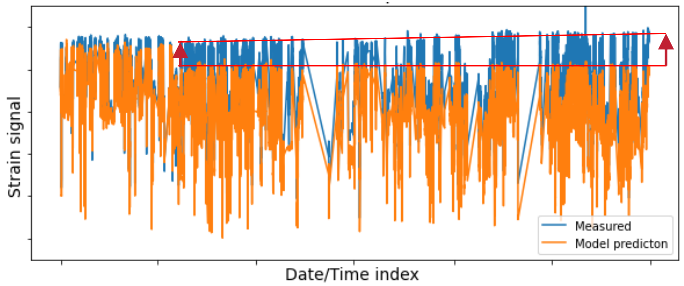
Further work

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

- Lower error thresholds can be set to flag even lower systematic error if inputs remain consistent

- There is possibility to flag and quantify sensor drifts with self consistent inputs





RAVE RESEARCH AT ALPHA VENTUS

Thank you

Feedback and questions?

<u>Contact Details:</u>

Nick Hansen
Tel.: +49 (0)4421 4808-837
E-Mail: Nick.Hansen@ul.com

Muhammad Omer Khan
E-Mail: Omer.Khan@ul.com

Projektträger Jülich
Forschungszentrum Jülich

Supported by:

Federal Ministry
for Economic Affairs
and Climate Action

on the basis of a decision
by the German Bundestag

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

RAVE
RESEARCH AT ALPHA VENTUS

References

[1] E. Luna, “Measurements & Uncertainties,” DeAnza College, Accessed: Jan. 20, 2022. [Online]. Available: www.deanza.edu/faculty/lunaeduardo/documents/MeasurementsUncertainties2A.pdf.

[2] “Quality Control of RAVE Measurements from AV00, AV04, AV05, AV07-AV12 and FINO1,” Bundesamt für Seeschifffahrt und Hydrographie, Oct. 2020. Accessed: Nov. 25, 2021. [Online]. Available: https://www.bsh.de/DE/PUBLIKATIONEN/_Anlagen/Downloads/Projekte/RAVE/RAVE-quality-control-of-RAVE.pdf.

[3] S. Aminikhanghahi and D. J. Cook, “A survey of methods for time series change point detection,” Knowledge and Information Systems, vol. 51, no. 2, pp. 339–367, Sep. 2016, doi: 10.1007/s10115-016-0987-z.

[4] B. C. Geiger and G. Kubin, “Relative information loss in the PCA,” IEEE Xplore, pp. 562–566, Sep. 2012, doi: 10.1109/ITW.2012.6404738.

[5] R. Orozco, S. Sheng, and C. Phillips, “Diagnostic Models for Wind Turbine Gearbox Components Using SCADA Time Series Data,” presented at the IEEE International Conference on Prognostics and Health Management, Seattle, Washington, Jun. 2018, Accessed: Nov. 25, 2021. [Online]. Available: www.nrel.gov/docs/fy18osti/71166.pdf.

[6] A. Venu, N. Hansen, and M. Schmager, “Data Quality Management in the RAVE project, introducing machine learning to the process,” Jan. 2021, Accessed: Jan. 26, 2022. [Online]. Available: www.rave-offshore.de/files/downloads/konferenz/Workshop-2021/Ses1_4_RAVE2021_MachineLearning_BSH_DNV-GL_UL.pdf.

Appendix A – Data preparation

Sensors Selection
SCADA, Environmental [mean, min, max, std.]
2 Nacelle, 2 Tower top accelerometers [mean, min, max, std.]

Data cleaning
→ Total features 265
→ Datapoints 52560

1. Sensors range test

2. Timestamps drop
→ Target signal missing, or range test failed
→ Remaining datapoints 37887

3. Removed features
→ Availability below 50%
→ Remaining features 221

4. Timestamps drop
→ 50% Features missing in each timestamp
→ Remaining datapoints 24749

5. Removed features
→ Availability below 80%
→ Standard deviation = ~0
→ Remaining features 183

6. Timestamps drop
→ Any missing feature
→ Remaining datapoints 19169

Data sets preparation
Sample and split
→ Train 70%, Validation ~20%, Test ~10%
Data scaling
→ Normalized train dataset
→ Validation and test dataset transformed to normalized scale

Appendix A – Data preparation

Model input space

Feature selection

→ Embedded methods [1]

→ Lasso CV, Random forest, Extra trees, Extra tree, and Decision tree

→ Selection criteria

→ Cumulative sum importance of sensors determined

→ Sensors cumulative sum 76% to 100% importance with increment of 5% importance

→ Transformed to Principal components (All components used)

→ Best metrics from basic neural network (Validation & Test dataset)

→ Selected

→ Lasso CV

→ Sensors cumulative sum importance < 96%

→ 26 Sensors, 53 features

Feature engineering

→ Principal component analysis (PCA) [2]

→ Eigen vectors and eigen values

→ Correlated to uncorrelated features

→ Train data set: Fit and transformed

→ Validation and test dataset: Transformed

→ 99.5% Variance components

1. S. Biswas, M. Bordoloi, and B. Purkayastha, "Review on Feature Selection and Classification using Neuro-Fuzzy Approaches," International Journal of Applied Evolutionary Computation, vol. 7, no. 4, pp. 28–44, Oct. 2016, doi: 10.4018/ijaec.2016100102
2. S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, "An Overview of Principal Component Analysis," Journal of Signal and Information Processing, vol. 04, no. 03, pp. 173–175, 2013, doi: 10.4236/jsip.2013.43b031.

Selected features

Sensor	Features
M7_PB_WirkleistungGenerator	Min, Max
M7_PB_BlindleistungGenerator	Mean, max
M7_PB_FrequenzNetz	Max
M7_PB_Generatorstrom_Effwert_L1	Mean, min
M7_PB_Generatorstrom_Effwert_L2	Min
M7_PB_Generatorstrom_Effwert_L3	Mean, min
M7_PB_Generatorspannung_Effwert_L1_L2	Mean, min
M7_PB_NetzspannungTrafoOS_seitig_Effwert_L3	Min, max, std
M7_PB_NetzstromTrafoOS_seitig_Effwert_L1	Mean, min, max
M7_PB_BlindleistungNetzTrafoOS_seitig	Max, std
M7_PB_WEAStatus	Mean, max
M7_PB_Generatordrehzahl	Min
M7_PB_RotorPosition	Mean, min, std
M7_PB_PitchwinkelBlatt2	Mean, max
M7_PB_Gondelposition	Mean, max

Sensor	Features
M7_PB_BeschleunigungGondellaengs	Std
M7_B_D56v_x	Mean, min, max, std
M7_B_D56v_y	Mean, max
M7_B_D56h_x	Mean, std
M7_B_D56h_z	Mean, max, std
M7_PB_Windrichtung1relativ	Min, max
M7_PB_Windrichtung2relativ	Min
M7_PB_Windgeschwindigkeit1	Mean, max
M7_PB_Windgeschwindigkeit2	Mean, min, std
F1_dir_90_	Mean, std
F1_v_50__mast_corrected	Mean