

# CHAIN

# AFSLUTNINGSSSEMINAR

## 4. OKTOBER 2021

Shuna R. Nazari

AI specialist

Sammen kommer vi #**ForanDigitalt**





# What have we been working with:

ML Modelling:

- Prognosis and Forecast
- Anomaly detection
- Prepossessing



# **Prognose and forecast**



# Prognose and forecast

What is it and what is it used for ?

- Mapping of water consumption patterns
- To predict future events
- Prediction of measurements based on historical data
  - Historical measurements, day, date, time of day, seson..

Investigations of models in collaboration with Kampstrup

- SARIMAX
- Gradient Boosting
- Seasonal model



# Prognose and forecast

Models: SARIMAX , Gradient Boosting (GB) and Seasonal model

Pros:

SARIMAX: Accurate, unlimited  
forecasts

GB: Fast and Accurate

Seasonal: Robust

Cons:

SARIMAX: Takes long time, needs  
continuous timeseries

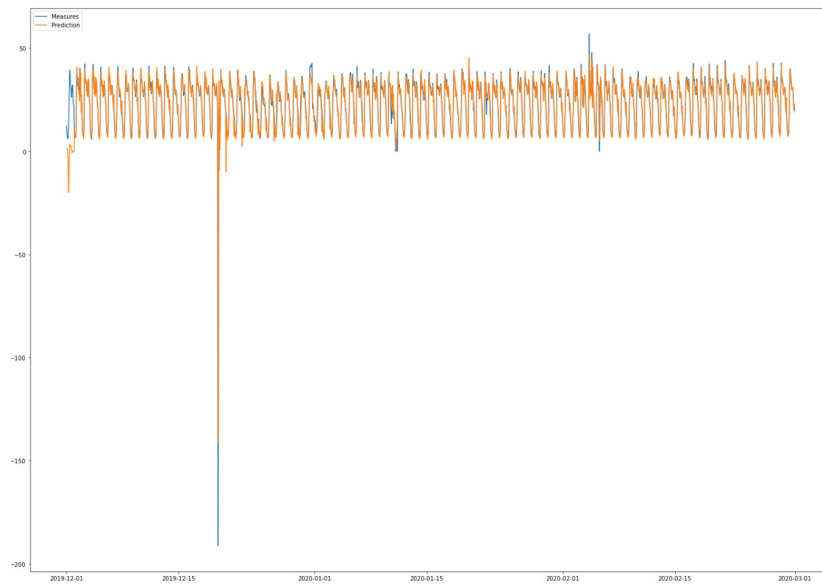
GB: Training on in- and output, may be  
affected by that

Seasonal: Can not handle new  
behavior

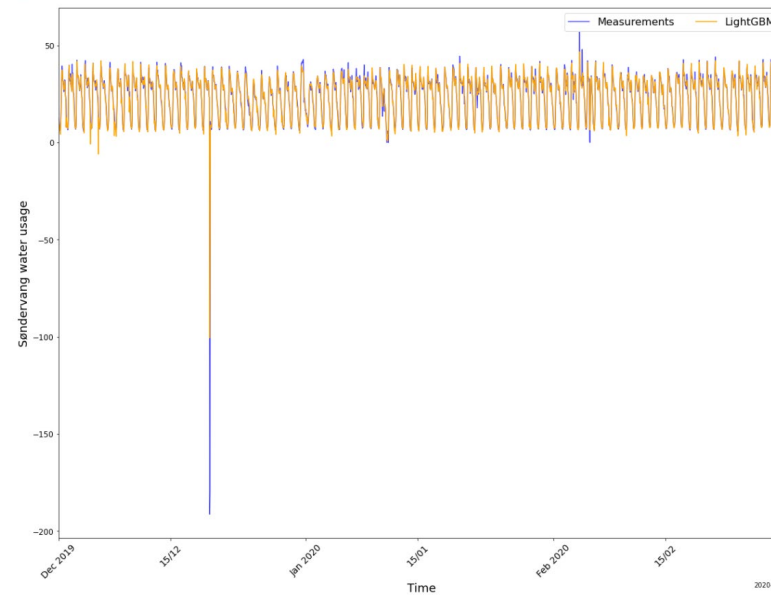
# Prognose and forecast

SARIMAX and GB can easily get affected by outliers/anomalies. Preprocessing is important!

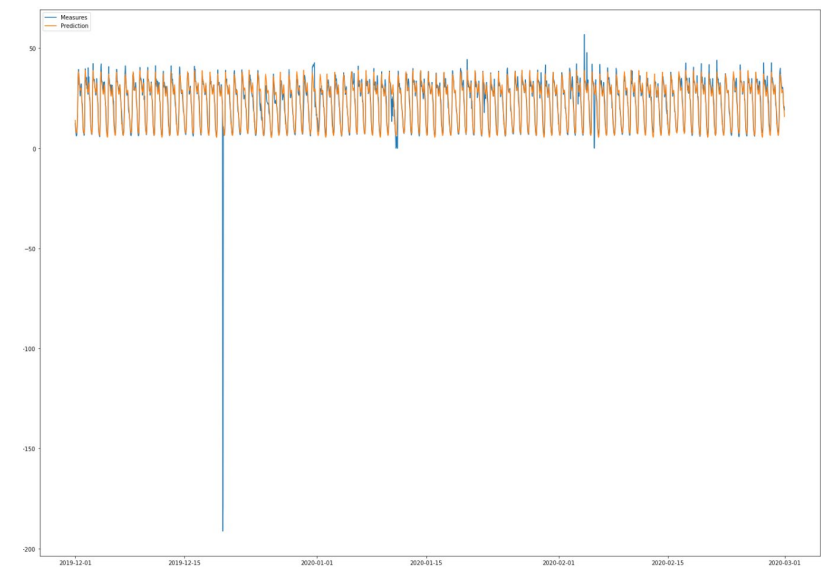
## SARIMAX Train



## Gradient Boosting Train

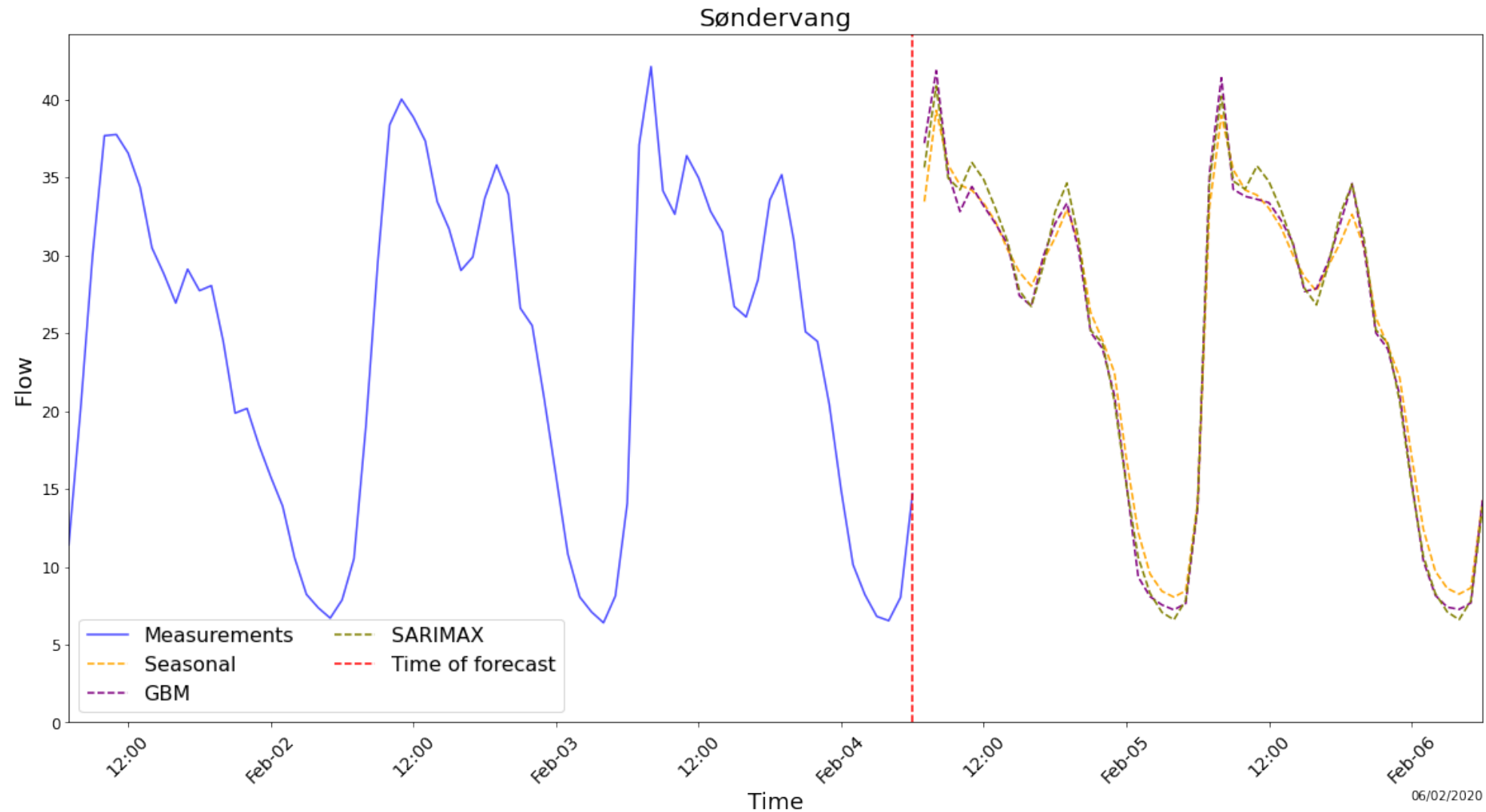


## Seasonal Train



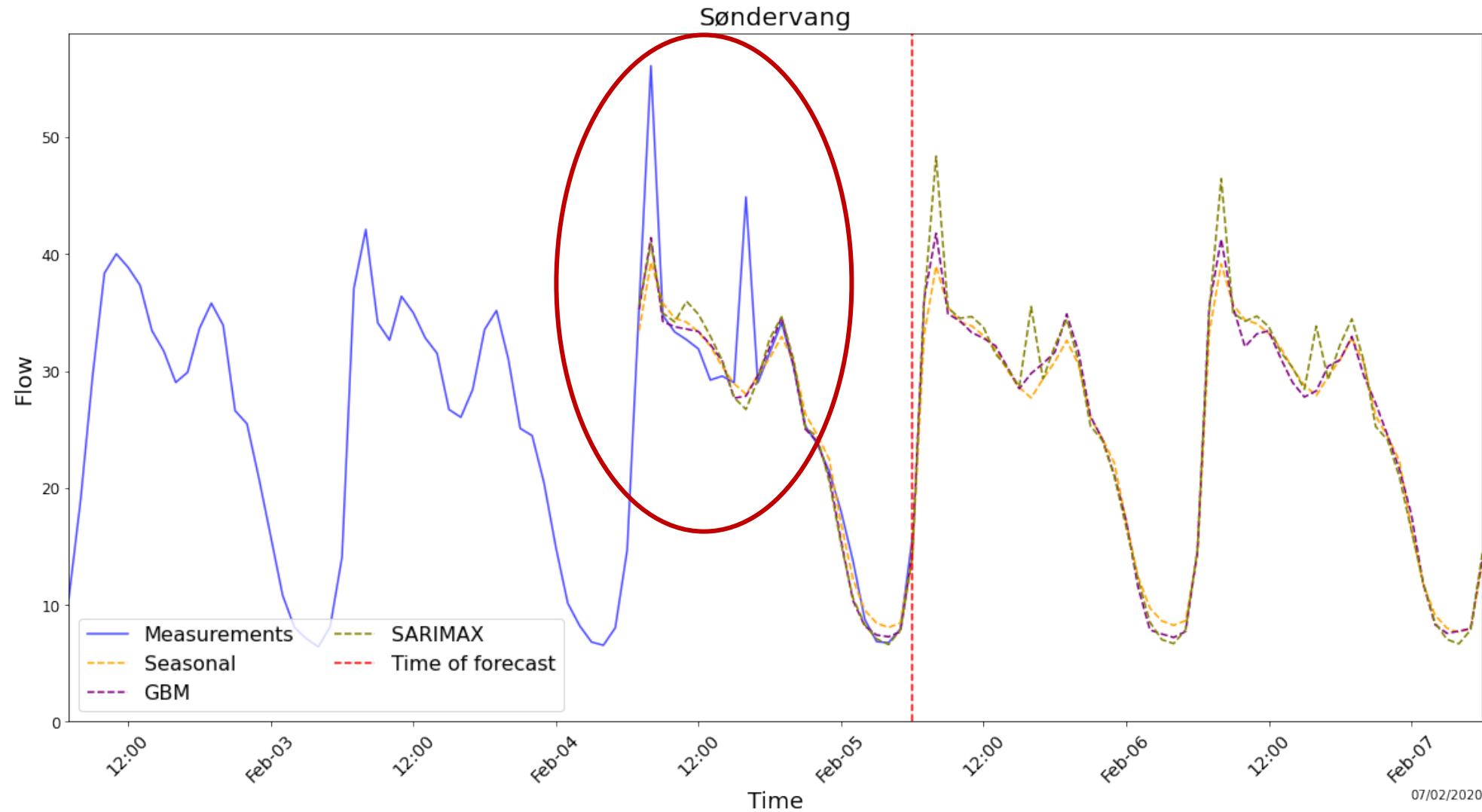
All units  
in  $m^3/h$

# Example single time of forecast



All units  
in  $m^3/h$

# Example for single time of forecast







# **Anomaly detection**

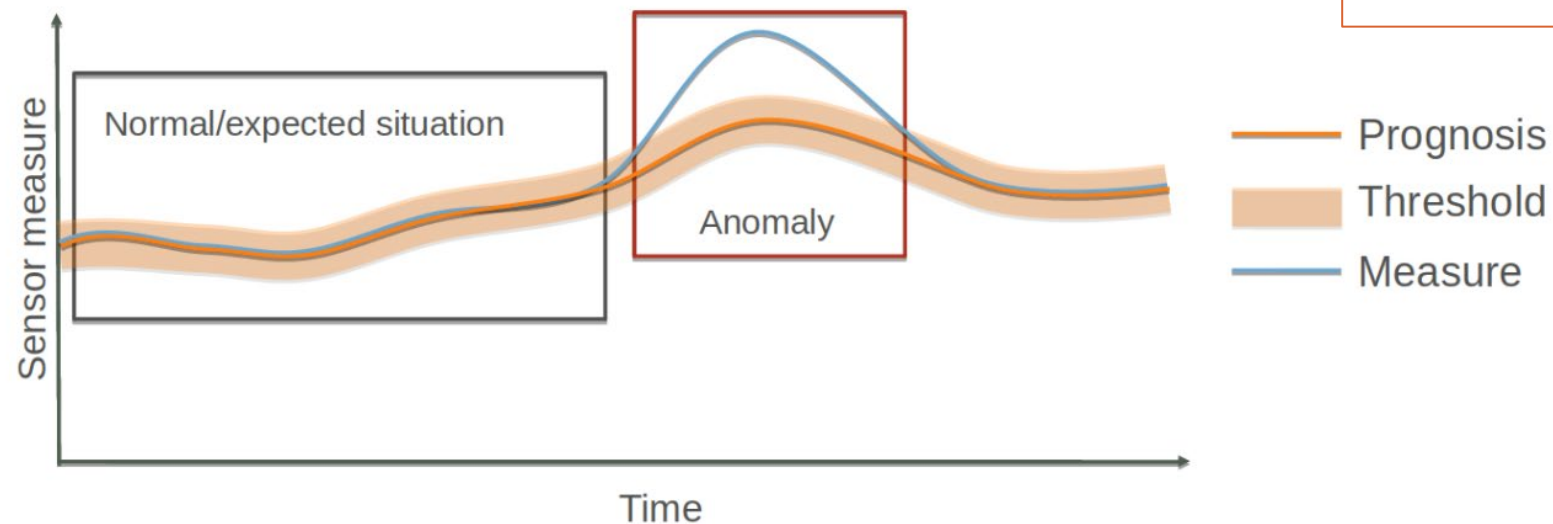
# Anomaly detection

To find events that are unusual

- Looks at the measurements and Prognosis
- Different ways find anomalies
- Preprocessing of data

Reasons for anomalies:

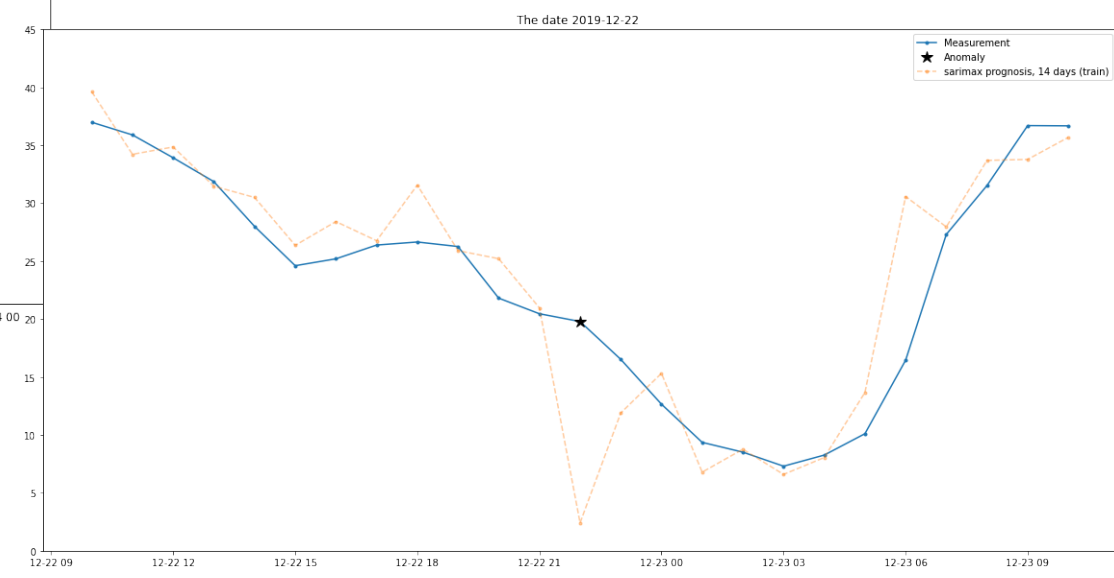
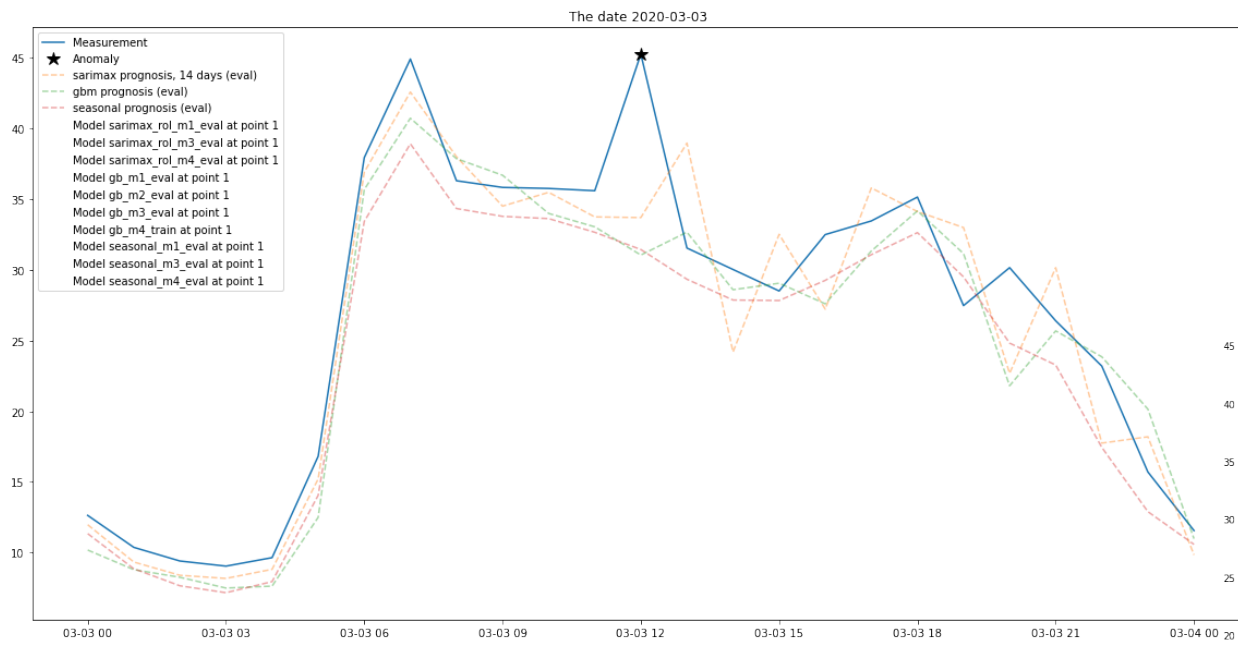
- Wrong prognosis
- Sensor errors not removed
- Unusual amount of flow (high/low)
- Unusual amount of flow (high/low) at a specific time





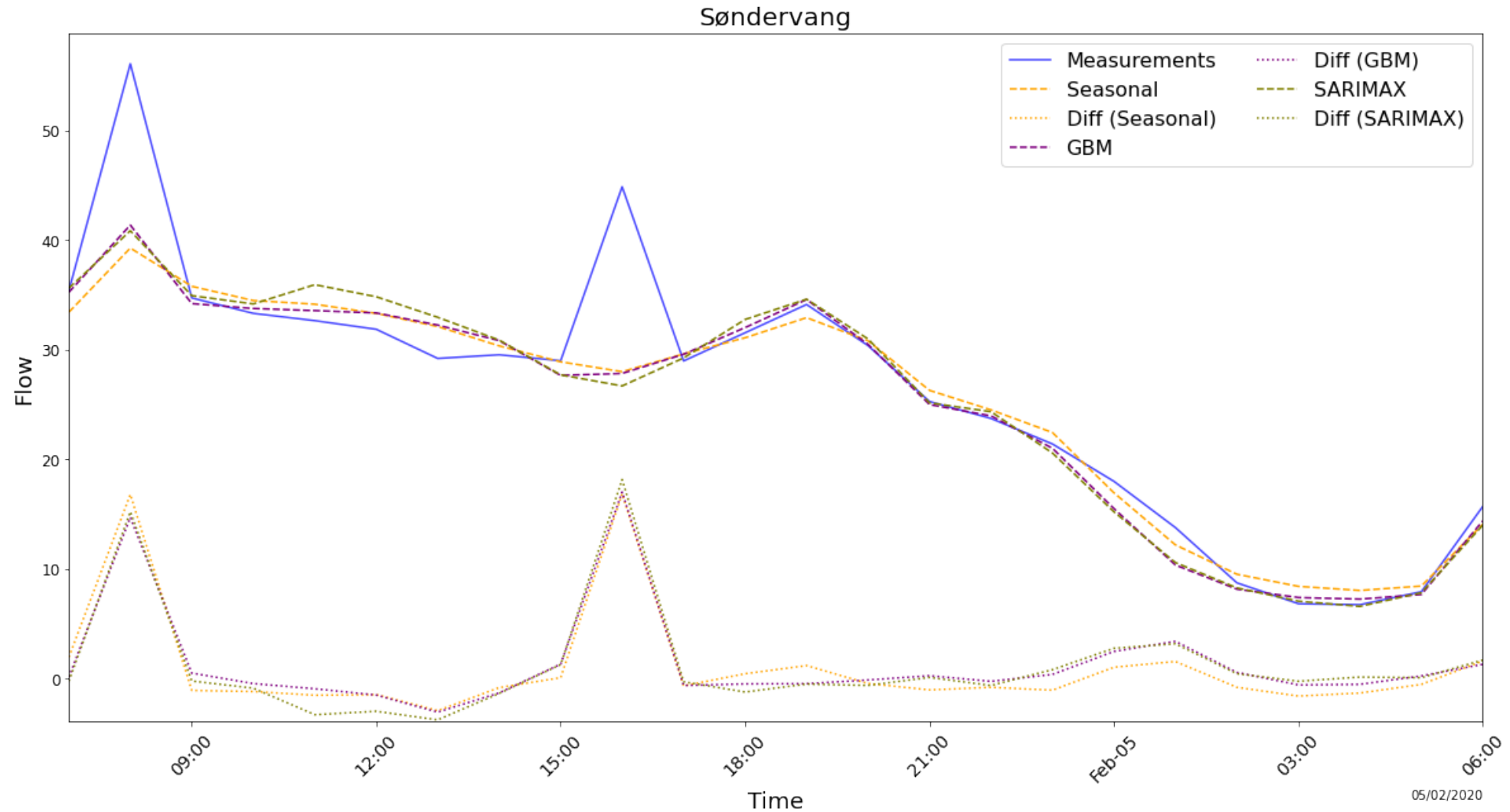
# Exampler of anomalies

All units  
in  $m^3/h$



All units  
in  $m^3/h$

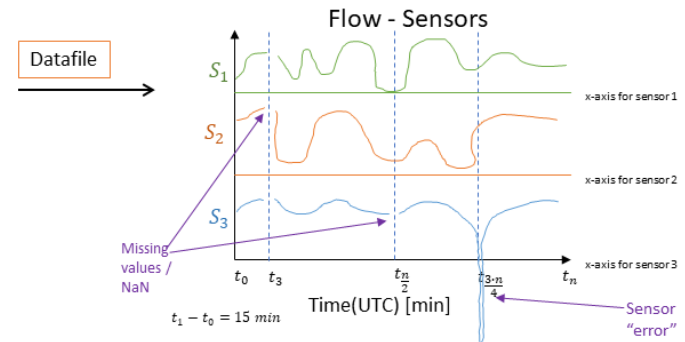
# Anomalis via differences





# Preprocessing

# Preprocessing



To CET / CEST

Continuous index, Missing values to NaN

Time index	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>
t <sub>0</sub>	0,8	0,9	0,6
t <sub>1</sub>	1	1,2	0,7
t <sub>2</sub>	1,5	1,4	1
t <sub>3</sub>	0,5	1,3	0,8
⋮	⋮	⋮	⋮
t <sub>n/2</sub>	0,2	0,4	NaN
⋮	⋮	⋮	⋮
t <sub>3-n/4</sub>	1	1,2	-50
⋮	⋮	⋮	⋮
t <sub>n</sub>	1	1,5	1,2

Fill missing

Seasonal model

1 hour resolution

Fill missing (interpolation)

Time index	d <sub>1</sub>
t <sub>0</sub>	2,2
t <sub>3</sub>	3,13
⋮	⋮
t <sub>n/2</sub>	1,8
⋮	⋮
t <sub>n</sub>	3

Cleaning data/ remove sensor "error"

Convert to district

Time index	S <sub>1</sub>	S <sub>2</sub>	S <sub>3</sub>
t <sub>0</sub>	0,8	0,9	0,6
⋮	⋮	⋮	⋮
t <sub>3-n/4</sub>	1	1,2	NaN
⋮	⋮	⋮	⋮
t <sub>n</sub>	1	1,5	1,2

Time index	d <sub>1</sub>
t <sub>0</sub>	2,2
t <sub>1</sub>	2,9
t <sub>2</sub>	3,9
t <sub>3</sub>	NaN
⋮	⋮
t <sub>n/2-3</sub>	NaN
⋮	⋮
t <sub>n/2-2</sub>	NaN
⋮	⋮
t <sub>n/2-1</sub>	NaN
⋮	⋮
t <sub>n/2</sub>	NaN
⋮	⋮
t <sub>3-n/4</sub>	NaN
⋮	⋮
t <sub>n</sub>	3,7