navigance

Chemical process models – from first principle to hybrid models

Models and why they matter

First principle models: putting knowledge to work

Empirical models: where data speaks

Hybrid models: the best of both worlds

Author: Sebastian Werner, 2020

MODELS AND WHY THEY MATTER

Models are a vital tool for both the chemical industry specifically and the engineering sciences as a whole. In essence, they're mathematical representations of the relationships between variables in (real) systems, which describe and help operators understand the behavior of a plant or a process.

They come in different shapes and sizes. They're built using various approaches and techniques. They offer different capabilities and benefits. And, of course, they have their own limitations, too. So it's important to understand these differences in order to build and use models that will provide the greatest operational advantage.

So what are the types most commonly used in chemical process engineering? And what are their advantages and disadvantages in answering key questions such as where and how to optimize?

Contrasting approaches

One feature that distinguishes different model types is the level of knowledge about the underlying physical or chemical phenomena in the chemical process needed to develop them.

There are many contrasting criteria: linear vs non-linear, discrete vs continuous, static vs dynamic, explicit vs implicit, and deterministic vs probabilistic. More specifically – and at opposing ends of the spectrum – there are first principle and empirical models.

First principle or 'knowledge-based' models start from the basis of established science, are generally accepted and have been extensively verified. They use explicit, recognized correlations in science and engineering, including chemical engineering. A well-known example is the Arrhenius equation, which describes the temperature dependence of different reaction rates.

In contrast, empirical models are based purely on arbitrary mathematical correlations. These describe relationships in the available data from any observed process or dataset.

In principle, the results of these approaches could be compared with measurements such as flow, composition, temperature, pressure, and other variables to paint a real picture of a plant's behavior. Collecting such measurements is a whole other challenge, of course.

Let's look at each model type in some more detail.

FIRST PRINCIPLE MODELS: PUTTING KNOWLEDGE TO WORK

First principle models are built on a fundamental understanding of underlying 'ab initio' physio-chemical phenomena such as mass transfer, heat transfer and mass flow. Often, they're also based on the explicit relationships in a particular unit operation within a chemical process. And they can be used to connect different unit operations by mass and heat balance in a process flowsheet.

First principle modelling has been used extensively to help design new plants. In many cases, a flow-sheet model combining first principles with correlations such as heat transfer, porosity and vapor pressure can be constructed to describe the full extent of the plant. But this can take many years (even decades) of work, lots of resources and substantial investment.

These models are often also the starting point for revamping plants, debottlenecking, or supporting and optimizing day-to-day process operations. Here, though, observed 'real world' data is often needed to "fill in the blanks" where parameters are unknown or missing.

The need for knowledge

First principle models depend on a lot of information, from the properties of chemicals and mixtures involved in a process to data about reactions (such as kinetics) and thermodynamics.

They also need a comprehensive, real-world knowledge of the underlying correlations, which demands expertise from different disciplines, including equipment manufacturers, process licensors and catalyst suppliers, to accurately describe a chemical operation.

Some of the information required can be found in readily available chemical engineering literature. Other, more specific data must be gathered from experiments designed with the particular process to be described in mind. This presents a challenge, though.
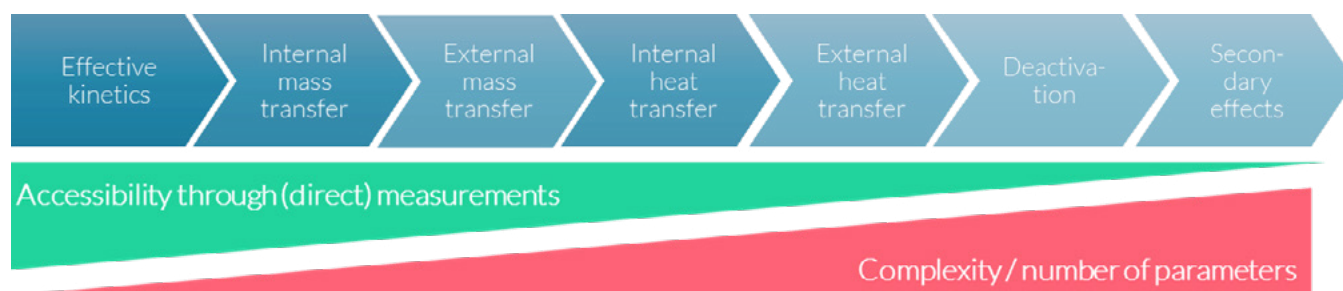
In real industrial scenarios, observed behavior can deviate markedly from the typical correlations we might expect to see (for instance, where chemical mixtures form azeotropes). The reason for this is that our understanding of certain phenomena – such as non-ideal mixtures, or heat and mass transfer in porous systems – is limited.

As a result, the chemical engineer's traditional workaround has been to rely on dimensionless number approaches, such as tabulated heat-transfer correlations or vapor-liquid equilibria (cf. VDI Wärmeatlas, Dechema Datenbank). And the workarounds don't stop there.

The problem with parameters

All first principle models rely on adequate parameters, applicable to the specific process they're describing or the problem they exist to solve. Getting them is easier said than done.

One of the big challenges in accessing the information needed is that, in looking beyond base effects to the secondary effects that also influence the chemical process, accessibility through (direct) measurements decreases as the complexity and number of parameters involved increases.

In search of an easier answer

We should note that pretty much anything can be modelled in this way, and deliver a number of reasonable qualities, if enough time, budget and energy is invested in doing so. But, with its reliance on engineering handbooks, the validity of the first principle approach will always be limited.

Knowledge-based models need to deliver on several fronts if they're to help design new plants, revamp old ones, reduce bottlenecks or drive process efficiency. They must have ample scope for extrapolation. Provide a smooth, continuous, and differentiable foundation for optimization. And, in environments where major changes in processes, equipment and catalysts are likely, they must adapt easily and be constantly reviewed to keep them up to date.

Accurately modelling real-world problems from the bottom up can get complicated quickly, as accessing the right measured parameters gets more and more challenging.

How accessibility of information decreases the greater the complexity and number of parameters in a chemical process

Accurately modelling real-world problems from the bottom up can get complicated quickly, as accessing the right measured parameters gets more and more challenging.

Despite decades of research to try to close these gaps, uncertainty remains and phenomena observed in real-world industrial systems still cannot be sufficiently described. It's easy to end up with thousands of model equations, often coupled with partial differential equations, that cannot be solved analytically. In many cases even solving them numerically is a major challenge. So "fiddle factors" have come into play.

These are parameters used to fit model outputs to the observed data, or to design and safety margins in a plant's design. Essentially, they compensate for unknowns or uncertainties in the models and, arguably, inadequacies in the first principle approach as a whole. But such approximations can leave significant inaccuracies in the final result, meaning all effects cannot be confidently captured to their full extent.

Is all of the traditional effort of first principle really needed in order to optimize a chemical process? After all, an existing plant is defined by properties such as its installed equipment or catalysts. And its behavior can be described by data measured within it.

As such, the extrapolation capabilities usually needed to design a new plant wouldn't be required to model or describe the process, at least within the scope of existing data. And surely data captured from the process is the asset that offers the greatest scope to learn what's needed to optimize?

Let's now look at the other extreme: empirical models.

EMPIRICAL MODELS: WHERE DATA SPEAKS

Pure statistical models are used to describe observed data and behavior by using arbitrary functions to find a good fit between the output of the model and the observations. These models can be applied even when little to no physical or chemical engineering understanding is available. But caution is needed.

One characteristic of chemical processes is the very complex interplay between variables and influencing factors. Some, such as feed flows and reactor temperature, can be controlled directly or indirectly. Other outside factors, such as ambient humidity and temperature, can influence the process but not be influenced themselves.

By looking only into the data from a chemical plant, without applying any domain knowledge to what you're seeing, it's all too easy to draw incorrect conclusions based on what's known as 'spurious correlations.' This is where at least two events or variables show the same or a similar trend, even though there's no likely link (or 'causal relationship') between them. It might be

Finding 'Spurious Correlations' by comparing Miss America's age with murders by steam, hot vapours and hot objects
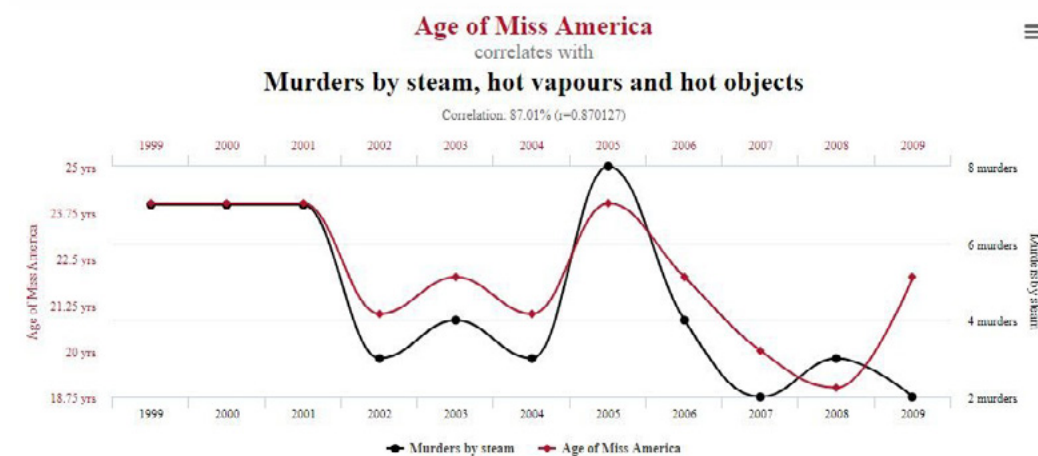
Statistical shortcomings

The point this serves is to show datasets alone don't necessarily reveal if relationships between trends are due to plausible correlations, let alone give indications on the cause-and-effect reasoning.

To make things trickier, in chemical processes residence time plays an important role too. Between input (let's say a feed flow) and output (e.g. a product flow), a certain time lag exists. If the flow of feed in the inlet is changed it could be hours before the product flow sees a corresponding change. This may be because of hold-up in the plant, slow equilibration of chemical reactions, temperatures and so on. Without specific domain and process knowledge, this also makes finding relationships between trends more challenging.

Relying on statistical parameters alone can be very misleading. They may look right yet not accurately capture basic relationships – a phenomenon explored well by both Anscombe's quartet and the Datasaurus Dozen. What's more, there's no escaping the fact that getting statistics right demands huge volumes of data. The less you know, the more of it you need.

Look carefully at your data before drawing conclusions. Correlations might be pure coincidence.

Source: Tyler Vigen, Spurious Correlations (data sources: Wikipedia; Centers for Disease Control & Prevention). Used here with thanks.

due to pure coincidence. Or it could be the result of a third, unseen factor.

A good example comes when comparing the age of Miss America winners with murders carried out using steam, hot vapors, and hot objects over a 10-year period. It may sound ridiculous – and it is – but it's just one of several extreme but hilarious illustrations gathered by Tyler Vigen in his Spurious Correlations series.

So while statistical models have their uses, they don't reflect long-standing chemical engineering knowledge, may actively contradict or misdirect, and can't keep up with the reality of operating a chemical process. Drawbacks that a number of chemical producers using them have reported, and which can't be resolved by applying advanced analytics or machine learning techniques, even with large sets of data.

Fortunately, there is another way...

HYBRID MODELS: THE BEST OF BOTH WORLDS

Some of the disadvantages and limitations of the first principle approach and empirical models we've discussed can be overcome using hybrid models instead.

These come in various shapes and sizes too. What they have in common is they still rely on first principle domain knowledge to one degree or another, and combine that with statistical methods where the knowledge needed is lacking or doesn't sufficiently describe reality.

They combine the best of both approaches, while also bringing fresh benefits of their own to the table. And they're not a new concept.

The use of data-driven neural networks and hybrid models to dynamically model chemical reaction systems was being described 20 years ago1. Using them successfully, though, is more of a skill. It's important to identify upfront the hybrid models genuinely capable of delivering the goals they're designed to achieve – such as process optimization. (And, of course, know their own limitations.)

In today's chemical plant operations, optimization is guided mostly by operators' and engineers' knowledge and experience. If time and budget permit, this may also be complemented by basic and periodic analysis of whatever data is available.

What if you could optimize on a continuous and ongoing basis without investing such time and resources? You can. Digitalization and the latest technologies are already offering chemical producers the scope to make great advances in this area.

We make models do more

Navigance takes a fresh approach to find untapped potential in your chemical processes. We've developed next-generation hybrid process models that can adapt easily to your plant's historic and real-time data, to generate reliable, prescriptive advice that helps you continuously optimize your operations.

How do these hybrid models compare to their first principle counterparts? They still use established first principle techniques, physico-chemical relationships and engineering principles as a basis. This ensures, for example, that the mass balance is closed and cause-effect relationships in the chemical process, such a residence times, are correctly captured.

As a starting point, we use a process-specific base model constructed using chemical engineering know-how. Then we review your plant, its setup, your available sensors and measurements, and any operational constraints. And we identify the relevant process data to measure and optimize performance.

It's this process data that brings our hybrid process models to life. By carefully integrating unknowns using arbitrary functions derived from the data, our models ensure a good fit for all data they observe.

We take great care to make sure all relevant effects in your plant are captured and considered, then apply advanced machine learning techniques. This creates AI-enhanced, data-driven, hybrid process models that, in contrast to rigid and inflexible first principle models, can learn intuitively and adapt quickly to changing conditions.

Using intelligent algorithms to consider all data observed from a continuous stream, they generate invaluable insights. So you can respond quickly to a host of factors that might affect your processes – from varying load scenarios to deactivating catalysts.

All this means you can act quickly and with confidence on real-time recommendations for process control variables to reach your optimization goals and most pressing KPIs.

IT'S TIME TO OPTIMIZE YOUR OPERATION.

We already have a Navigance base solution that's ready to go.

It works with any technology license and catalyst. And we can't wait to show what it could mean for you.

So let's talk about your plant's setup and needs and tailor it to suit.

TALK TO NAVIGANCE TODAY

Email: info@navigance.com

Phone: +49 89 5110126

Navigance GmbH

Rundfunkplatz 2

80335 Munich

Germany

FIND OUT MORE

www.navigance.com

[Footnotes and references]

[1] Dynamic Modeling of Chemical Reaction Systems with Neural Networks and Hybrid Models in Chemical Engineering & Technology by Hans-Jörg Zander, Roland Dittmeyer, Josef Wagenhuber (8 July 1999) https://doi.org/10.1002/(SICI)1521-4125(199907)22:7%3C571::AID-CEAT571%3E3.0.CO;2-5

And see other publications such as Molga and Westerterp (1997), Zander et al (1999) and Molga and Cherbanski (1999)