Interreg North-West Europe

Co-funded by the European Union

BEPROACT

INTERNATIONAL MEUSE SYMPOSIUM

# Predicting water quality using Machine Learning:

## Introduction of a ML-model for WFD monitoring


**Joost van Dijk**

Nelen & Schuurmans

Rijkswaterstaat
*Ministerie van Infrastructuur en Milieu*


9 September 2026

# Introduction

- **Water Framework Directive (WFD):**
  - European Guideline since 2000
    - Update in 2026
  - Good qualitative and quantitative status, reducing and removing pollution and ensuring enough water to support wildlife and human needs.
  - Applies to surface and groundwaters
  - Protect and restore water bodies on both good chemical as good ecological status

# Monitoring network - Rijkswaterstaat

- Water quantity monitoring
- Water quality monitoring

Rijkswaterstaat
*Ministerie van Infrastructuur en Milieu*

# Monitoring network - Rijkswaterstaat

## Water quantity (LMW)

- Water level [***MSL***]
- Water temperature [***degrees***]
- Wave height [***MSL***]
- Salinity [***g/kg***]
- Discharge (***m3/s***)
- Flow velocity (***m/s***)
- Tide measurements [***MSL***]

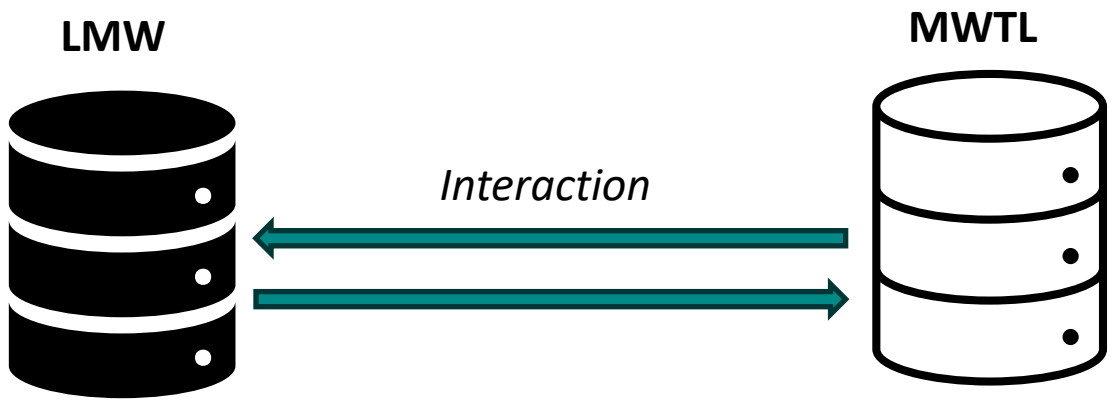
## Water quality (MWTL)

- Surface water (dissolved fraction)
- Sediment
- Suspended matter

Rijkswaterstaat
*Ministerie van Infrastructuur en Milieu*

# Goal: from insights to impact

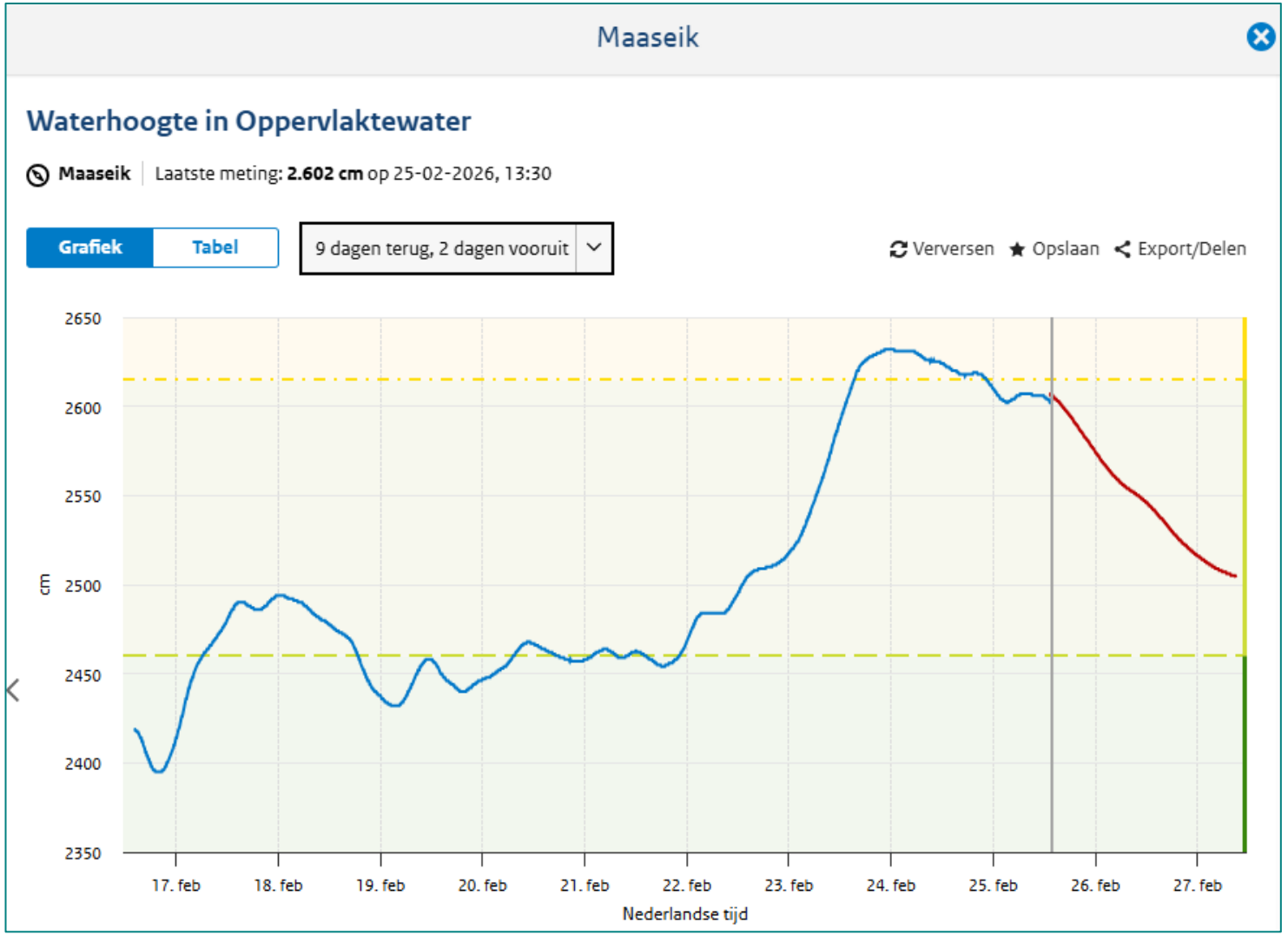
- **Valorisation** of data already collected for regular operational purposes
- **Smart integration** of data leads to monitoring cost reduction and enhanced insight into asset performances
- **Water quality predictions** based on high frequent water quantity data, and regular forecasting of water quantity




Rijkswaterstaat
*Ministerie van Infrastructuur en Milieu*

# How?

- Training of **Machine Learning models** to predict mass concentrations of (dissolved) substances or suspended matter in surface water with water quantity parameters as input:
  - **Water quantity**: high frequency timeseries data (every 10 minutes)
  - **Water quality**: low frequency and labour-intensive measurement (monthly)



# Why Machine Learning models?

Machine Learning models ingest big datasets and turn it into meaningful patterns and relationships, allowing water quality to be predicted with greater accuracy.

## Partial Least Squares (PLS)

Statistic modelling technique to find relations between multiple explanatory variables (e.g. water level, discharge, temperature, oxygen saturation) and an unknown variable (water quality)

## Elastic Net Regression

Regularized regression technique that selects the most relevant explanatory variables from a large dataset and uses them to predict an unknown variable (water quality).

# What: Predictions of mass concentrations based on $\log K_{oc}$

$\log K_{oc}$ indicates a substance affinity for organic matter and therefore its mobility in water.

## High $\log K_{oc}$

*"Strong sorption to organic matter, low mobility"*

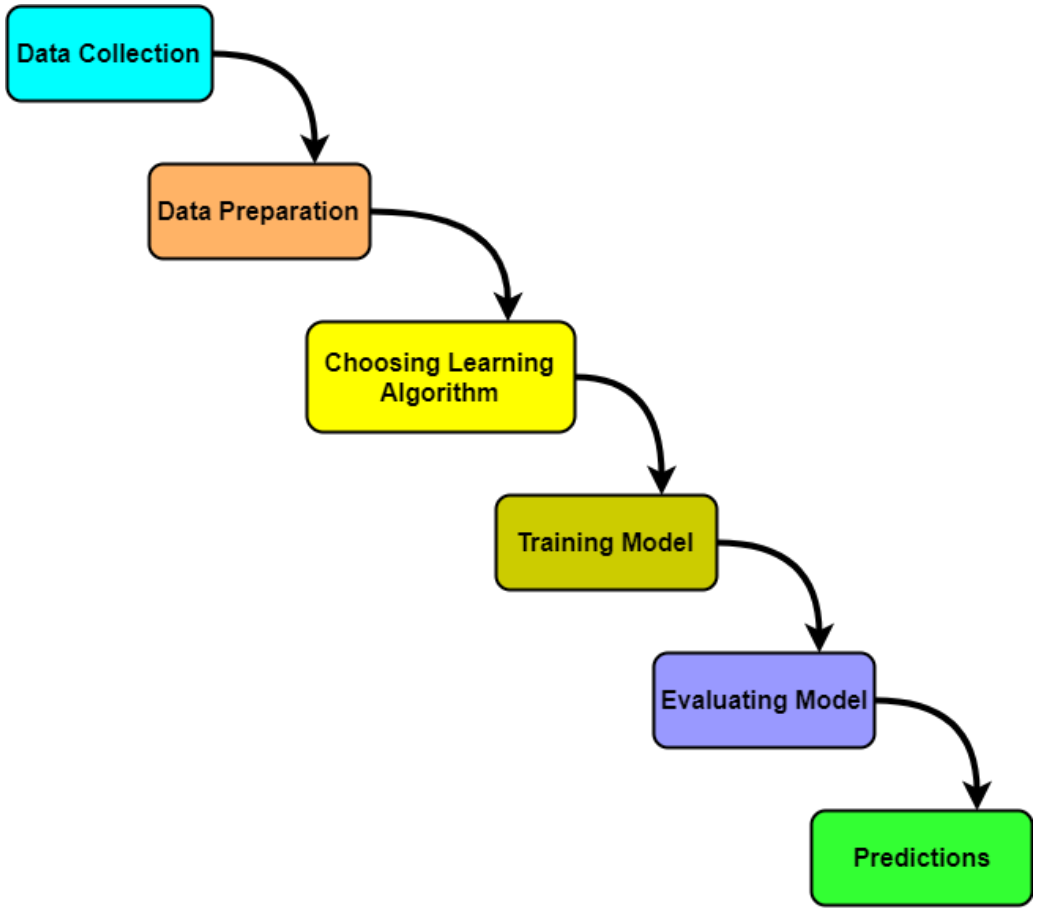
- Man-made organic chemicals, such as PCBs
- Environmental pollutants, such as dioxins
- Pyrethroids pesticides, such as cypermethrin
- Combustion products (PAHs), such as benzo(b)fluoranthene

## Low $\log K_{oc}$

*"Weak sorption to organic matter, high mobility"*

- Industrial substances, such as PFAS
- Nutrients, such as nitrate
- Herbicides, such as nicosulfuron or atrazine
- Pharmaceuticals, such as diclofenac

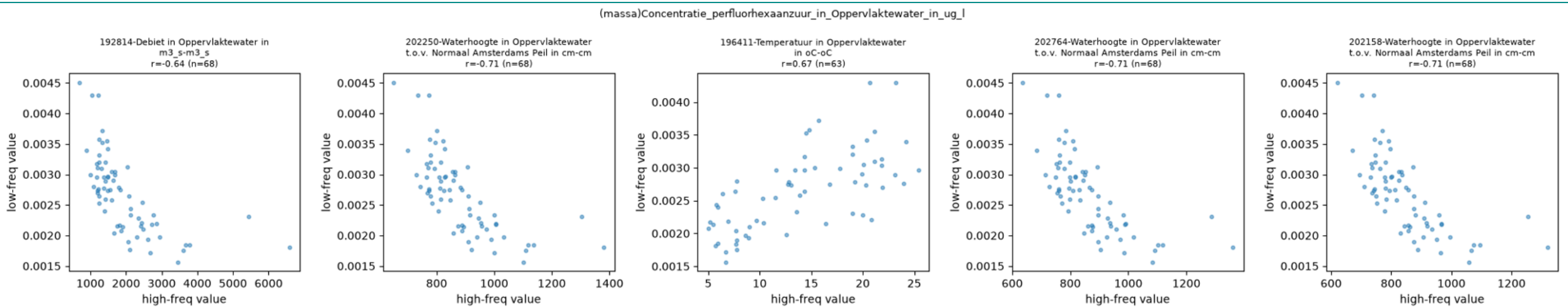
# Water quality predictions at Lobith (Ponton) - Rhine




1) **Data collection**: retrieving high-frequency time series of temperature, discharge, water level, seasonality.
2) **Data preparation**: time series quality check, followed by aggregation and correlation analysis.
3) **Choosing Learning Algorithm**: PLS-model where multiple explanatory variables define an unknown variable.
4) **Training model**: Splits input data in a training, testing and validation set. Feeds all features to PLS-regression model and picks best scores on selected objectives.
5) **Evaluating model**: Scores the model by calculating error metrics.
6) **Predictions**: Plots predictions against measurement from MWTL.

Rijkswaterstaat
*Ministerie van Infrastructuur en Milieu*

# Water quality correlations at Lobith (Ponton) - Rhine




## Correlation graphs of a PFAS-variant to water quantity parameters

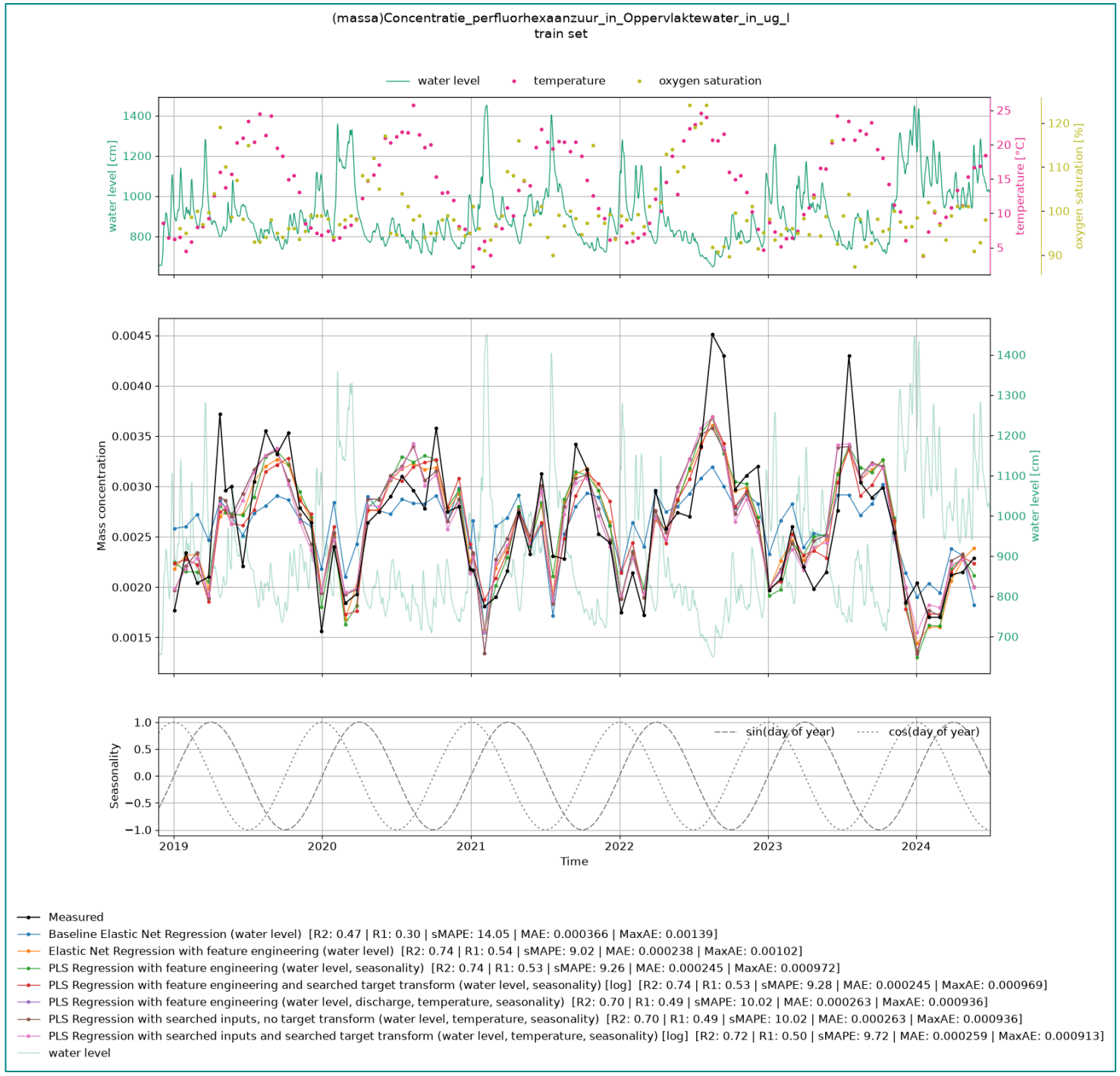
- Perfluorohexanoic acid (synthetic organic sulfur) belongs to the PFAS-family.
- Low logKoc values between 1,6 – 2,4 (liters/kg)

Rijkswaterstaat
*Ministerie van Infrastructuur en Milieu*

# Results

## Perfluorohexanoic acid (PFHxA):

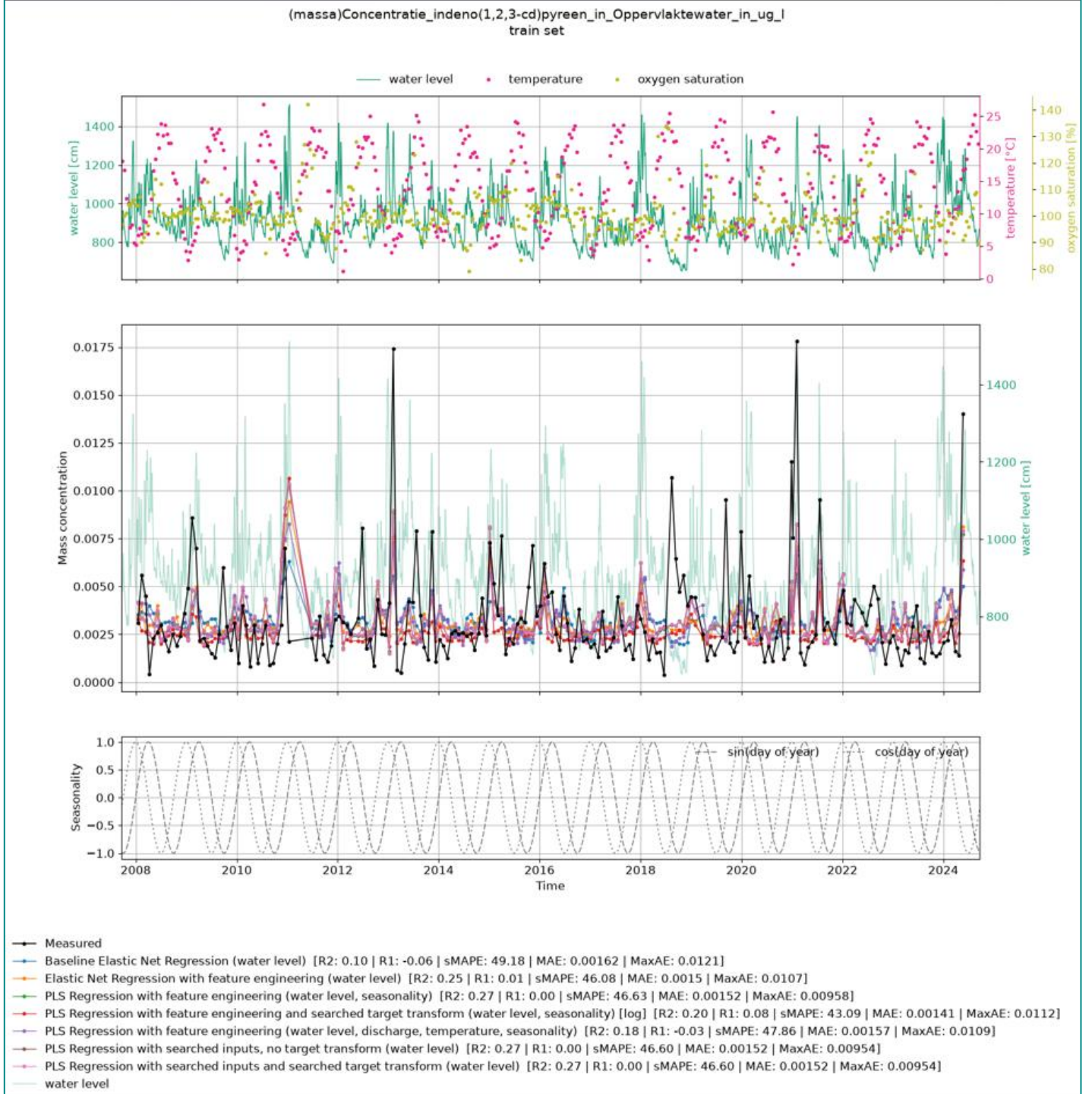
- Priority substance (PFAS)
- High mobility, low log$K_{OC}$
- Industrial substance
- Higher concentrations during lower water levels
- Lower concentrations during higher water levels



# Results

## Indeno(1,2,3-cd)pyrene:

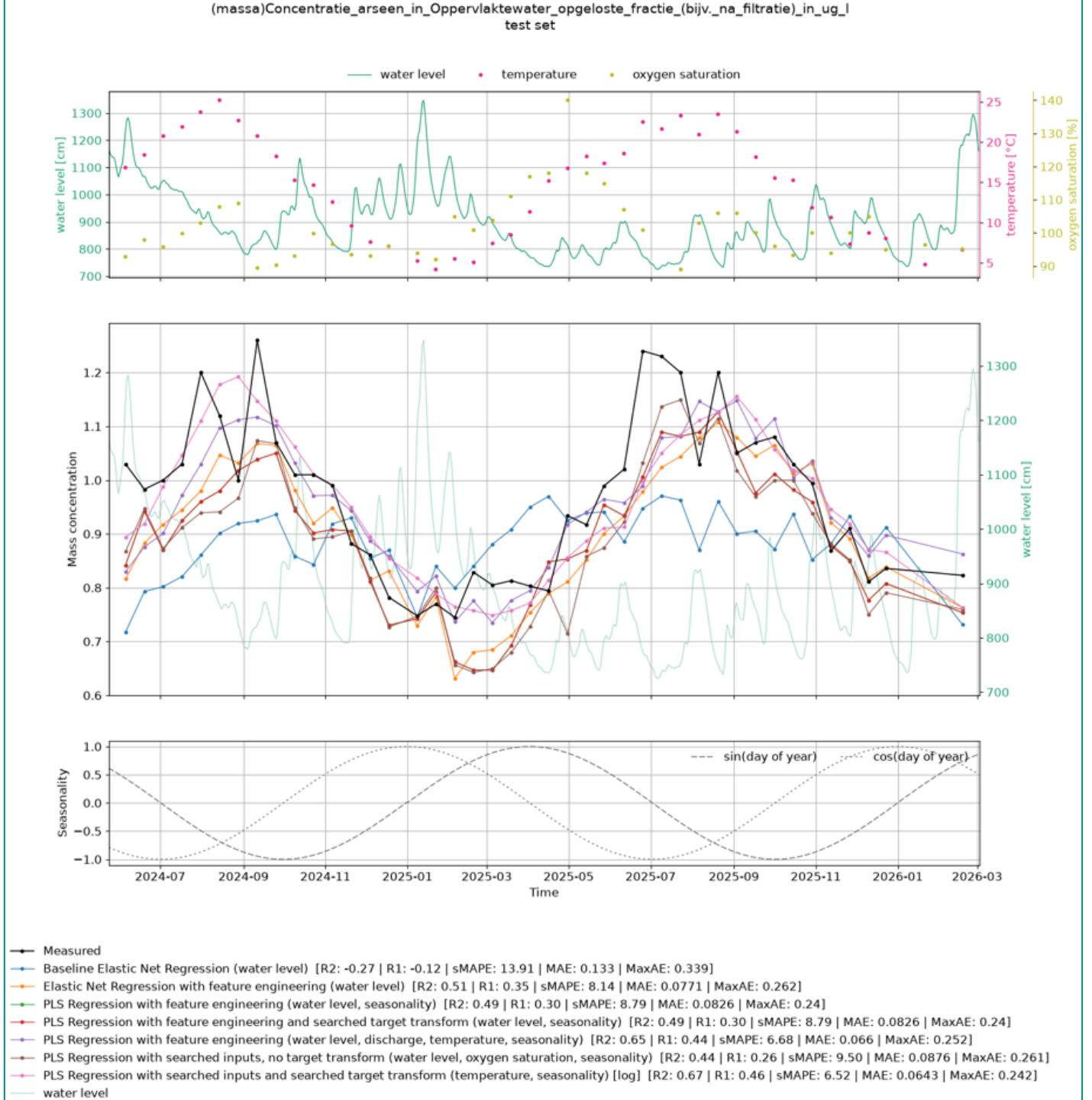
- Priority substance (PAH)
- Low mobility, high log$K_{OC}$
- Combustion product
- Higher concentrations during higher water levels
- Lower concentrations during lower water levels



# Results

## Arsenic (non-priority substance):

- Heavy metal
- Bioaccumulative, persistent
- Low mobility
- Higher concentrations during lower water levels
- Lower concentrations during higher water levels



# Preliminary conclusions

- Promising method for using a ML-model:
  - to predict intermediate concentrations between actual field measurements
  - to predict water quality based on regular forecasting water quantity measurements
- The choice of a ML-model should be tailored to specific water quality parameters and study area.
- Additional information from ML-model potentially supports policy officers in decision-making and annual reporting for WFD-monitoring

# Limitations and disclaimer of machine learning models

- Model outputs are predictions, not measurements
- Predictions contain uncertainty and its performance may vary over time
- Predictions should support, not replace, expert judgement
- Regular retraining and validation are required
- Predictions are reliable only within the range represented in the training data (e.g. no outliers or extreme events or industrial discharges)

# Outlook

- Testing of the ML-models on different locations, such as the Schelde, Meuse and other MWTL locations
- Further differentiation in solvability of prioritized substances
- Predictions of water quality in the Rhine based on regular water quantity forecasts of Rijkswaterstaat in Q4 of 2026

# Colofon


- Joost van Dijk
- joost.vandijk@nelen-schuurmans.nl