Machine Learning Based Control System Design for BSM1 Wastewater Treatment Plant

Yu Yang, Chengyan Ye

Presenter: Chengyan Ye

General overview of the BSM1 plant

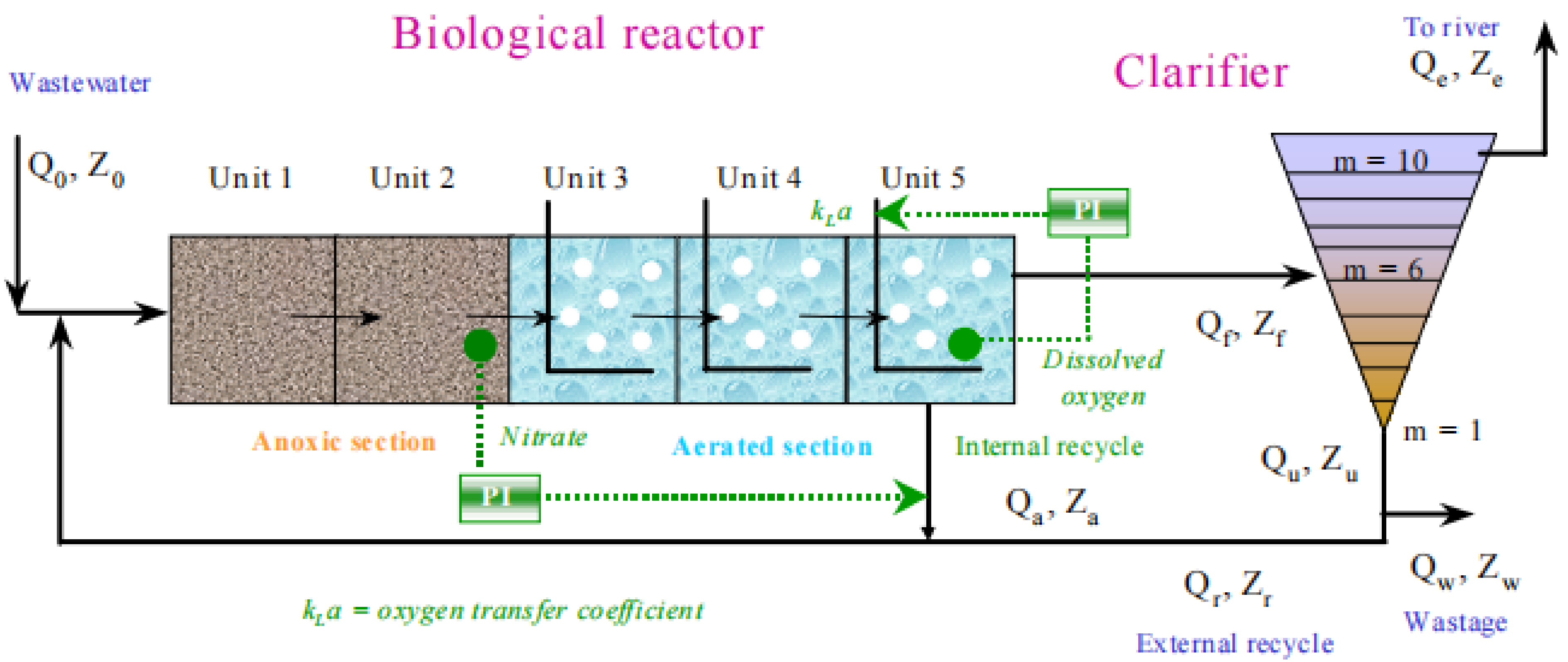
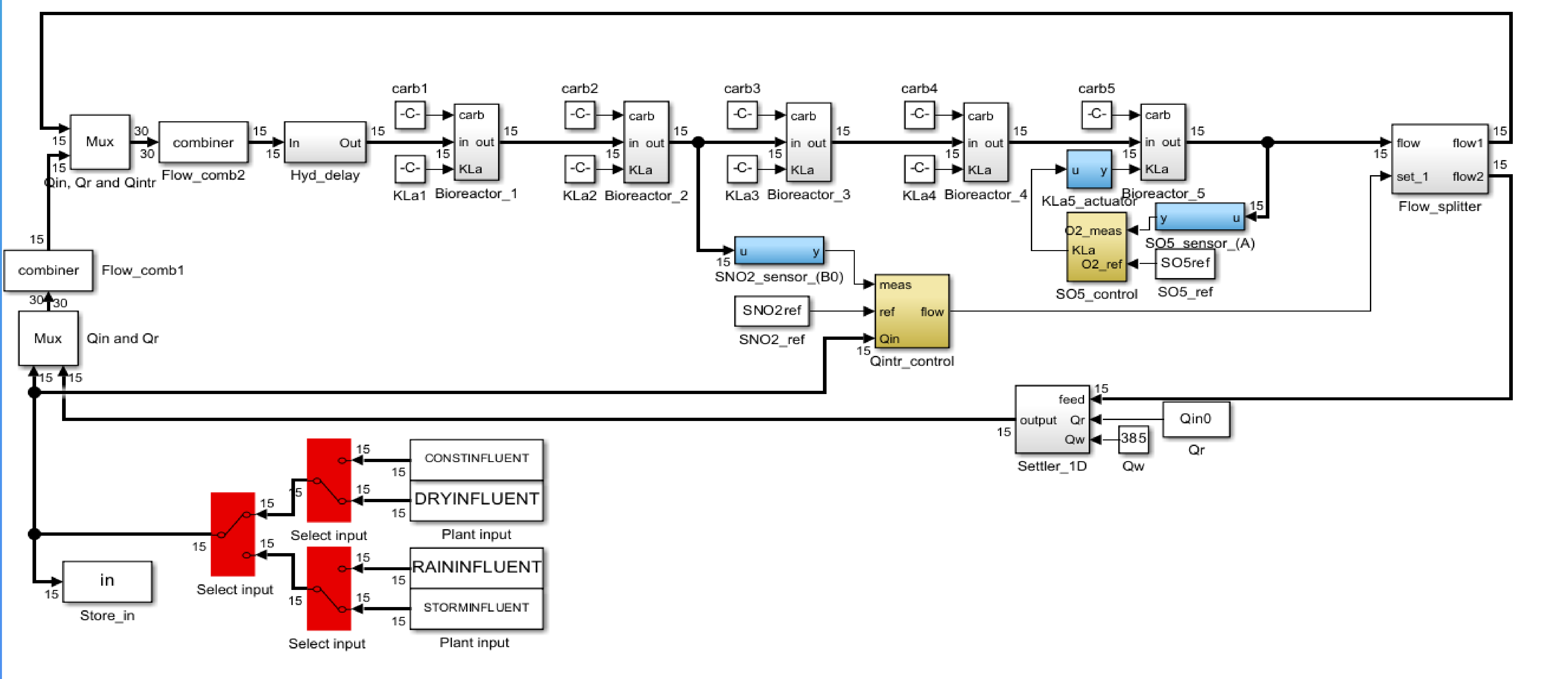


Figure 1: General overview of the BSM1 plant

BSM1 MATLAB SIMULINK OUTLINE



Effluent quality violations

Variable	Value
N_{tot}	<18 g N.m^{-3}
COD_{tot}	<100 g COD.m^{-3}
S_{NH}	<4 g N.m^{-3}
TSS	<30 g SS.m^{-3}
BOD_5	<10 g BOD.m^{-3}

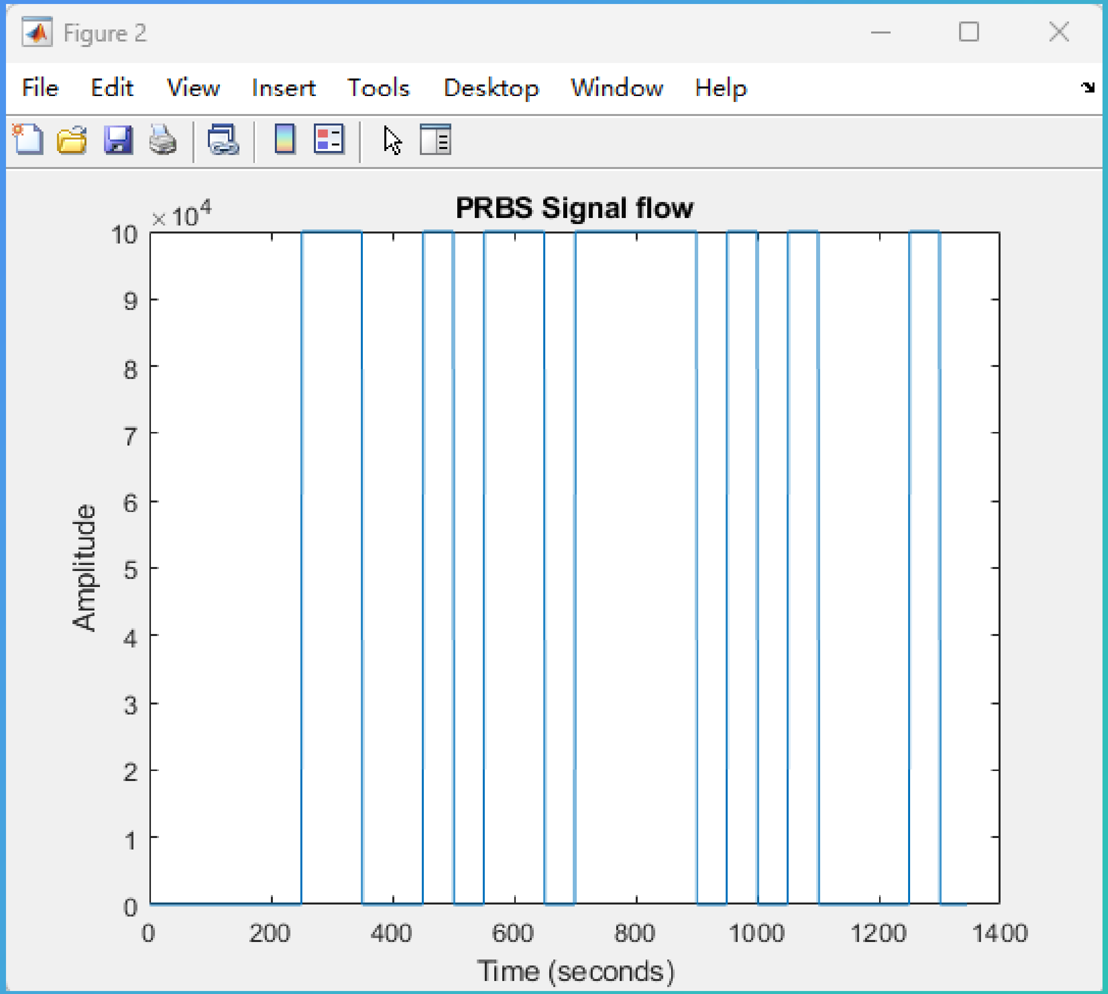
Effluent quality limits

```
Effluent violations
-------------------
95% percentile for effluent SNH (Ammonia95) = 10.6066 g N/m3
95% percentile for effluent TN (TN95) = 20.1673 g N/m3
95% percentile for effluent TSS (TSS95) = 20.9932 g SS/m3

The maximum effluent total nitrogen level (18 mg N/l) was violated
during 2.1979 days, i.e. 31.3988% of the operating time.
The limit was violated at 9 different occasions.

The maximum effluent ammonia nitrogen level (4 mg N/l) was violated
during 4.6667 days, i.e. 66.6667% of the operating time.
The limit was violated at 8 different occasions.
```

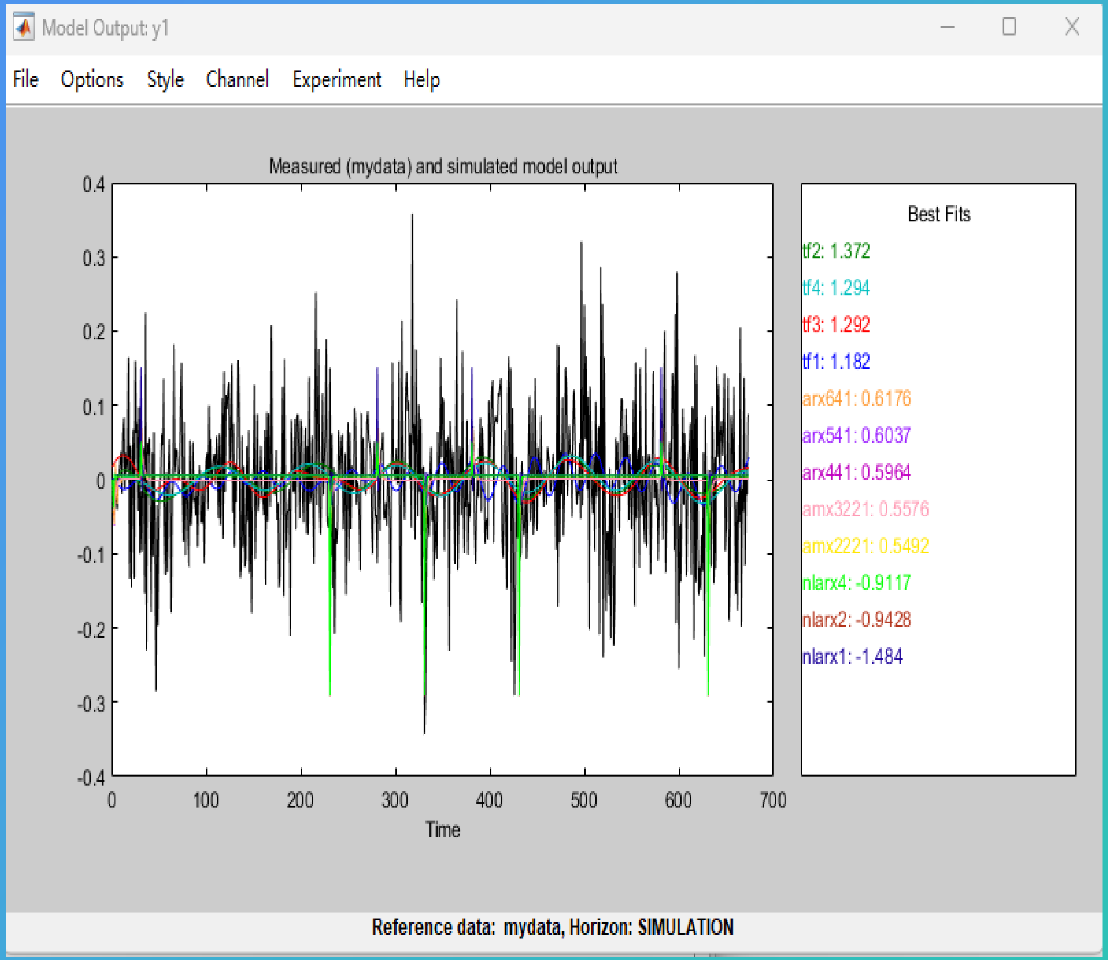
- Problem: Three (Dry, Rain, Storm) scenarios have violations on effluent quality
- Goal: Improve the effluent quality (lower the violations) by using better control system design



Data Driven Modeling using PRBS signal

- **Pseudorandom Binary Sequence(PRBS)**
  - Binary Sequence (two values)
  - Random
  - Commonly using for testing and evaluating systems

Dryinfluent (SO5)



Bad fitting with traditional data driven methods

- Input:
  - Internal Recycle Flow Rate
  - KLA5 (oxygen transfer coefficient)
- Output:
  - NO3-N(Nitrate) in the Second Anoxic Compartment (SNO2)
  - Dissolved Oxygen in the Last Aerated Compartment (SO5)
- Best fit range: 0-3 %

Machine Learning Based Control

- We model the control system rather than modelling the process
- Methodology
  - Generate 15 rain and 15 storm scenarios
  - Generate random combinations of PI parameters and setpoints
  - Run BSM1 simulation and collect the data
  - Build the regression model
  - Use the regression model to search the optimal combinations of PI parameters and setpoints

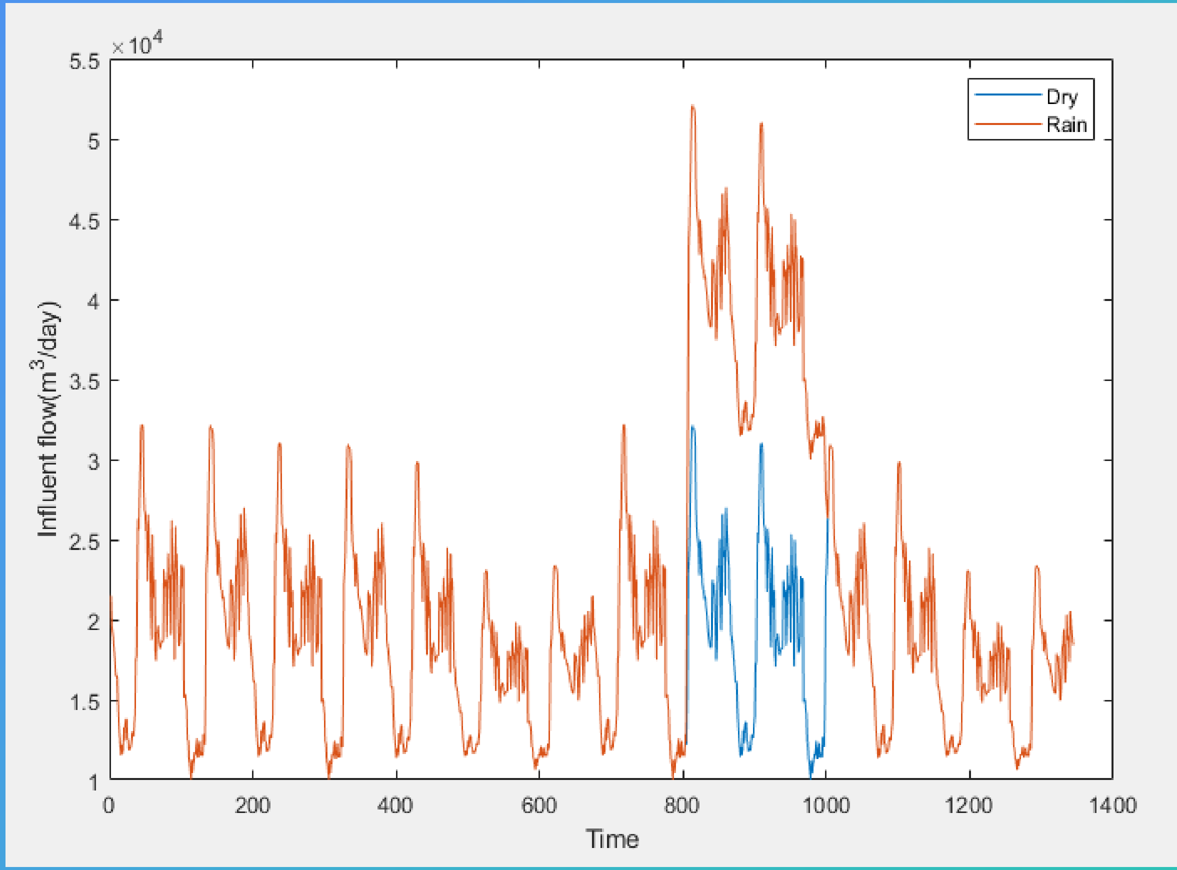
Generating Scenarios

Three dynamic influent (Dry, Rain, and Storm)

Rain influent: Dry influent + 1 rain event

Storm influent: Dry influent + 1 storm event + 1 rain event

Following the pattern, we modify when the event starts, how long it lasts, and the magnitude of the peak and trough.



Generating Scenarios

- Pre-control phase: 0-7s (first week)
- Control action: 7s - 14s (second week)
- Influent soluble concentrations (S_S, S_{NH}, S_{ND}, S_I)
- Influent particulate concentrations (X_I, X_{BH}, X_{BA}, X_S, X_{ND})
- Methodology of generating rain scenarios
  - Modify the influent flow
  - Modify the load concentrations follow the dilution calculation

Simulation and data collection

1

Generate random combinations of PI parameters and setpoints

2

Run simulations of each combination of PI parameters and setpoints (total 9 variables) with 30 scenarios

3

Collect the jointed violation time of each simulation

4

Take the average violation time of 30 scenarios with each combination of PI parameters and setpoints

Machine Learning Based Regression model

Predictors: combination of PI parameters and setpoints (total 9 variables)

Response: the average violation time

Using the **machine learning regression** model, we can estimate the relationship between Predictors and Response

Once we built the model, we can use it to make prediction of new preditors

Modelling result

Violation time **default PI performance** (no sensornoise)

Dry: 68.6429%

Rain: 73.3631%

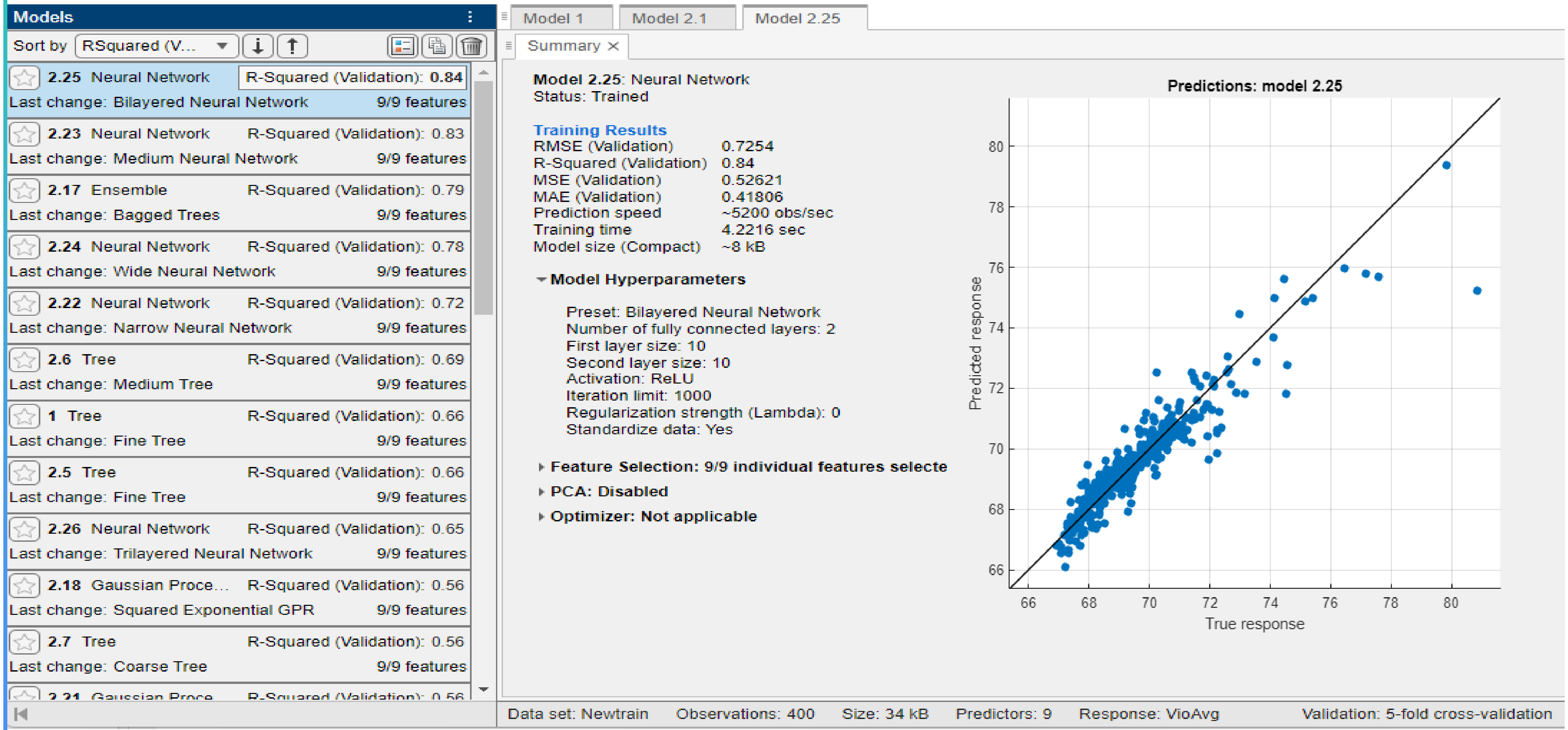
Storm: 71.7262%

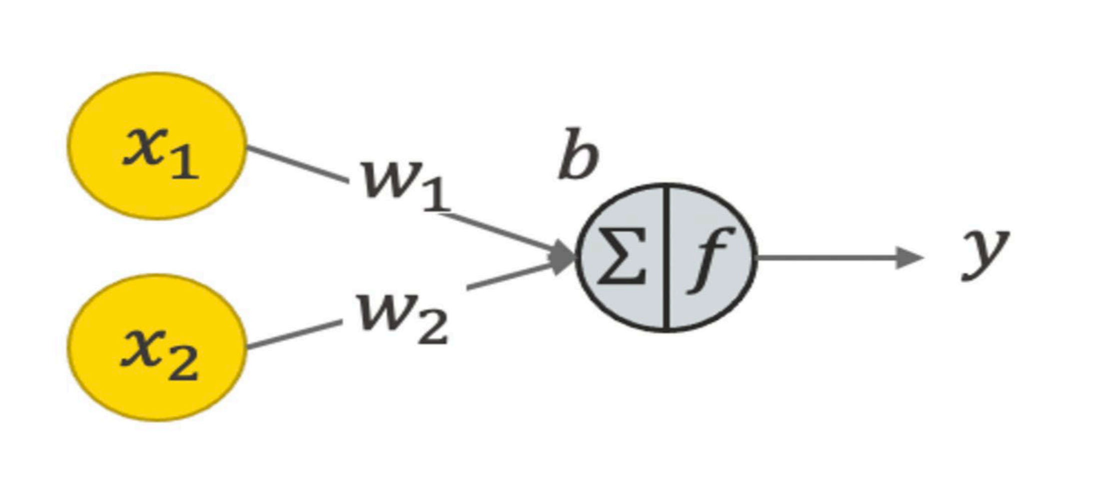
Simulation result

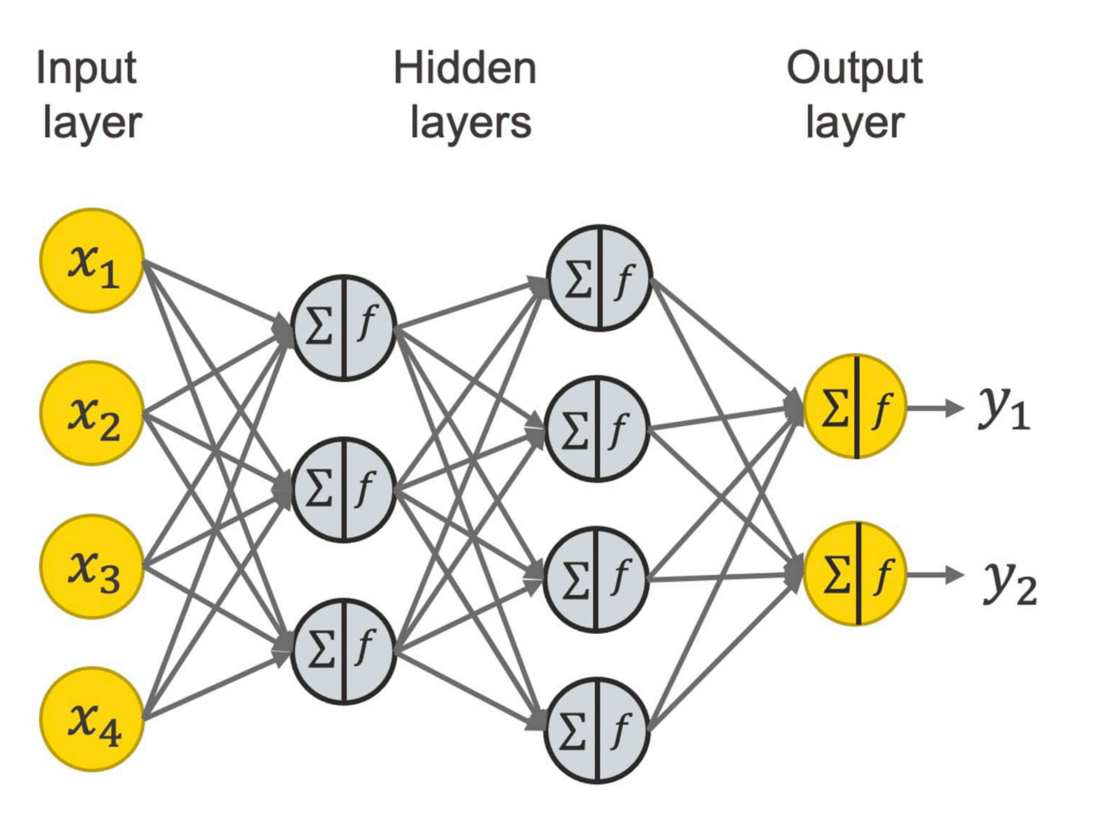
400 sets data

Range: 66.9296% to 80.8681%

Machine Learning Based Modelling result







Neural network

- Input layer: Predictors (9 variables)
- First layer: 10 neurons
- Second layer: 10 neurons
- Output layer: Response
- Each neuron has its own set of weights and a bias
- Each neuron output can be defined by 2 steps
  - $x_1 * w_1 + x_2 * w_2 + b$
  - An activation function (ReLU)

$$\mathrm{ReLU}(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

- Use the previous layer values to calculate its output

Continuing work

Generate	Generate more combination of PI parameters and setpoints (total 9 variables)
Use	Use the regression model to make predictions
Search	Search the optimal combination with the lowest value of response

References

- (1) Alex, Jens & Benedetti, Lorenzo & Copp, Jb & Gernaey, Krist & Jeppsson, Ulf & Nopens, Ingmar & Pons, Marie-Noelle & Rieger, Leiv & Rosen, Christian & Steyer, J-P. (2008). Benchmark Simulation Model no. 1 (BSM1). Report by the IWA Taskgroup on Benchmarking of Control Strategies for WWTPs.
- (2) Vanhooren H., Nguyen K. (1996) Development of a simulation protocol for evaluation of respirometry-based control strategies. Report University of Gent and University of Ottawa.
- (3) Melcher, K. (n.d.). *A friendly introduction to [deep] neural networks*. KNIME. https://www.knime.com/blog/a-friendly-introduction-to-deep-neural-networks