GHENT UNIVERSITY

WRRmod 2021

Integrated **hybrid modelling** of water resource recovery facilities by coupling **mechanistic** and **data-driven** ODE models

W. Quaghebeur[1,2], E. Torfs[1], B. De Baets[2], I. Nopens[1]

[1] BIOMATH, Department of Data Analysis and Mathematical Modelling, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, 9000 Ghent, Belgium
[2] KERMIT, Department of Data Analysis and Mathematical Modelling, Faculty of Bioscience Engineering, Ghent University, Coupure links 653, 9000 Ghent, Belgium

Problem

Water treatment plant operators want to apply models online as a *digital twin*. To be useful, these models need to be able to make accurate predictions and extrapolate to unknown data. The currently used techniques either **lack accuracy** (mechanistic models) or **fail to extrapolate** (data-driven models).

Mechanistic models

- Encorporate physical and empirical knowledge
  (e.g. mass balances, conversion rates, …)
- E.g. Activated Sludge Models (ASM)
- Formulated as differential equations

$$\frac{\mathrm{d}\boldsymbol{X}_t}{\mathrm{dt}} = f(\boldsymbol{X}_t)$$

\+ Interpretable
\+ Can extrapolate to unseen data
\- **Domain knowledge** is **incomplete** and a **simplification** of the underlying system
  E.g., CSTR instead of mixing, simplifying microbial communities, …

Data-driven models

- Search for relationships in available data
- Different *machine learning* techniques, e.g. artificial neural networks, random forests
- Here formulated as neural differential equations

$$\frac{\mathrm{d}\boldsymbol{X}_t}{\mathrm{dt}} = nn(\boldsymbol{X}_t)$$

\- Black-box
\- Fail **to extrapolate** to unseen data
\+ No domain knowledge needed
\+ Leverage all available data
\+ Accurate

Hybrid models

Combine a mechanistic and data-driven model

$$\frac{\mathrm{d}\boldsymbol{X}_t}{\mathrm{dt}} = \underbrace{f(\boldsymbol{X}_t)}_{\text{mechanistic}} + \underbrace{nn(\boldsymbol{X}_t)}_{\text{data-driven}}$$

The strengths of both models are combined:

- The data-driven part fills in the gaps in domain knowledge of the mechanistic model
- The mechanistic model fills in the gaps in the dataset of the data-driven model

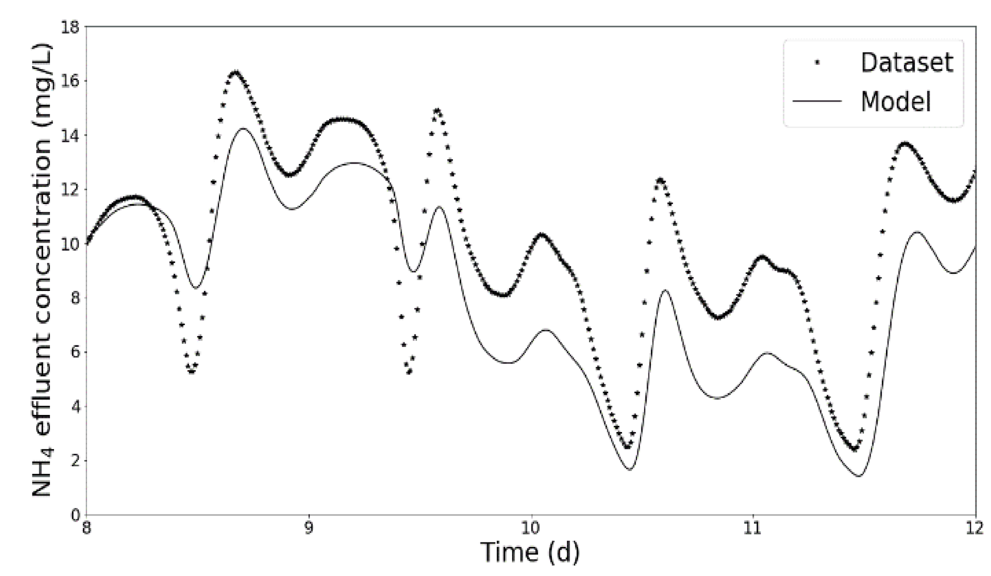
- Hybrid models can make **accurate predictions,** without complete domain knowledge or a representative dataset.
- The data-driven part learns the **dynamics not incorporated** by the mechanistic part.

Simulation study

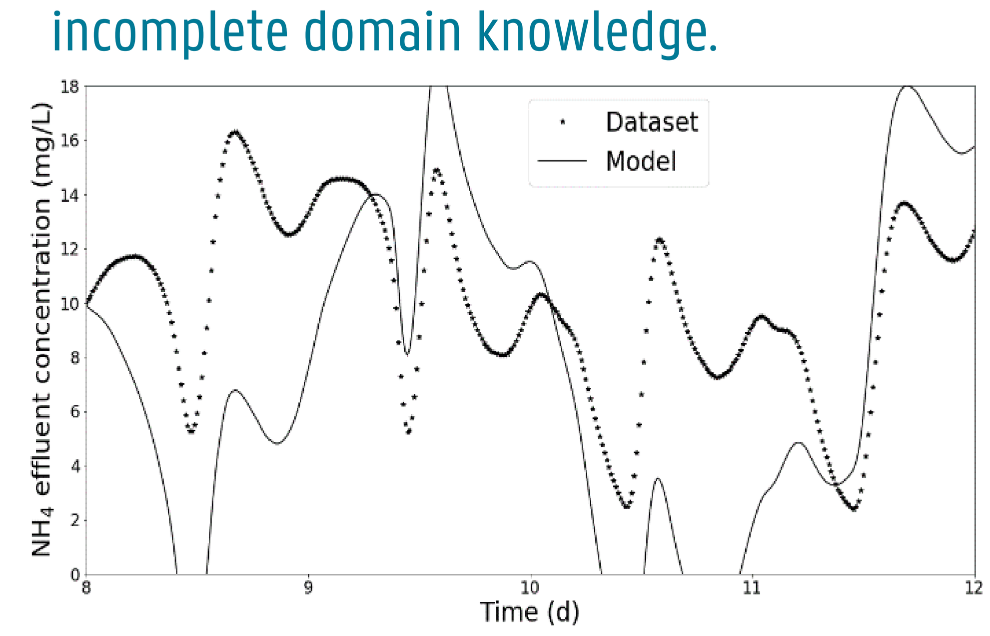
1. Generate data of 7 days of **dry weather** using BSM1
2. Train three models using this dataset
   - A **mechanistic** models with incomplete domain knowledge
   - A **data-driven** model
   - A **hybrid** model with incomplete domain knowledge

   To simulate **incomplete domain knowledge**, the mechanistic (sub)models do not account for anoxic growth in heterotrophs
3. Simulate 7 days of **wet weather**

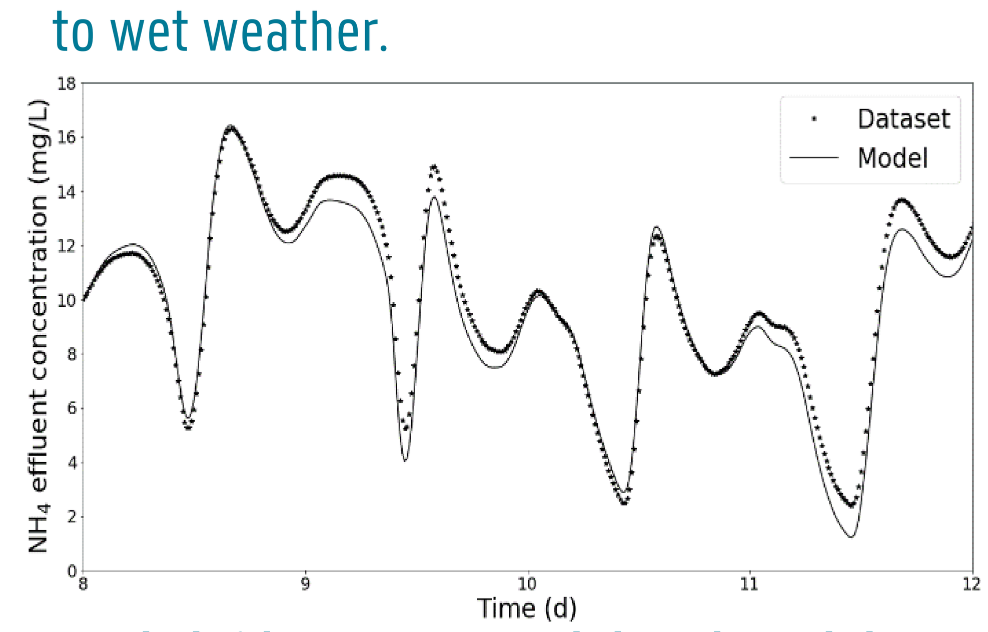
Results



The **mechanistic** model predicts the general trends, but **lacks accuracy** due to incomplete domain knowledge.



The **data-driven** model **fails to extrapolate** to wet weather.



The **hybrid** model **accurately extrapolates** to wet weather.

ward.quaghebeur@ugent.be

BIOMATH KERMIT CAPTURE fwo Grant n° 3S79219