CoPro: Estimation of grey-box models with constraints

Project:
Improved energy and resource efficiency by better coordination of production in the process industries

COORDINATED PR
FOR BETTER RESOUR

The goal of the CoPro project was to develop and to demonstrate methods and tools for process monitoring and optimal dynamic planning, scheduling and control of plants, industrial sites and clusters under dynamic market conditions, to provide decision support to operators and managers and to progress to automated closed-loop solutions to achieve an optimally energy and resource efficient production.

CoPro brought together 17 partners from 8 EU countries, including 5 industrial end users and 6 technology providing SMEs. The project developed solutions for the **plant-wide optimisation of large plants, for balancing production and consumption in industrial parks for industrial symbiosis**, and addressed **power plant scheduling** and **demand-side response**. It further developed online data analytics for **anomaly detection**, and **decision support** for plant operators and managers. The solutions can be integrated into the IT infrastructure of the plants via an **integration platform** that supports the connection to different IT systems. CoPro developed **model libraries** for the efficient development of advanced optimisation-based solutions and techniques and software for **hybrid modelling** and **model management**.

The developments of CoPro were motivated by and applied to challenging use cases from different sectors of the process industries:

- (Petro-)chemical production;
- Cellulose fiber production;
- Production, formulation and packaging of consumer goods;
- Sterilisation and packaging of food.

CoPro demonstrated that significant savings of energy and resources are possible by using advanced technologies for monitoring, decision support, optimisation, and planning and scheduling.

Sector:

Cement

Ceramics

Chemicals

Engineering

Minerals

Non-ferrous metails

Steel

Water

Summary:

The Problem

- Process models are the cornerstone for the successful deployment of advanced control and real-time optimisation routines.
- Models often need to be customised with plant data to fit the actual processes, but fully replacing process knowledge by so-called deep learning is not a good idea.
- Data-driven models often match plants just at the current or past operation but predict unphysical responses far from these.

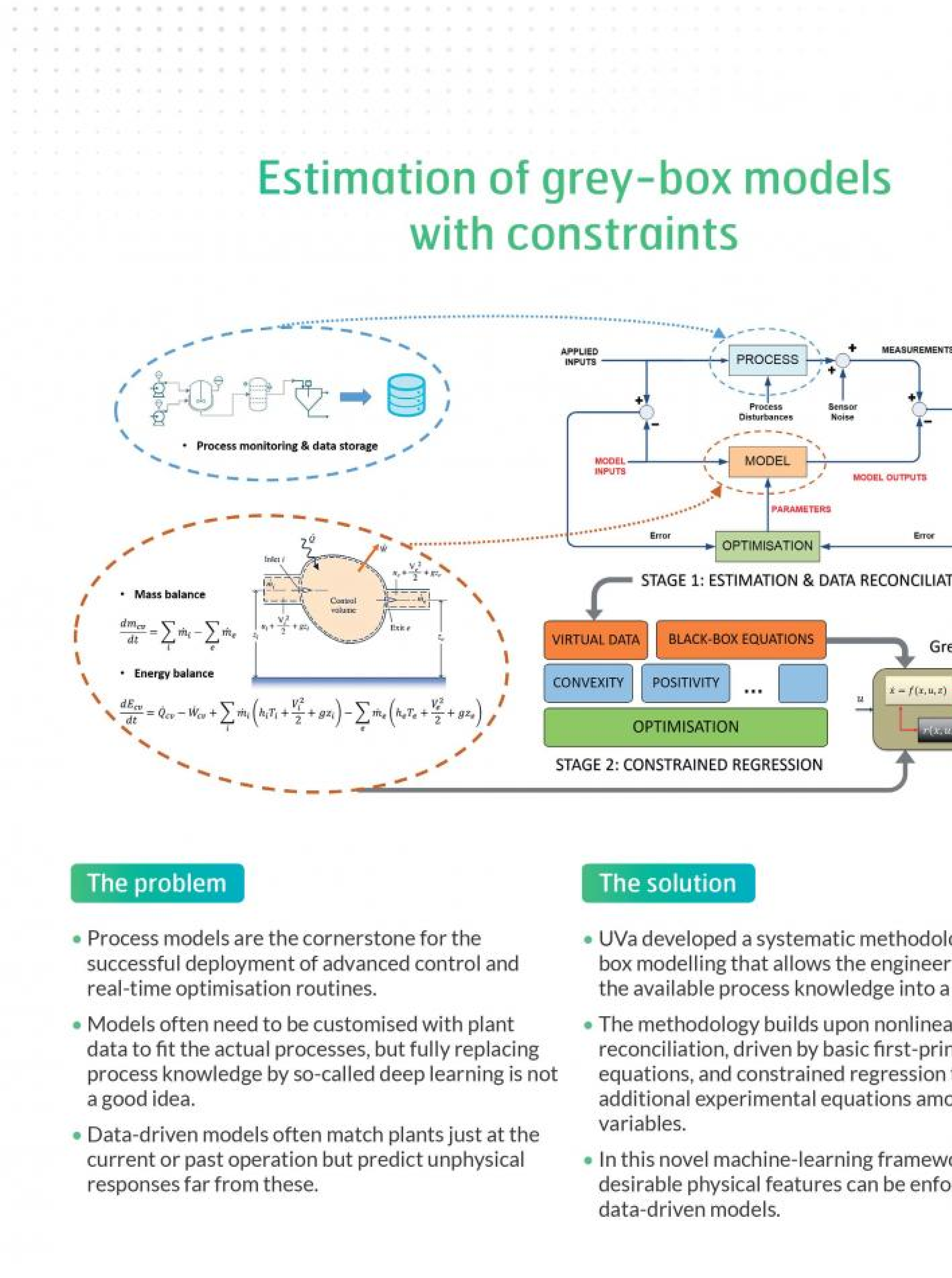
The Solution

- UVa developed a systematic methodology for grey-box modelling that allows the engineer to transfer the available process knowledge into a model.
- The methodology builds upon nonlinear data reconciliation, driven by basic first-principles equations, and constrained regression to seek for additional experimental equations among process variables.
- In this novel machine-learning framework, several desirable physical features can be enforced on the data-driven models.

Theme:
Plant-wide monitoring - SPIRE02-2016
Keywords:
Hybrid modelling; Data reconciliation; Constrained regression; Process-knowledge transfer; Physical coherence; Machine learning; Virtual measurements; Improved extrapolation; Systematic methodology; Pulp & Paper
Type:
Software
Poster

Resources

Link:
Technology Short Description: Estimation of grey-box models with constraints



Estimation of grey-box models with constraints

The problem

Pure machine learning is not enough

Recent computational advances opened the door to the development of computer-aided systems which support plant operators to solve complex decision problems in real time. However, **traditional rigorous physical-chemical models used in process design are not always suitable** for their usage within real-time decision support systems because of their excessive computational complexity or because they cannot be fitted to the actual plant data well enough.

This limitation can be overcome to some extent thanks to the increased amount of data due to the industrial advances in digitalisation, and the new techniques for machine learning. However, **machine learning relies on data of good quality and in large quantity,** and data from industrial process systems often lacks both: plants usually operate around the same set-points and tight production constraints prevent running experiments for data collection in other regions of the operational space. Hence, using **prediction models that were exclusively built from historical data may lead to unreliable models,** even if the data has been pre-treated and standard regularisation techniques are used in the fitting problem.

The solution

A two-stage modelling methodology

In the scientific process-systems community, the agreed way to approach the above problem is to build so-called grey-box models. These are hybrid models that are partly built from first principles and partly from plant data, with the aim of combining the advantages of both approaches: physical coherence and accurate match with the observed plant outputs.

The systematic methodology for grey-box process modelling proposed by UVa is structured into two stages: 1) estimation and 2) constrained regression.
The first stage starts from a set of basic first-principle laws of the process (e.g. mass and energy balances) and a large set of process data recorded from sensors, usually pre-treated to exclude plant stops and other operation outliers. Then, a **dynamic data reconciliation** problem is formulated (with this incomplete model) and solved via nonlinear programming. Numerical integrators and nonlinear optimisation algorithms are the involved tools in this stage. The outcome is a set of estimates of all variables over time which is as coherent as possible with the basic physics of the process.

Note that the model is still incomplete at this point, i.e., more inputs than the actual are needed to compute the outputs. Therefore, a **constrained regression** problem is stated and solved in a second stage, in order to find the unknown relationships between some process variables from experimental data, hence completing the model. Here, two main features make the difference:
- The data for regression is a combination of real data recorded from sensors with the virtual data for the rest of the variables that are generated in the estimation stage.
- Unconstrained regression is extended to include physical knowledge on the candidate models, such as (local) bounds on the outputs (e.g. positivity), constraints on the model slope or curvature, etc.

If candidate experimental sub-models are of polynomial structure, this constrained regression problem can be handled efficiently via sum-of-squares (SOS) programming (convex optimisation).

Absolute LS regression error for modelling a heat-transfer coefficient.

Method	*Training*	*Validation*	*Total*	*Fit deterioration*
Exponential regularised	*13.448*	*14.282*	*27.73*	*-*
SOS constrained	*14.751*	*13.362*	*28.113*	*1.36%*
First principles	*-*	*-*	*37.361*	*25.78%*

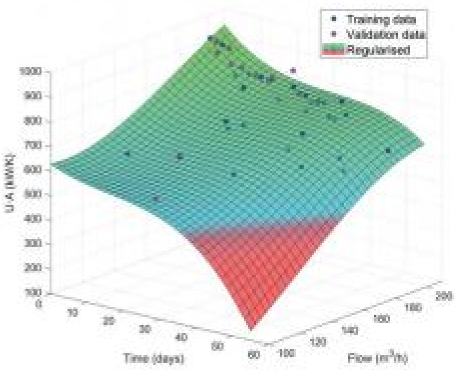


Fig. 1: Unreliable model prediction

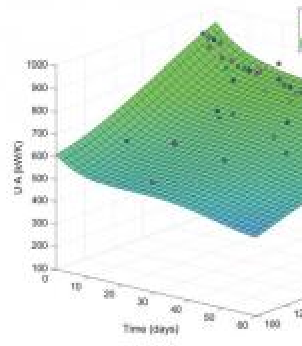


Fig. 2: Physically coherent predic

The key advantage of the SO is that these additional const only enforced on the recorde but in the whole input-outpu of interest for prediction. In predictions in between data importantly, for extrapolatio accordance with the known p process.

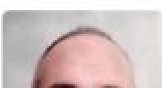
The summary

Coherent and reliab grey-box models

The two-stage systematic m proposed by UVa provides a where the knowledge that th have about the process is exp in the data-driven modelling objective is to avoid wrong p contradict the known physic which would invalidate the s provided by decision-suppor are based on such models.

The SOS-constrained regres well-behaved polynomial mo presence of scarce measurem a limiting factor in other mac techniques. It can be combin standard regularisation appr balance model complexity w of the fit to the data in the ob

The developers



Dr. José Luis Pitarch
Supervision and Process Control research group,
Systems Engineering and Automatic Control Dpt.

Prof. Dr. César de Prada
Supervision and Process Control res
Systems Engineering and Automatic