Prediction of vessel propulsion power using machine learning on AIS data, ship performance measurements and weather data

DNV GL Group Technology and Research

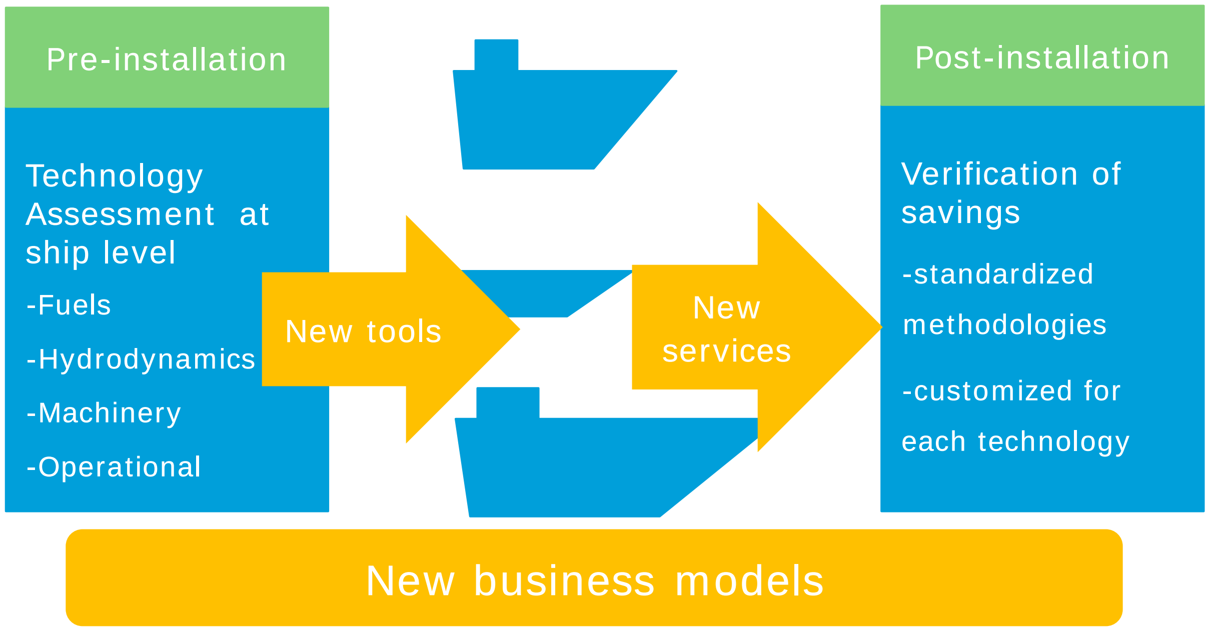
Qin Liang, Hans Anton Tvete, Hendrik W. Brinks

14 November 2019

Agenda

- Background and Objective
- Methods
- Data and defined scenarios
- Results
- Summary and next steps

VERDE (VERification for Decarbonization)
Project funded by Research Council of Norway

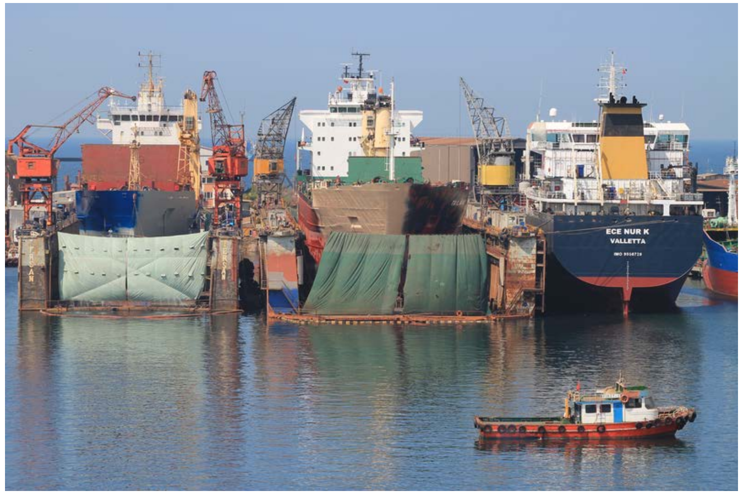


Backgroud and Objective

- Increasingly pressure on shipping for reducing GHG emissions
- New methods and more data are available
- The current AIS-based consumption and emission model need improvement

- Evaluate the performance of different machine learning models against ship measurement data and existing physical AIS-based models.

$$P_{Operation} = \left(\frac{V_{Operation}}{V_{Design-max}}\right)^3 P_{Design-max}$$

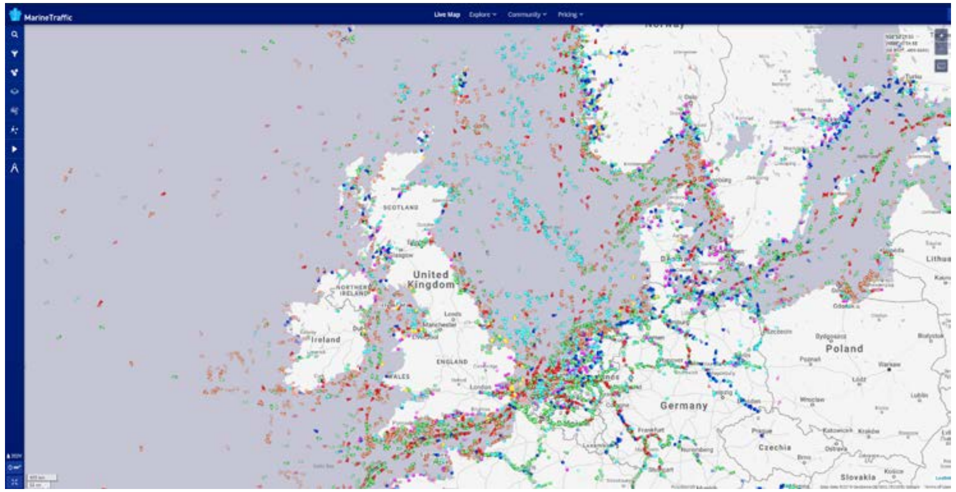


Methods

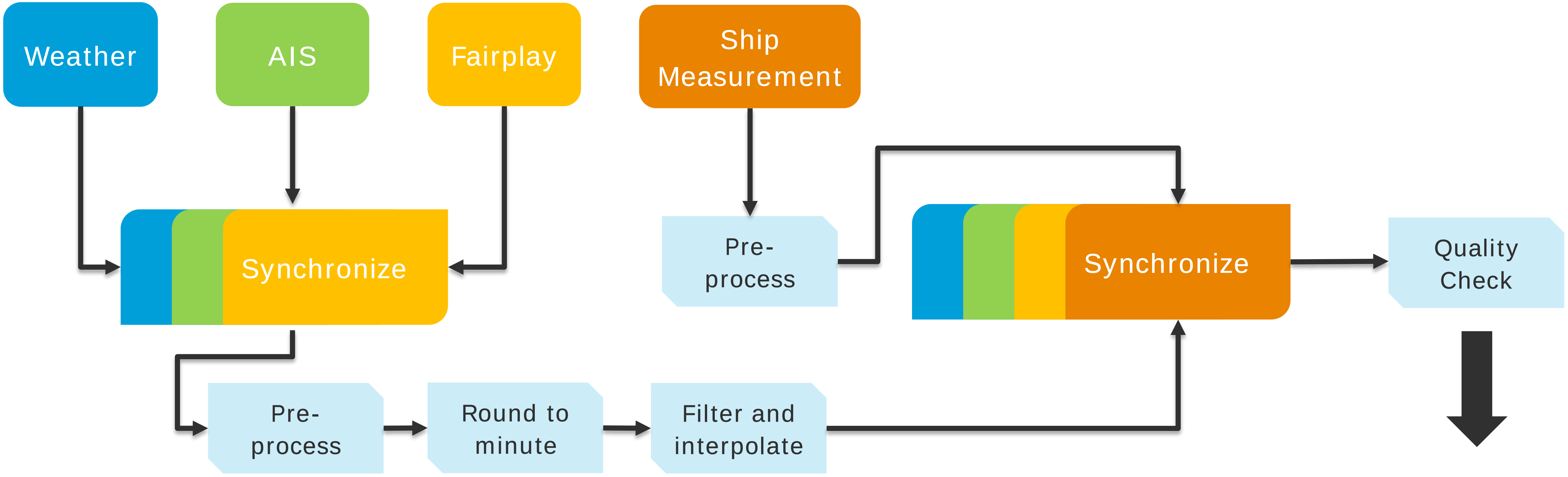
- A regression problem
- Traditional machine learning methods:
  - Decision Tree
  - Random forest
  - Support Vector Regression (Support Vector Machine)
- Deep learning methods:
  - Multilayer perceptron

Data adapted

- Automatic identification system (AIS) data
  - Heading, Speed over ground, Course over ground, etg.
- IHS Fairplay data
  - Design draught, moulded depth, moulded breadth, etg.
- Ship measurements data
  - Propulsion power
- Weather data
  - Significant wave height, wave direction, wave speed, wind speed, etg.

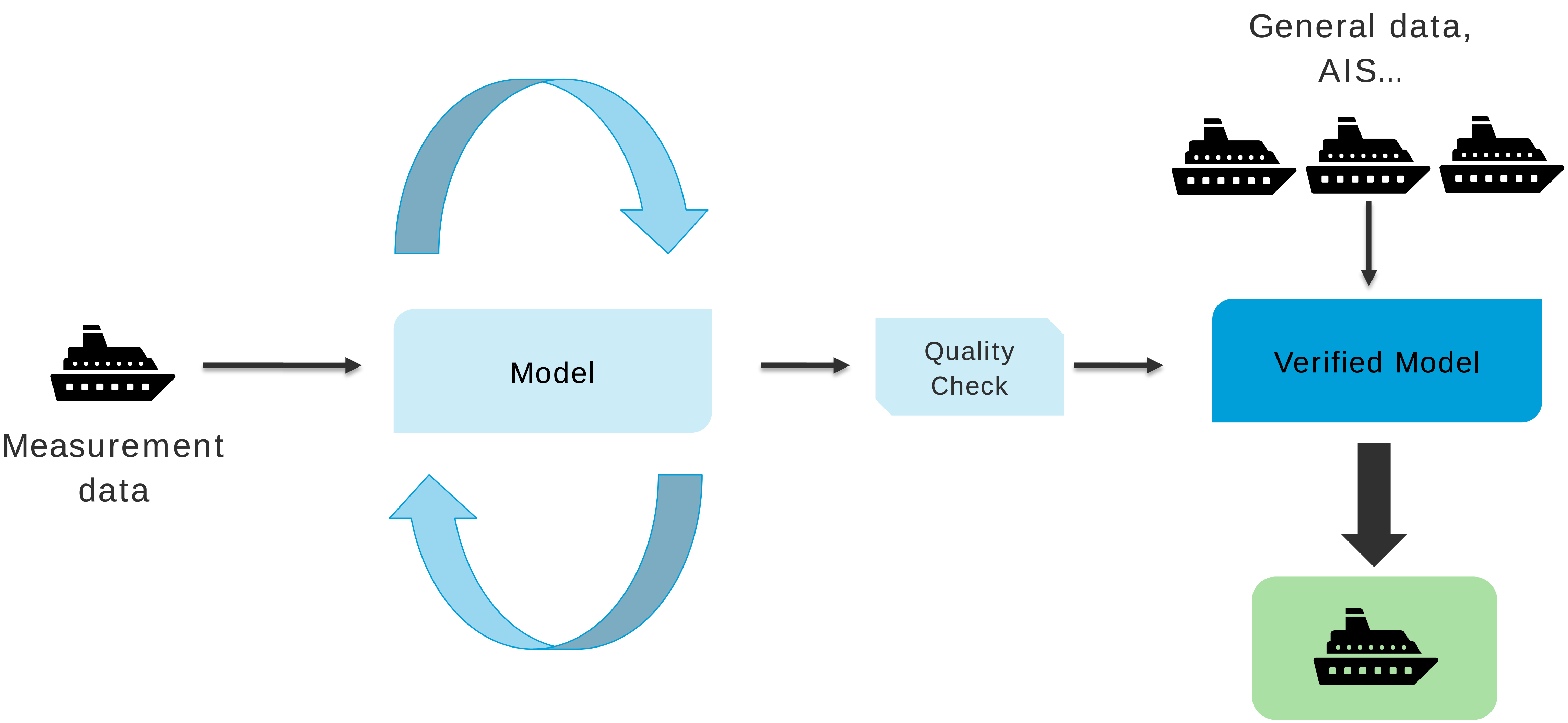


Data preparation



- Data preparation
- Data quality check and feature selection
- Scenario 1 and 2

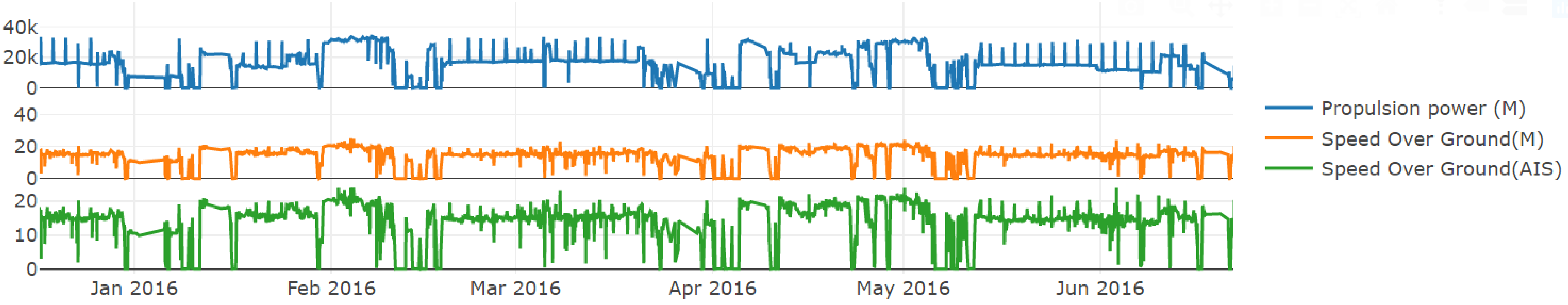
Data flow


General data,
AIS...
Measurement
data
Model
Quality
Check
Verified Model

Defined scenario 1

Scenario 1		
Features	Unit	Correlation
Speed over ground	knots	0.89
Significant wave height	m	0.39
Wave direction to vessel	degrees	0.04
Wind speed	m/s	0.21
Wind direction to vessel	degrees	0.04

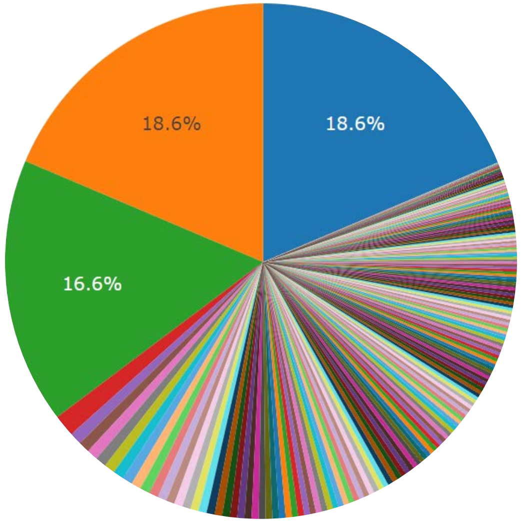
- To explore vessel operation and environment.
- 1 vessel with most data and satisfactory synchronization result selected
- 75,000 measurements over a period of 50 months



Defined scenario 2

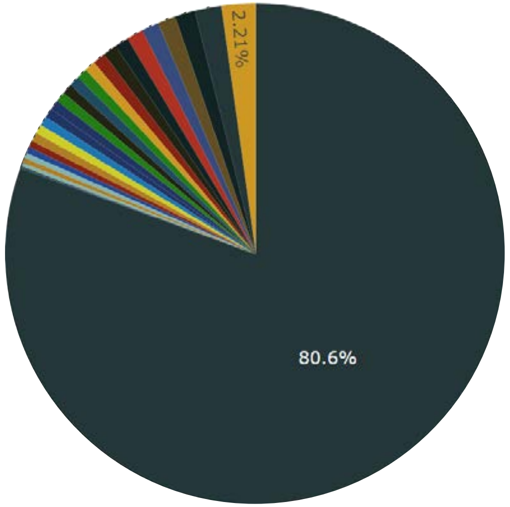
- Focused on the general prediction of the container vessels
- 238 vessels data, 210 training and validation, 28 test
- 400000 sampling points

Scenario 2		
Features	Unit	Correlation
Speed over ground	knots	0.77
Speed through water	knots	0.77
Trim	m	-0.19
Draft fore	m	0.49
Draft aft	m	0.52
Design draught	m	0.42
Moulded depth	m	0.49
Moulded breadth	m	0.49
Length between perpendiculars	m	0.52



Training data

320,000 records

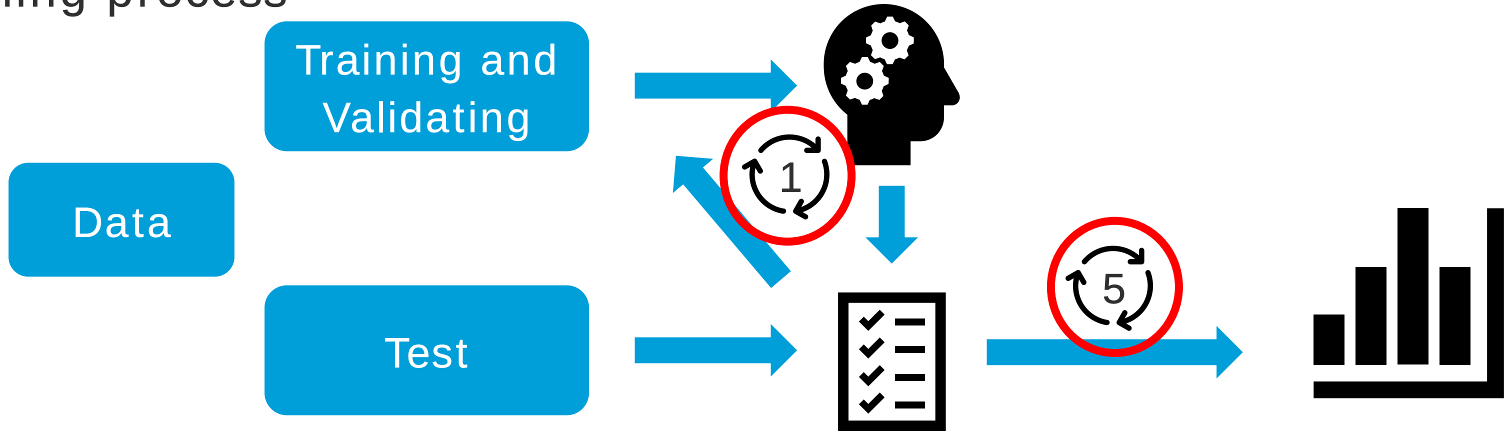


Test data

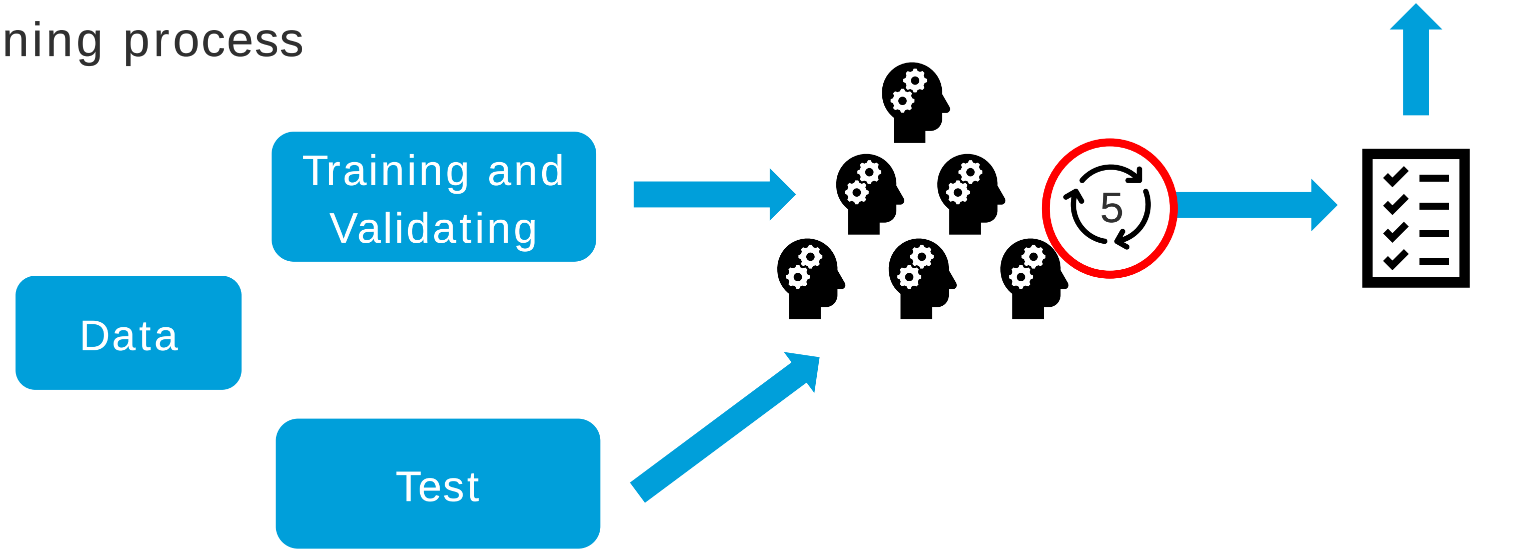
80,000 records

Training process

- Traditional ml training process



- Deep learning training process



Training process

Table 3. MLP models for both scenarios. The model name refers to number of neurons in each layer.

Model	Model Name	Layers	S1 parameters	S2 parameters
1	M_1000	1	7001	12001
2	M_100	1	701	1201
3	M_50	1	351	601
4	M_30	1	211	361
5	M_20	1	141	241
6	M_10	1	71	121
7	M_5	1	36	61
8	M_500_100	2	53201	55701
9	M_100_50	2	5701	6201
10	M_50_30	2	1861	2111
11	M_10_5	2	121	171
12	M_50_20_10	3	1541	1791
13	M_10_8_4	3	189	239
14	M_10_8_6_4	4	235	285

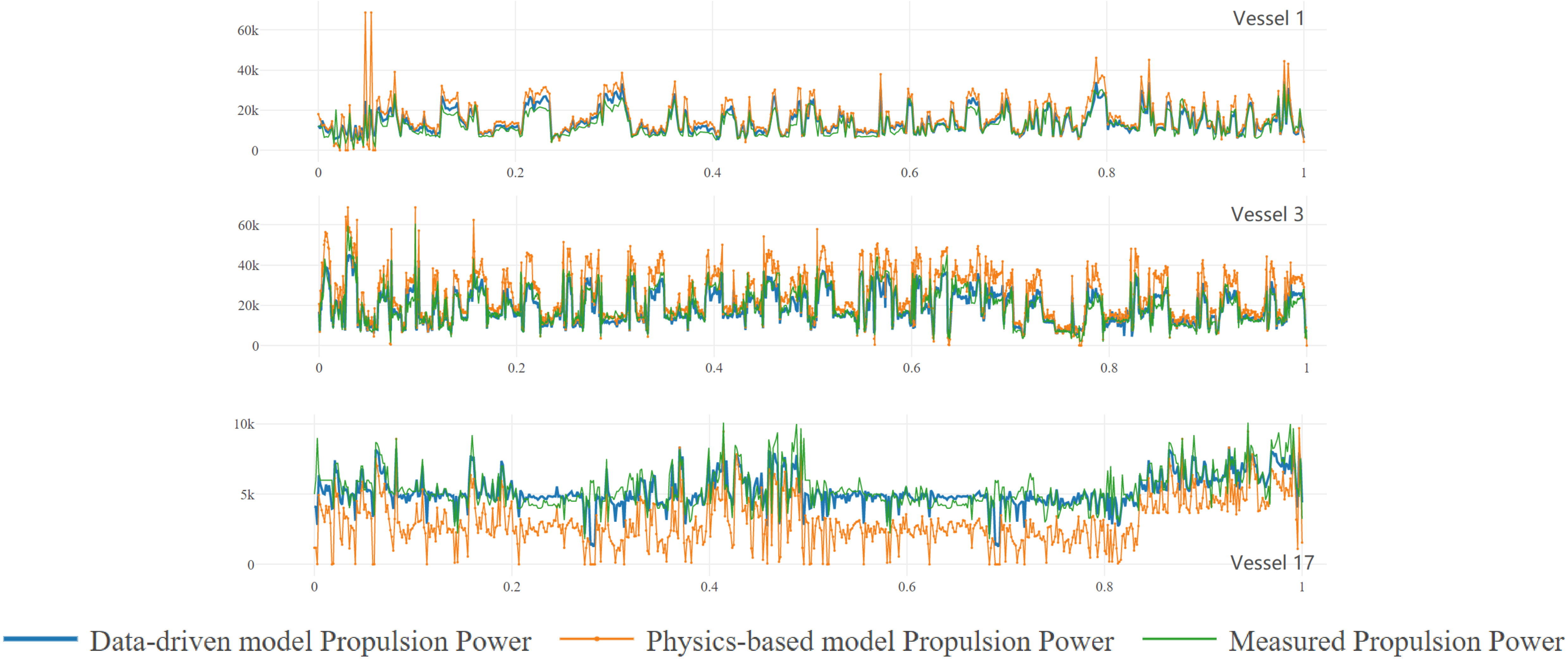
Results

- Traditional machine learning models

Model name	Scenario 1 R^2 score	Scenario 2 R^2 score
Adaboost Decision Tree	0.759	0.835
Gradient Boosting Decision Tree	0.772	0.847
Random Forest	0.764	0.840
SVR-RBF	0.710	0.766

- Deep learning models

Results – Scenario 2



Results

- Data manipulation result on test data

	Scenario 1	Scenario 2
Physics-based	47.93%	68.59%
Original speed	78.31%	83.83%
Speed by 2nd order	78.69%	83.93%
Speed by 3rd order	78.53%	84.93%

$$P_{Operation} = \left(\frac{V_{Operation}}{V_{Design-max}}\right)^3 P_{Design-max}$$

- Prediction results of scenario 1 MLP model on other vessels

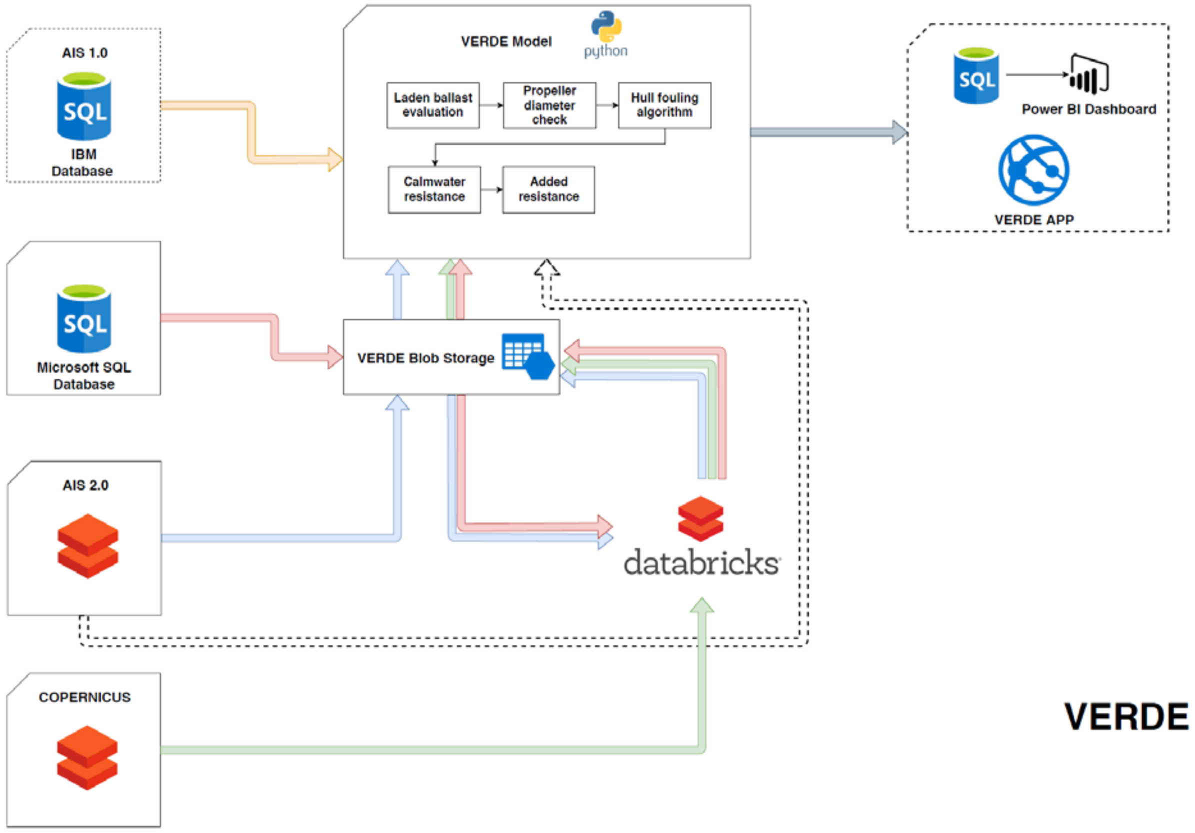
	Physics-based	Trained MLP
1.Target vessel (Container)	48%	78%
2.New container vessel	61%	72%
3.New passenger vessel	54%	82%
4.New general cargo vessel	72%	-58%

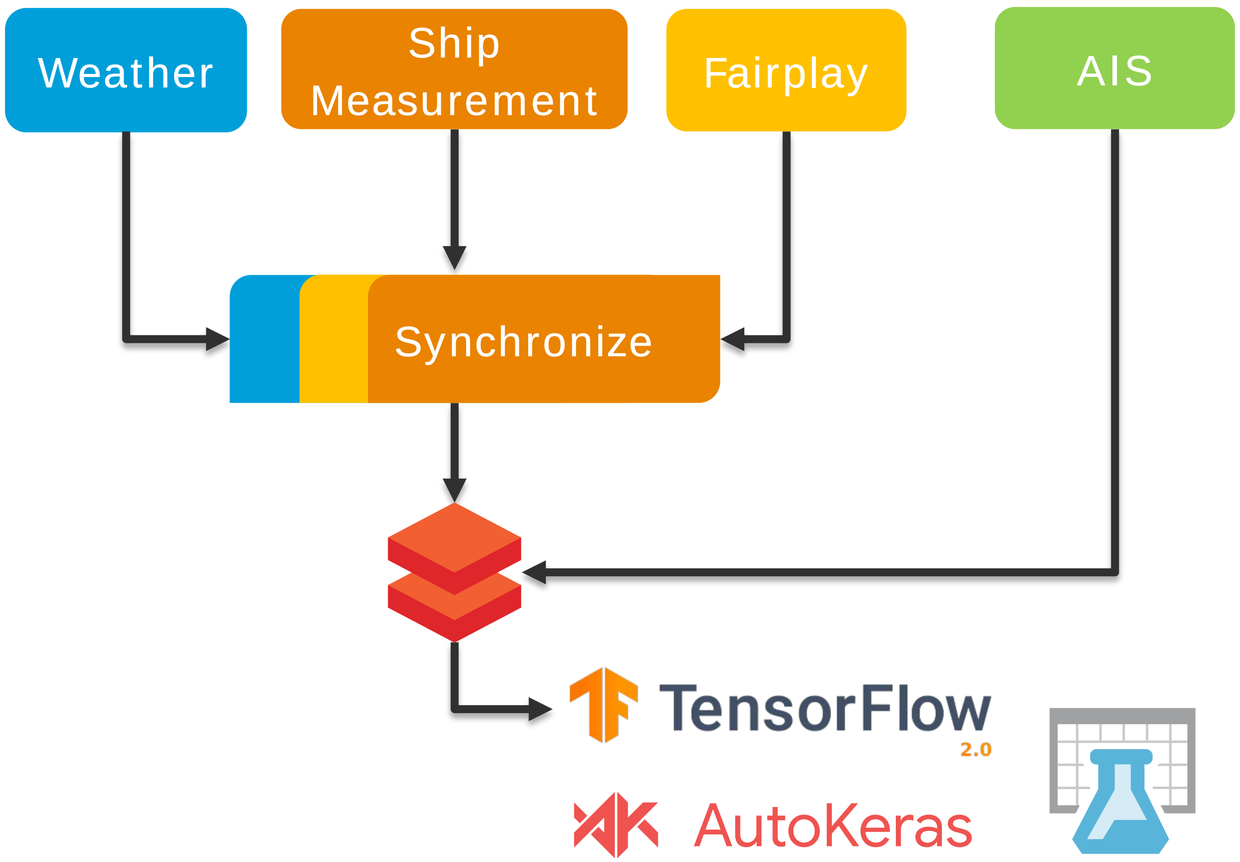
Summary

- Physics-based and data-driven models were compared under two scenarios. In both scenarios, data-driven models showed better result than the physics-based models.
- Data-driven models strongly rely on the data it has been trained by.
- The distribution of the data affects the model performance.
- Machine learning fill the gap due to lack of knowledge for feature engineering.
- The neural network does not need to be complex to provide better performance.
- Both traditional machine learning and deep learning models perform well.

Next Steps

- Vessel specific physics-based model in progress
- Scenario 3





Thank you for your kind attention.

https://www.dnvgl.com/technology-innovation/index.html

Qin Liang
Qin.Liang@dnvgl.com
+47 91240179

www.dnvgl.com

SAFER, SMARTER, GREENER