Automation of Quality Control for Global Ocean Data

Dr. Serdar Demirel, Dr. Sebastian Mieruch,
Prof. Dr. Reiner Schlitzer

*Alfred-Wegener-Institut,
Bremerhaven, Germany*

- Team name: **PARROTs** -- dee**P** le**AR**ning fo**R** quality control of **O**cean da**T**a

- Application: **Salacia** -- Roman mythology, the goddess of salt water and wife of Neptune

➢ **What is our dataset ? Where it is coming from ?**

➢ **Reseach vessels like Polarstern take ocean profiles using CTD sondes, which measure**

Properties:

➢ **pressure (depth),**
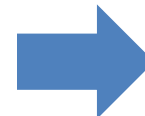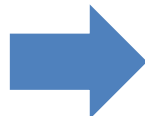
➢ **temperature,**

➢ **salinity, and some others**
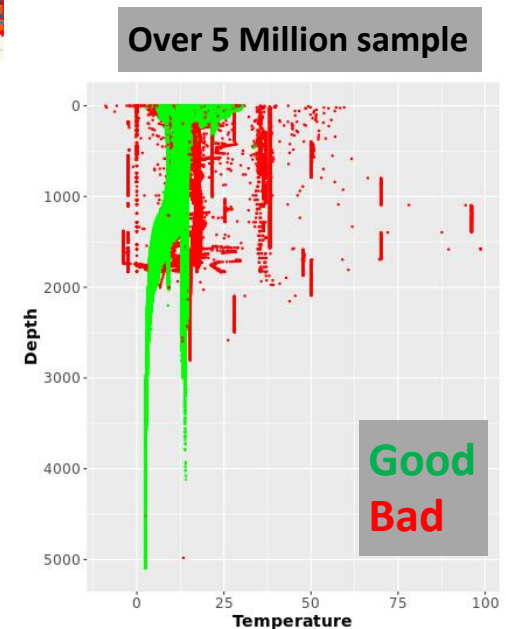
The source of our dataset

- A project that more than 35 countries and more than 100 data centers is included
- Data from year 1800 to now
- Approx. 2 Million data sets (ca. 9 Million profiles with many samples each)
- **All data quality controlled but how ?**
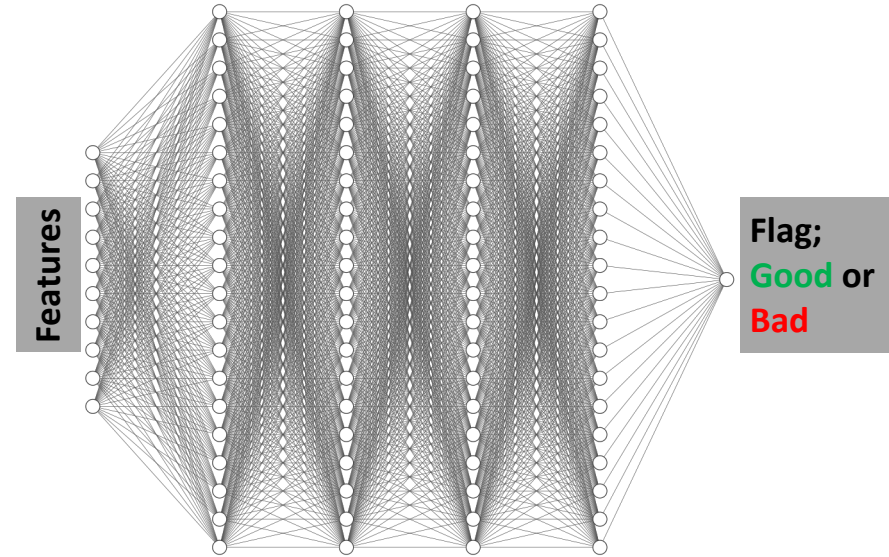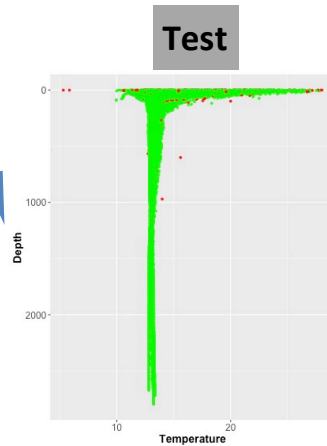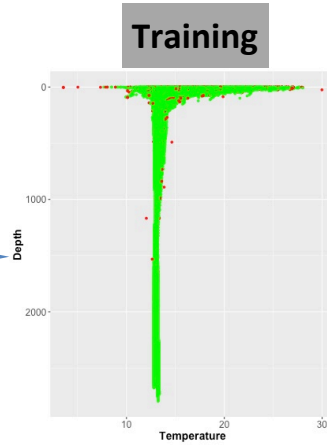

Over 5 Million sample

Experts

Good
Bad
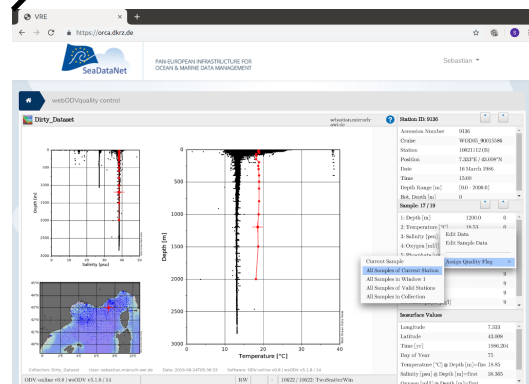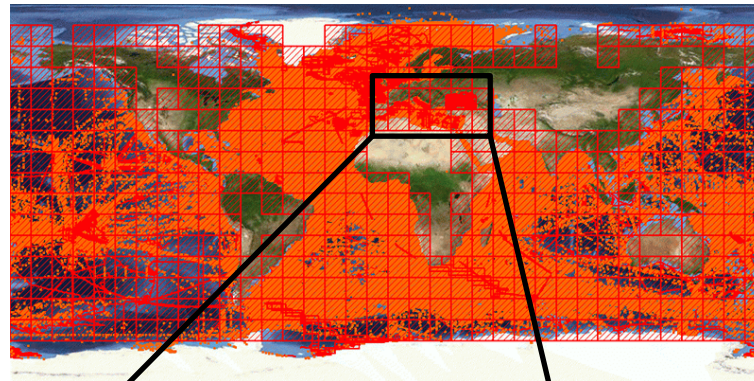
using the Ocean Data View (ODV) software

Preliminary studies

- Separation of chosen section: 75% for training and 25% for test

Training

Test

Good
Bad
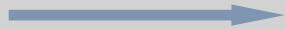
Features

Flag;
Good or
Bad

Features:
- Depth,
- Temperature,
- Latitude,
- Longitude,
- and some others

Preliminary studies

Testing of trained Neural Network

Test Data ⟶ **Trained Neural Network** ⟶ **Predictions**

?
=

Good
Bad

Good
Bad
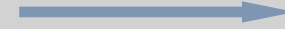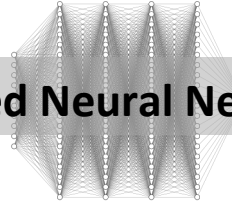
Question: Can Neural Networks mimic the decisions of an expert ?

➢ **As summary**

   ➢ **Why Salacia need to be used ?**

      ➢ The experts spend too much time to analyze every single sample

      ➢ Automation of Quality Control is a necessity

   ➢ **Future of Salacia**

      ➢ Development of a skillful and reliable AI system

      ➢ Huge impact on the Ocean Science community

      ➢ Covering the World Ocean

➢ **Thank you for your attention**