Photo: Siemens

From theory to results

Data-driven modelling and computer simulation can be used to help improve process control performance, leading to more efficient and sustainable operations in oilseed crushing plants and edible oil refineries

Tomáš Svoboda

The processing and refining of oilseeds – such as soyabeans, canola or sunflowerseeds – into vegetable oils is a complex technological process.

The oilseeds must first be prepared through cleaning, conditioning, dehulling and flaking. The next step is the extraction of crude oil, either by mechanical pressing or solvent extraction (typically using hexane), or a combination of both. The meal remaining after hexane extraction is a valuable material for animal feed production and must be desolventised, toasted, dried and cooled. The extracted oil then undergoes refining, which typically includes degumming, neutralisation, bleaching and deodorisation.

These diverse unit operations require carefully designed and maintained control strategies to ensure both product quality and process efficiency.

There is also significant variability in the chemical and physical parameters of seeds from the field. These differences are not always identified by routine intake analysis and may later manifest as unexpected process behaviour, posing challenges for operators.

Traditional control relies heavily on Proportional Integral Derivative (PID) controllers, a workhorse of industrial automation.

PID is a widely used control strategy that adjusts a process output (such as flow or temperature) to match a desired value (setpoint) by reacting to the difference (error) between them.

For example, think of a home's thermostat. When the temperature drops below the desired setting, the heating turns on. When it gets too hot, it turns off.

PID controllers continuously adjust process variables like flow, temperature or pressure to maintain a desired setpoint.

While reliable and widely used, PID loops have limitations: they react to changes but do not anticipate them. In systems with long delays or complex interactions, PID alone may fall short. This is where Model Predictive Control (MPC) enters the picture.

MPC is a type of Advanced Process Control (APC) strategy which uses a dynamic model of a process to forecast how it will respond in the near future. Based on these predictions, it optimises control actions in advance.

To continue with the thermostat analogy – unlike a thermostat that reacts after you're cold, MPC knows a cold front is coming and starts heating early to keep you comfortable.

MPC allows companies to operate closer to constraints, coordinate multiple variables and proactively handle disturbances. This is possible because a model-based computational simulation of a process has the ability to provide virtual transparency into the process unit – allowing for the prediction of unmeasured or difficult-to-measure parameters such as moisture content, chemical composition, particle size distribution or viscosity.

This enables the formulation of control objectives that not only ensure the final product consistently meets specifications, but also minimises cost factors such as steam consumption, retention time or catalyst dosage.

Siemens Digital Industries has an APC suite that offers a range of MPC options that can be configured based on application needs.

It has two primary options: Linear MPC and Nonlinear MPC.

Linear MPC

Linear Model Predictive Control (L-MPC) is often the first step toward smarter control. It is best suited for processes that are relatively stable and predictable but still require the coordination of multiple variables.

In oilseed processing, one good example is the Desolventiser-Toaster-Dryer-Cooler tower. This unit must carefully manage steam input, air flows and residence time to achieve the right combination of solvent removal, deactivation of anti-nutritional factors and preservation of protein quality. Too little heating and the meal quality suffers; too much, and valuable amino acids degrade.

L-MPC helps strike that balance. It uses a simplified mathematical model of how the system reacts to changes in input – such as steam or air – and finds the optimal way to adjust them over time. Because it is computationally lightweight, it can be run directly on the automation system.

Nonlinear MPC

Some processes, however, are far from simple. Chemical refining, especially degumming and neutralisation, is a good example. In degumming, crude oil is treated with water and acid to remove gums, which are phospholipids that must be separated to improve oil quality. This is followed by neutralisation, where caustic soda is used to remove free fatty acids. Both steps influence the formation of soapstock, a by-product that can trap significant amounts of oil if not properly controlled. What makes this process tricky is that gums and soaps behave like surfactants. They can form emulsions and micelles, small droplets that are difficult to separate and cause valuable oil losses. Small improvements in control can lead to noticeable gains in oil yield and chemical savings.

This is where Nonlinear MPC (NL-MPC) comes in. Unlike its linear counterpart, NL-MPC is built on a detailed process model that includes chemical reactions, mass transfer and non-linear interactions between variables.

Siemens is actively developing such models using the gPROMS platform, combining known system properties with parameters estimated from real data. Once validated, the model becomes a digital twin — a predictive replica of the real process. It is then used in a NL-MPC application that can adjust dosing rates, ▶

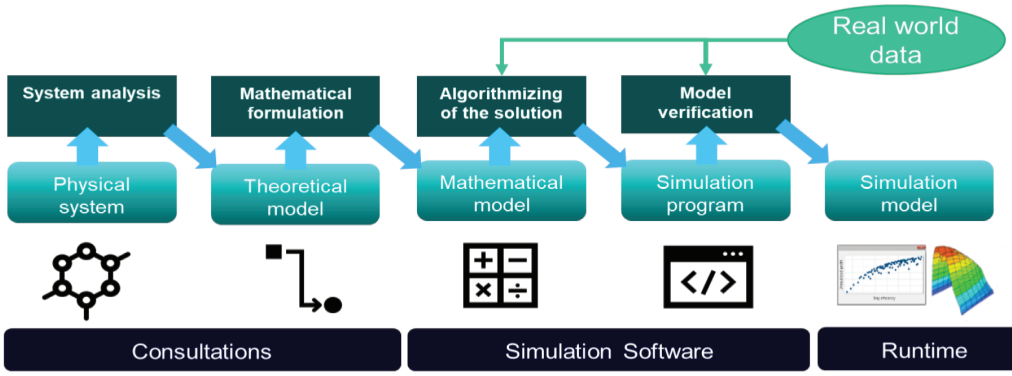
What is a data-driven approach?

Data-driven control uses historical or real-time process data to model and improve operations, without needing detailed physical equations. This approach includes:

Machine learning
Algorithms learn patterns in data to predict outcomes like product quality, energy use or equipment behaviour.

Deep learning
A subtype of machine learning that uses complex neural networks to model large, nonlinear systems with many variables.

Reinforcement learning
A type of AI where an agent learns to make decisions by interacting with the process (real or simulated) and improves based on feedback, like trial-and-error learning with rewards.



Mechanistic simulation model development workflow.

Table: Potential benefits from Advanced Process Control deployment

Unit Operation	Potential Improvement in KPI
Flaking	1% increase in extraction yield due to more consistent flake thickness
Solvent Extraction	0.5–1% increase in oil yield; up to 2% reduction in solvent loss and energy usage
Desolventizing / Toasting	Up to 4% energy savings; improved meal digestibility (1–2% increase in protein solubility)
Degumming / Neutralization	Up to 3% reduction in caustic or acid use; 0.5–2% higher refined oil yield
Bleaching	Up to 6% reduction in bleaching earth consumption; improved oil color consistency
Deodorization	1–5% reduction in steam usage; stabilization in product quality (FFA, odor)

Figure 1: Potential benefits from Advance Process Control (APC) deployment

Source: Siemens

mixing conditions and temperature in real time, all while accounting for variability in feedstock and other disturbances. By doing so, the controller helps reduce reagent consumption and increase refined oil yield.

Data-driven, hybrid approaches

In some parts of edible oil processing, the underlying mechanisms are too complex, variable or poorly understood to be described fully by first-principles or mechanistic models.

In these cases, data-driven modelling offers a practical alternative. Instead of trying to model the physics or chemistry directly, data-driven models use historical or real-time plant data to understand and predict system behaviour (*see box above*).

Machine learning or deep learning algorithms can be trained to estimate key process variables or product quality indicators, even when their physical dependencies are unclear.

For example, in semi-continuous deodorisation, where feedstock quality and operating conditions vary, machine learning models can anticipate changes in free fatty acid (FFA) content.

When connected to a supervisory control layer, such models can help operators adjust parameters proactively, improving consistency and throughput without detailed mechanistic insight.

In more advanced scenarios, reinforcement learning agents can be trained to interact with a process – real or simulated – and learn optimal control strategies over time, guided by a reward function such as energy efficiency or yield.

While promising, the success of data-driven approaches depends heavily on the availability and quality of data.

Plants must have reliable instrumentation, good data storage practices and sufficient historical coverage to build accurate models. This can be a limiting factor in older facilities.

When some parts of the process are well understood – such as heat transfer or flow behaviour – but others remain opaque, a hybrid modelling approach can be used.

Here, known physical relationships are described using equations, while less-understood phenomena are modelled using ML techniques. This method is particularly effective for advanced toasting, where physics-based models describe moisture and temperature evolution, but neural networks are used to predict complex outputs like amino acid degradation.

A refined form of this approach is the use of Physics-Informed Neural Networks. These AI models are trained to make predictions that remain consistent with known governing equations.

Siemens has explored this method in VRX-DTDC applications to improve the balance between energy use and nutritional value in soyabean meal processing.

Conclusion

The edible oil industry presents numerous opportunities to unlock the potential of model-based control to significantly improve process efficiency, product quality and energy usage.

To help companies implement advanced control solutions, Siemens is moving beyond a traditional product-based business model towards an integrated, value-driven approach.

It is currently undergoing a strategic transformation – from a focus on traditional industrial technologies towards a more digitally-driven approach – centred around the expansion of its Xcelerator platform. As part of this move, Siemens strategically integrates software companies into its organisation with specialised domain expertise and established solutions.

Notable examples include the integration of PSE, the company behind the widely recognised process simulation framework gPROMS, in 2018; the acquisition of computational chemistry company Culgi in 2019; and the recent acquisition of Dotmatics, a provider of R&D software for the life sciences sector. •

Tomáš Svoboda is a technical account manager at Siemens