Data-driven digital twins

Where statistics meets physics

- A digital twin simulates how a physical system will perform throughout its lifecycle.
- Modelling such systems involves many uncertainties: the modeller must guess the structure of an equation to describe the phenomenon and fit it to inaccurate measurements (training data) to find the model parameters.
- The equation is then used to represent the system and predict the behaviour of new (unknown) data, and to enable various control and optimisation strategies.
- Associate professor Jan Kloppenborg Møller, Dr Goran Goranović and their colleagues at the Technical University of Denmark (DTU) and the Danish company Grundfos have developed a digital-twin methodology – Stochastic Greybox Modelling and Control – that accounts for uncertainties and random changes.

A digital twin is a virtual representation of a physical process or product. It is used to simulate how its physical counterpart will perform throughout its lifecycle. Digital twins can evolve with the real-time flow of data from a real-world system, helping developers understand and control the performance of the physical process or product. They also help developers overcome process uncertainties and deal with unforeseen fluctuations as they occur.

Associate professor Jan Kloppenborg Møller, Dr Goran Goranović and Professor Henrik Madsen from the Technical University of Denmark (DTU) and Dr Per Brath from Grundfos (currently at Danfoss) have developed a digital-twin methodology called the *Stochastic Greybox Modelling and Control* that accounts for uncertainties and random changes. They demonstrate its suitability for industrial use with an example application to fluid ultrafiltration.

Membrane separation technologies
Membrane separation involves passing a fluid through a membrane to remove unwanted contaminants. In industry, membrane separation technologies are used to remove unwanted particles from liquids, including water, blood, milk, fruit juice, and wine. Membrane separation processes are categorised according to the size of membrane pores and the particles they reject

(reverse osmosis <1 nm, nanofiltration 1–2 nm, ultrafiltration 2–100 nm, and microfiltration 100 nm–10 μm).

Uncertainties, such as unknown membrane surface forces, inconsistent pore size, and the nature of the liquid passing through the membrane mean that both ultrafiltration and microfiltration lack rigorous theoretical description. More knowledge is needed to account for the various assumptions and conditions used to control these processes.

Modelling with uncertainties
The research team explain how modelling physical systems involves different types of uncertainties. First, the modeller must guess the structure of an equation (eg, linear or quadratic) to describe the phenomenon. Second, they fit the equation to training data to get the best model parameters. Third, the equation is used to predict the behaviour

of new (unknown) data. Once the modeller chooses the best equation or most accurate model, it is used to represent the system, often controlling its performance in terms of minimal cost or optimal energy use.

A common problem in membrane separation, for example, is the accumulation of particles on the membrane surface clogging up its pores. This flow-retarding build-up increases with pressure but reduces with the crossflow. In their study, the researchers show how their digital twin model can control the filtration process using separate pressure and crossflow pumps while minimising the energy used.

Statistical blackbox vs physical whitebox
Two basic types of mathematical models can be used to describe data. Blackbox models, such as neural networks, rely on statistical

measures and data trends. These can lack insight and are prone to overfitting (when the model's parameters fit too closely to the training data).

Whitebox models contain only a few physically interpretable parameters, eg, Ohm's law (used to calculate the relationship between voltage, current, and resistance in an electrical circuit). They can be applied to a variety of phenomena but can be time-consuming and impractical to implement in complex physical systems, such as chemical process plants, or when real-time random uncertainties hamper the determination of model parameters (eg, renewable electricity productions).
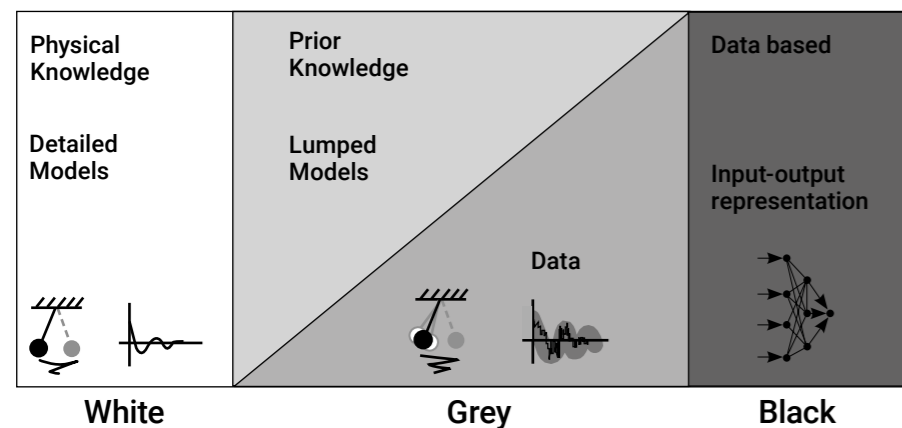
Stochastic greybox models
Møller, Goranović, Brath and Madsen chose to use a combination of these two models, known as greybox modelling. Combining physics with statistics, these models have a simplified physical equation structure. Using stochastic differential equations (differential equations in which one or more of the terms is a stochastic process – a series of random numbers), they include a diffusion term for

> **Data-driven digital twins use stochastic greybox models to describe hidden states of a system, ie the uncertain states that are not measured directly.**
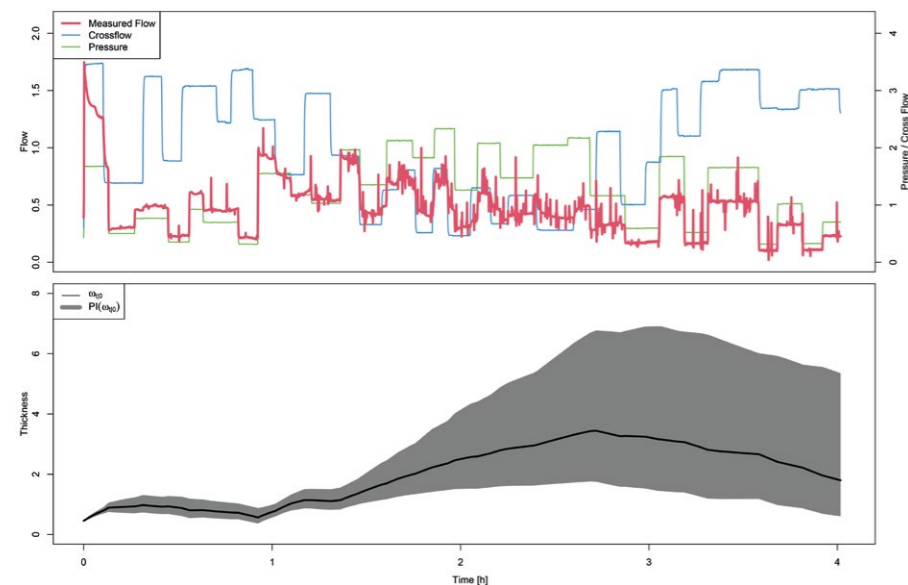
Top: Example of a measured variable (flux through a membrane) and input control variables causing it (pressure and crossflow). Bottom: Example of a hidden (unmeasurable) state: prediction of the mean value and the uncertainty margin of the accumulated filtrate (cake) at the membrane yielding the measured flux above.

Types of models: the researchers distinguish detailed physical whitebox models, simplified combined physical and data-driven (stochastic) greybox models, and purely data-driven blackbox models.

Comparison of the optimal controls with two experimental series

	V	$\overline{\Delta P}$	$\overline{Q}$	$\int Q^3$
Fixed ctrl.	3.00	2.50	2.35	53.74
Adaptive ctrl.	2.97	2.42	1.73	35.64
Ser. 9	3.02	1.93	2.86	105.40
Ser. 20	2.96	1.72	2.45	69.08

V—collected volume (integrated flux), $\overline{\Delta P}$—average pressure, $\overline{Q}$—average cross-flow, $\int Q^3$—consumed energy.

Control scenarios provide significant energy savings compared to uncontrolled process sequences.

> **The digital twins optimise real-time operations with programmable time-dependent inputs in a variety of control scenarios.**

uncertainty quantification. These models are called stochastic greybox models, as their outcomes are based on random probability.

Stochastic greybox models can describe the hidden states of a system that can't

be measured directly. For example, in membrane separation the thickness of the layer of dirt that accumulates on the membrane (essentially, the clogging) is modelled as a state. The flowrate through the membrane is a function of this state.

Modelling this fluctuating state, rather than just the flowrate, offers more insight into the system's behaviour, and controlling it helps control the system.

One problem facing the researchers is that the actual state is unknown, so they must estimate it based on noisy measurements of the functions of the state, ie the flowrate.

Identifying models and parameters

Modelling an unknown physical system requires diverse training data to establish accurate equation structures and model parameters. In physical sciences, simpler linear sequences of inputs are often used, but this is limiting. Instead, the researchers use statistical distributions to randomise the input, for example when inputting pressure sequences. Although this requires a more thorough statistical analysis and a programmable experimental set-up, the models that they identify are much more reliable.

To fit the greybox models to data and find the best parameters, the team applies advanced statistical hypothesis testing combining the Kalman filtering (apportioning of the uncertainties) with maximum likelihood estimation (minimising the overall error score). Care must be taken varying the number of parameters to enhance the likelihood, as this can overfit the model, so they measure the opposing tendencies of simplicity versus accuracy using Akaike Information Criterion (a numerical assessment of the goodness of a model that penalises the use of more parameters).

Real-time control

Once the researchers identify the best stochastic greybox model, it can be used to predict outcomes in new situations. They can optimise real-time operations with programmable time-dependent inputs in a variety of control scenarios, such as optimising pressure and dirt-removing crossflow for minimal energy usage during operations. Each new experimental run has a unique realisation (outcome of a stochastic process). The diffusion term ensures that the stochastic realisations are within the model's stochastic prediction margins, and the control steps are adjusted accordingly, so the greybox model itself is modified.

The researchers refer to their stochastic greybox models and control as the data-driven digital twins of the physical processes. In addition to the industrial water ultrafiltration application, they are currently developing digital twins for CO_2 capture as part of the Green Twins project and indoor climate optimisations for their new book, *Statistical Modelling of Occupant Behaviour*.

Details

Assoc professor Jan Kloppenborg Møller

e: jkmo@dtu.dk

Dr Goran Goranović

e: gogo@dtu.dk

Dr Per Brath

e: per.brath@danfoss.com

Professor Henrik Madsen

e: hmad_dtu@gmail.com

w: inno-ccus.dk/project/green-twins-data-driven-digital-twin-platform-to-reduce-co2-in-industrial-processes

Bio
Jan Kloppenborg Møller is associate professor in stochastic dynamical systems at the Technical University of Denmark (DTU). His research is concentrated on modelling and forecasting of (continuous or discrete time) stochastic dynamical systems, eg, ecosystems, urban drainage, wastewater treatment, wind and solar power forecast, and occupancy behaviour.

Goran Goranović, PhD, is senior researcher at DTU and physicist in theoretical modelling, and former associate professor from the University of Southern Denmark. His main research concerns fundamental theory and applied simulations of flows, chemical reaction kinetics, optimisation, and control of energy/CO_2 consumption of flow processes, including data-driven digital twins for optimal real-time CO_2 reduction in industry (Green Twins).

Per Brath, PhD, is a senior control engineer currently working at Danfoss Drives A/S. He holds an industrial PhD in control engineering and has been working in developing advanced control systems for tech departments in Danish companies Vestas Winds, Grundfos and Danfoss. He worked as an associate professor at Aarhus School of Marine and Technical Engineering.

Henrik Madsen is professor of mathematical statistics at DTU and Head

of Dynamical Systems group (DynSys) at DTU Compute. His research focuses on stochastic dynamical systems: time series analysis, greybox modelling, optimisation, and control. The main applications are energy systems, biostatistics, process modelling, finance, and indoor climate. Madsen received a knighthood from Her Majesty the Queen of Denmark (2016) and an honorary doctoral degree from Lund University (2017).

Further reading
Møller, J K, Goranović, G et al, (2022) A data-driven digital twin for water ultrafiltration. *Communications Engineering*, 1, 23.

Møller, J K, et al, (2024) *Statistical Modelling of Occupant Behaviour*. Chapman & Hall/CRC. ISBN 9781032334608