Predict Oxygen Depletion using Machine Learning

Design Project - SIE 2019
Authors: Joëlle Perreten and Maëlle Romero
Supervisor: Prof. Johny Wüest, EPFL and Florence Cuttat, DHI Group

Problem Statement

The metalimnion of the Inner Danish waters is prone to oxygen depletion under given environmental conditions such as strong stratification in autumn. **Hypoxia** is legally defined as **concentration of dissolved oxygen (O_2) below 4 mg/L** and can lead to the reduction of benthic species and the loss of biodiversity.

Data from the Århus Bay, Denmark consist of:
- Time series from 2000-2017 with **724** dates of measurement
- Measurements along a vertical water profile of **17 m** to the seafloor
- Data for **oxygen content**, **temperature** and **salinity**

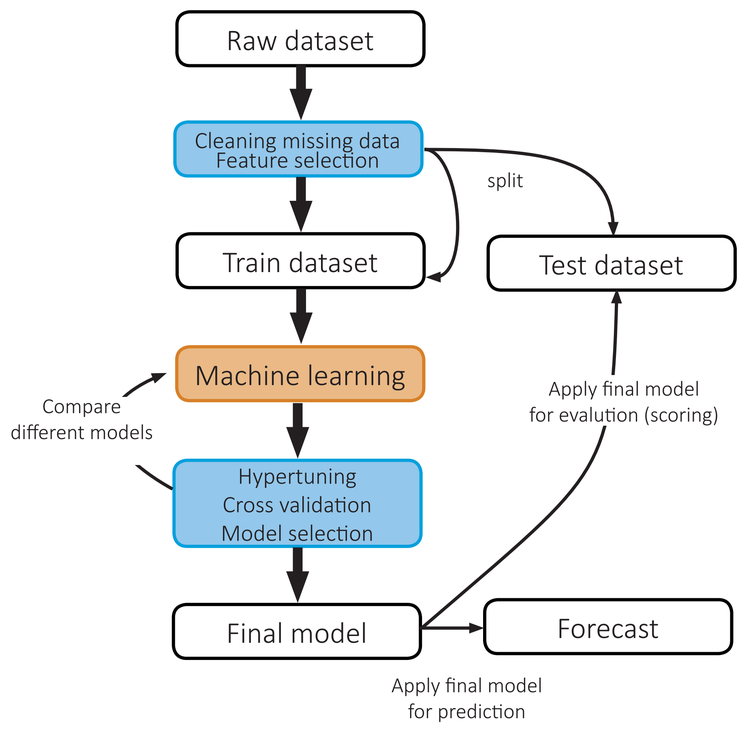
Method

Objectives

- Develop a machine learning model to predict the O_2 concentration
- Assess the reliability and accuracy of the model
- Predict O_2 concentration in the future based on the official forecast

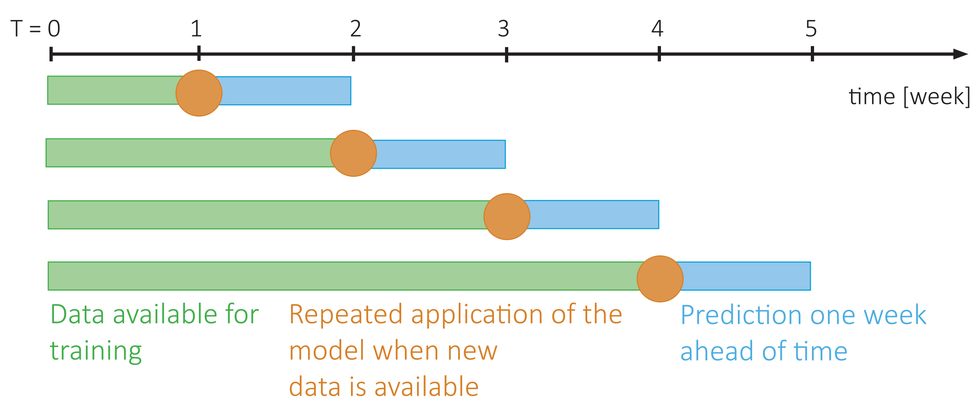
Machine Learning

The measured data is split in a train and test dataset. The train set is used to develop the machine learning model. The accuracy of the prediction is estimated by applying the model to the test set.



Expanding Window for Forecast

The best machine learning model allows to predict the O_2 concentration ahead of time. When the official forecast has new data of temperature and salinity available, the machine learning model is executed again and predicts the O_2 content for the future.



Results

Feature Selection

- Temperature explains **45 %** of O_2
- Salinity explains **26 %** of O_2 content
- Month and depth are important

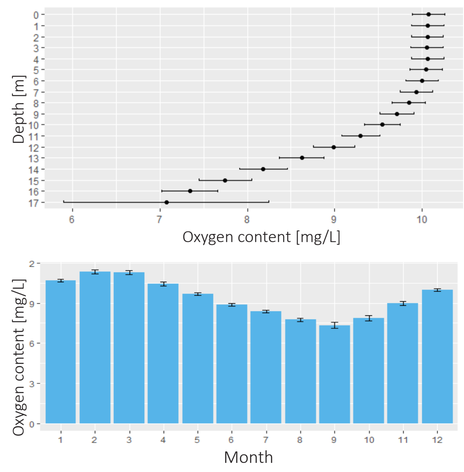
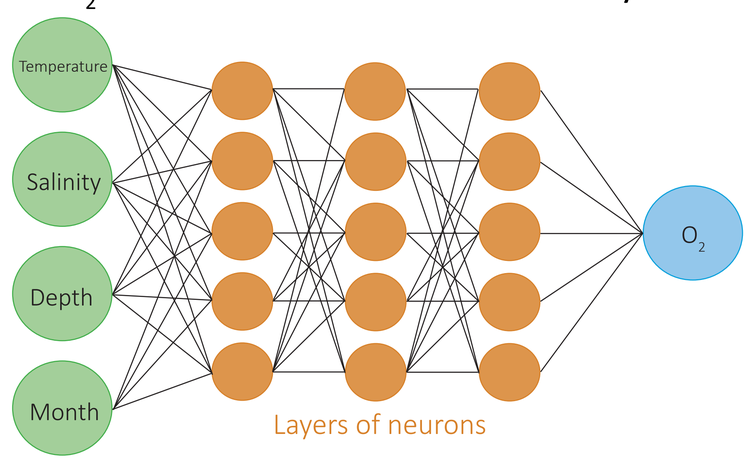


Figure 1: The 95 % confidence intervals show that O_2 contents are significantly different between months and gradually depend on the depth.

Model Selection

A **neural network** is the most appropriate machine learning model to predict the O_2 content and consists of **1000** layers of neurons.



The scores (R^2) compare the measured and predicted O_2 contents:
- for the training set a score of **83.42 %** is achieved
- for the testing set a score of **81.20 %** is achieved
- the root mean squared error is **0.8120 mg/L**

Forecast

The O_2 content predictions **1 week ahead** of time are similar to the measured values. The two hypoxia events are successfully predicted.

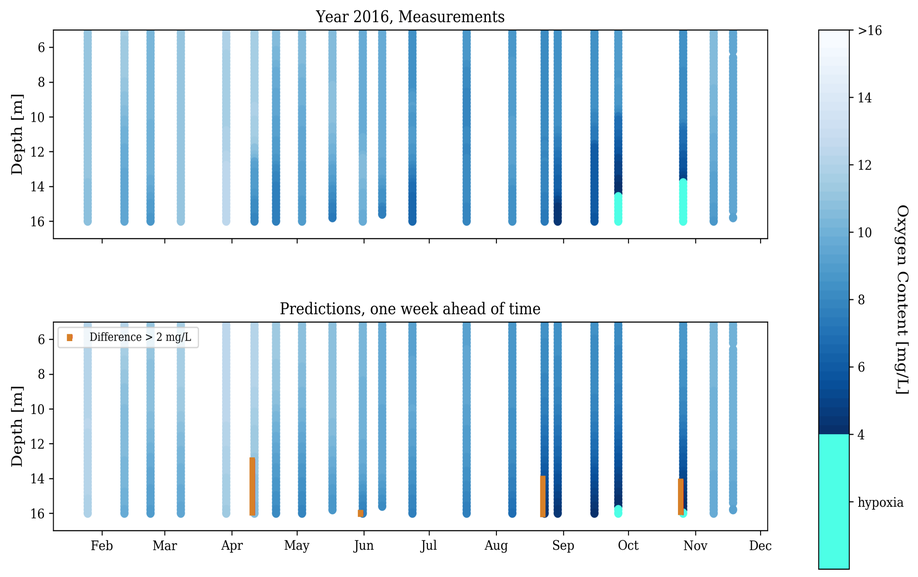


Figure 2: The measured O_2 concentration in the year 2016. Hypoxia events are in light blue.

Figure 3: Predictions for the next week done with the model. Inprecise predictions are red.

Conclusions

+ Excellent prediction of O_2 content troughout the year in the water column
+ Good performance in prediction of hypoxia in autumn
+ Possibilty to predict ahead of time on the basis of the official forecast
+ Can be used as alternative or complementary to the official forecast

- Machine learning models are black boxes that do not relate to the physical reality
- Extensive tuning of models is time consuming

References

Lund-Hansen, L. C., Skyum, P., & Christiansen, C. (1996). Modes of Stratification in a Semi-enclosed Bay at the North Sea–Baltic Sea Transition. Estuarine, *Coastal and Shelf Science*, 42(1), 45–54.| Müller, A. C., C, M. A., & Guido, S. (2017). Introduction to Machine Learning with Python: A Guide for Data Scientists. *O'Reilly Media, Inc.* | Raschka, S., & Mirjalili, V. (2017). Machine Learning and Deep Learning with Python, scikit-learn and TensorFlow. *Packt Publishing*. Birmingham, 2017.