



Energy Data Science: Data Handling, Processing & Analysis

Klaus Lichtenegger
 Valentin Kaisermayer
 Uwe Poms
 Tagwa Sidahmed
 Markus Gölles

Increased Automation through
 Digitalisation and AI @ Subarea 2.2

**BEST – Bioenergy and
 Sustainable
 Technologies GmbH**

Head Office Graz
 Inffeldgasse 21b
 A 8010 Graz

P +43 5 02378-9201
 office@best-research.eu
 www.best-research.eu

Introduction: the Age of Data

The 21st century has already been nicknamed “the century of data”. Big data analytics, machine learning and Artificial Intelligence (AI) are pillars of the field of **data science**. Such approaches, while certainly to be handled with care, have revolutionized complete domains, from material research to medical diagnosis to translation services and to software development. Also for complex energy systems, a lot can be gained by establishing digital services that make good use of the data available.

Data at its BEST!

Data has always been important in science, process and control engineering, and many traditional, well-established methods are extremely powerful. Machine learning and AI, however, promise to expand the range of applications and find useful patterns in large amounts of data. At BEST, our approach is to bring together the best of both worlds, by using, on the one hand, physics-based models and sound domain knowledge, acquired during decades of dedicated research & development, and, on the other hand, state-of-the-art approaches from machine learning and AI.

Data platform

In complex energy systems, possibly hundreds of sensors are present, with sampling times up to 1 second or even smaller. The amount of data gathered that way quickly adds up and requires special infrastructure – in particular, if insights and value should be extracted from the data close to real time. This requires a well-constructed data pipeline, as exemplarily illustrated in Fig. 1.

Some Applications

Data Preprocessing and Feature Engineering: Often sensor data is incomplete, faulty, contradictory, ... and thus very hard to understand and analyse. Advanced statistical imputation techniques, data reconciliation methods and sensor fusion (virtual sensors) often allow us to “measure what can be measured and make measurable what is not so” (G. Galilei).

Data analysis and Visualisation: Sometimes, many sensor signals have to be examined in parallel in order to identify interesting patterns. Often, however, many signals are strongly correlated. In such cases, as shown in Fig. 2, the variables can be combined to a few meaningful features that allow an analysis of the whole system.

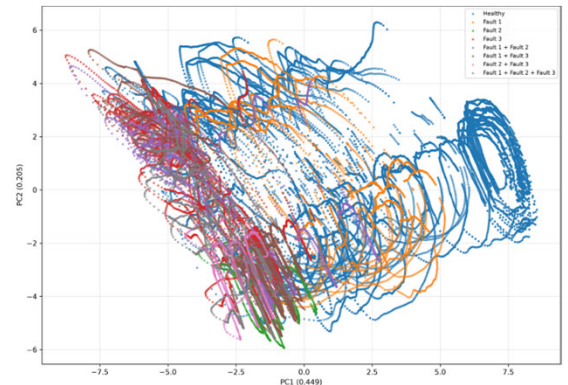


Fig. 2: Principal component analysis (PCA) for a domestic solar thermal plant. Via PCA, 33 variables are recombinced to just two new ones that still explain >65% of the variance.

Fault Detection and Predictive Maintenance: Model-based checks, combined with AI-based pattern recognition allow for fast detection of faults and anomalies. This permits fast response and thus a minimisation of losses. Prediction of expected failure time also allow to optimize maintenance schedules.

Collaboration and Services

BEST is a research, development & innovation partner for energy data activities, from **designing** a suitable **data handling system** / data platform to developing soft sensors for interesting quantities, establishing **advanced monitoring solutions** for anomaly detection and implementing a **predictive maintenance** scheme.

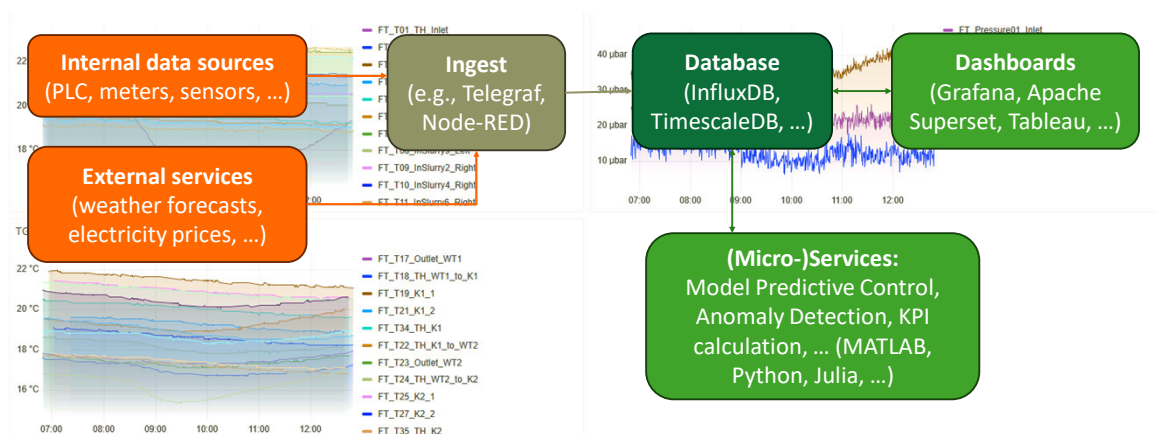


Fig. 1: Elements of a data processing pipeline, with a real-world example for dashboards (from the ATLAS data platform for the Syngas Platform Vienna) in the background