DNV·GL

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

Data Quality Management in the RAVE-Project

Introducing Machine Learning to the Process

RAVE Workshop 2021
28. Januar 2021, Hamburg

BSH, DNVGL, UL International

Agenda

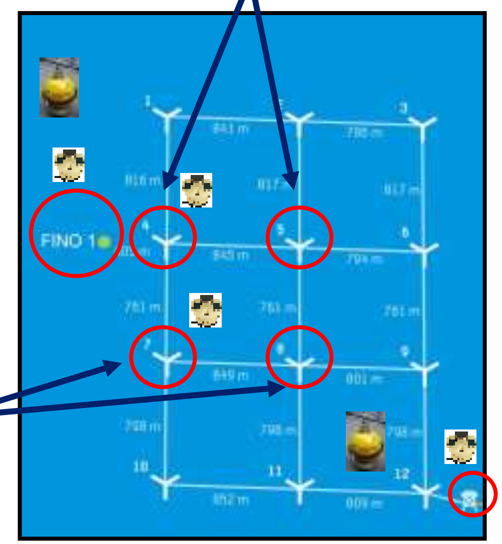
- Introduction
- Limitations/Challenges of ADQC
- Can this be solved with ML model?
- Feature Selection & Data Compression
- Case Study – First results
- Accuracy and Performance of the model
- Other applications
- Future works

Introduction

- The research initiative RAVE carries out research and development work on the offshore test field alpha ventus
- RAVE is funded by the Federal Ministry for Economic Affairs and Energy (BMWi) and coordinated by the Fraunhofer Institute for Wind Energy Systems (IWES)
- In more than 30 research projects, more than 60 partners from science and industry have been working on a wide range of research questions since 2008
- The financial support from the BMWi so far amounted to more than 50 million euros

Wind Farm Outlook

- Commission Date: 2009
- 45 Km North of Borkum
- 30 m water depth
- 12 Wind turbines
  - 6 AREVA WIND M5000
  - 6 Senvion 5M



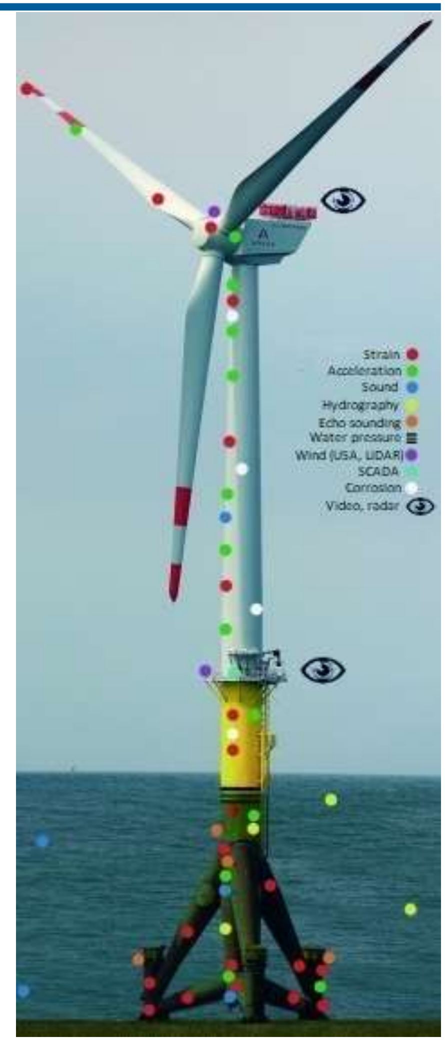
Introduction

<u>Available Measurements</u>

- Controller signals
- Accelerometers and multiple strain gauges at tower, blades and support structure
- Environmental measurements (Atmosphere, Wind and Sea State)
- Other critical structural measurements
- Other electrical signals

<u>**Continuous time series data since 2010**</u>

- The data are stored and published via RAVE-Database
- The RAVE-Database is developed, hosted and administrated by BSH,
  - **Most importantly open to the public**
    → <u>*serviceportal.bsh.de*</u>

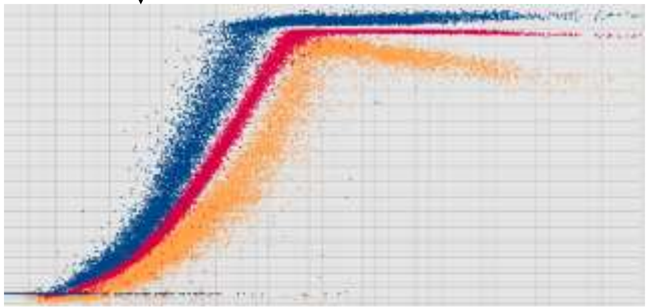


BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

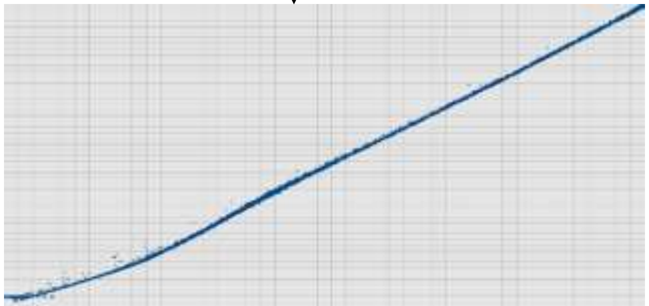
Introduction - Recap

Standard data quality control of measurement data

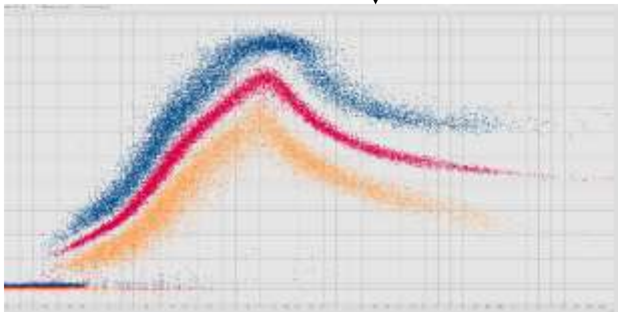
High frequency Raw data signals → Calibrated data signals → Descriptive statistics → Manual/visual check of statistical data → Flagging abnormal data signals



Wind Speed vs Electrical Power

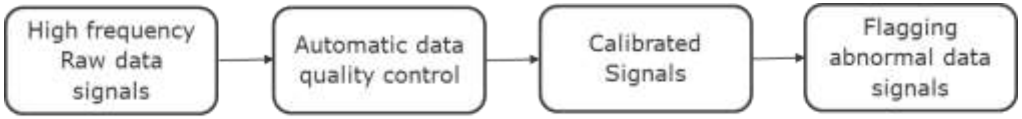


Main shaft torsion vs Electrical Power



Tower tilt moment vs Electrical Power

Introduction – Automatic Data Quality Control



Detailed Flag

Flags : 001000/1 ← Master Flag

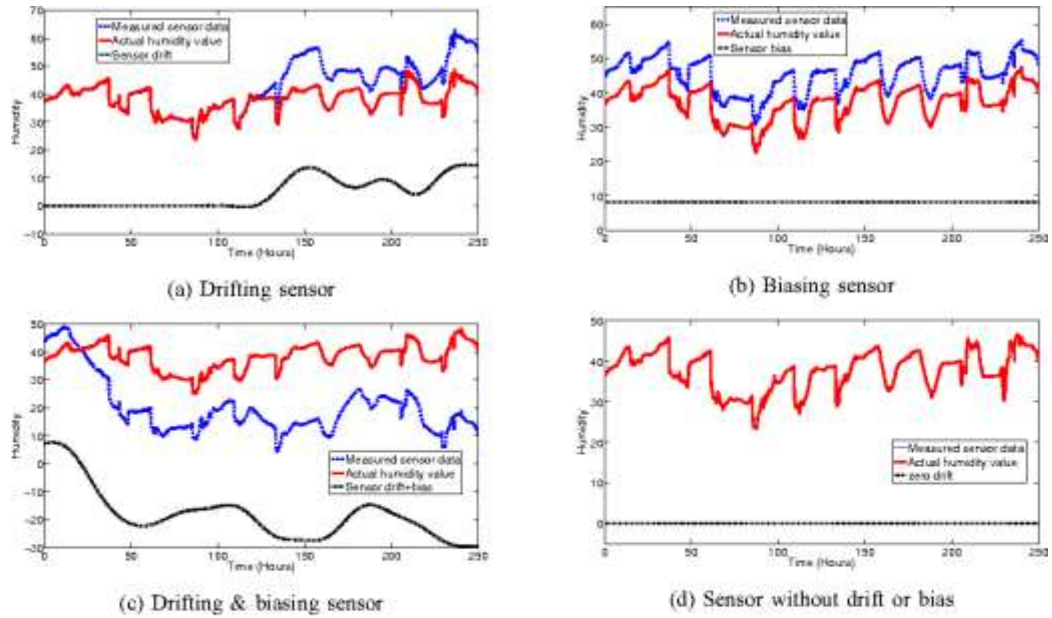
Objective

- Control the data collected from RAVE wind farm
- Plausibility check on raw signals (0.2 to 50 HZ signals)
- Automating the control and flagging process
- Independent to sensor and measurement system
- Minimal input parameters (Robust model)
- Save time and operational cost
- High quality data for future applications

Position	Test Type	Meaning	Thresholds	Description
1	Length	Reduced data length	N_{crit}%	Data of length of some value N_{crit} deviating from N 100%
2	Flat Line	Constant Signal	N/A	All values the same (e.g. bad if sensor is strain gauge, Ok/Check if machine data)
3	Flat Line	Partially Constant	t_{crit}	Constant values for a period of > t_{crit} seconds (e.g. signal dropouts)
4	Pre-defined Limits	Measurement Range	$\sum(x_i > x_{crit}) > 0$	At least one value outside the measurement range (e.g. ±10 V)
5	Spike	Spike events exceeded	n_{crit}	Number of spikes found in signal exceeds critical value.
6	Spike	Low Correlation	r_{crit}	Despiked signal poorly correlated with uncorrected signal.
7	Visual/ Qualitative	Qualitative assessment	N/A	Data assessed manually (e.g. poor correlation with wind speed).
8–16	-	*- Spare -*	-	Further tests included here.

Limitations/Challenges of ADQC

- Detects only 70% of the commonly occurring events
- Time & environmental sensitive events are not detected
- Not using the historically available cleaned data
- No data filling/replacement method available
- No additional advantages



(a) Drifting sensor

(b) Biasing sensor

(c) Drifting & biasing sensor

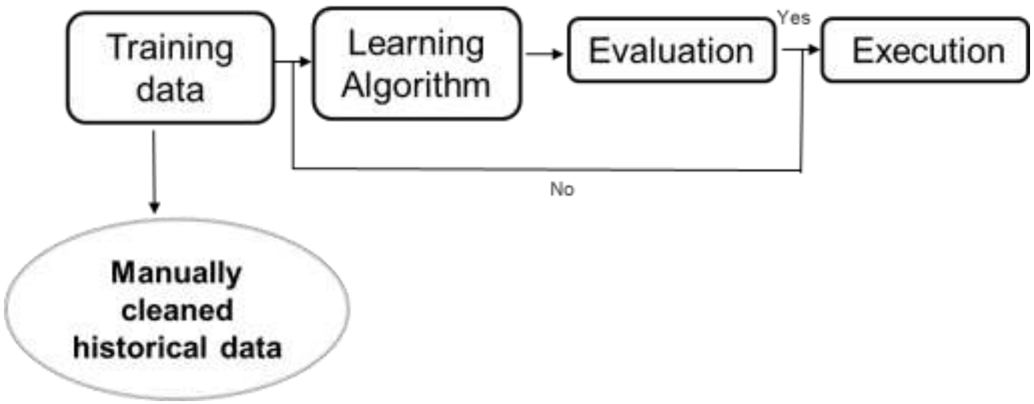
(d) Sensor without drift or bias

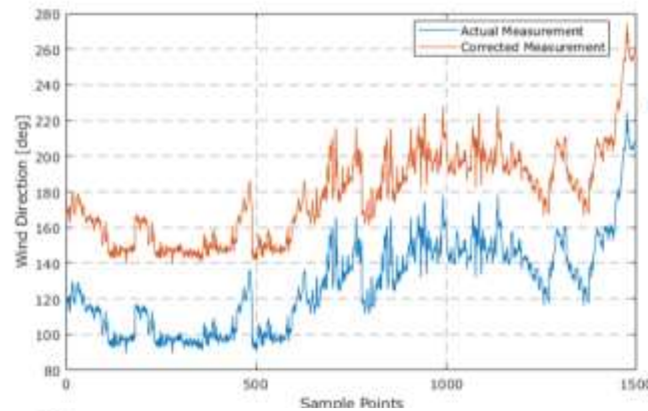
ADQC Output – **000000/0**

No Events Found

Can this be solved using Machine Learning?

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

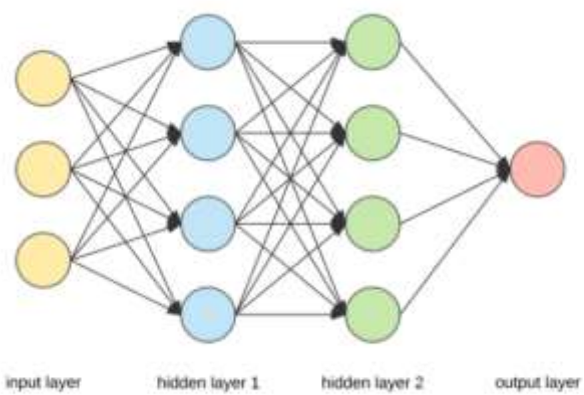


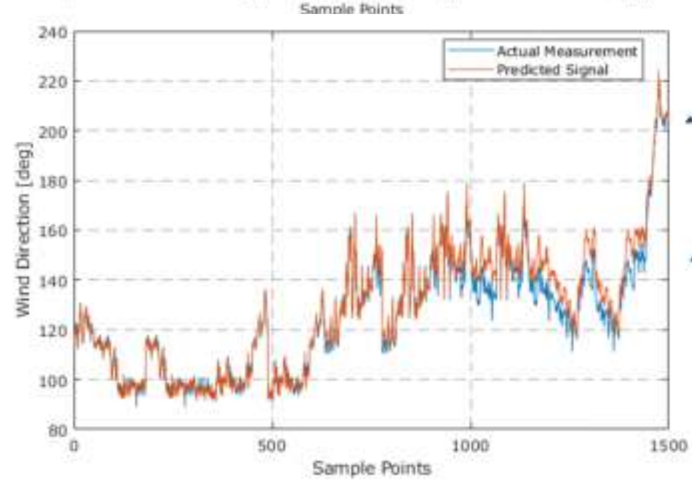


50 degrees offset

ADQC Output
000000 – No events found

Deep Neural Networks





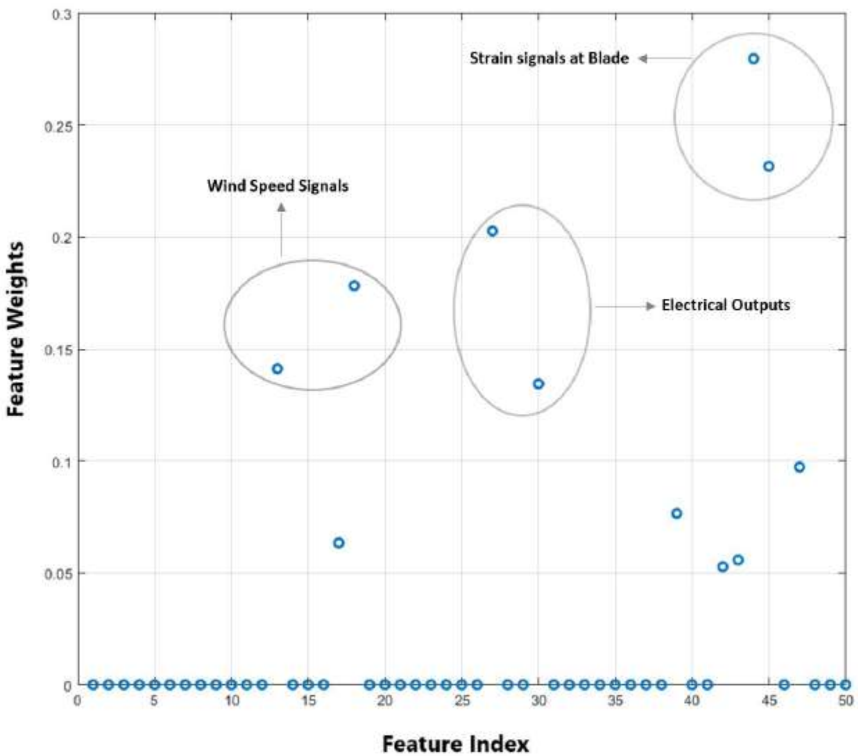
ADQC Output
000000 – No events found

RAVE-ML
0 - Pass

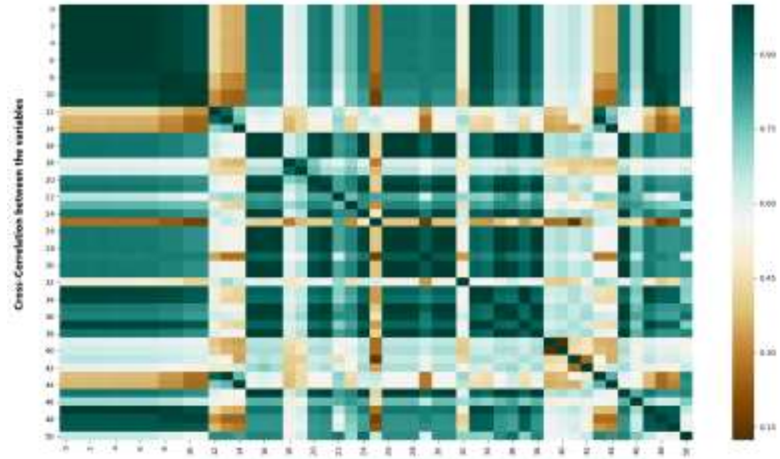
35

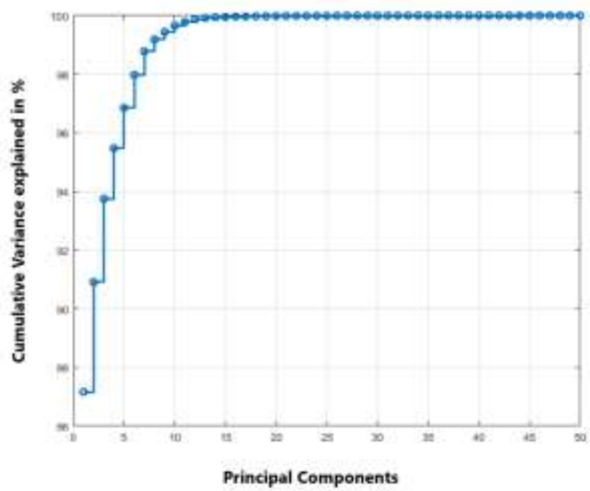
Feature Engineering/Data Compression

Neighbourhood Component Analysis



Cross Correlation





Principal Component Analysis

Feature Selection

Feature engineering

Taking the most valuable features into account

Dimensionality reduction

by principal component analysis (PCA)

→

Avoidance of collinearity

Principal components (compressed data)

In consideration of the model's robustness the features have been reduced to SCADA data and accelerometers:

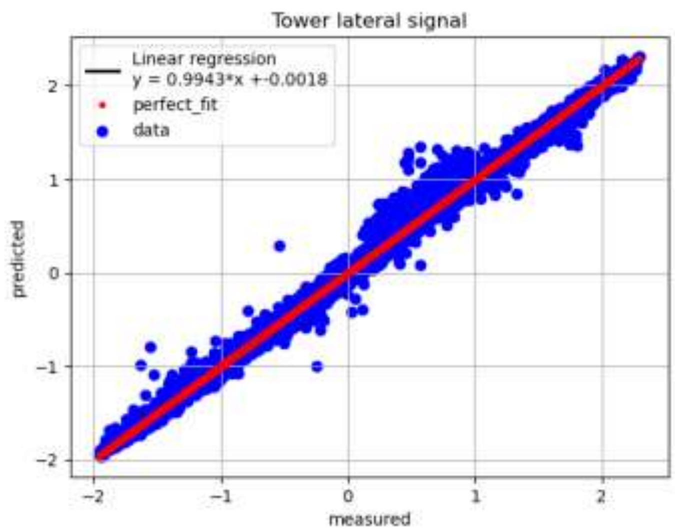
- Pitch angle
- Generator speed
- Yaw angle
- Wind speed
- Electrical power

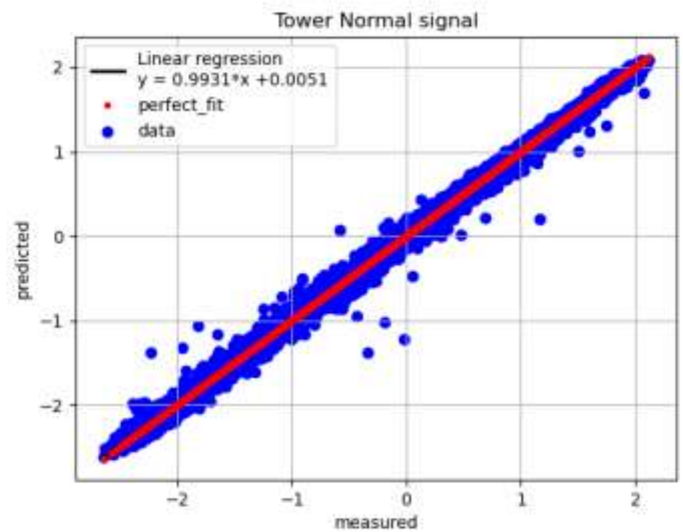
- Accelerometer
  - Support structure
  - Upper tower
- Temperature
- Humidity

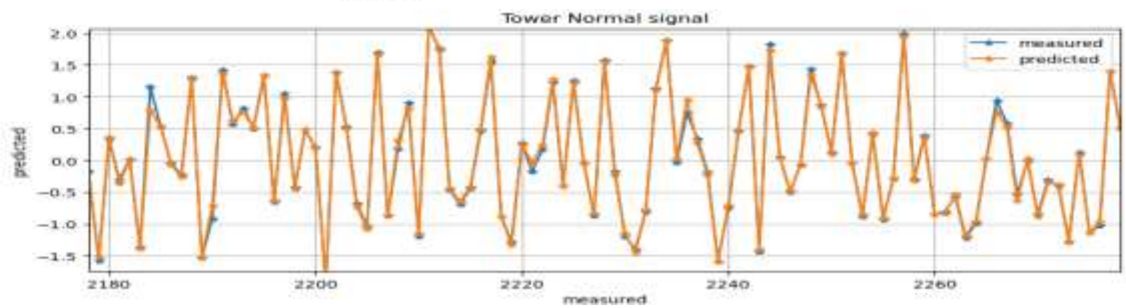
Case Study – First Results

Able to estimate all signals (Structural, acceleration, controller , etc..)

Example 1 : Tower signals



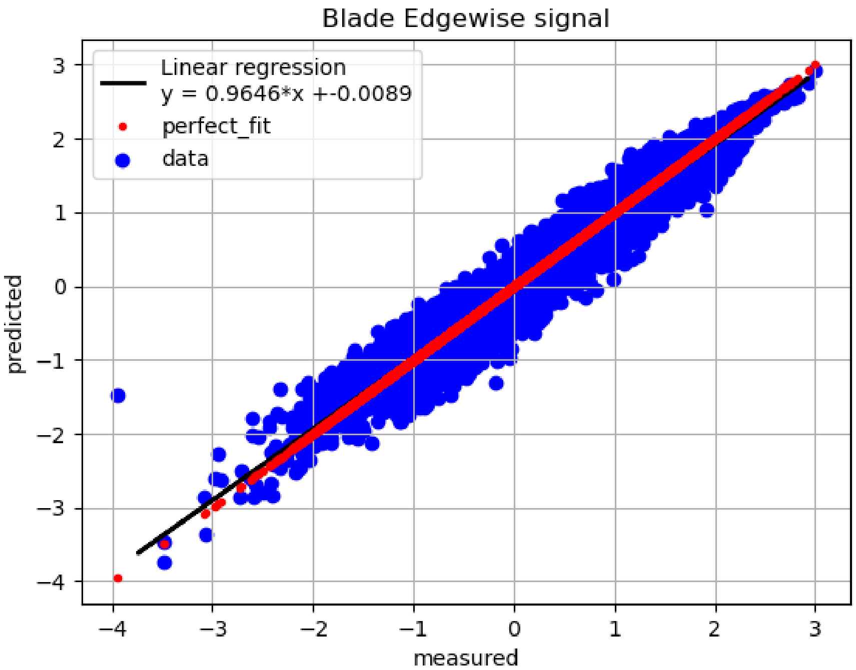


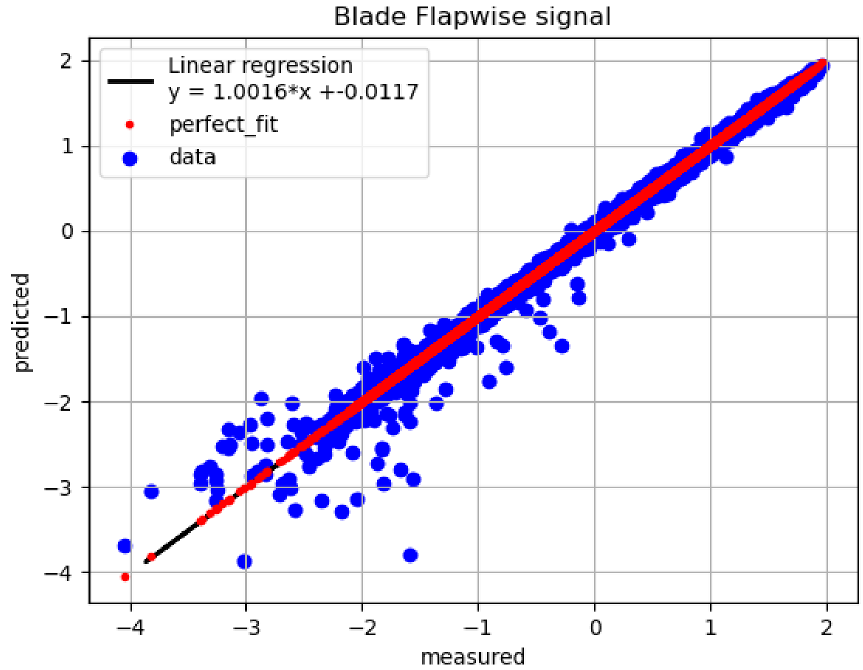


Case Study – First Results

Able to estimate all signals (Structural, acceleration, controller , etc..)

Example 1 : Tower signals

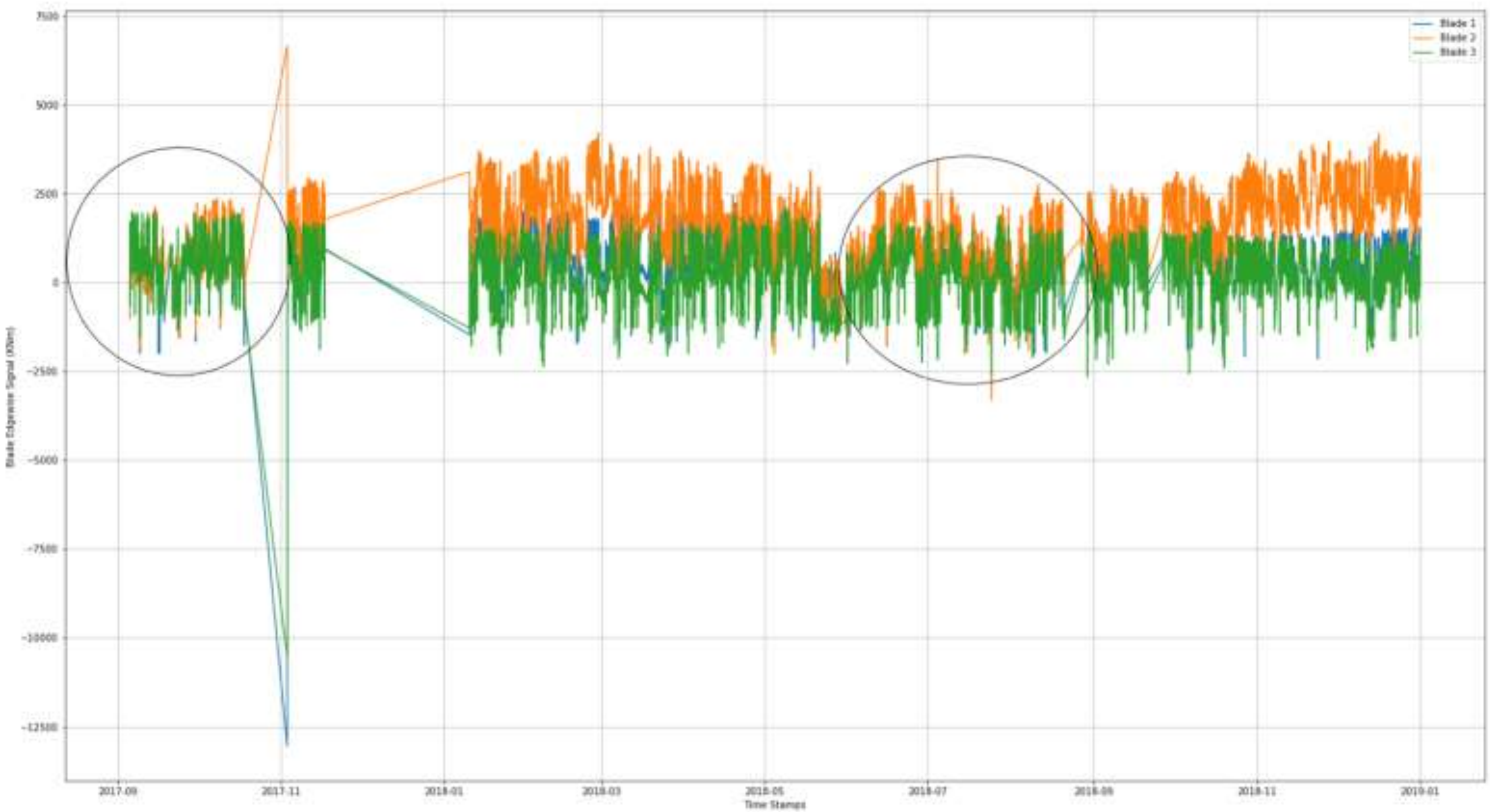




Case Study – First Results

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

Detection of sensor drift in the blade signals due to temperature change



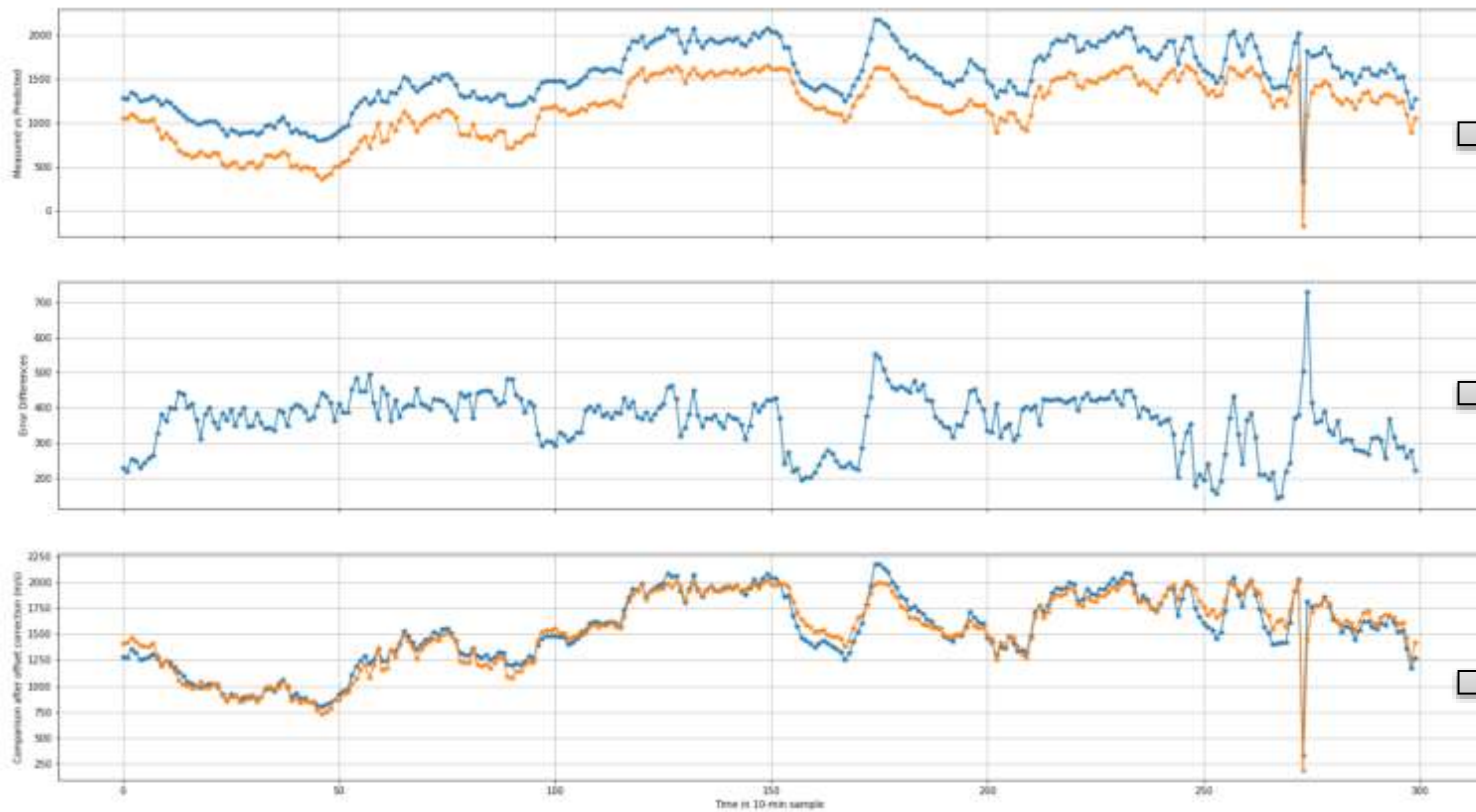
- ➢ **Sensor installed and calibrated in Autumn (Black circles)**
- ➢ **Drifting problem in the other seasons**

DNV·GL

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

Case Study – First Results

Detection of sensor drift in the blade signals due to temperature change



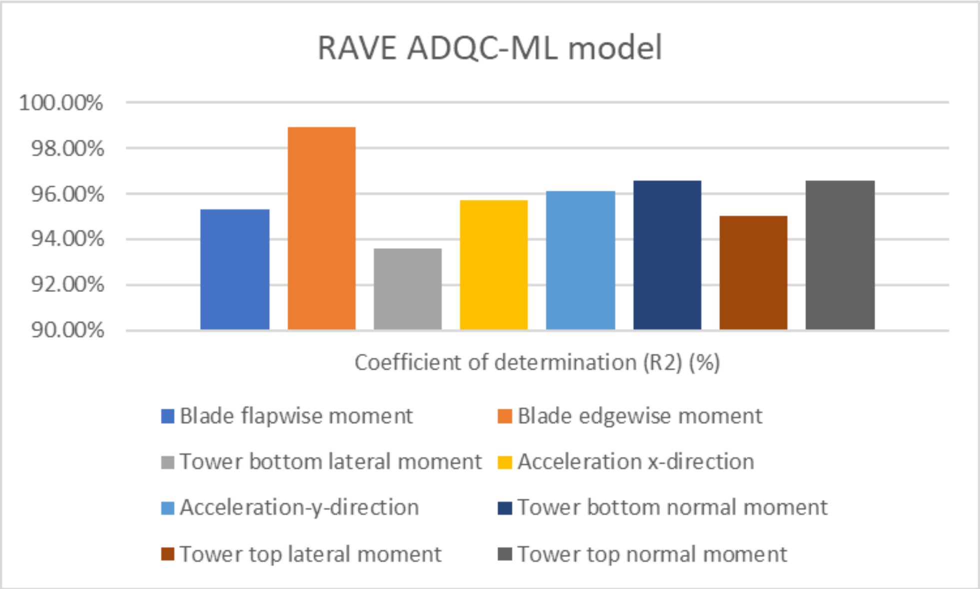
Measured vs Estimated

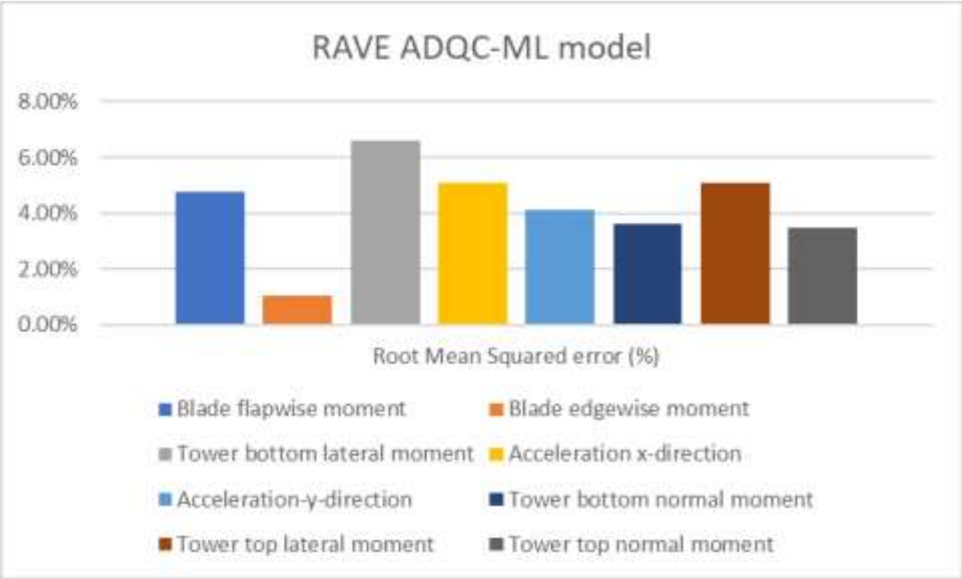
Error difference

Correcting the measured signal based on prediction

Accuracy/Performance of the model

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE





$$R^2 = 1 - \frac{Variance\ (DEL_{Estimated} - DEL_{Real})}{Variance(DEL_{Real})}$$

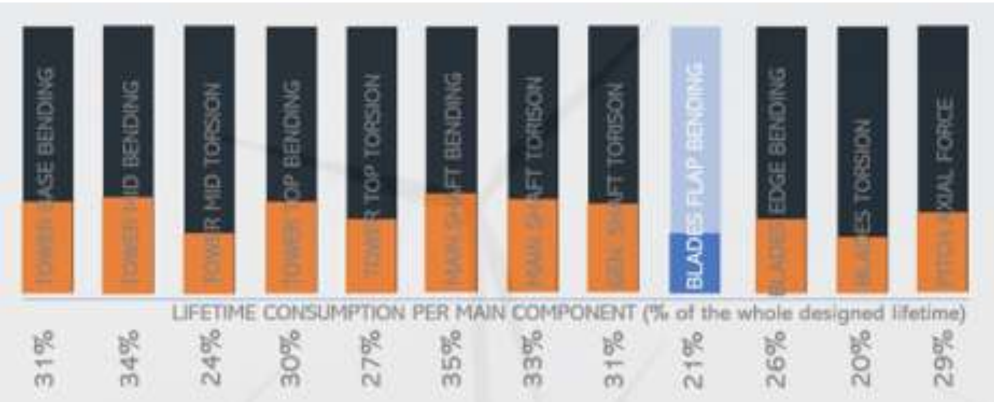
$$RMSE = \sum(DEL_{Estimated} - DEL_{Real})\text{^}2$$

- Quantification of important signals are shown here
- Accuracy/performance range applicable to most of the signals available in RAVE project

Other applications

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

Reliable Lifetime Estimation



Wind farm optimization

Model built based on one turbine

Transferred to other turbines

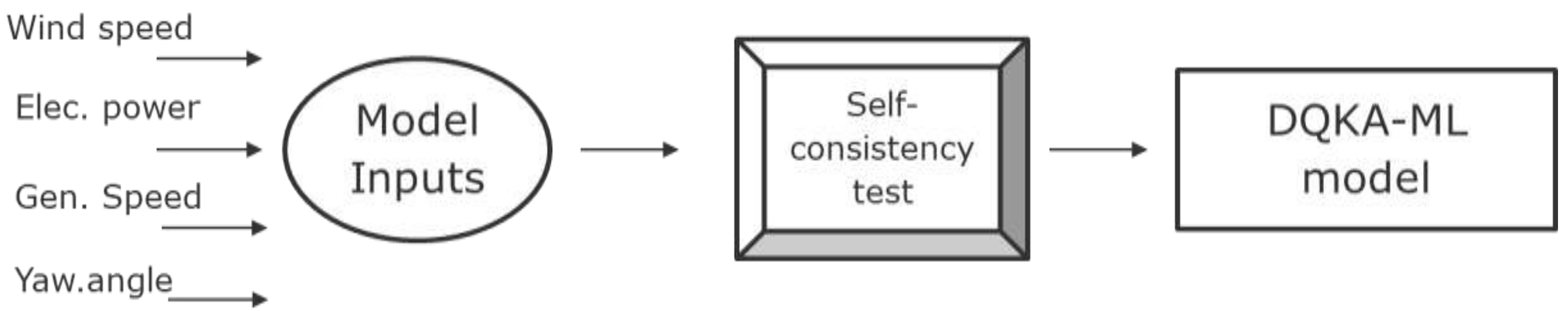
Measurement Data fulfilment/extrapolation

Turbine 1

Turbine 2

Future Work

➢ **Self consistency test for model's input variables**



➢ **Flagging strategies… how can be compare estimated and measured data to flag ?**

➢ **How can we quantify the uncertainties ?**

➢ **Providing calibration factors along with the flags**

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

Thank you for your attention !!!!

Questions ??

Anish Venu
Data Scientist - DNVGL
Anish.Venu@dnvgl.com

Nick Hansen
Project Engineer – UL Intern.
Nick.Hansen@ul.com

Marten Schmager
Scientist - BSH
Marten.Schmager@bsh.de

Contact Details

Kai Herklotz
Tel.: +49 (0)40 3190-3230
E-Mail: Kai.Herklotz@bsh.de

Hans-Peter Link
Tel.: +49 (0)4856 901-46
E-Mail: Hans-Peter.Link@dnvgl.com

Tom Neumann
Tel.: +49 (0)4421 4808-814
E-Mail: Thomas.Neumann@ul.com

BUNDESAMT FÜR SEESCHIFFFAHRT UND HYDROGRAPHIE

RAVE RESEARCH AT ALPHA VENTUS

Projektträger Jülich
Forschungszentrum Jülich

Bundesministerium
für Wirtschaft
und Energie