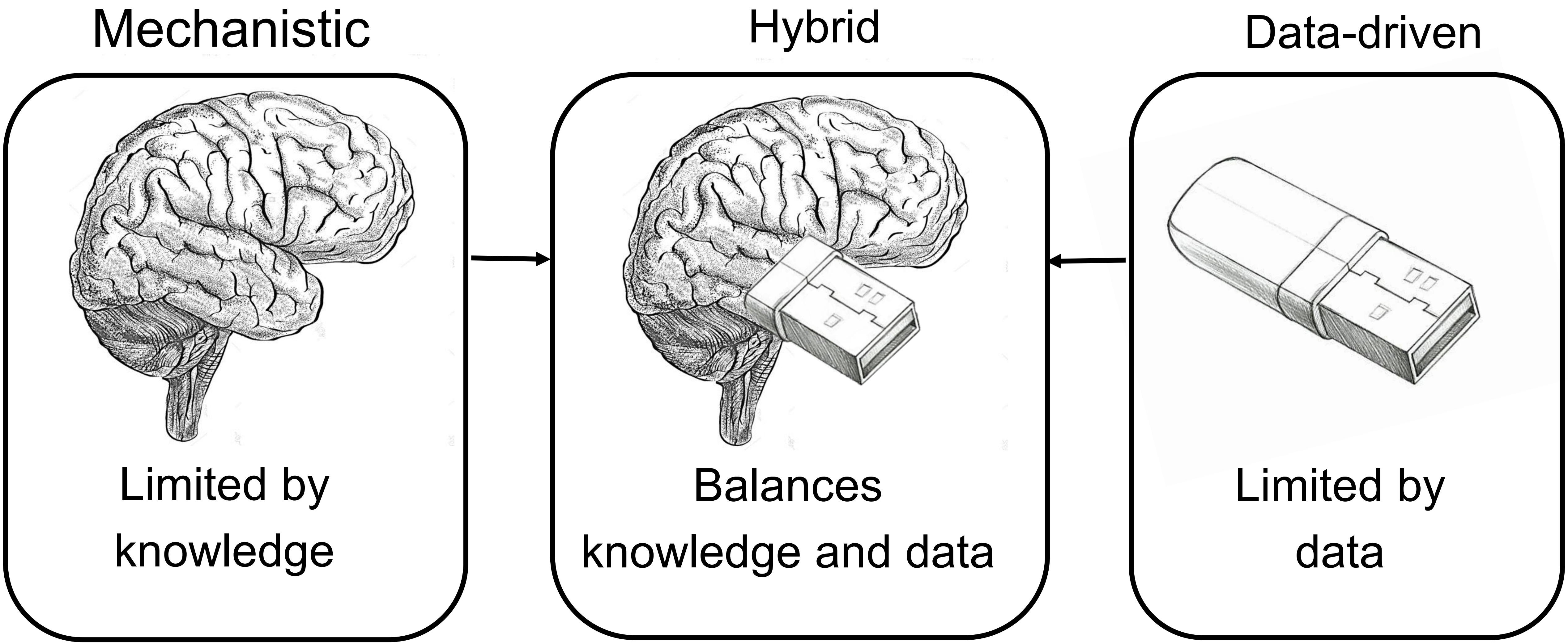
How to integrate domain knowledge with data-driven models

Elena Torfs

UNIVERSITÉ LAVAL

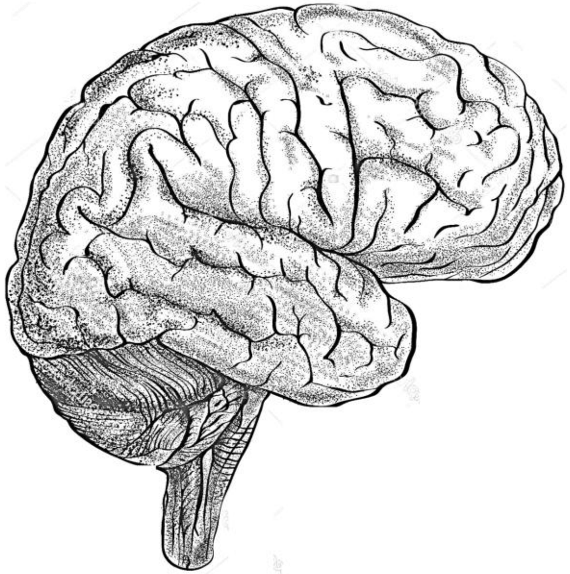
modelEAU



Schneider et al., 2022. Hybrid modelling in the water and wastewater sector: quo vadis?. Water Science and Technology. 85(9): 2503-2524.

Scientific Machine Learning

Mechanistic



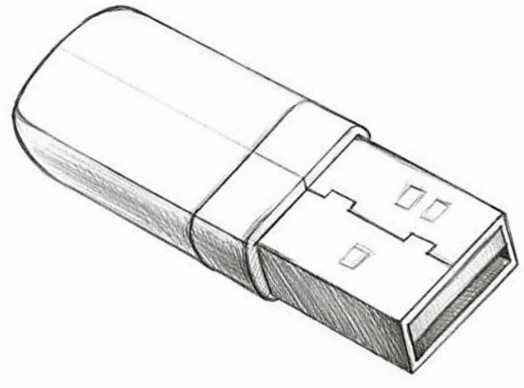
Domain knowledge
Scientific computing

SciML

Scientific Machine Learning

- Hybrid modelling
- Surrogate models
- Equation discovery
- ML–enhanced data assimilation
- Reinforcement learning with physical constraints
- Physics Informed NN
- ….

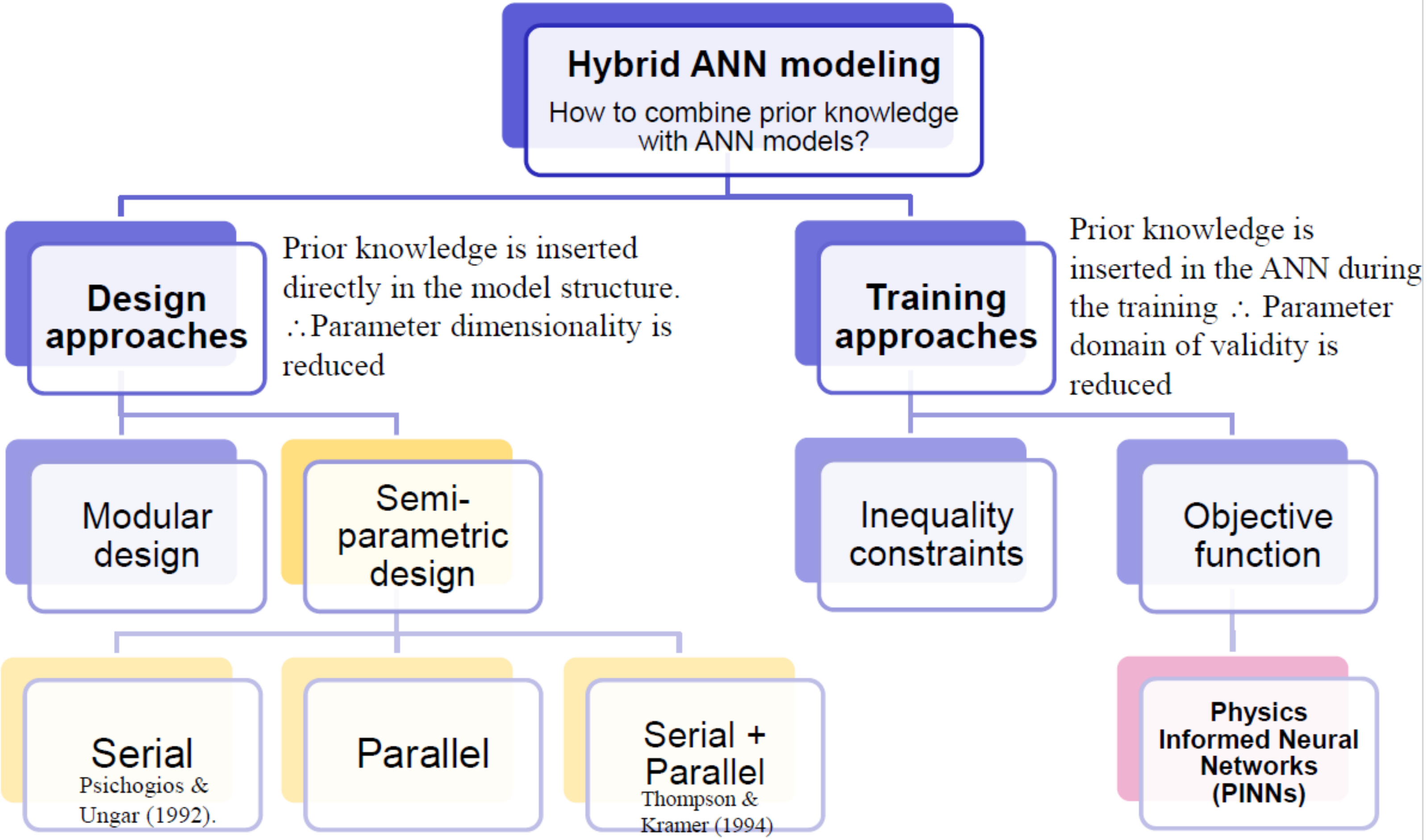
Data-driven



Machine Learning

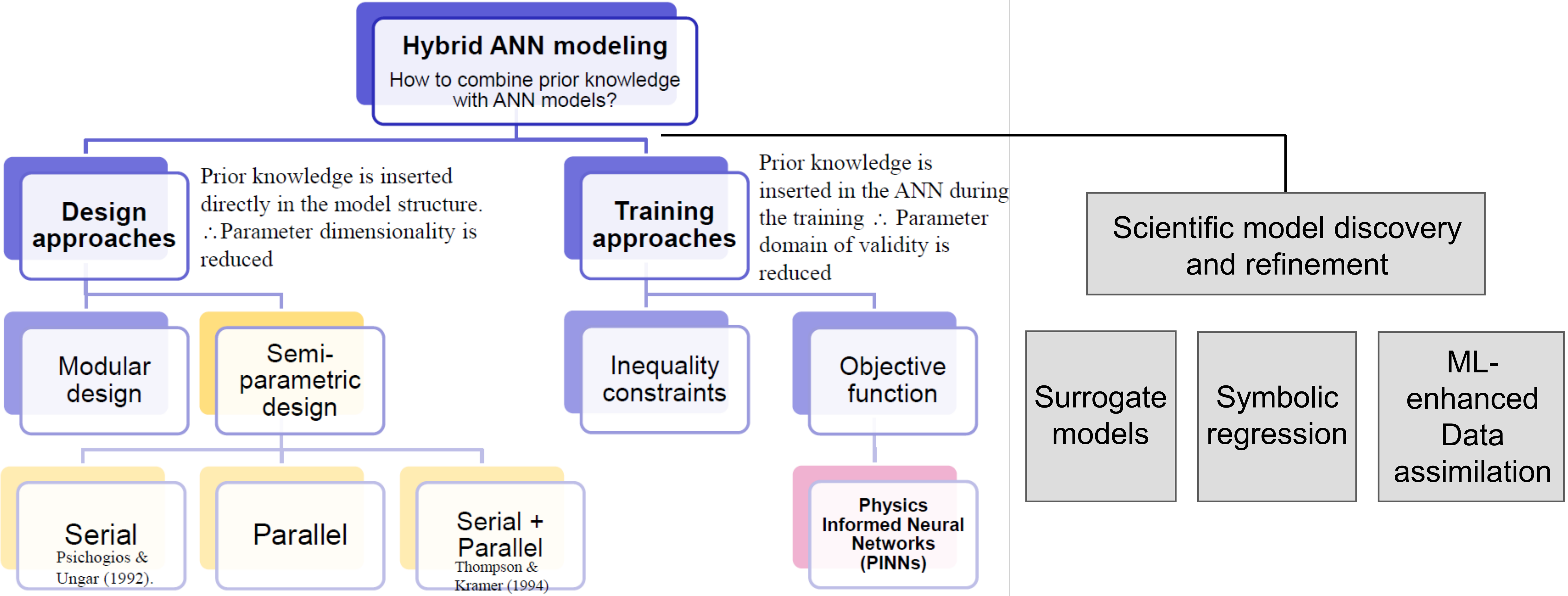
Scientific Machine Learning

Thompson, M. L., & Kramer, M. A. (1994) **Modeling chemical processes using prior knowledge and neural networks**. AIChE Journal, 40(8), 1328–1340. https://doi.org/10.1002/aic.690400806



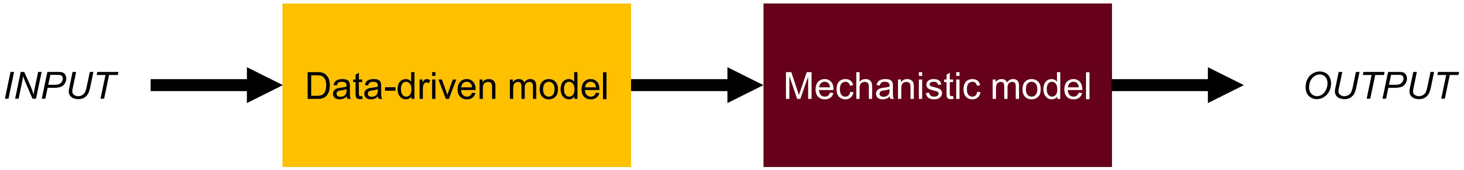
Scientific Machine Learning

Thompson, M. L., & Kramer, M. A. (1994) **Modeling chemical processes using prior knowledge and neural networks**. AIChE Journal, 40(8), 1328–1340. https://doi.org/10.1002/aic.690400806

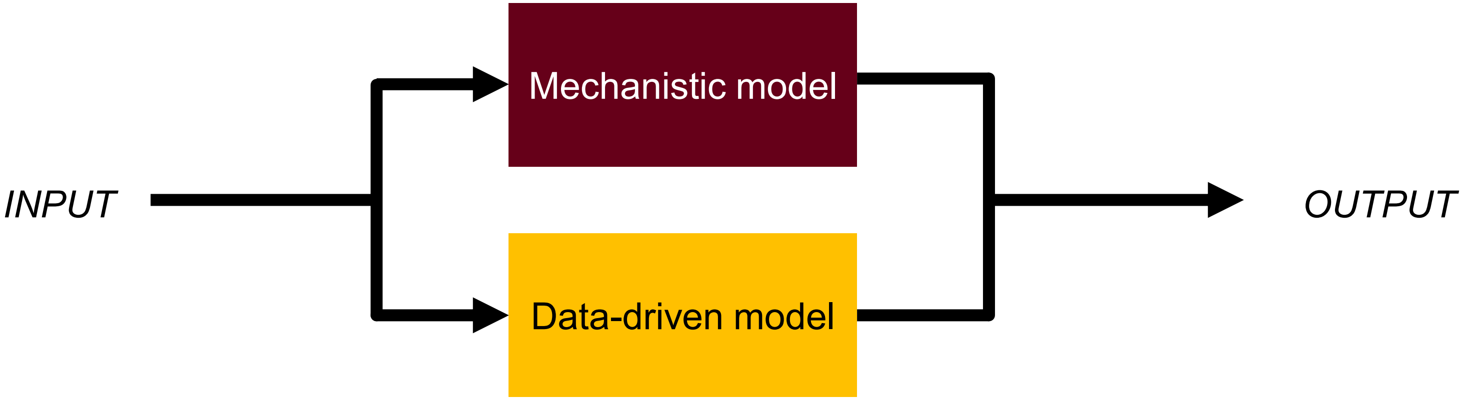


Hybrid Modelling Architectures

Serial



Parallel



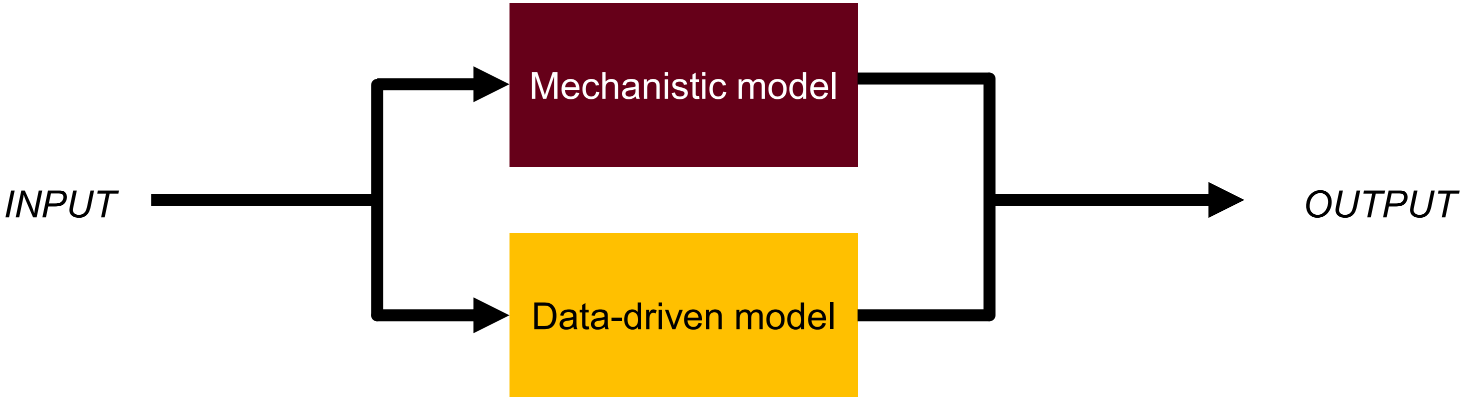
Hybrid modelling working group initiative to collect, review and publish hybrid modelling examples

Sources: Schneider et al. 2022, Quaghebeur et al. 2022

Hybrid Modelling Architectures

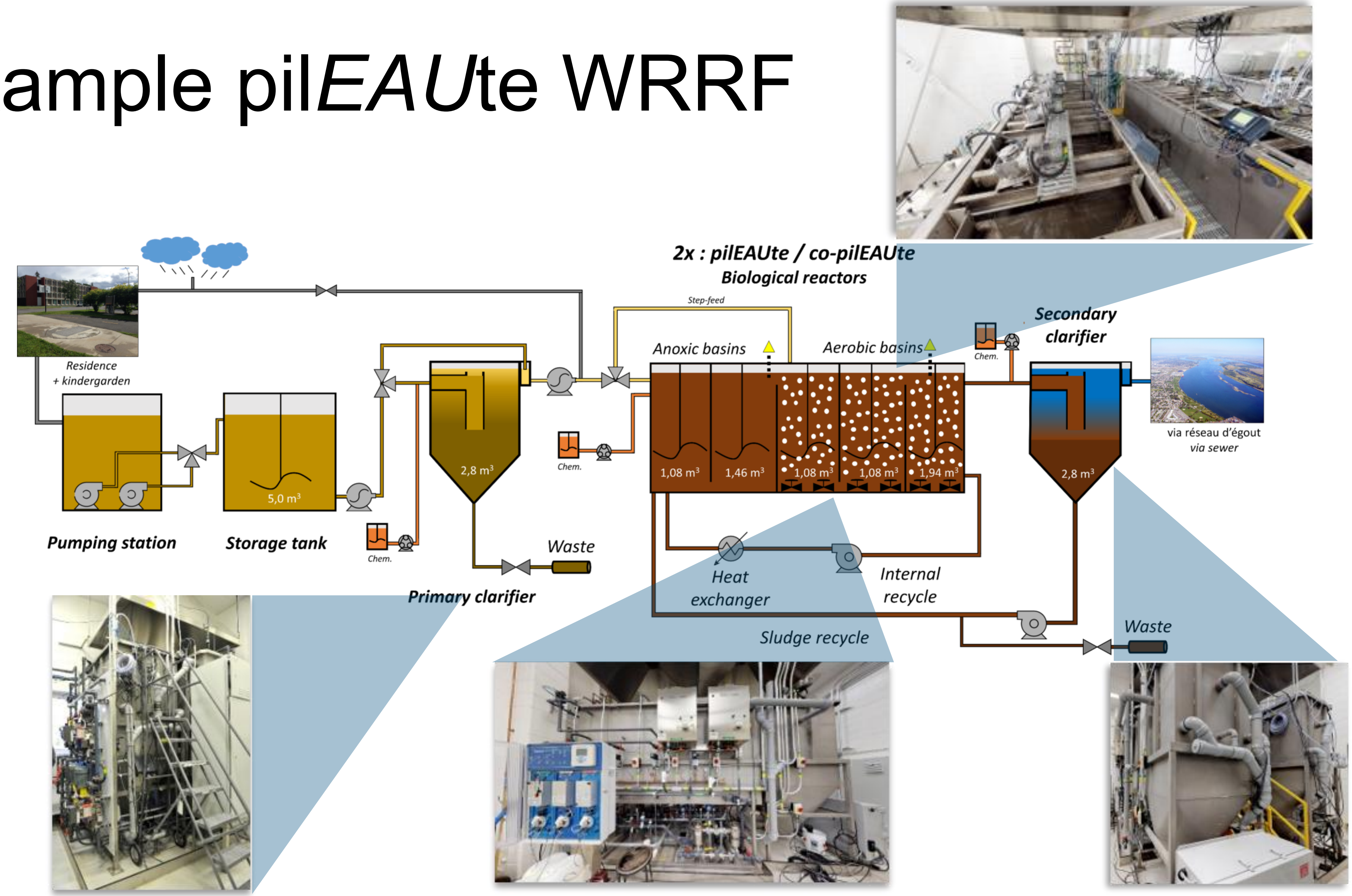
→ e.g. interesting in case of incomplete knowledge or oversimplification and / or for accuracy gains

Parallel

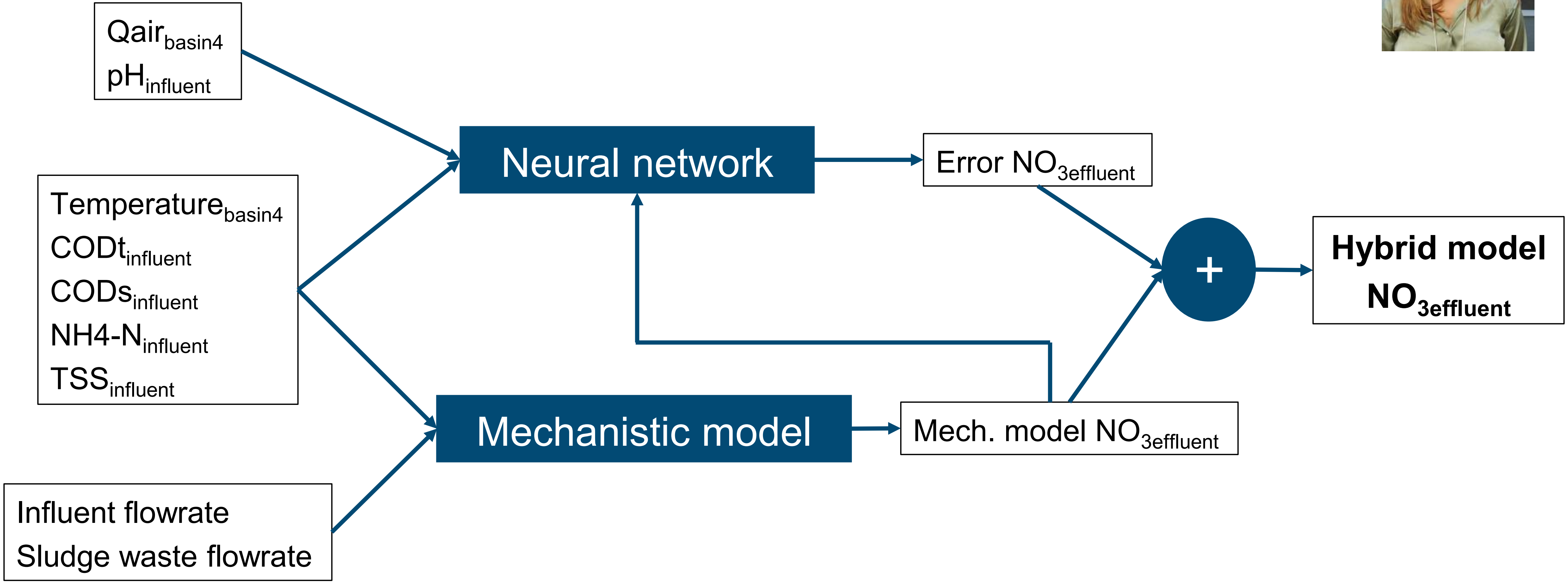


Sources: Schneider et al. 2022, Quaghebeur et al. 2022

Example pil*EAU*te WRRF

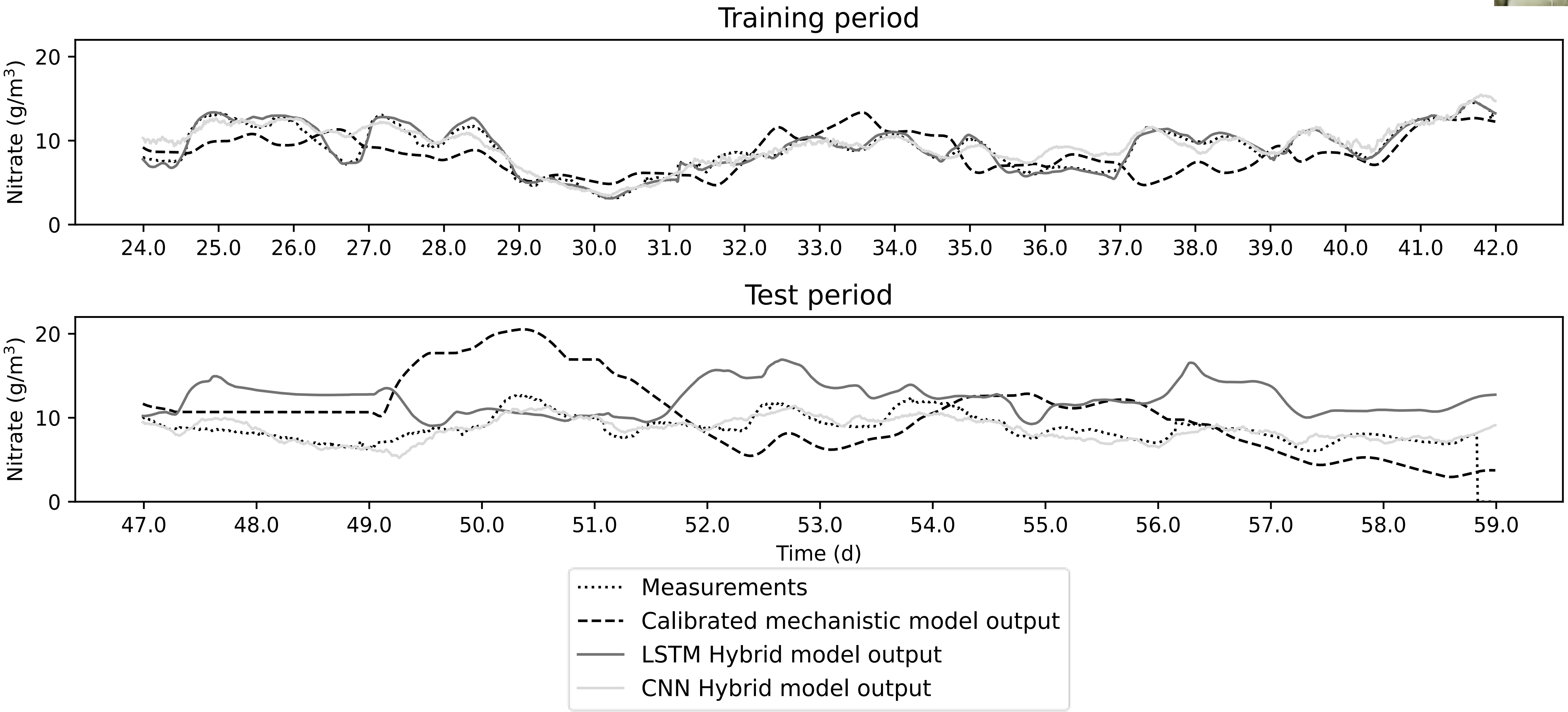


Parallel hybrid model for effluent nitrate



Verhaeghe et al. (2024). Towards good modelling practice for parallel hybrid models for wastewater treatment processes. *Water Science and Technology*.

Parallel hybrid model for effluent nitrate

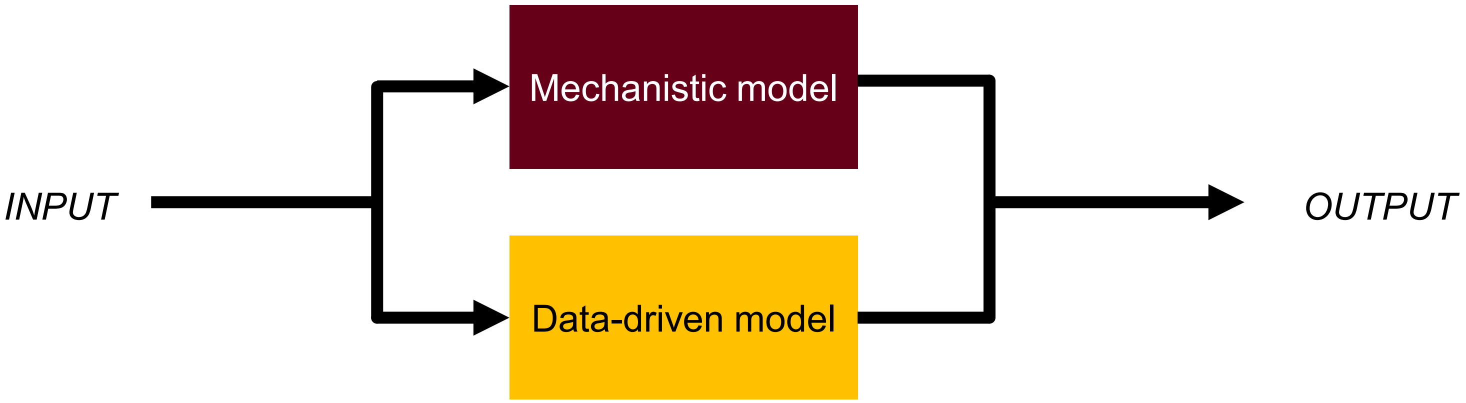


Verhaeghe et al. (2024). Towards good modelling practice for parallel hybrid models for wastewater treatment processes. *Water Science and Technology*.

Hybrid Modelling Architectures

→ e.g. interesting in case of incomplete knowledge or oversimplification and / or for accuracy gains

Parallel

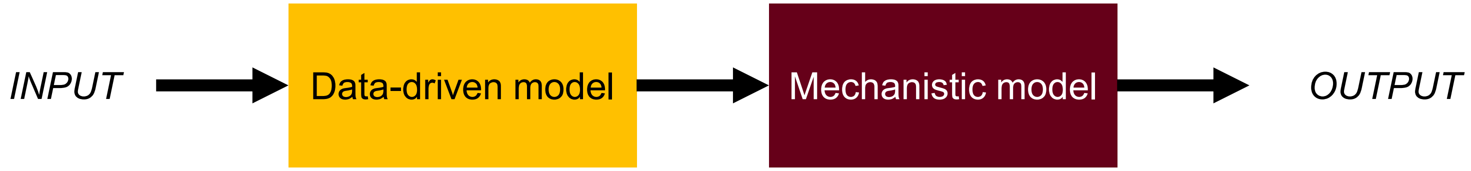


Challenges:
-Mass balance isn't closed
-Balancing the calibration effort

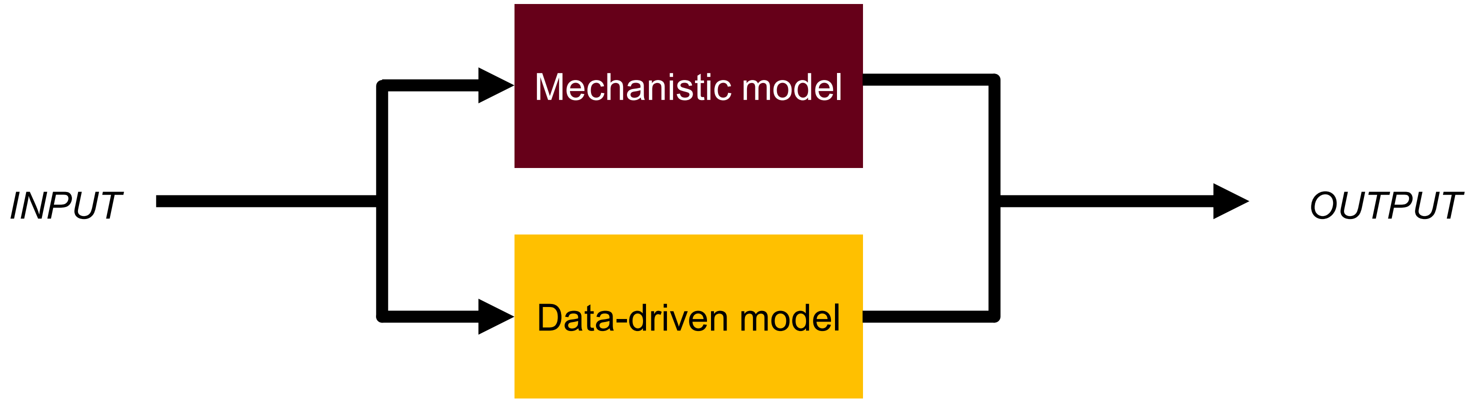
Sources: Schneider et al. 2022, Quaghebeur et al. 2022

Hybrid Modelling Architectures

Serial



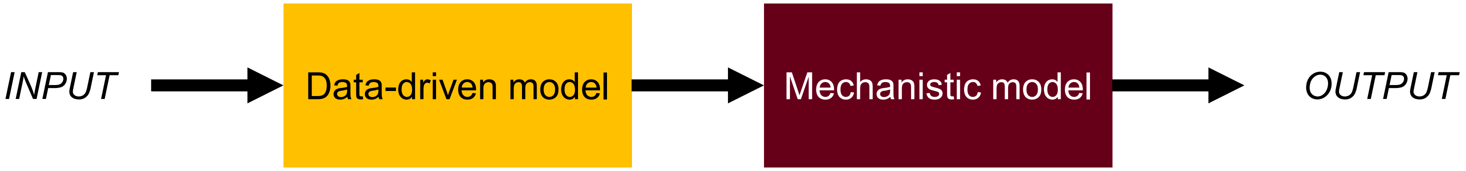
Parallel



Sources: Schneider et al. 2022, Quaghebeur et al. 2022

Hybrid Modelling Architectures

Serial



→ e.g. interesting if overall mechanistic model structure holds, but sub-processes are insufficiently defined

Sources: Schneider et al. 2022, Quaghebeur et al. 2022

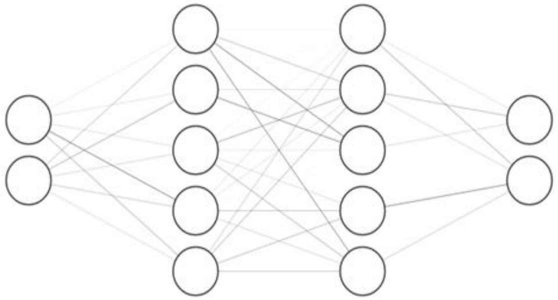
Example reverse osmosis

Gaublomme et al. (2023). A hybrid modelling approach for reverse osmosis processes including fouling. *Desalination*.

Example serial hybrid model for RO

LSTM model predicts fouling

T
κ
Q_f
R
Ca, f
CIP



$R_{m,a}$

$$R_{m,a} = \frac{1}{L_p\,\eta}$$

$$B = f(R_{m,a}, w1)$$
$$h = f(R_{m,a}, w2)$$

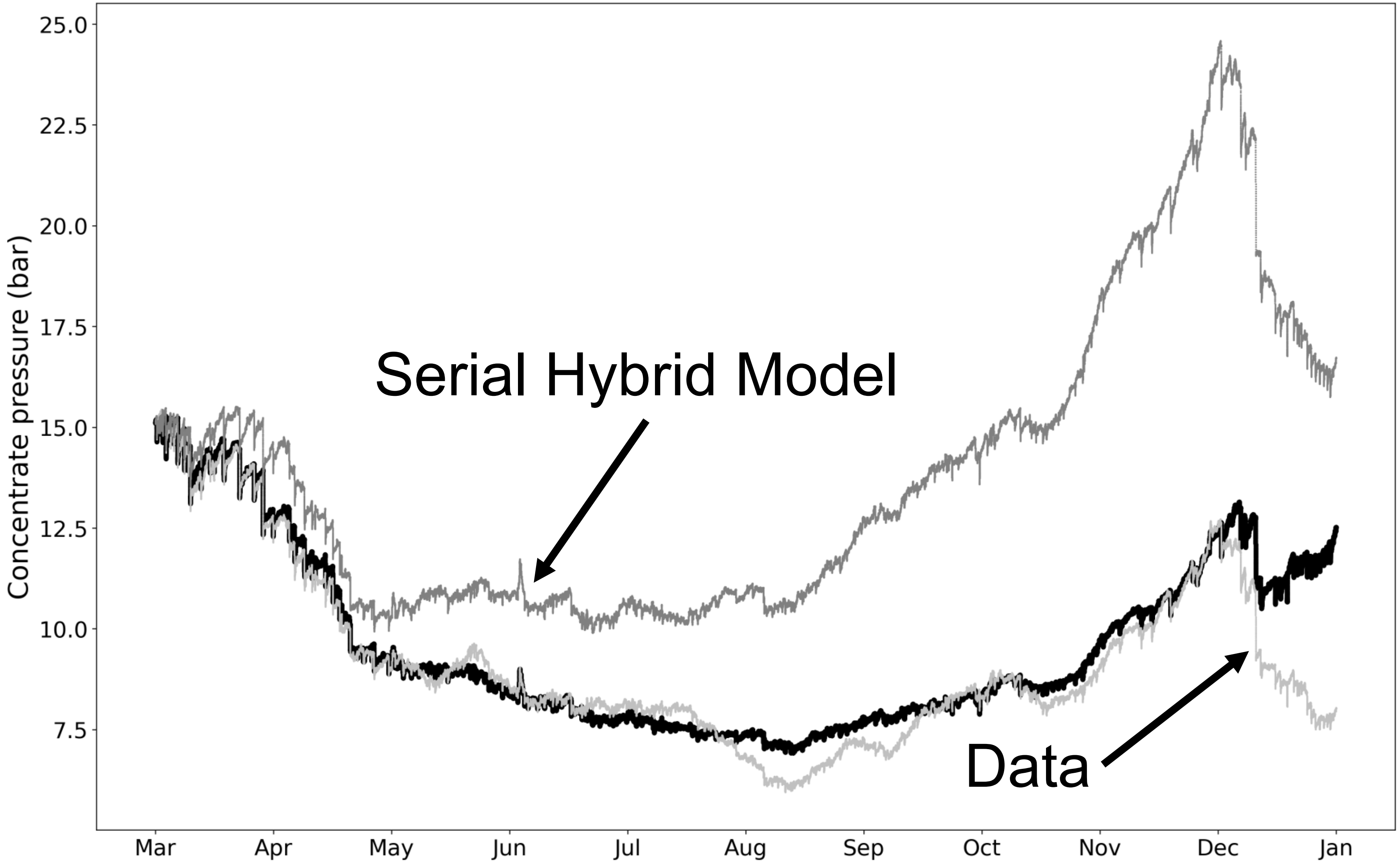
Mechanistic model

$$\frac{dx}{dt} = f(x, L_p, B, h)$$

C_c
C_p
P_c
Q_c
Q_p

Gaublomme et al. (2023). A hybrid modelling approach for reverse osmosis processes including fouling. *Desalination*.

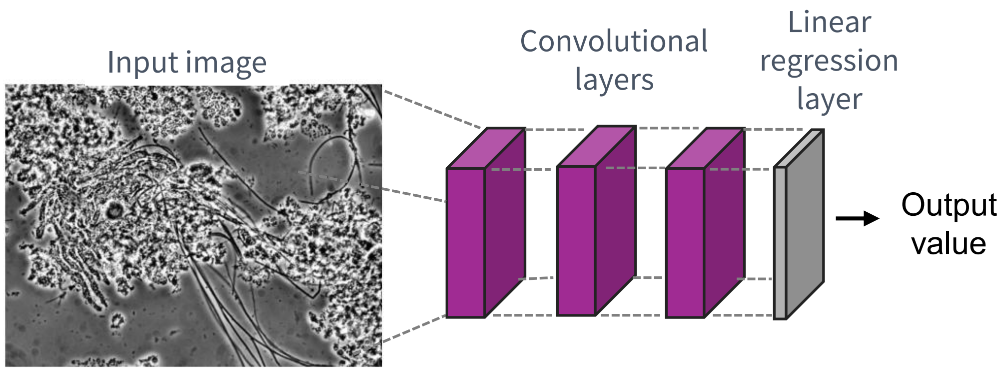
Example serial hybrid model for RO



Gaublomme et al. (2023). A hybrid modelling approach for reverse osmosis processes including fouling. *Desalination*.

Example Serial hybrid model with image data

Data-driven model



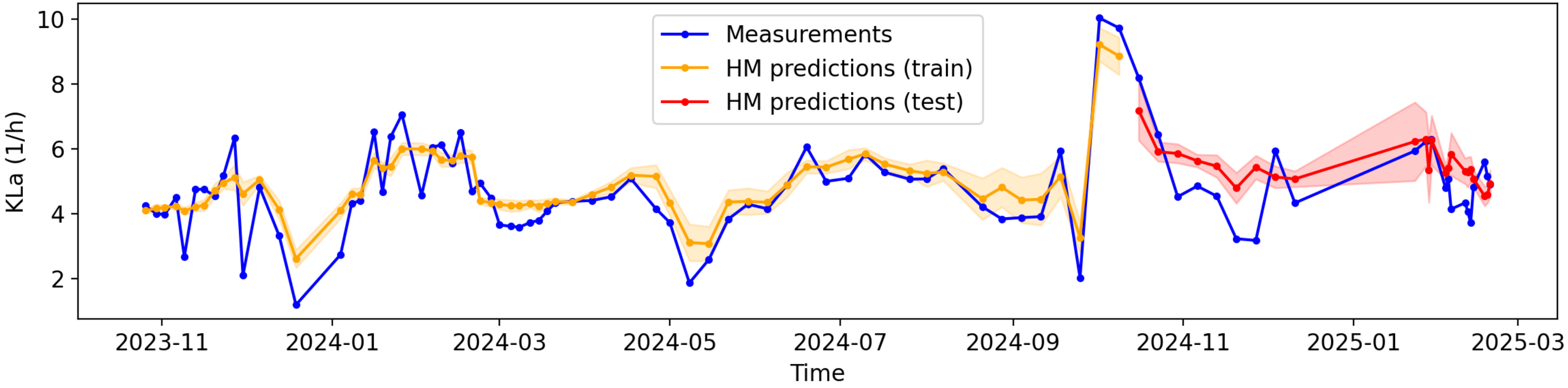
P=KLa

Mechanistic model

$$\frac{dX}{dt} = f(X, U, P)$$

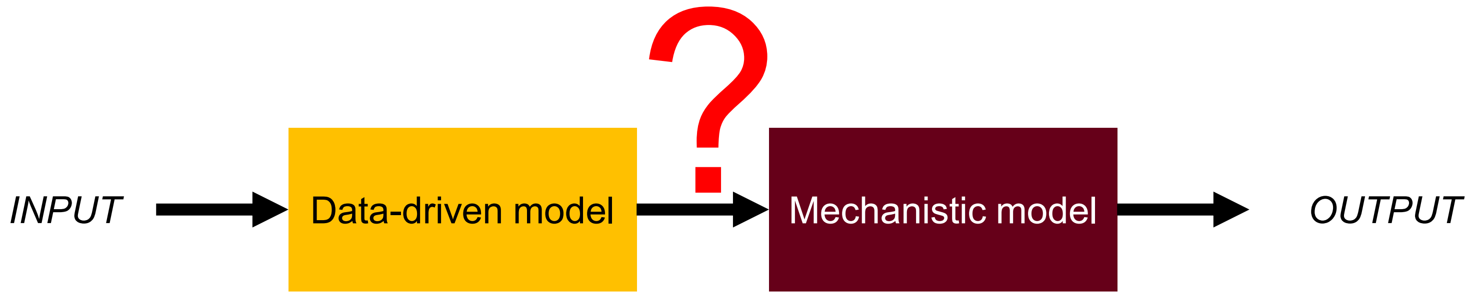
Borzooei et al. (2024). Evaluation of Activated Sludge Settling Characteristics from Microscopy Images with Deep Convolutional Neural Networks and Transfer Learning. *Journal of Water Process Engineering*

Example Serial hybrid model with image data



Hybrid Modelling Architectures

Serial



Challenge: how to train serial hybrid models?

Option 1: Two step approach

1. Dynamic parameter estimation of parameter of interest in mechanistic model
2. Train data-driven model to dynamic parameter estimates

Option 2: Train both components together – **Universal differential equations**

Sources: Schneider et al. 2022, Quaghebeur et al. 2022

Hybrid neural differential equations
Universal differential equations

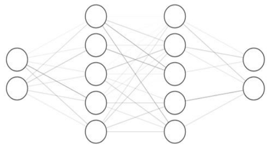
Integrated hybrid model

Mechanistic component

Data-driven component

$$\frac{dx}{dt} = f(x) + nn(x)$$

$$= \mu \frac{x}{K+x} x +$$



Quaghebeur et al., 2022. Hybrid differential equations: Integrating mechanistic and data-driven techniques for modelling of water systems. Water Research. 213: 118-166.

Hybrid neural differential equations
Universal differential equations

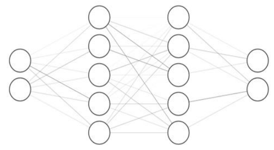
Integrated hybrid model

Mechanistic component

Data-driven component

$$\frac{dx}{dt} = f(x) + nn(x)$$

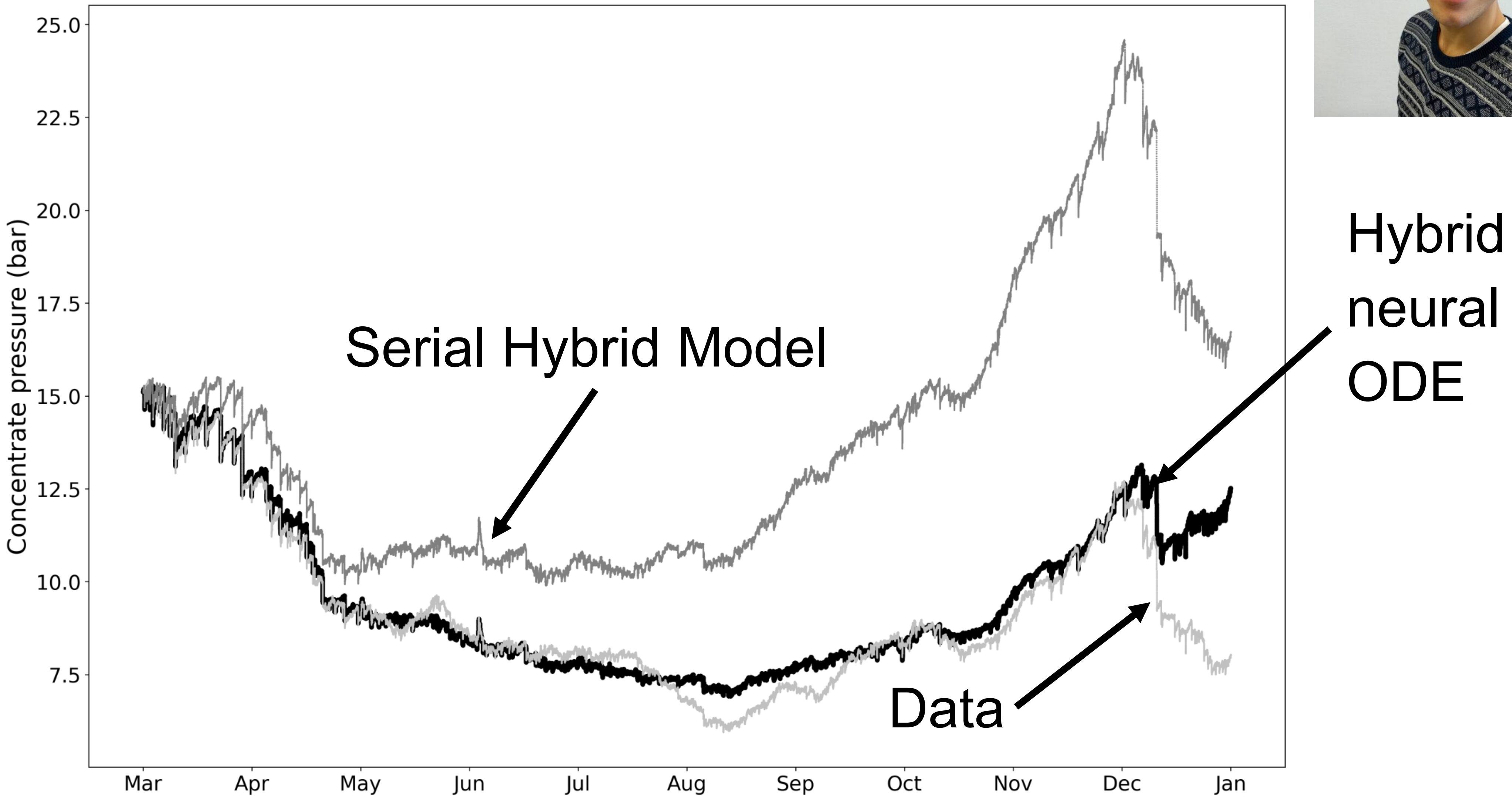
$$= \mu \frac{x}{K + x} x + $$



Challenge: how to train Universal Differential Equations?

Quaghebeur et al., 2022. Hybrid differential equations: Integrating mechanistic and data-driven techniques for modelling of water systems. Water Research. 213: 118-166.

Example universal differential equation for RO



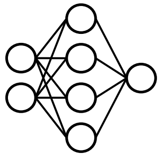
Gaublomme et al. (2023). A hybrid modelling approach for reverse osmosis processes including fouling. *Desalination*.

Universal Differential Equations for N_2O

eawag aquatic research

Universal Differential Equations (UDE)

$$\frac{dX}{dt} = \frac{Q}{V}(X_{in} - X) +$$



- Extend mechanistic models with learned biokinetic reactions
- Adheres to physical principles, e.g. mass conservation
- Learned process can be extracted and interpreted

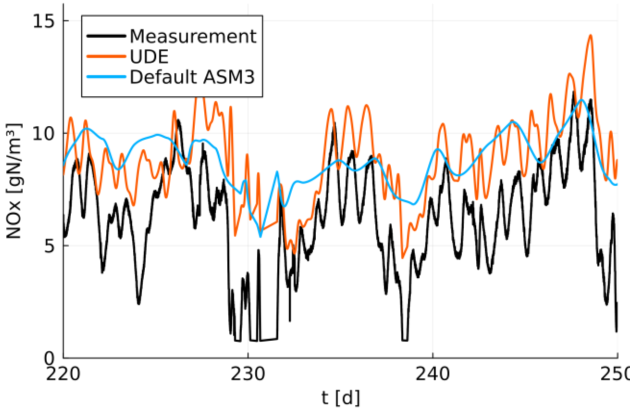
The Case of N_2O Emissions

- Process not fully known
- We have a multi-year, multi-plant dataset

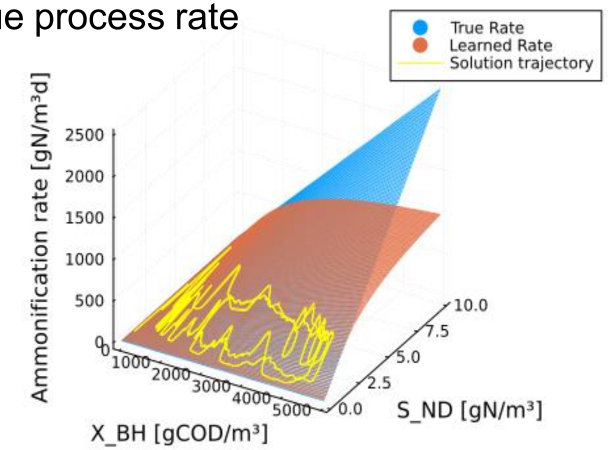
→ Promising conditions for learning N_2O process rate to enhance process knowledge

Examples

Re-learning denitrification process on plant data



Re-learning ammonification rate in synthetic activated sludge system: Comparison of learned vs true process rate



Florian Wenk

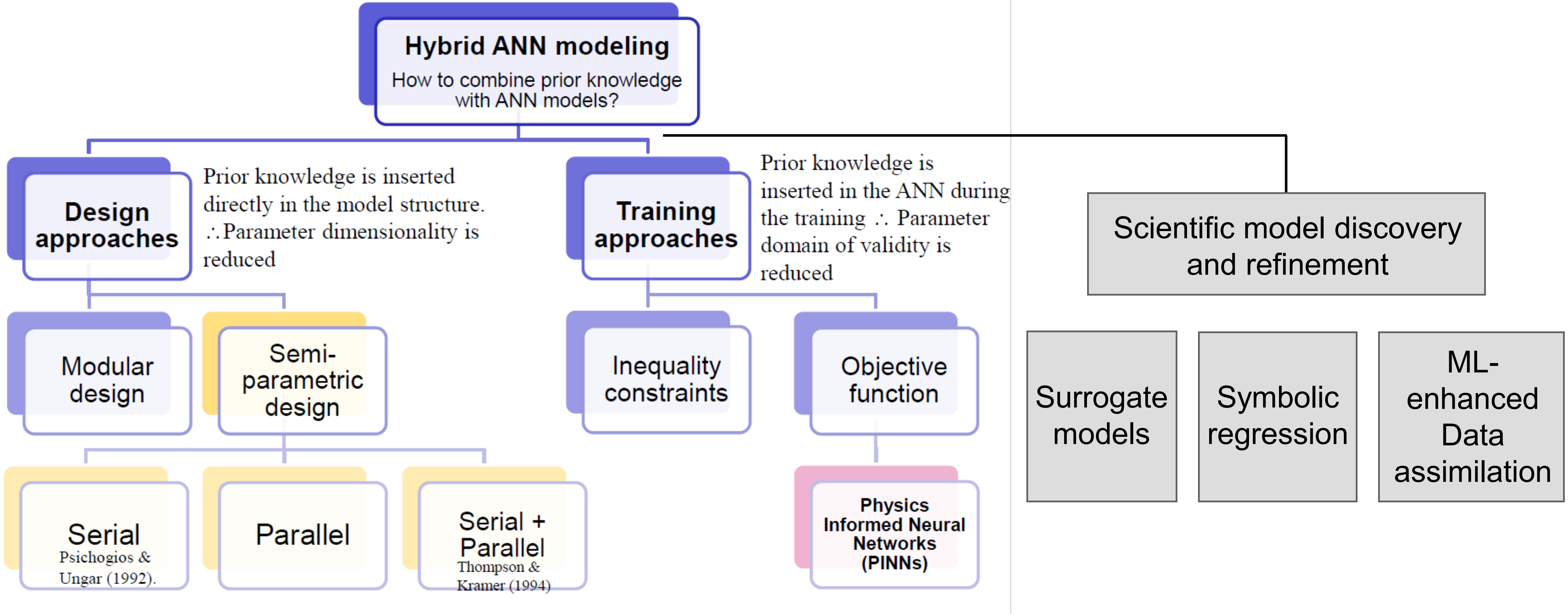
Andreas Froemelt

Our Project: MAD4WW

Our Software-Package

Scientific Machine Learning

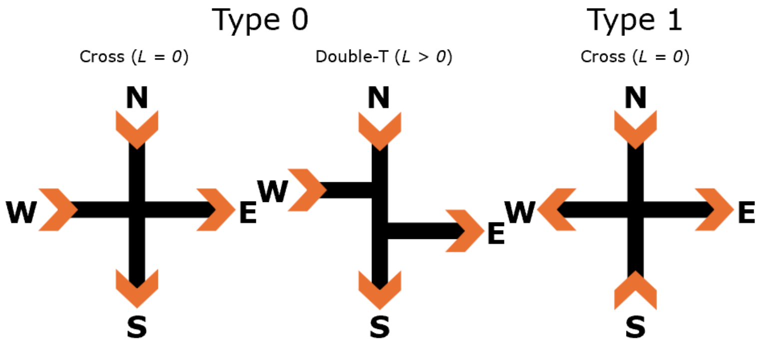
Thompson, M. L., & Kramer, M. A. (1994) **Modeling chemical processes using prior knowledge and neural networks**. AIChE Journal, 40(8), 1328–1340. https://doi.org/10.1002/aic.690400806



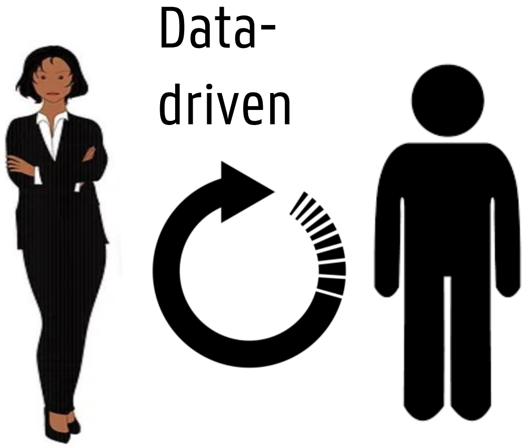
Surrogate modelling

Example surrogate model for junctions in drinking water networks

CFD model



Objective: predict outlet concentration ratio



Neural network

(6 inputs, 5 hidden layers)

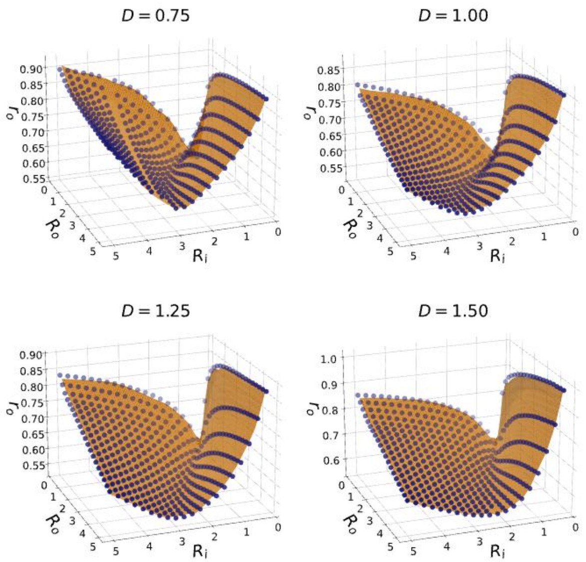


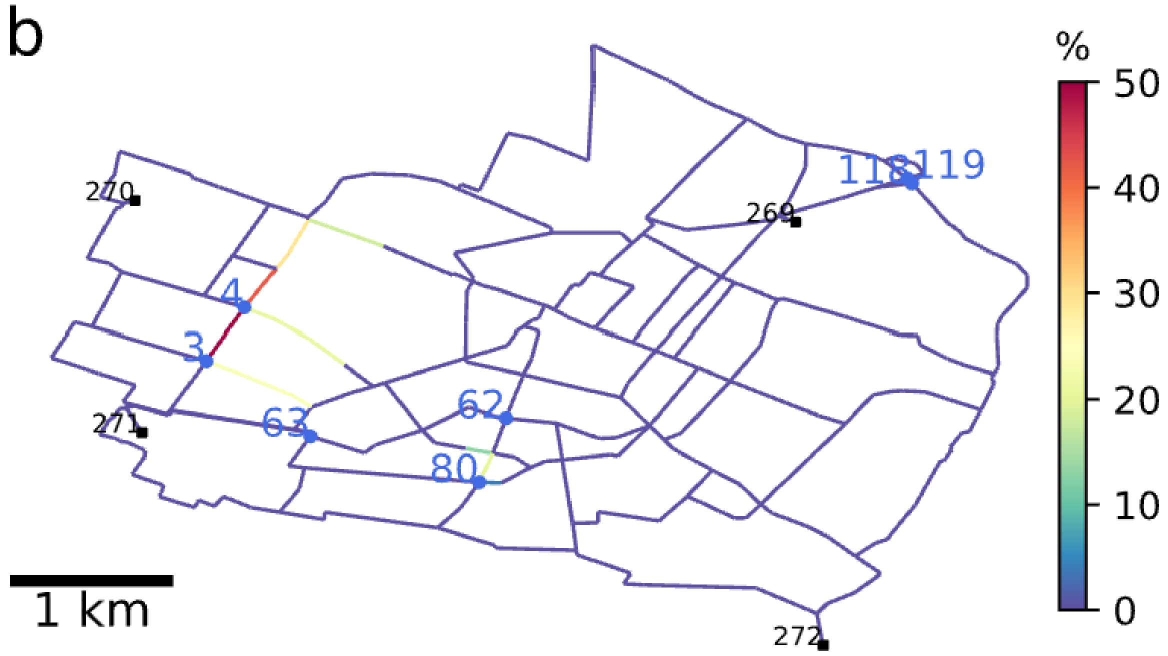
Table 1. Parametric space.

Parameter	Description	Range	Step size
t	Junction type	0 - 1	1
r_i	Inlet concentration ratio	0 - 1	0.1
R_i	Inlet Reynolds number ratio	0.25 – 5, 0.25-12	0.25, 1
R_o	Outlet Reynolds number ratio	0.25 – 5, 0.25-12	0.25, 1
D	Diameter ratio	0.5 - 1.5	0.25
L	Dimensionless separation distance	0 - 10	2

Gallo et al., 2025. Efficient Junction Mixing Prediction in Drinking Water Distribution Networks Using Surrogate Models. *Journal of Water Process Engineering*

Example surrogate model for junctions in drinking water networks

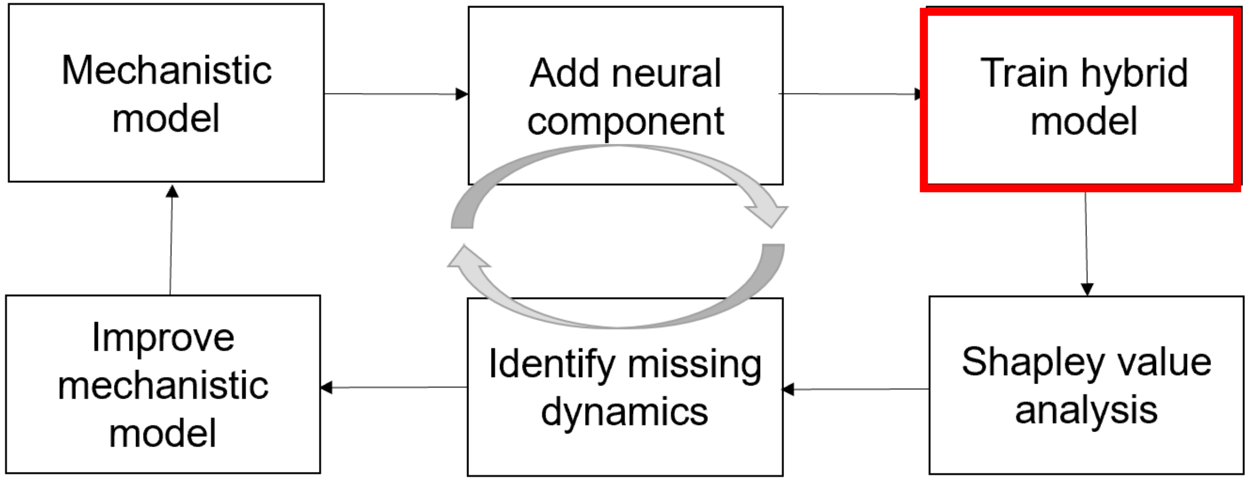
Percentual differences in water quality (solute concentrations) between EPANET simulations with complete mixing in nodes and simulations performed using a surrogate model-derived mixing model.



Gallo et al., 2025. Efficient Junction Mixing Prediction in Drinking Water Distribution Networks Using Surrogate Models. *Journal of Water Process Engineering*

Equation discovery

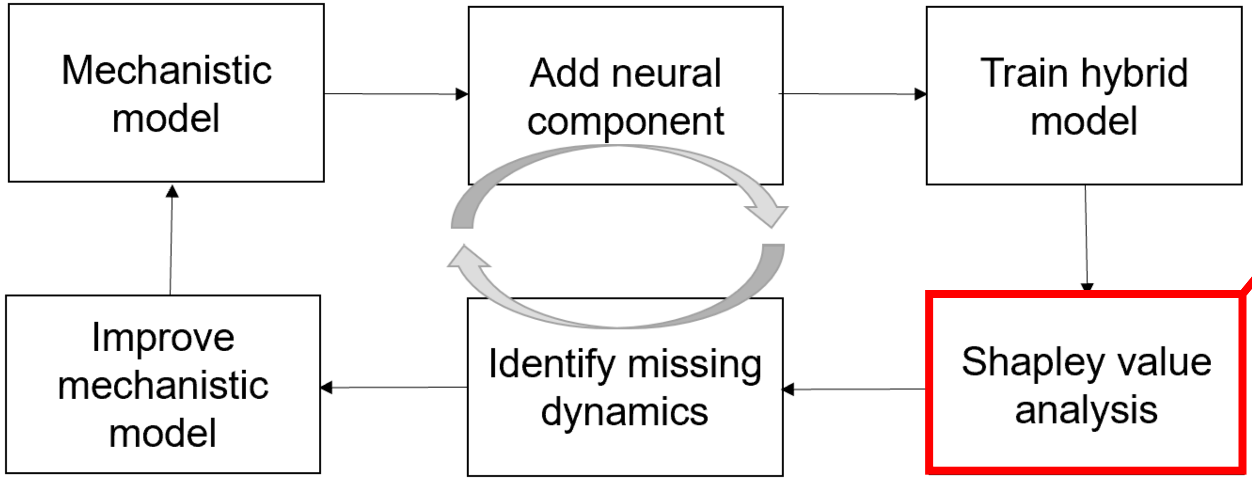
$$\frac{dx}{dt} = \boxed{f(x)} + \boxed{nn(x)}$$



Quaghebeur et al., 2025. Uncovering unknown dynamics in water resource recovery facilities with neural differential equations and Shapley value analysis. *Journal of Water Process Engineering*

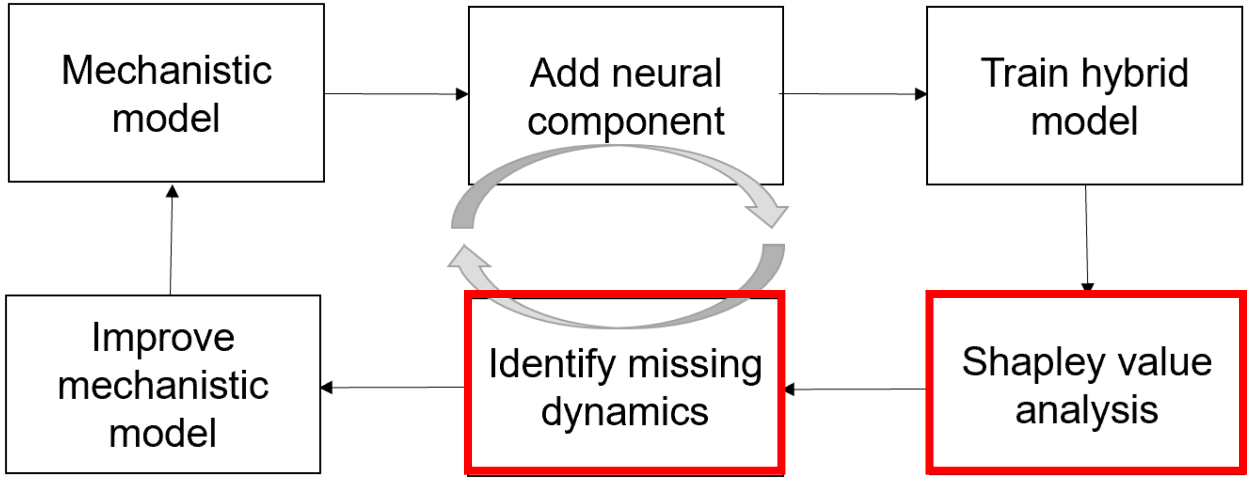
Equation discovery

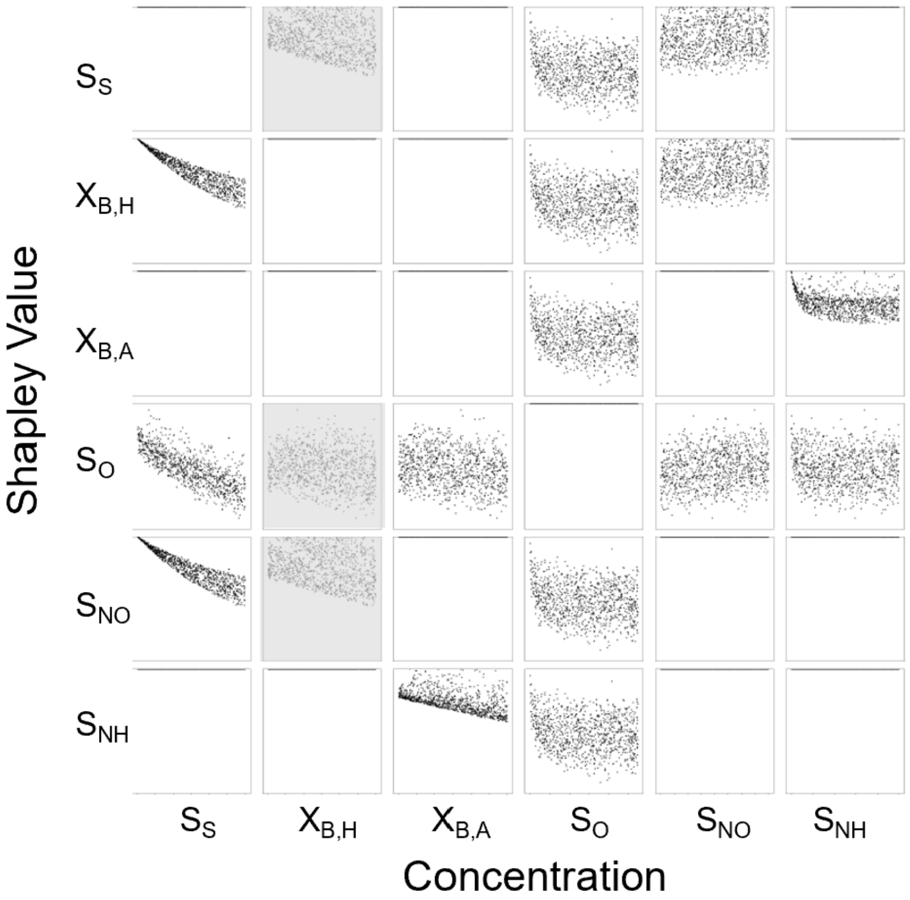
$$\frac{dx}{dt} = \boxed{f(x)} + \boxed{nn(x)}$$



Quaghebeur et al., 2025. Uncovering unknown dynamics in water resource recovery facilities with neural differential equations and Shapley value analysis. *Journal of Water Process Engineering*

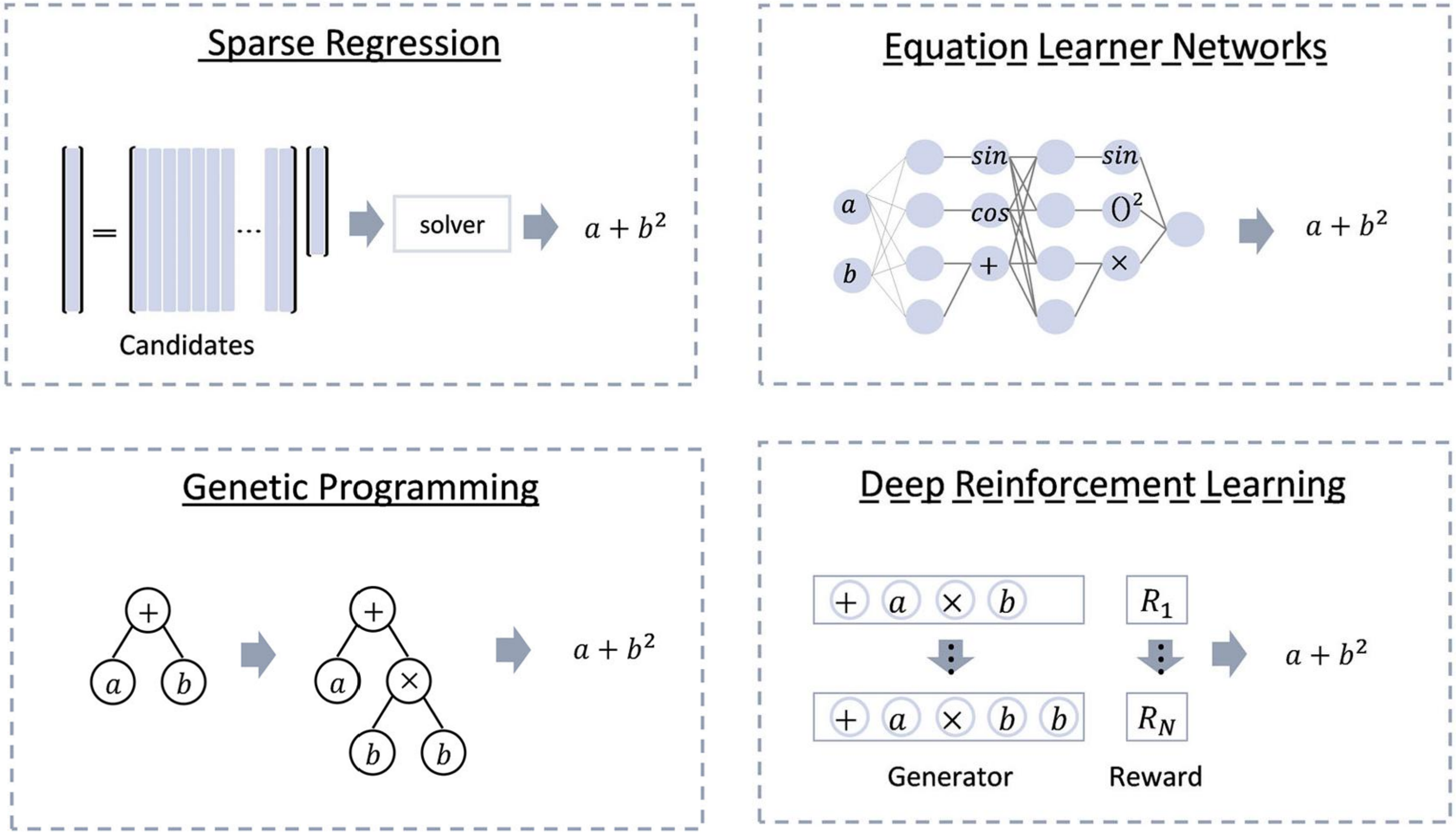
Equation discovery





Quaghebeur et al., 2025. Uncovering unknown dynamics in water resource recovery facilities with neural differential equations and Shapley value analysis. *Journal of Water Process Engineering*

Equation discovery



Song et al., 2024. Towards data-driven discovery of governing equations in geosciences. *Communications Earth & Environment*

Take aways and perspectives

- SciML
  - Groups many methodologies to integrate domain knowledge with machine learning
  - Obtain scalable, domain-aware, robust, reliable, and interpretable learning

- Hybrid modelling is maturing for applications in the water sector but other SciML methods are underexplored

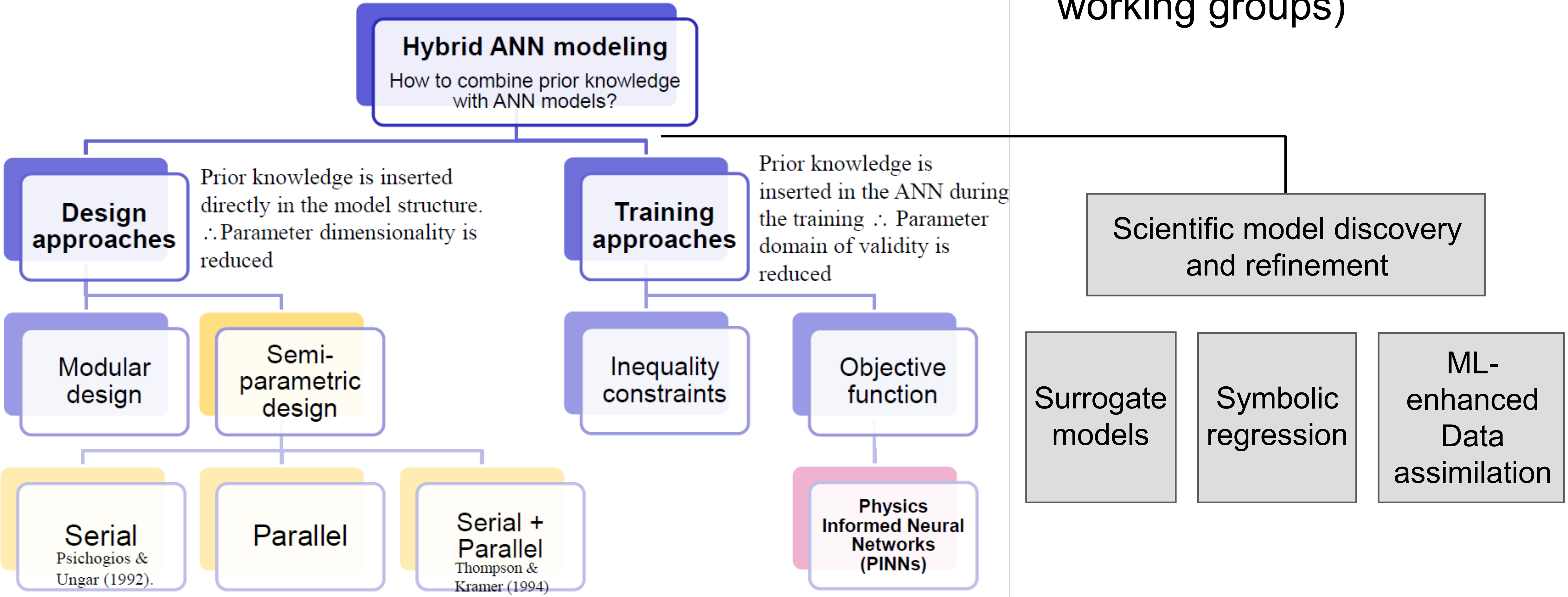
- Each method comes with challenges

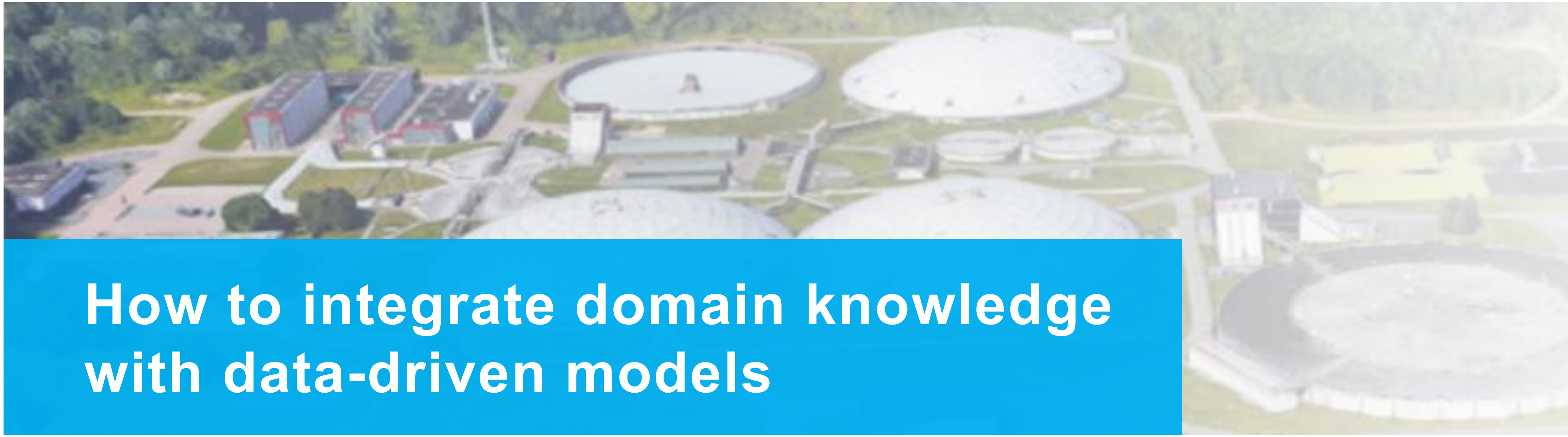
- <u>Common challenge</u> is model/method selection
  - First efforts for GMP ungoing in a combined effort of the GMP and HM working groups

Scientific Machine Learning

Thompson, M. L., & Kramer, M. A. (1994) **Modeling chemical processes using prior knowledge and neural networks**. AIChE Journal, 40(8), 1328–1340. https://doi.org/10.1002/aic.690400806

- <u>Common challenge</u> is model/method selection
- First efforts for GMP ungoing (GMP and HM working groups)





How to integrate domain knowledge with data-driven models

Elena Torfs