Developing a new tool for hybrid process modelling

K.M. Nauta, A. Silva, C.C. Pantelides
Process Systems Enterprise Ltd.
www.psenterprise.com

Purpose

PSE's gPROMS platform is a state-of-the-art tool for first-principles modelling with an established customer base in the process industries. One of the goals of PSE's involvement in the COPRO project is to enhance gPROMS to additionally facilitate and support *hybrid modelling*.

When developing process models two approaches can be followed:

1. First-principles modelling, whereby prior knowledge of the process, in the form of laws of mass and energy conservation, intrinsic kinetics and correlations for hydrodynamics, are combined to yield a model that is predictive over a wide operating range.
2. Data-driven modelling, whereby measurement data is combined with regression and machine learning to automatically yield models that best capture the input-output behaviour of the data.

These two approaches can be combined to yield *hybrid* or *grey-box* models. This allows the most appropriate approach to be used for each aspect of the system's behaviour. For example, while the behaviour of ideal unit operations is well-understood and can be represented well using first-principles modelling, the fouling or degradation of a real-world piece of equipment might be poorly understood. Here, adding a data-driven part to the first-principles model to predict this part of the behaviour would be an efficient way to model the unit.

Our goal is to develop a *hybrid modelling tool* that allows users to seamlessly combine both approaches in order to speed up model development.

Figure 1. Heat exchangers are well-understood unit operations and most first-principles modelling tools have accurate representations of these models in their model libraries.

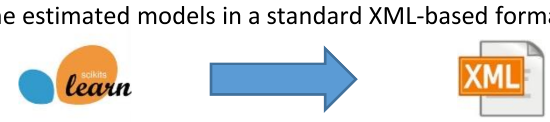
Figure 2. Fouling on a heat exchanger for a particular material might be poorly understood. Here plant data can be used to develop a data-driven model that captures this fouling behaviour.

Approach

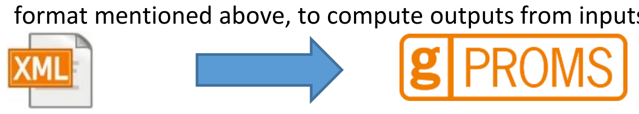
A prototype tool is being developed to facilitate and support hybrid modelling in the gPROMS platform. The tool is tested on COPRO industrial case studies and other case studies originating from PSE's customer base.

The hybrid modelling tool has the following elements:

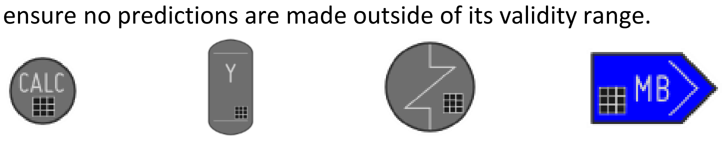
1. Python code built on top of scikit-learn to facilitate estimation of data-driven models in accordance with recommended workflows. Serialization (storage on disk) of the estimated models in a standard XML-based format.



2. A facility within the gPROMS platform that uses given data-driven models, in the XML-based format mentioned above, to compute outputs from inputs.



3. A library of standard unit operation models for flowsheeting in gPROMS ProcessBuilder. Each model has a data-driven component that predicts a key aspect of the performance of the unit. The data-driven component is monitored to ensure no predictions are made outside of its validity range.



4. A native C++ implementation of the ALAMO estimation algorithm (Cozad et al., *AIChE J.*, **60**, 2211-2227, 2014) for generation of data-driven models, integrated in the gPROMS platform.

Data-driven modelling approaches

The data-driven models need to accurately capture the behaviour of nonlinear phenomena, such as reaction, separation, hydrodynamics and fouling. They need to be able to cope with high-dimensional input and output data. Finally, their predicted outputs should satisfy constraints such as non-negativity, monotonicity, convexity. One promising approach to accomplishing these objectives involves using models that are linear in the estimated parameters with inputs selected from a feature space that is a constructed from nonlinear transformations of the original input space (see Fig. 3). Linearity in the parameters facilitates the use of estimation algorithms such as Partial Least Squares (PLS) or techniques that select a limited number of the most relevant features from the feature space such as ALAMO.

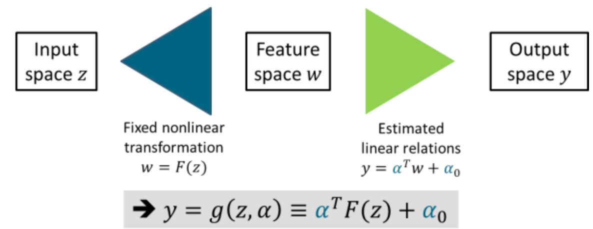


Figure 3. Data-driven models based on nonlinear transformations of the original input space. The models are linear in the parameters α.

We have successfully used this approach to construct data-driven models that accurately describe the behaviour of complex unit operations such as Solid-Oxide Fuel Cells (SOFCs) and olefins cracking furnaces.

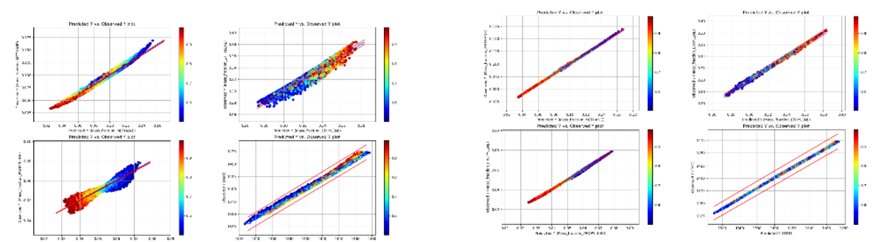


Figure 4. PLS model fitting of olefins furnace first-principles model. Fitting of a linear PLS model (left) does not yield sufficiently accurate results. Using a series of nonlinear polynomial transformations to generate higher-dimensional feature space followed by PLS model fitting (right) yields sufficiently accurate results.

Olefins Plant Case Study

Olefins plants produce ethylene, propylene from gas feedstock (typically ethane or propane) or naphtha. An olefins plant consists of a number of cracking furnaces with a cracked gas compression section and a downstream separation section. The gPROMS ProcessBuilder already includes a comprehensive modelling capability for steam furnaces including detailed modelling of cracking and coking kinetics.
In this case study, the detailed first-principles furnace model is replaced by data-driven models. Surrogate modelling is used to fit a Partial Least Squares (PLS) regression model to input/output data generated from the detailed first-principles model. This data-driven model is then used to replace the detailed first-principles model in the overall plant model. The result is a hybrid model where certain unit operations (the furnaces) are represented using a data-driven model and others (compression, separation) using first-principles models.
The hybrid model is used to optimise the operating point of the plant given the coking state of each furnace. The hybrid model has a significantly reduced number of equations (see Table 1). Nevertheless, the optimal point obtained closely matches that from the original first-principles plant model.

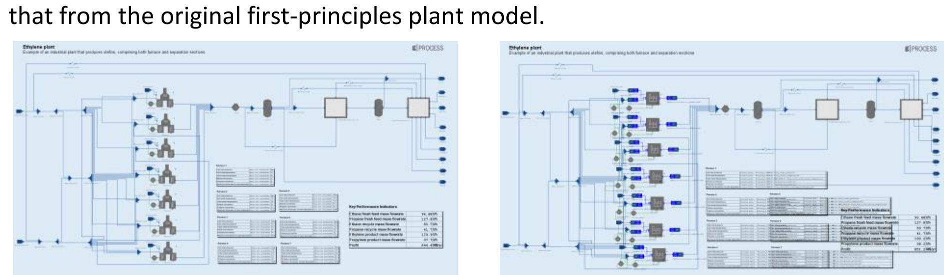


Figure 5. Olefins plant flowsheet with the original furnace models (left) and the data-driven models (right). Hybrid modelling can be done by drag-and-drop flowsheeting.

	Original plant model	Hybrid Model
Number of equations	112,865	9,132
Annual profit (optimised)	844.4 M$/yr	851.2 M$/yr

Table 1. Comparison of original and hybrid olefins plant model number of equations and annual profits at the optimal solution points.

SPIRE Sustainable Process Industry through Resource and Energy Efficiency

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723575