Sustainable Water Management

What can we learn by Data Exploration and Modelling

Motivation

- Personal: Understanding time-series better

- get

Location

Petrignano

Data Source:
https://www.kaggle.com/c/acea-water-prediction/overview

Data

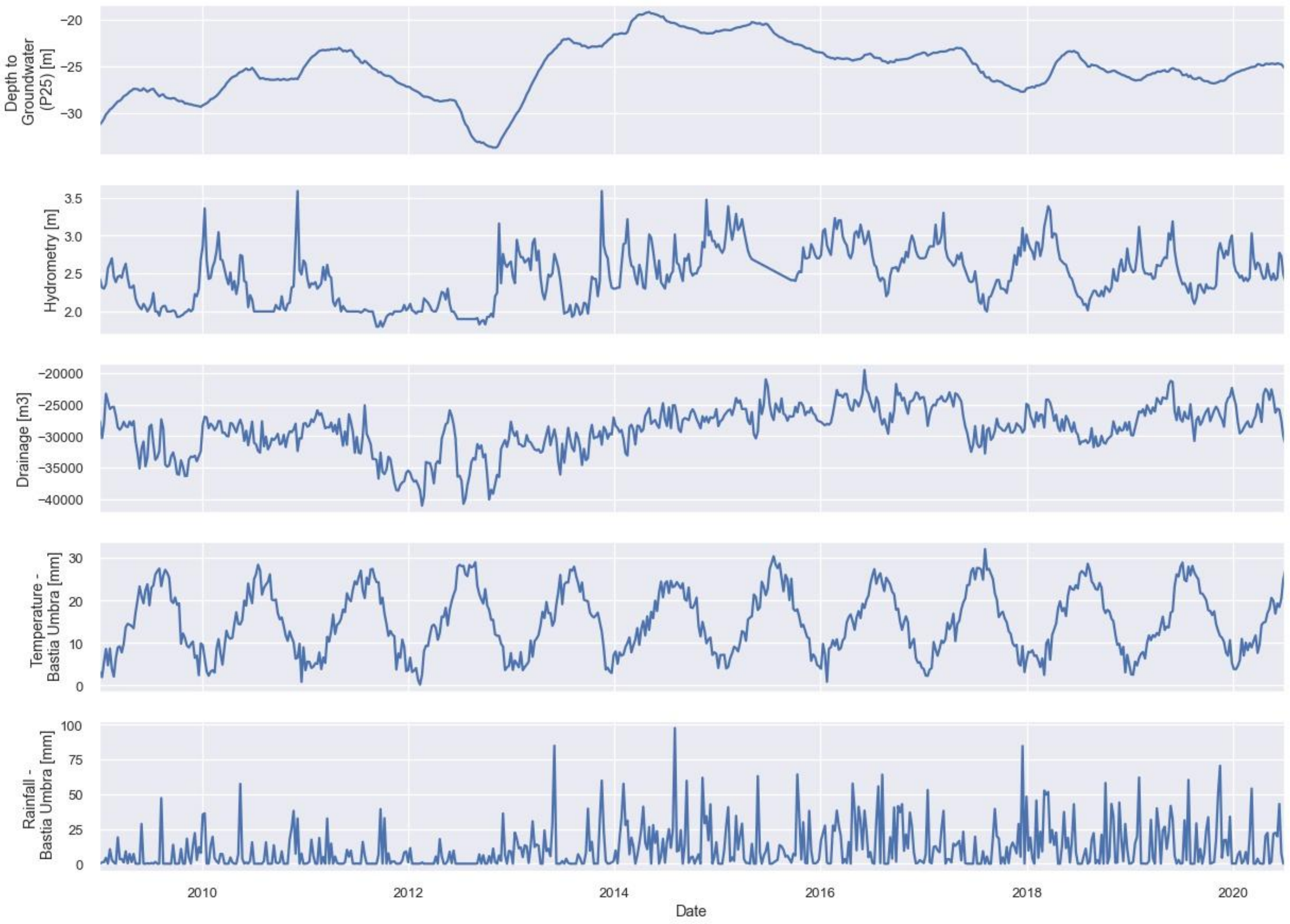
Daily measurements

Y: Depth to Groundwater [m]

X: Drainage, Hydrometry, Rainfall, Temperatur

→ **Downsampled to weekly data to reduce noise**

Data Source: https://www.kaggle.com/c/acea-water-prediction/overview



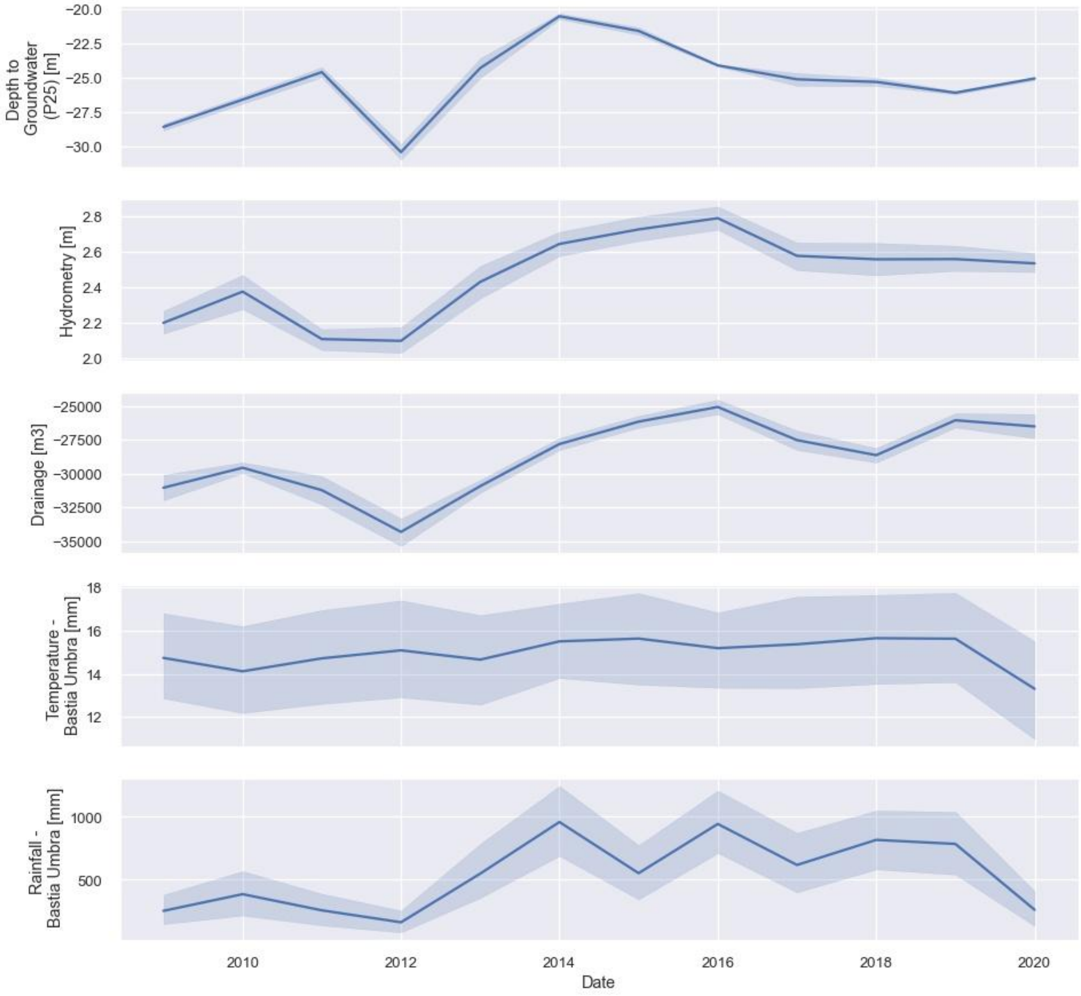
Trends in Data

Depth to Groundwater is following:

- Drainage (Depleting)
- Hydrometry (Refilling)
- Rainfall (Refilling)

Negative Spike in Groundwater level follows drought period and an increased demand of water

Recovery of Groundwater due to refilling (Rainfall, River water) and reduction of water extraction (water saving?)

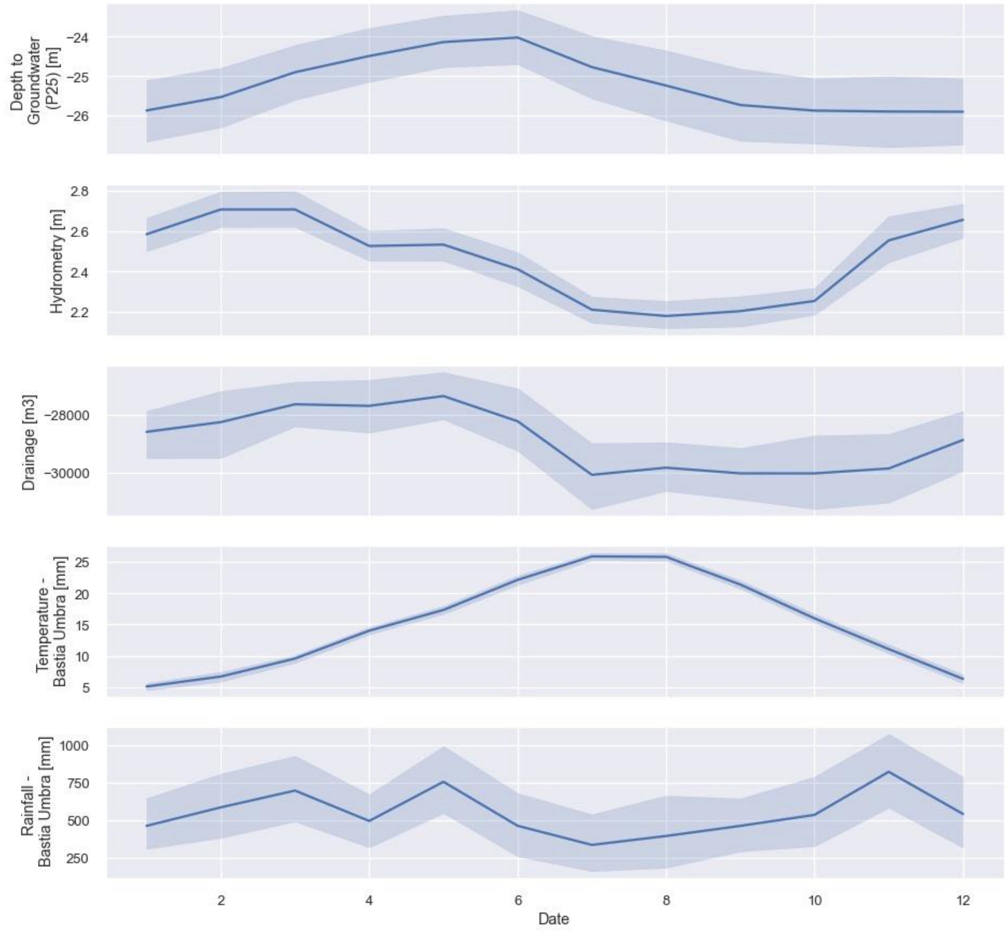


Seasonality in Data

Depth to Groundwater:
- Refilling in Spring
- Depleting in Summer
- Min: Nov / Max: June

Lagging rel. Groundwater Level:

- Hydrometry: 3 Month ahead
- Drainage: **2 Month ahead**
- Temperature: 2 Month behind



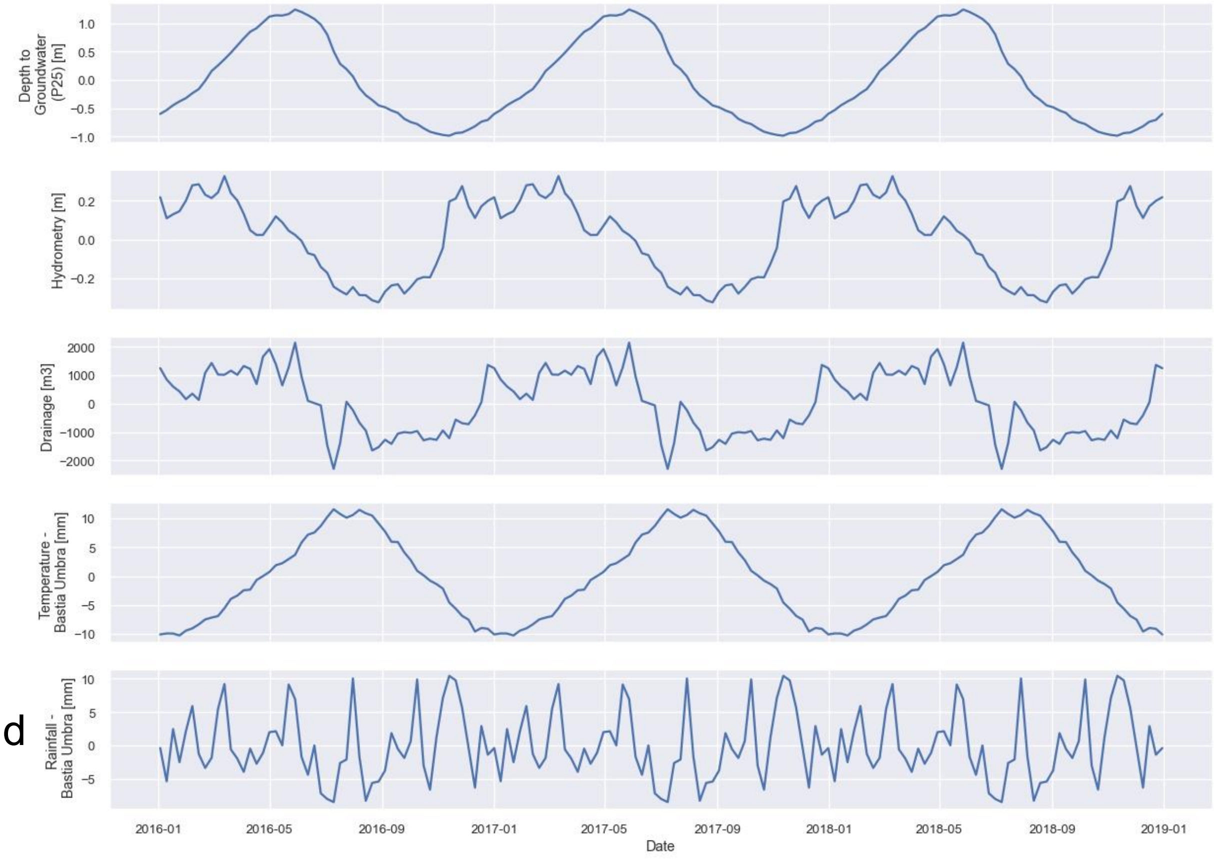
Data - decomposed seasonality

Decomposition (statsmodels):

- To get seasonality subtract trend with yearly moving average

Lagging rel. Groundwater Level:

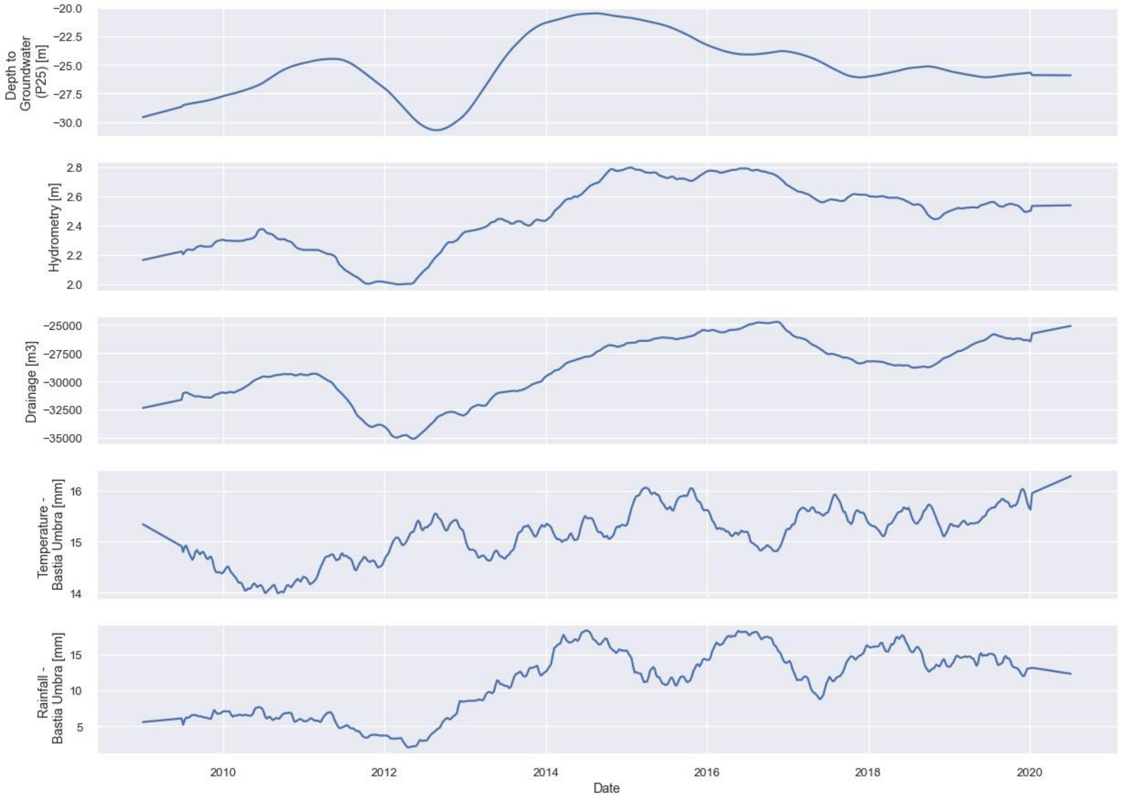
- Hydrometry: 3 Month ahead
- Drainage: **2 Month ahead**
- Temperature: 2 Month behind



Data - decomposed yearly trend

Lagging rel. Groundwater Level:

- Hydrometry: 6 Month ahead
- Drainage: **4 Month ahead**
- Rainfall: 6 Month ahead



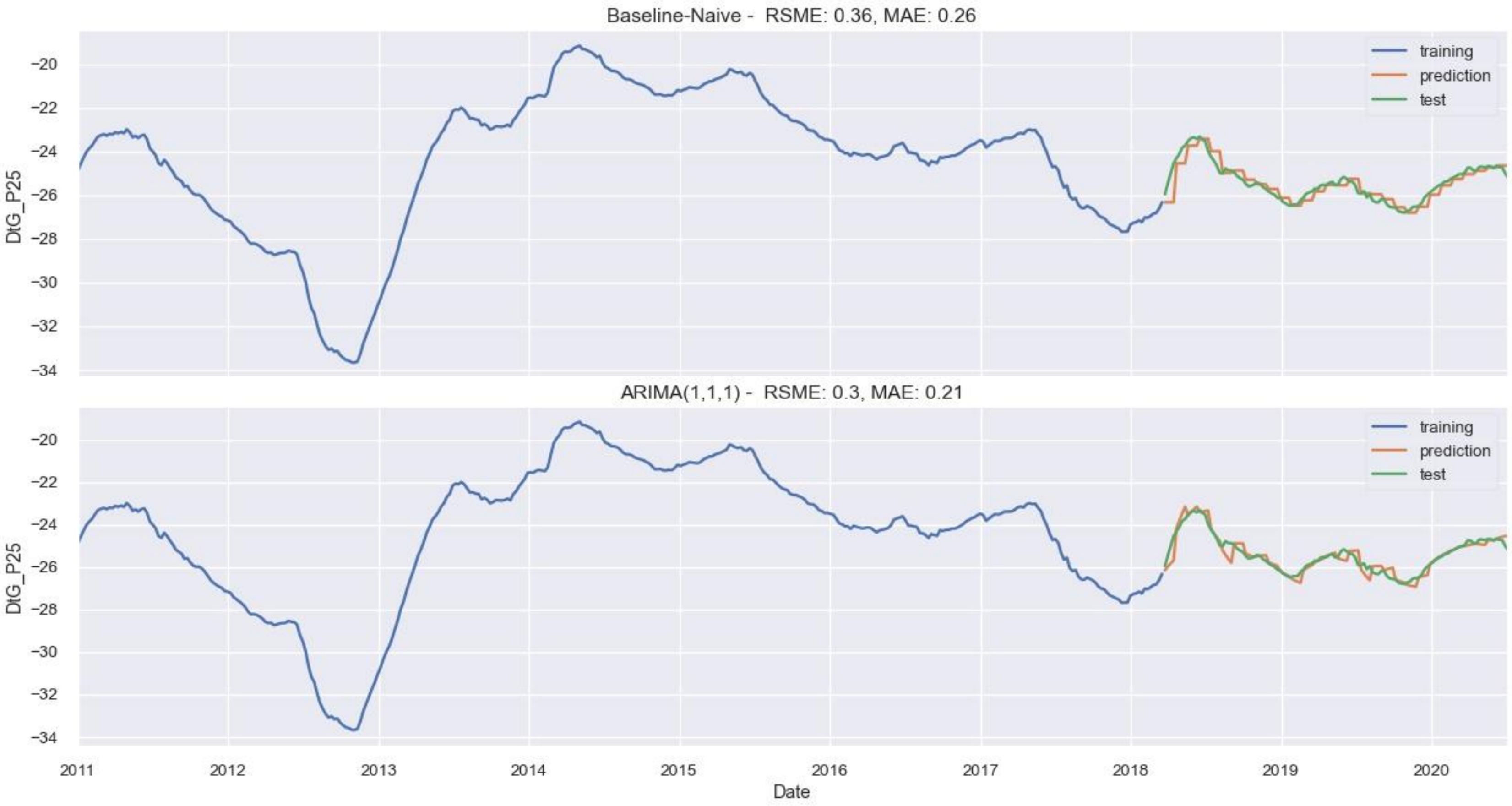
Predicting - Water level

- Aim: Predicting Depth to Groundwater [m] for the next 4 weeks
  → Walk-forward prediction with 4 weeks steps
- Train: 481 weeks (ca. 9.3 years), Test: 120 weeks (ca. 2.3 years)

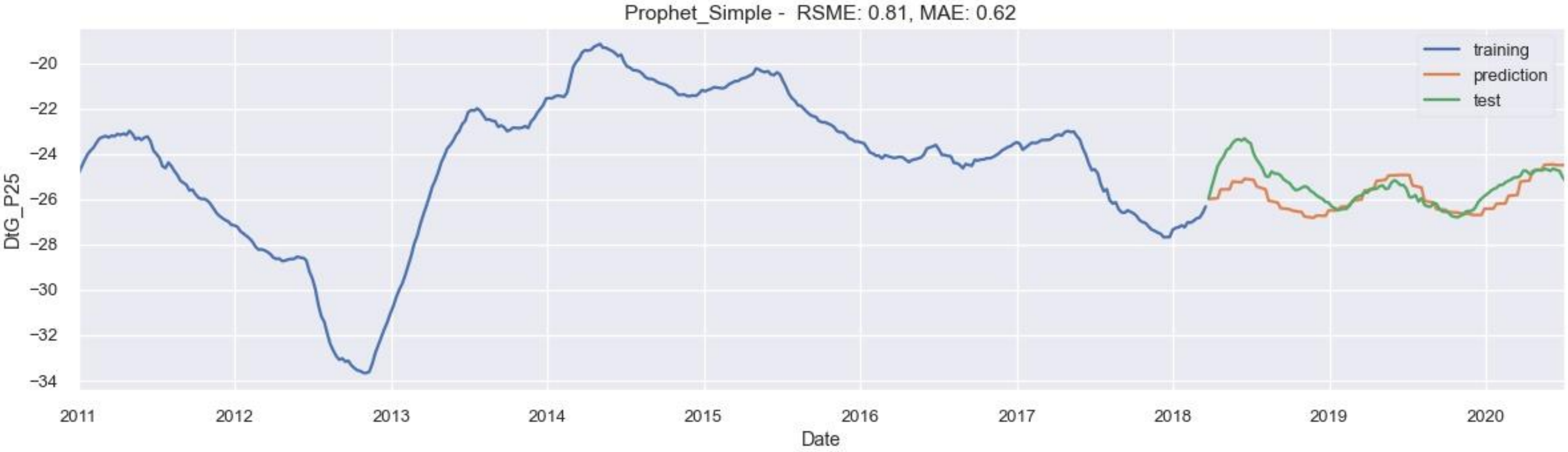
- Models - univariate:
    - Naive Approach
    - ARIMA
    - Facebook Prophet

- Model - multivariate:
    - Catboost Regressor

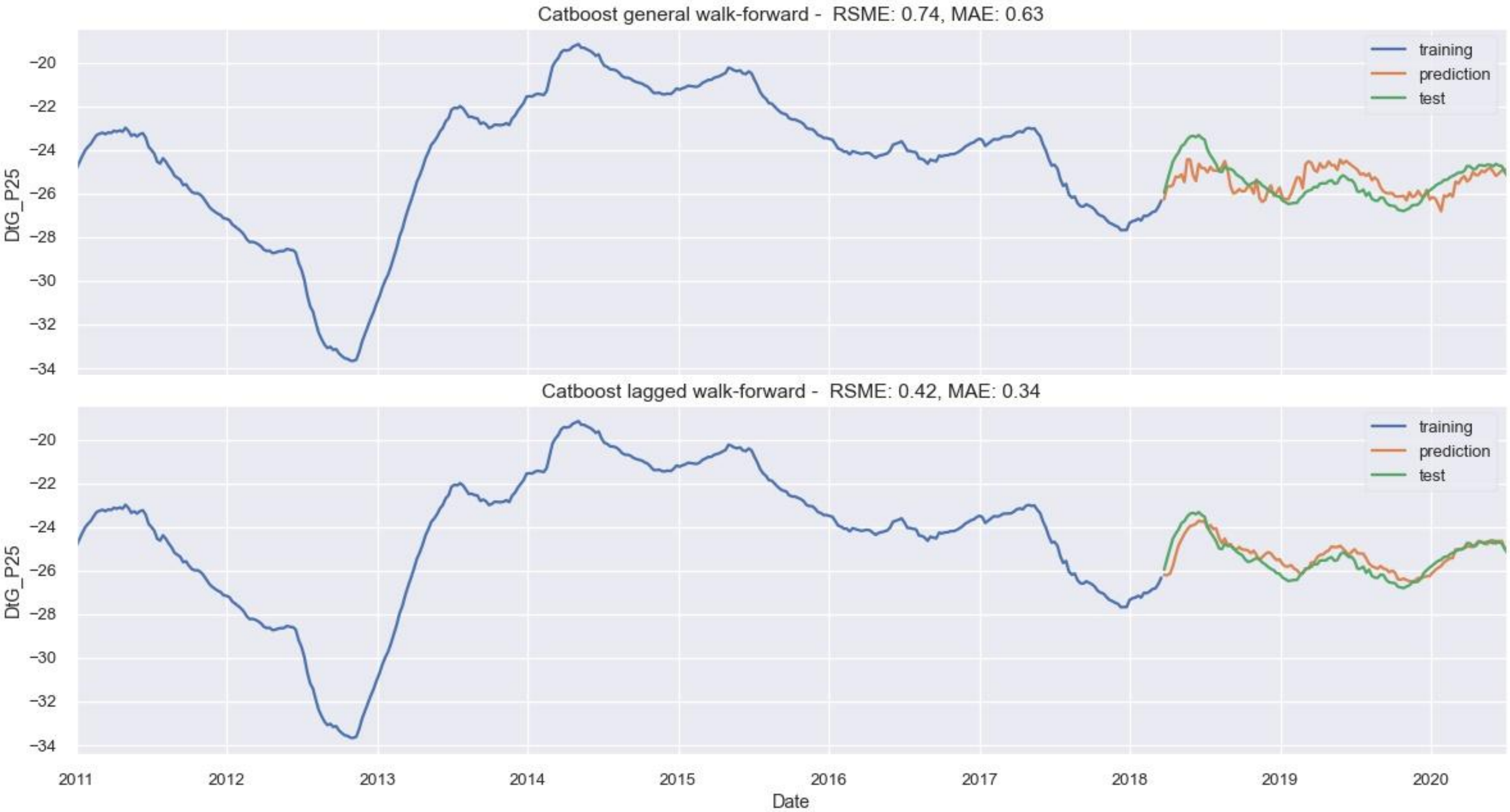
Model - Naive / ARIMA



Model - Prophet univariate

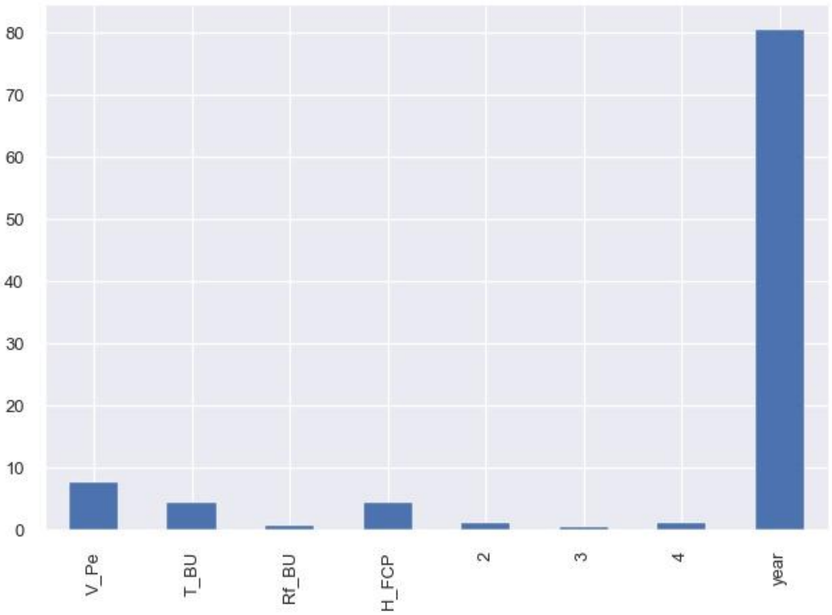


Models - multivariate

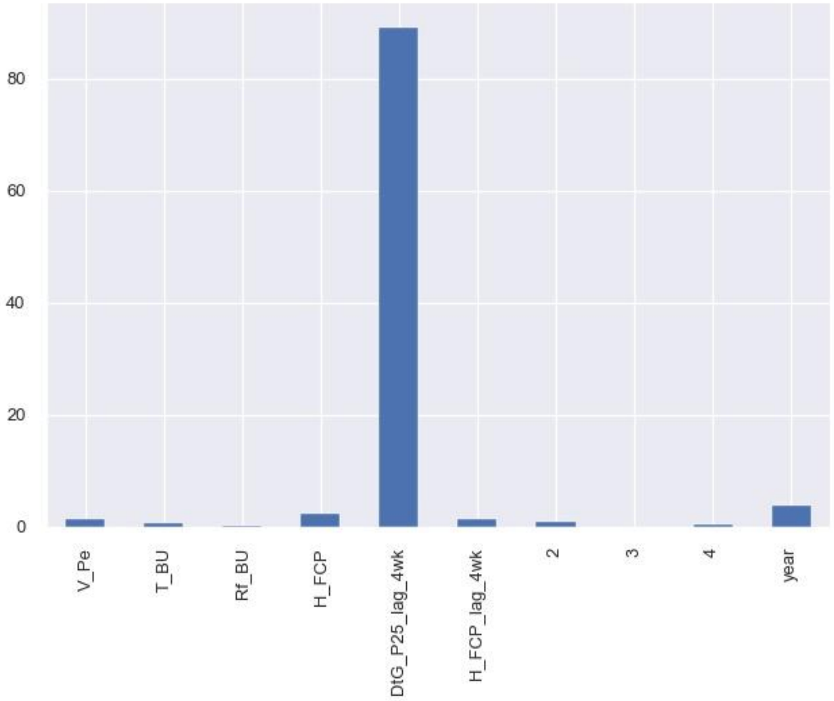


Models - multivariate

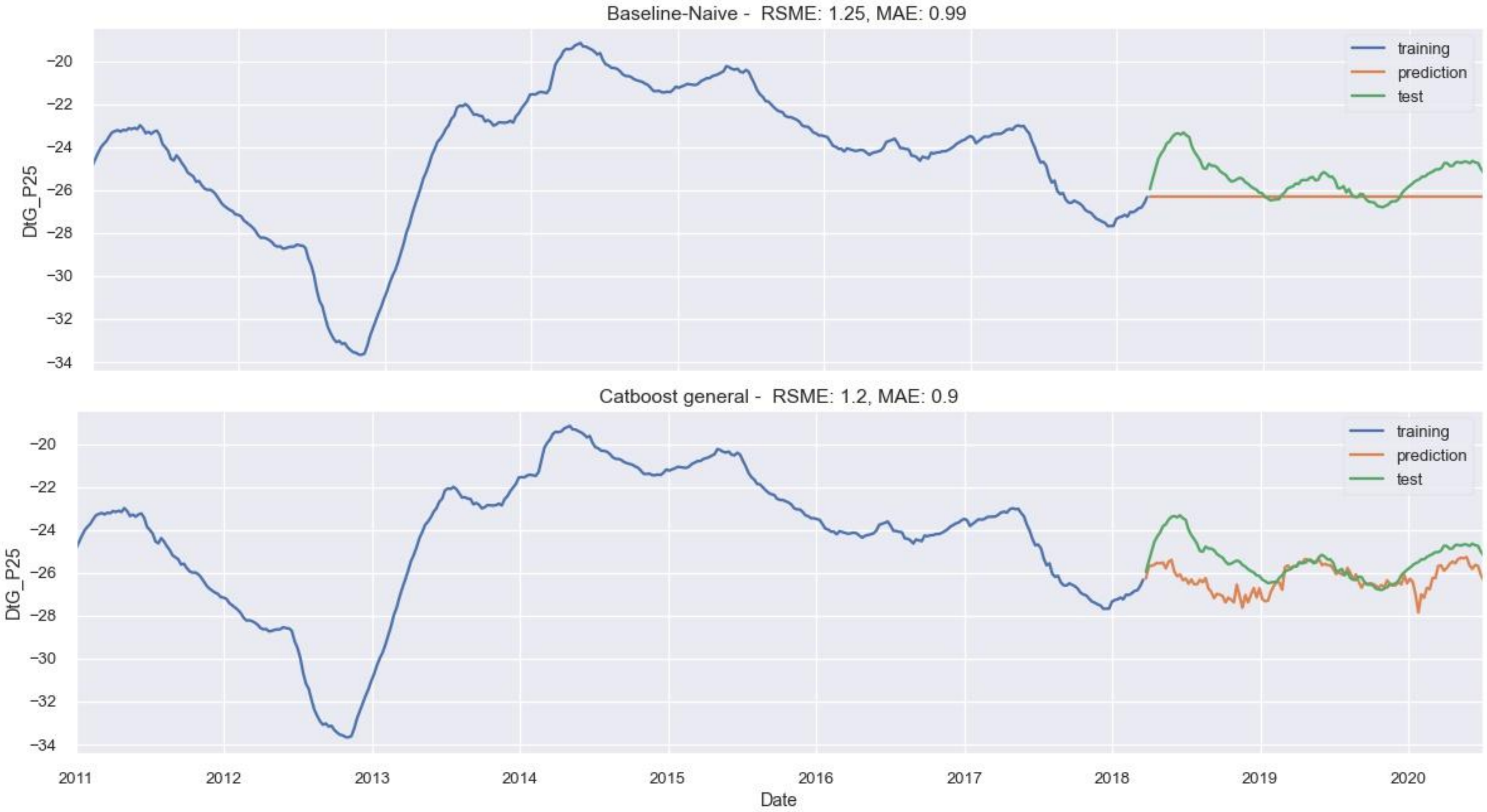
Feature Importance - Catboost general



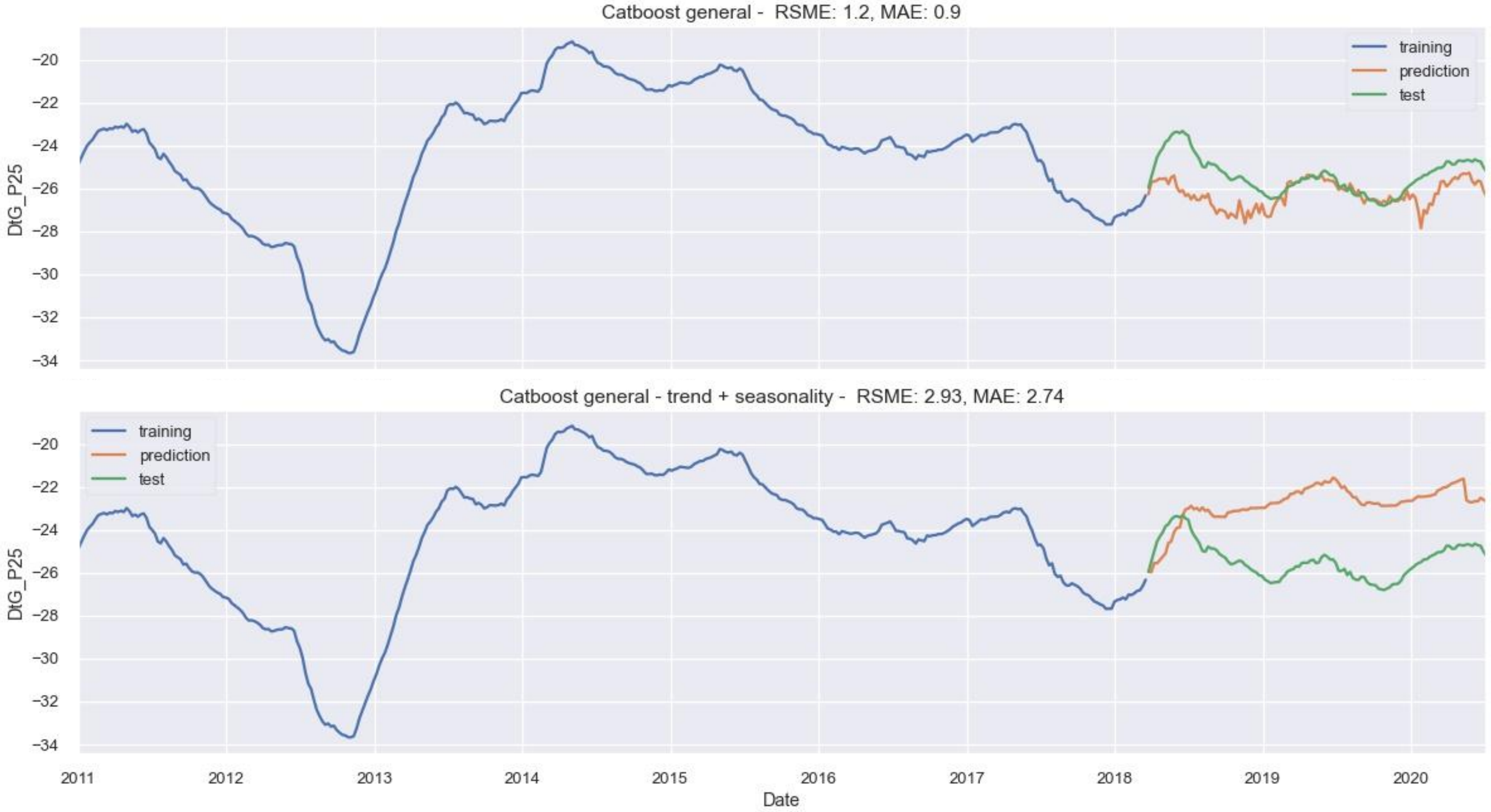
Feature Importance - Catboost 4 weeks lag



Models (without retraining model)



Models (trying trend and seasonality of features)



Conclusion

Groundwater level is best model by knowing previous time-steps (lags) or using walk-forward prediction

Solely prediction on X-features is possible but is not precise

Important Features:

1. Drainage
2. Temperature
3. Hydrometry / Rainfall

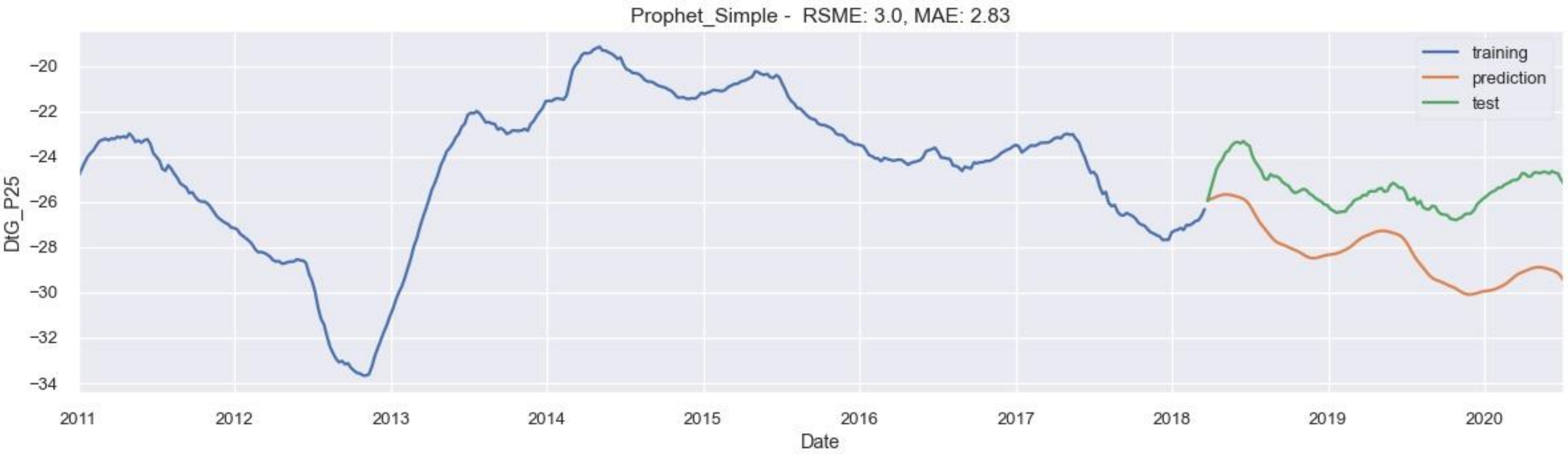
Further work: Find a better way to teach the model feature dependence inferred from EDA / Try LSTM

Model	**RMSE**
Naive	0.36
ARIMA	0.30
Prophet	0.62
Catboost	0.74
Catboost with 4w lag	0.42

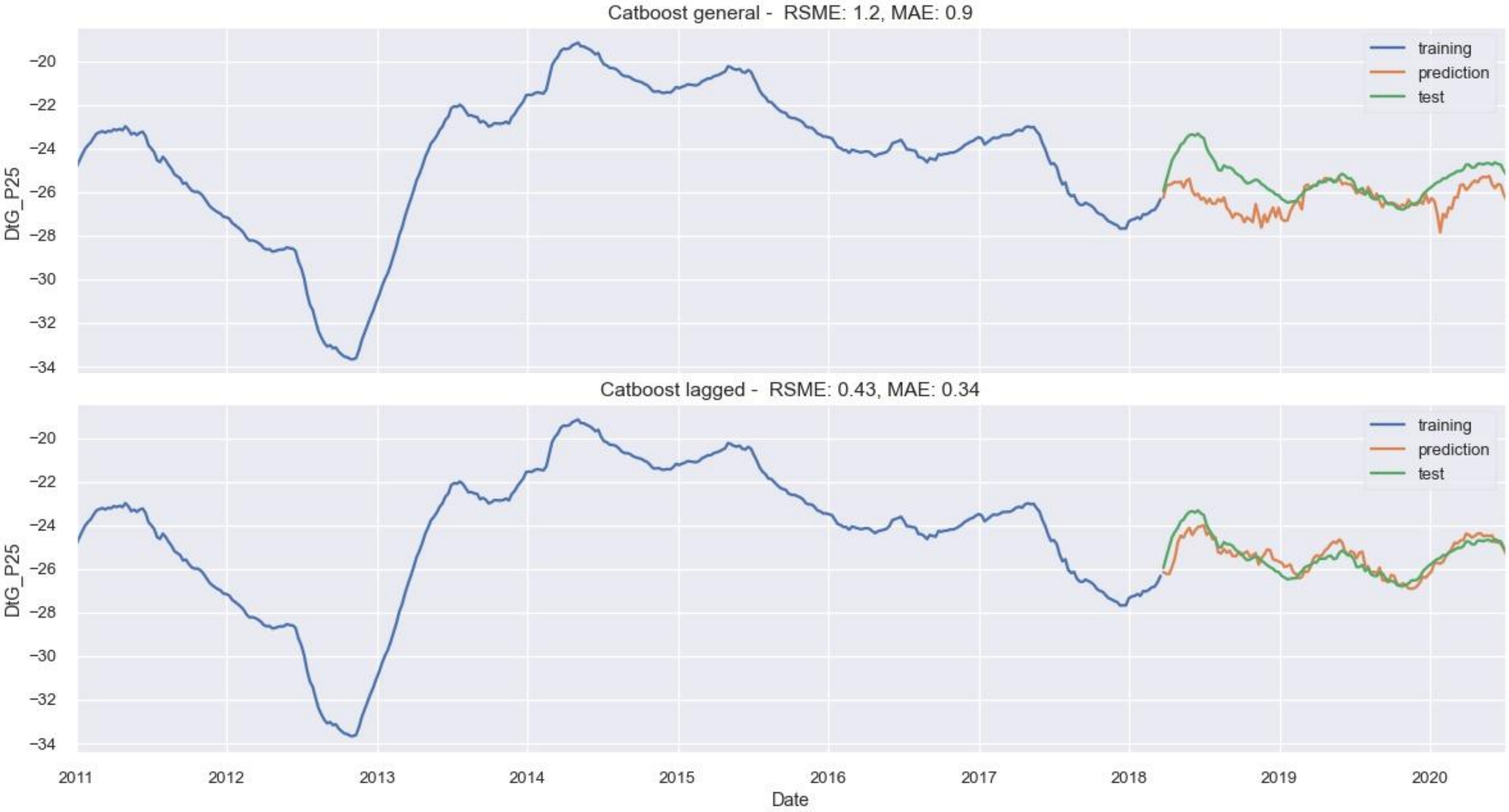
A big thanks to SPICED Academy

- Especially to Tom, Gesa, Stefan, Kristian, Ugur
- My fellow a-star-anises
- Spiced Academy

Model - univariate



Models - multivariate



Data

Daily measurements

Y: Depth to Groundwater [m]

X: Drainage, Hydrometry, Rainfall, Temperatur

Missing values: Imputed with interpolation

Only one Y target used: P25