THE ONLINE USE OF FIRST-PRINCIPLES MODELS IN PROCESS OPERATIONS: REVIEW, CURRENT STATUS & FUTURE NEEDS

C.C. Pantelides
Process Systems Enterprise/Imperial College London
London W6 7HA, United Kingdom

J.G. Renfro
Honeywell Process Solutions
Houston, TX 77042

Abstract

The online use of First-Principles Models (FPMs) to support process operations has been practised in the chemical and petroleum industry for over 40 years. FPMs can encapsulate a large amount of process knowledge and many companies have realized significant value from the use of these models in online model based applications (OMBAs). Such applications include real-time optimization, model predictive control, data reconciliation, virtual sensors, and process performance monitoring to name a few. The sophistication of both the FPM models and applications based on them has increased over time. At some points in the evolution certain applications were not as successful due to issues related to sustainability, which includes model complexity, solvability, maintainability and tractability. Also, model development cost can be a factor in considering the type of model used in these applications. Hence many simplified and empirical model based online applications became preferred in some domains, even though the overall prediction quality of the FPM may be superior. This paper will review the past experiences, current status and future challenges related to FPM based online modeling applications. There are many areas where the issues related to FPMs can be addressed through proper model management, better software tools and improved technical approaches and work processes. It is hoped that this paper can serve as a basis to promote an understanding of the issues for researchers, modeling software vendors, modeling engineers, and application engineers and help to stimulate improvements in this area so that more usage and value of FPMs to support process operations is enabled in the future.

Keywords

Process operations, Process modeling, State estimation, Model-based control, First-principles models

Introduction

This paper is concerned with the usage of first-principles models in online computer applications to support industrial process operations. Below, we start by defining some key concepts.

First-Principle Models (FPMs)

We are interested in mathematical models which establish a quantitative description of the effects of the process inputs (including both deliberate control actions and external disturbances) on the process outputs and key performance indicators (KPIs) characterizing the technical and economic performance of the process, and its safety, operability and environmental impact. In this context, we loosely define FPMs to be those based on fundamental engineering, physics and chemistry principles, in contrast, for example, to empirical mathematical or statistical correlations between input and output variables derived from plant or other data (e.g. models based on step/impulse responses, neural nets, time series, or other forms of data regression). Of course, most non-trivial models do contain some empirical components since they involve quantities that cannot be fundamentally calculated. However, our FPMs will include, at a minimum, mass and energy balances; they will also usually employ a physics and/or chemistry-based description of some or all of the terms (e.g. flux or source terms) that appear in these balances in terms of fundamental thermodynamics, kinetics, fluid mechanics and so on.

Inevitably, the above considerations allow a spectrum of models of varying rigor to fall under the FPM description. However, a common characteristic of all but the most trivial of these models is their nonlinearity which can be quite high for certain classes of applications such as those that involve chemical reactions. On one hand, the nonlinearity of such processes provides at least part of the justification for using FPMs in terms of their ability to accurately describe behavior over wider ranges of operation than is possible using simpler alternatives such as linear models. On the other hand, it also means that it is often impossible to guarantee that the models will always be solvable within the desirable timeframe, or even at all. Thus, the issue of robustness of model-based calculations and the tractability of the model is of central importance.

A second common characteristic of the FPMs of interest to this paper is their dynamic nature. Although certain types of online calculations (e.g. the so-called "Real-Time Optimization" used to determine optimal steady-state set-points for continuous processes) are based on steady-state models, most of the applications considered here require a description of the transient process behavior.

FPMs are implemented in modeling and programming languages, and delivered in deployment forms for embedding or direct execution. In these deployment forms, FPMs can be a part of a distributed application or can be embedded in the controller/estimator application directly.

Online Model-Based Applications (OMBAs)

The online applications of interest to this paper are various types of model-based calculations which involve some form of direct data connection to the process plant and/or its control system. At a minimum, we assume that OMBAs receive as inputs plant measurements made available by online sensors; these may be complemented by additional offline measurements (e.g. results of laboratory analysis) where available.

From the point of view of their functionality, OMBAs can be categorized into several distinct types:

(a) **Plant monitoring**: these provide information to the plant operators about the current state of the plant, that is generally superior in accuracy and/or extent to what can be provided directly by the sensors.
(b) **Plant forecasting**: potentially, OMBAs may also provide information on the future state of the plant, given its current state and, possibly, one or more hypothetical scenarios involving future actions or events.
(c) **Open-loop decision support**: these take the form of actions (e.g. changes to the plant control system set-points) that are suggested to the plant operators as means of achieving certain objectives or outcomes. This is a purely advisory function, with the actual implementation of these actions being left at the discretion of the operators.
(d) **Closed-loop control & automation**: these are similar to the above, except that the proposed actions are implemented automatically on the plant, either directly or via its control system, without any operator intervention.

OMBAs may be executed continuously (e.g. in synchronization with real-time control systems) or at scheduled times (e.g. at regular time intervals as, for example, in the case of daily production data reconciliation). Alternatively, they may be triggered on demand (e.g. by operator action) or under certain conditions (e.g. suspected abnormal situations).

Objectives and Structure of This Paper

Recent years have witnessed significant advances in a number of online model-based applications, such as model-predictive control and state estimation. It is not the intention of this paper to provide a detailed review of technical developments in each of these specific areas. Instead, our main focus is on the issues arising from the use of such applications in conjunction with FPMs, and on the implications and demands that this poses on FPMs, for example in the context of their transitioning from offline to online use and keeping them up-to-date in the face of slow changes in the underlying processes.

We also consider how different OMBAs can be synthesized within a coherent framework for online process management, and how to ensure that the most appropriate model is used for each OMBA while still ensuring model consistency across different applications in a sustainable manner that minimizes the effort required for model maintenance. At the same time, it is hoped to identify the needs that are not addressed by current technology.

In general, FPMs potentially contain a high degree of process knowledge and can provide a valuable tool for operations. However, it is important to manage their complexity and to ensure that OMBAs based on them are designed with sustainability considerations taken into account right from the start. In this regard, we would like to give an update on the past experiences, status and what the challenges are for expanding the use and value of this class of models over the coming decade.

The next section provides a brief historical perspective looking at the successes and failures of using FPMs in past online applications; this is followed by a discussion of the closely-related issue of sustainability of OMBAs. We then consider in detail aspects relating to the development and calibration of FPMs. We subsequently discuss different types of OMBAs and their requirements as far as FPMs are concerned, with particular emphasis on the challenges this poses for modeling technology. Experiences from a case study are presented to provide an example of the application of several of the above aspects, and to introduce some additional practical considerations. We conclude by outlining some key challenges for the future.

Historical Perspective

FPMs were first used in support of process control applications in the 1960s-1970s. This coincides with the initial use of digital computers to support process operations. In some cases these models were set up to generate gain information for steady state optimizers based on linear programming (LP) or sequential linear programming (SLP). During this time most were custom models or derived from modeling systems developed by the operating company. Many factors contributed to making the use of FPMs a complicated exercise. The modeling tools were not very powerful, computers were difficult to program and applications required experts to maintain them. In particular, most models used in this context were in "black box" form and their outputs were often subject to significant numerical noise, which prevented the calculation of accurate derivative information by input perturbation. In spite of these difficulties, these initial model-based applications delivered value.

In the 1980s, more progress was made in optimization and equation oriented (EO) approaches to first principles modeling. EO approaches greatly facilitate the development of models of complex unit operations in a manner that is independent of the solution method. Also, since there is no fixed directionality in model solution, the same model can be used for different applications, thereby reducing the overall cost of maintenance. Moreover, due to the availability of partial derivative information and the use of solution methods with superlinear or quadratic convergence rates, the EO approach tends to offer better computational efficiency. This is especially true when the solution is "hot-started" from previous solutions, as is often the case in online applications. Overall the EO approach offers a more natural and numerically suitable interface for optimization problems. Other approaches to using FPMs for optimization support in operations also appeared this decade, based on LP/SLP or even mixed-integer linear programming (MILP, see Nath, 1986). However, despite this progress in process modeling and simulation, tools for model development targeted at online applications lagged behind those available for process design and simulation. This may have been, at least partially, due to the perceived size of the market for online applications in comparison to that for offline modeling, and the fact that modeling for online applications tended to be done by skilled specialists rather than general practitioners.

In the 1990s, a large number of online FPM-based steady state optimization applications were developed and delivered in the petrochemical and refining industry. However, in general they lacked sustainability due to their complexity, need for good support expertise and lack of available tools. Partly because of this, competing approaches developed that could deliver much of the benefits in a more cost-effective, reliable and maintainable manner. Consequentially, the use of FPMs embedded in the optimization application endured a setback, but the FPM model class was still active and delivering value in other architectures for process optimization and control.

During this same time, the use of model predictive control was spreading and achieving success in the process industries. In some cases, the use of the linear empirical dynamic models was inadequate to achieve the performance required. Some turned to FPMs to compute gain information to be used to update the MPC controller model. Others used dynamic FPMs to do step testing to generate data for an identification package for controller development. FPMs were also used to supplement process measurements in control applications in the form of soft (or virtual) sensors, effectively using a computed model output to represent an unknown measurement. The practice of using FPMs for gain updating, identification support and soft sensors has been successful and continues today, but in the past this usage was not supported in a unified or standardized way by the modeling software vendors.

In the mid-1990s, approaches to using nonlinear dynamic FPMs were developed to address some challenging process control problems (Young et al. 2001). For these applications, an EO approach is also required to get the reliability and computation times required for an MPC application. Once again, the dynamic modeling tools to support visual flowsheet modeling that implemented an EO solution approach were not readily available since most dynamic modeling system vendors were targeting other domains and EO modeling approaches were not supported. As a result, custom models or 1980s-style modeling packages were used for the model development, thereby continuing the need for expertise to develop and maintain the models. Also, model development time was a concern, which tended to bias service companies offering OMBAs towards simple modeling approaches. In addition to nonlinear control, a few custom applications for nonlinear state estimation based on FPMs were also published during this time frame (Froisy et al., 1999).

In the 2000s, a number of applications for nonlinear model predictive control were developed in the polymers area, where the extra complexity of using a nonlinear first principles model can be justified due to performance expectations. A limited number of industrial custom state estimation applications based on FPMs, targeted both at performance monitoring and controller support, continued to be implemented (Hedengren et al., 2007).

Other novel uses of FPMs have developed to support process control recently. This includes using online models to compute controller limits (e.g., approach to critical process values) and computing future values of controller feed forward variables. In addition to providing gain information, FPMs can also be used to generate a "reduced complexity" state space or "control relevant" model. This is one way to balance preserving the prediction capability while reducing the complexity to a manageable level for online applications.

Model Complexity and OMBA Sustainability

The experience gathered to date and summarized above indicates that FPMs have the potential to deliver more value than empirical approaches in many OMBAs. As we look back on the issues encountered in steady state RTO applications and initial attempts at nonlinear model predictive control and state estimation, the barriers to increased usage of FPMs can be traced to a few key issues.

First is *application sustainability* i.e. the ability to maintain applications in a cost-effective way such that they continue to produce value over time. There are tradeoffs in sustainability vs. the rigor added by the FPM to the application. For example, MPC applications embedding linear empirical dynamic models have delivered huge benefits while still achieving acceptable sustainability. The identification technology and tools have improved to make this process more reliable. Even in cases where linear models are found to be inadequate to meet ongoing performance expectations, this can often be addressed in a practical and satisfactory manner by manual model adjustments, variable transformations and some model updating. Overall, using more rigorous FPMs would bring too much complexity and little benefit to be competitive in control applications for which linear or almost linear models already offer adequate representations of process behavior. However, this still leaves a wide class of processes whose control can be better addressed by FPMs, especially where highly nonlinear effects are involved, as is the case, for example, with chemical reactors and polymerization processes.

In general, if the underlying model is too complex, it will not be easy to modify or maintain an OMBA, and some or all of the value derived from it may be lost. However, it is worth noting that sustainability is not always a monotonically decreasing function of the degree of modeling rigor. If a model is too simple, there are associated maintenance issues with refitting, retuning and the OMBA value may be lost. This is the case in many applications of linear MPC.

The second major barrier to the increased use of FPMs in OMBAs is the *cost of model development*. Developing an FPM from scratch can be a challenging and time consuming task, and being able to re-use existing models (or part thereof) developed in the context of other offline or online applications can significantly affect the economic feasibility of developing a new OMBA. We discuss this and related issues in more detail below.

In summary, as illustrated schematically in Figure 1, the OMBA designer needs to seek an optimal level of modeling rigor that delivers maximum sustainability and cost effectiveness. The position of this optimum point is a strong function of the technology available, and developments over the next decade must aim to shift this to the right. We should also aim to make this optimal level of model rigor easier to identify with less guesswork.

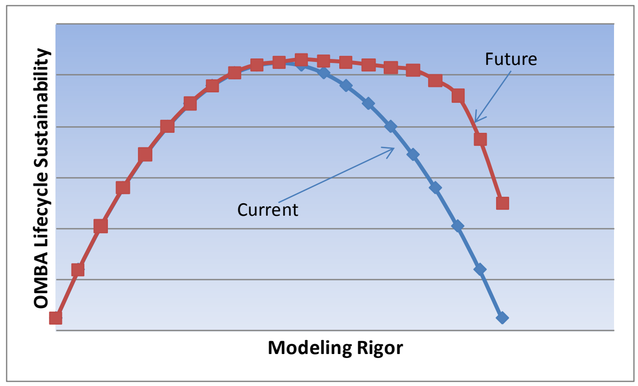

Figure 1: Modeling rigor vs. sustainability trade-offs

FPMs: From Offline to Online Applications

FPMs for Offline Model-Based Applications

Although the use of FPMs for online applications is still rather limited, this is by no means true in the case of offline computer-based computations. FPMs have been used for process and equipment design with significant success for more than 50 years now (see Sargent, 1958 for some early but relatively sophisticated examples). First-principles modeling is now used across the process lifecycle from basic R&D (e.g. catalyst development and design; design of new equipment) to engineering (e.g. design of integrated processes and their control systems), and ultimately to process operations for applications such as operational optimization, troubleshooting and debottlenecking; operator training simulators based on very detailed FPMs are being developed and successfully deployed in industrial applications, and are generally maintained over long periods of time. Significant advances have been achieved in industrial practice in recent years by (a) using high-fidelity models of individual unit operations which can significantly increase the confidence in the designs derived from them, (b) combining these into models of integrated processes, thereby drawing a system envelope that is wide enough to take account of all important interactions and to represent the true process objectives and constraints in a meaningful manner, and (c) coupling these integrated process models with formal optimization techniques which allow a much more complete, effective and efficient exploration of the design space than is possible using manual trial-and-error simulations. A compelling account of such an approach applied to the development of a new process has recently been reported by Martín Rodríguez *et al*. (2010).

Early offline applications used to employ several different models for the same physical object. For example, it was (and still often is) customary to use relatively detailed models for design of complex equipment items (e.g. reactors), but then to rely on rather simple approximate models (e.g. based on user-specified conversions, or complete chemical equilibrium assumptions) for representing these same units within a process flowsheet, with the relationship between these different models often being rather ill-defined (e.g. the detailed reactor model would be used to compute a conversion at a nominal set of operating conditions, with this conversion subsequently becoming a fixed parameter in the simplified model). However, with increasing computer power, there is much less incentive to employ such approximations within any given system envelope (e.g. plant or plant section). There is also a better understanding of the benefit of using accurate predictive models throughout the process lifecycle in terms of improved confidence in design and reduction of the need for extensive adjustments at the time of plant commissioning.

From the point of view of modeling technology, the above considerations have provided the motivation for moving from process simulators to multipurpose process modeling environments in which the process models are independent of the specific applications for which they are being used. The use of such environments offers significant advantages in terms of consistency across different applications, and also simplifies model maintenance in correcting errors or incorporating improved physical understanding.

A key point of the above discussion is that increasingly sophisticated FPMs are used to capture and deploy knowledge and understanding throughout the process lifecycle; and that the required (often non-negligible) investment of time and money is leveraged across multiple offline applications. Similar considerations are also important as far as online applications are concerned. In particular, if a FPM already exists, having been developed and tested for offline usage, then this removes, or at least mitigates, one of the largest obstacles to employing FPMs for online applications.

FPM Development for Model-Based Applications

The development of a FPM that is "fit-for-purpose" for the successful delivery of any model-based application is a non-trivial undertaking which typically involves a number of steps. The availability of a mathematical description that relates the system outputs of interest (including process KPIs) to the system inputs taking account of all the relevant physical phenomena is, of course, an essential first step. The next step is to implement this model in a suitable modeling platform. As has already been mentioned, this is generally easier in EO environments as they take much of the responsibility for the solution of a model away from the user, in contrast to sequential or simultaneous modular (SM) environments where the user has to code both the model equations and their solution. Even so, model implementation usually involves more than a simple transcription of the mathematical equations into a modeling language, as certain considerations that promote numerical stability and robustness need to be taken into account in choosing an appropriate formulation of the model equations.

The next step typically involves obtaining a first solution of the model equations for a nominal set of values of the input variables and model parameters. This is necessary in order to test the correct implementation of the model, and usually involves the solution of a set of nonlinear algebraic equations at the initial point of a dynamic trajectory. SM environments have a distinct advantage in this context as they employ hand-coded solution procedures which can be tailored to the model under consideration, in particular by including code for generating good initial guesses for the model variables. On the other hand, this step is often much more difficult for EO environments which have to rely on general-purpose numerical solvers. The latter may not be able to converge from the (often very poor) initial guesses available, especially if the model is not written in an optimal manner or is not even entirely correct – a common situation during early model development. In the authors' experience, up to 20% of the total project effort may be expended at this step. More generally, in the past the potential lack of robustness of solution from poor initial guesses has been a significant reason for preferring SM approaches to EO ones in practical industrial applications (see, for example, Cox *et al*., 2006). However, recent technological developments may significantly shift this balance. For example, the concept of Model Initialization Procedures (MIPs) recently introduced in gPROMS® allows the model developer to express complex initialization procedures as sequences of model transformations described in a purely declarative manner; these are executed automatically by the modeling tool using a robust proprietary algorithm as and when required.

Calibration of First-Principles Models

Once sufficient tests of the model's correctness are carried out, the next step is usually the validation of the model against experimental data and the estimation of any model parameters (e.g. kinetic constants) which cannot be predicted from first principles or estimated from available correlations. There are well established procedures for this activity (see, for example, Asprey and Macchietto, 2000), and in principle they are applicable to experimental data from a variety of sources, including laboratory rigs, pilot plants or full-scale plants. In the context of FPMs for OMBAs, the obvious choice is often to use plant data for this purpose. However, this may not lead to optimal, or even acceptable, parameter values because plant measurements, even when complemented with the results of additional process testing, do not always contain sufficient information to allow accurate estimation of parameters. Moreover, the observed plant unit behavior is usually the outcome of the interplay of several complex physical phenomena, each with its own parameters, and is influenced by, often unmeasured, interactions with upstream and downstream equipment and the environment. Finally, the values of some model parameters may actually vary during plant operation; for example, reaction kinetic parameters may change due to catalyst deactivation, and heat transfer coefficients due to fouling or coking.

In view of the above, it is preferable to distinguish three different activities:

1. FPM validation and parameter estimation: This is an offline activity that makes use of as much information as possible from well-designed laboratory- or pilot-plant scale experiments to obtain the most accurate characterization of the fundamental physics of the process. Examples of parameters determined during this step include binary interaction parameters (usually estimated from vapor-liquid equilibrium data), mass transfer and heat transfer coefficients (for clean equipment surfaces), and most kinetic parameters (usually estimated from specially-designed pilot plant experiments or, potentially, carefully prepared historical process data). These parameters can be assumed to remain constant during plant operation and most rarely, if ever, change. They often have a strong effect on the model sensitivity information and it would not be appropriate to adjust them with online process data.

2. *FPM-to-plant calibration*: This is also an offline activity and involves a set of (usually one-off) adjustments that may be necessary to match the FPM's predictions to the behavior of the plant within the accuracy required for supporting OMBAs. Such adjustments involve a strong empirical element, and aim to address deficiencies of the original FPM arising from lack of fundamental knowledge, or inability to handle the modeling complexity. Nevertheless, for this exercise to be successful, it is important to have at least some *qualitative* understanding of where and why adjustments are necessary. For example, in the case of polymerization reactors, they may reflect deviations of residence time distribution from that assumed by the FPM; or in the case of thermal cracking furnaces, they may account for variations of radiation-related factors from those predicted by the standard correlations already embedded in the FPM.

3. *FPM online calibration*: This is an online activity involving the continual adjustment of the model to take account of changes to the plant itself taking place during its operation. Examples include catalyst activity which can, for example, increase due to interaction with other components, decrease due to deactivation related to poisons, sintering or other effects, or be affected by issues relating to catalyst feeders or catalyst preparation; heat transfer coefficients which deteriorate due to coking or fouling; efficiencies of compressor and other equipment which may degrade over time; or, more generally, changes in various (largely empirical) biases relating to temperature, pressure, flow or energy (see section on *Online FPM Calibration Approaches* below). From the process control point of view, all of these can be viewed as "unmeasured" disturbance variables (UDVs) which are normally observable with process data, have well understood effects on model sensitivity, and are expected to vary with time during normal process operations. In this respect, they are similar to, for example, unmeasured compositions of impurities in feed streams.

The changes to these disturbance variables usually take place over relatively long periods of time, and may be viewed as slow process dynamics. It is possible to include descriptions of some of these phenomena in the FPM itself, and in fact to estimate the corresponding parameters (e.g. catalyst deactivation kinetics) using carefully-designed laboratory experiments during the offline calibration step. However, even in these cases, online calibration may still need to be applied to account for the inaccuracies that are almost always present in such models, and also for the fact that the actual state of the plant at the start of the application of the OMBAs is usually not known.

Online FPM Calibration Approaches

For process control engineers, online calibration can be viewed as a feedback process, to bring the model predictions on target to the process measurements, the same way feedback is used in a control application to bring measured values to requested target values or within ranges.

Ideally, the FPM will already include appropriate parameter/disturbance variables which can be adjusted during online calibration. If this is not the case and the models cannot easily be modified, the options for calibration are limited. The simplest approach is to compute additive or multiplicative biases after the model solves to match the process outputs. Simple additive offsets have no effect on gains; consequently, this approach will not properly translate information in the measurements that indicate changes to process sensitivities. Use of multiplicative biases can be appropriate in some cases but these will change model sensitivity between the inputs and the biased outputs. However, appropriate use of external multiplicative biases requires good judgment – which implies there could be some useful tool development in this area.

In addition to the above basic output biasing strategies, there is a progression of more sophisticated online model calibration approaches that have been used in practice for OMBAs.

A dynamic generalization of the "square" parameter estimation problem used in the early RTO applications can be used. This approach requires the application of engineering knowledge to pair a measurement with an individual disturbance variable. There is a dynamic relationship between these pairs, with a tuning factor to control the speed of matching the current process measurement (a filtering capability). This can be viewed as a type of nonlinear observer and is a nonlinear generalization of the feedback strategy used in a number of linear MPC technologies. This square approach has an advantage of simplicity to understand, and allows updating both input/output disturbance variables. However, it can be more difficult to tune and lacks disturbance variable bounding capability, since it is not optimization based. It also only focuses on the current measurement and does not consider a history of measured values for the model online calibration problem.

An improvement to this approach for online FPM calibration is the formulation of a state/disturbance estimation problem (Hedengren et al., 2007). The two currently predominant approaches to state estimation are those based on Extended Kalman Filtering (EKF) and its variants, and Moving Horizon Estimation (MHE). A good comparison of these is discussed in Heseltine and Rawlings (2005). Recent developments in MHE and its application to a polymerization process have been described by Zavala and Biegler (2009b). MHE has the advantage that it can be applied directly to models based on (partial) differential-algebraic equations (DAEs, PDAEs, Index > 1 cases) without requiring elimination of algebraic equations, thereby preserving sparsity – an important factor for large-scale equation-based FPMs. Being optimization-based, MHE also naturally handles constraints without requiring any *a posteriori* clipping, unlike some EKF methods.

As described in the next section, an effective and efficient online FPM calibration procedure is at the heart of most OMBAs.

An Integrated View of OMBAs

Although OMBAs are often thought of as being synonymous to model-based control, in fact a wide variety of such applications can be constructed and deployed, leading to significant practical benefits. Despite their diversity, almost all OMBAs have two common pre-requisites: an up-to-date FPM reflecting the current plant condition; and a reliable estimate of the current state of the plant. As explained in the previous section, both of these can be produced via the online calibration of the FPM (marked as **[1]** in Figure 2[1]) by processing raw plant data.

In the remainder of this section, we review some of the most common OMBAs, with particular emphasis on aspects relating to FPMs. We then consider the demands placed on modeling technology by integrated OMBAs.

[1] Throughout this section, the notation ***[n]*** refers to a correspondingly marked block in Figure 2.

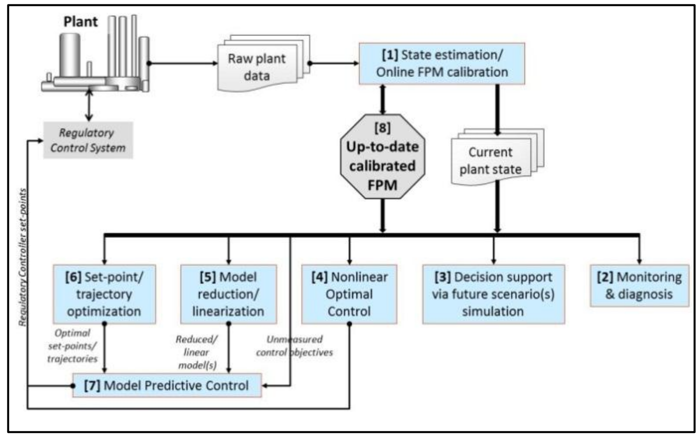

Figure 2: Integrated framework for OMBAs

Open-Loop Advisory OMBAs

Some OMBAs are of an open-loop nature, generating information and, potentially, advice in support of plant operations. One example is monitoring the current state of the plant **[2],** in particular in terms of KPIs that relate to plant throughput, product quality and safety and operability considerations, the values of which can in principle be obtained directly from the output of the state estimator provided they are included in the set of FPM variables. Based on the KPI values, it may also be both desirable and possible to carry out diagnosis of potential faults, e.g. to detect and locate possible blockages or leakages within a pipe network, based on estimates of various pressures. The corresponding analysis may be either rule-based or model-based, in the latter case making use of the FPM itself (see below). Differences in statistical and FPM-based approaches to this fault detection problem are also discussed by Yoon and MacGregor (2000).

Plant monitoring and diagnosis applications are primarily concerned with the *current* state of the plant. A different class of open-loop OMBAs **[3]** are those concerned with *future* plant behavior as it will evolve either naturally from its current state, or following additional future operator actions and/or external disturbances. Such applications are often concerned with the identification of potential risks and threats that may affect the plant at some future point in time given its current state (which, in itself, may be perfectly safe) in the event of one or more hypothetical scenarios taking place. In the simplest case, this involves dynamic simulations of the plant behavior under fully specified scenarios starting from the current state. More sophisticated applications of this type may involve the solution of optimization problems, e.g. aiming to identify the *worst-case* form of a particular scenario, or the *maximum margin* of safety of the process should a scenario materialize. For example, in the case of a flare network designed to protect plant equipment against over-pressurization, an important question at any point in time is the maximum capacity of the network to accept additional relief flows from one or more, currently inactive, sources. In any case, the main output of these OMBAs is information that can support the plant operators in their decisions.

Closed-Loop Control OMBAs

Closed-loop OMBAs interact directly with the plant, or more commonly its regulatory control layer. The most common class of such applications is that of model-predictive control (MPC, **[7]**) which has now reached a significant level of development and industrial deployment (cf. the early review by Garcia et al. 1989 and the more recent one by Lee, 2011).

Most MPC applications in industrial practice to date (Qin and Badgwell, 2003) make use of linear models of various types, the main reason being that such models bring efficiency, reliability and robustness to the MPC's optimization computations. Currently, these models are usually derived from plant operational or test data; however, they can also be obtained directly from FPMs via linearization to produce linear state-space or transfer function models. This approach **[5]** has the advantage of minimizing the cost and disruption to plant operations associated with MPC model identification, even in applications requiring multiple linear models, each constructed around a different operating point (including non-steady-state ones). Of course, given the requirement for an appropriate FPM, in practice this is attractive mainly in situations where the FPM has already been constructed in the context of other offline and/or online applications.

Nonlinear MPC (Allgöwer et al., 2004), based on either empirical models (e.g. artificial neural networks) or FPMs is also finding increasing application in industrial practice (Qin and Badgwell, 2003). Despite significant advances in the computational efficiency of the underlying optimization algorithms (see, for example, Zavala and Biegler, 2009a and Würth et al., 2009), the FPMs used in this context remain rather simple in comparison with those that are now routinely used for offline applications, and usually involve extensive use of simplifying physical assumptions. If a more detailed FPM is available, then it is often possible to use it to generate "pseudo-experimental" data, from which the simple FPM's parameters are fitted offline to ensure that the process KPIs are predicted to an appropriate degree of accuracy within restricted regions of interest. An alternative is to derive a simpler model from the detailed FPM by applying model reduction techniques (Marquardt, 2001, Backx et al., 2006). This approach is subject to a number of pitfalls, not least the fact that it often leads to reduced order models which, albeit smaller in size, are not significantly or sufficiently more efficient than the original FPM (see, for example, the results reported by Astrid, 2004 and van den Berg, 2005). Recent attempts to resolve this issue include the use of proper orthogonal decomposition in conjunction with (semi-) empirical ("grey box" or "black box") modeling (see, for example, Sun and Hahn, 2005, Romijn et al., 2008, Wattamwar et al., 2010). However, it is too early to judge whether these approaches will have a significant impact on industrial practice in the process industries.

An alternative approach to nonlinear MPC which aims to address the issues of computational efficiency and robustness is based on the use of explicit control laws derived from linear or nonlinear models using multi-parametric programming. An overview and perspective of this fast developing area, including its relation to first-principles modeling and model reduction, has been given by Pistikopoulos (2009).

In general, the MPC algorithms that have been considered so far in this section solve a constrained optimization problem that aims to maintain the process close to a given set-point or within setpoint ranges (for continuous processes meant to operate at steady state) or a dynamic trajectory (for batch/semi-batch processes, or continuous ones meant to exhibit a given transient behavior). These set-points or trajectories are themselves derived from the solution of optimization problems **[6]** with a wider (usually economics-based) view of the process objectives. The use of FPMs for steady-state optimizations of this kind, known as Real-Time Optimization (RTO), is already well established in industrial practice. More recent is the use of FPMs for dynamic optimizations (some use the term D-RTO). For example, van Brempt et al. (2003a, 2003b) and Dünnebier et al. (2005) have applied it to grade transitions in two industrial polymerization processes, and Mesbah et al. (2010) to batch crystallization processes. The recent work by Huang (2010) also addresses NMPC and dynamic optimization

The most sophisticated use of a dynamic FPM is using it as a part of a real-time dynamic optimization or nonlinear optimal control **[4]** application which merges model predictive control and dynamic economic optimization within a single-layer architecture. The nonlinear MPC approach described in Young et al. (2001) accommodates this strategy in the problem formulation, but to date most applications deployed in industrial practice use production maximization as the economic objective function, although wider objectives have been considered in the academic literature (see for example, Rolandi and Romagnoli, 2005, and Engel, 2007). In principle, this approach combines the best features of RTO (continuous economic optimization once control objectives are met) with model predictive control (keeping the process in the feasible operating ranges) simultaneously in one application. The competing economic and control objectives are best resolved with a multi-level optimization problem formulation, resolving control (feasibility) first, then optimality tracking (dynamic economic optimization).

Overall, increasingly sophisticated approaches to real-time dynamic optimization are bringing significant benefits to industrial practice, but advances in this area will progress at a slow rate until issues related to FPM development cost and OMBA sustainability are addressed. In the near term, strategies that leverage a well-calibrated

FPM with the closed loop identification tools that are implemented in the current generation of MPC products (as well as gain generation for updating MPC) will likely become more common.

Model Structures & Modeling Tools for OMBAs

The FPM **[8]** of the plant provides the foundation for all OMBAs, being used by them either directly or following some pre-defined simplification operation (e.g. linearization). In this sense, the FPM serves as a "master" or "reference" model (see Backx et al., 2006 and references therein) which captures all relevant knowledge on the plant's behavior and is kept up-to-date at all times. This potentially brings many advantages in terms of consistency across different applications, and reduced cost of maintenance.

For simple applications (e.g. those concerned with a single major unit operation and its associated peripheral equipment), the master model has a fixed structure. Thus, the underlying assumption is that such a model will be carefully constructed and tested manually using an appropriate modeling tool; and that, thereafter, it will remain fixed over a significant length of time (with the exception of any parametric changes involved in its online calibration).

However, this simple and rather attractive paradigm is not necessarily optimal or even feasible for more complex applications. Consider, for example, a flare network providing over-pressurization protection to an oil and gas production asset or to a refinery. The network would typically be connected to a large number of process vessels, each one potentially acting as a source of material to be flared; and it would consist of many hundreds of pipe segments, valves, vessels and other devices. However, at most points in time, only a small subset of the network is likely to be active in the sense of having material flowing in it. Therefore, basing any online application on a dynamic model of the complete network would not only introduce unnecessary computational cost, but may also significantly reduce robustness by having to deal with zero-flow situations and other related issues. On the other hand, it is not possible to predict and construct *a priori* all possible combination(s) of active sources that will occur during plant operation. Instead, what is required for the OMBA management system is a *horizontal model configuration* capability allowing it to determine the relevant system envelope in real time (e.g. based on information from source sensors installed on the plant), thereby configuring and utilizing the plant sub-model that is most appropriate at any particular point in time.

A further complication arises from the fact that it is not always possible to fix *a priori* an appropriate level of modeling detail to be used by a particular OMBA. As an example, consider an OMBA monitoring the safety of the flare network mentioned above against a number of risks. The primary purpose of the flare network is to allow process equipment to depressurize safely by releasing sufficient amounts of material into it; therefore, the minimum acceptable level of modeling detail for the various pipe components is one that determines the dynamic relationships between flow and pressure, thereby providing accurate predictions of back-pressures at the various relief valves. However, in some cases, another potentially serious risk is that of low-temperature metal embrittlement arising from the Joule-Thompson effect inherent in rapid depressurizations from high pressures. In this context, the OMBA would need to employ more detailed pipe models, typically taking account of spatially-distributed thermal dynamics. On the other hand, if *all* pipe components in the (active subset of the) network were to be modeled to this degree of detail, the resulting model would be computationally unsuitable for online application. Instead, what is required is (a) a *multi-level model* of each plant component (e.g. pipe segment) comprising a set of consistent models of varying degrees of detail, coupled with (b) a *vertical model configuration* capability which allows the OMBA management system to decide the appropriate level of detail for each component model and to configure the FPM accordingly based on plant sensor information and/or model predictions. For example, by default the OMBA would use relatively simple pipe models producing rough temperature predictions; however, if the temperatures in one or more pipe segments are predicted to be potentially too low, the OMBA would automatically refine the models for these specific segments to more detailed forms while still using the simpler models for all others.

For most OMBAs making use of increasingly complex FPMs, the efficient execution of the underlying algorithms is a key issue. As indicated by the brief review presented earlier in this paper, current efforts to address this efficiency issue focus primarily on improvements in the numerical methods. It is, however, also important to take advantage of the opportunities afforded by advances in computer hardware and middleware (e.g. in a distributed computing context). At the simplest level, this takes the form of the parallelization of the numerical algorithms or some of their most computationally expensive components (see Keeping and Pantelides, 1998, and Hartwich et al., 2010 for an example relating to dynamic optimization computations). More complex OMBAs, such as the safety-monitoring applications described earlier in this paper, often involve multiple computations which can naturally be carried out in parallel. The software architectures for such OMBAs need to launch and manage a number of parallel computations, ultimately collecting and synthesizing their results into a coherent set of messages to the operators.

In conclusion, the support of sophisticated OMBAs places new and complex demands on process modeling technology and their architectures. Whether these needs can be addressed by superimposing additional layer(s) of software on top of existing modeling tools, or whether completely new software architectures are necessary is, at present, an open question.

A Case Study – Polymer Process Control.

In order to properly assess future needs, it is useful to explore a case study where FPMs have been applied in an online application. There are lessons learned from this practical experience, and real benefits and challenges will emerge from this review.

In the 1990s, a number of attempts were made to use linear model predictive control for some challenging polymer process control problems. The basic objective in most problems was controlling polymer properties both while running on grade and during grade-to-grade transitions. Most of these approaches were unsuccessful due to the inadequacy of representing the process by linear dynamic models, even with model updating, and the general consensus among practitioners is that a nonlinear model is required to achieve the performance expectations. More specifically, approaches that use a linearized nonlinear model and update linear MPC can work well for on grade control, but may not give the best performance for transition control, because the control problem doesn't reflect the *process sensitivity change* over the prediction horizon. As a result, the moves to control during transition will inevitably be inconsistent until the proper CV/MV sensitivity is finally seen by the controller. A good review of the evolution of this approach is given in Young et al. (2001).

Polymer process control is an area where the extra rigor (and even complexity) of an FPM is justified. The process model equations are generally well known and can provide good predictions. An application of this approach, described by Allsford et al. (2008), can be more accurately described as nonlinear optimal control, since economic objectives can be designed into the controller to achieve simultaneous dynamic optimization and control.

The FPM form discussed here for this case study is a flowsheet based nonlinear dynamic differential/algebraic equation (DAE) model. The mass and energy balances are rigorous. Phase equilibrium calculations are rigorous in some cases and in others some empirical separations correlations are used. Enthalpies, k-values and densities are normally computed by an appropriate equation of state. The reactors are modeled rigorously and the polymer processes discussed in this case study have reactions/catalysts based on either free-radical or Ziegler-Natta kinetics. Flow/pressure relationships are not normally modeled since the dynamics of these relationships are very fast compared to the control cycle time. All regulatory controls, including PID and other calculations that impact process sensitivity prediction and dynamics are modeled. The model is almost always supplemented by some custom empirical modeling (in a modeling language) to compute special properties or other outputs that cannot be predicted by fundamental means (e.g., polymer hardness, fouling curves established in tables for operators, etc).

For these types of polymer control applications, examples of manipulated variables (MVs) are chain transfer agent flow rate, comonomer flow rate, catalyst flow rate, co-catalyst flow rate, reactor levels, and reactor temperature. Examples of controlled variables (CVs) include polymer melt index or viscosity, polymer density, polymer production rate, comonomer % incorporation, reactor pressure, approach to dew point, % solubles.

Offline Model Calibration

The offline model calibration step for the polymer process FPMs is primarily done as a steady-state multiple data set parameter estimation problem. For a given process type, a base set of kinetic parameters from the literature is used as a starting point. Process step testing is not a real option in most polymer processing units; hence model calibration relies heavily on historical plant data collected for the process operation for each of the polymer grades during different operating scenarios. Care is taken to cover all catalyst types, grades and get a wide variety of production rates and process variation for each product. This is the application engineer's way of trying to ensure that the right amount of excitation exists in the process data. Kinetic parameters for each catalyst are estimated by simultaneously processing all relevant data sets. Multiple catalysts can be configured in the model, and transitions between catalyst types automatically and naturally change to use the appropriate kinetics in the FPM as the different catalyst compositions are accounted for. Other empirical parameters can be estimated for separations correlations, or other empirical property calculations.

It is important for offline model calibration using process data to take account of the fact that the latter may contain the effects of unmeasured disturbances. In some cases, these can be accounted for in the parameter estimation problem via a "square" disturbance estimation strategy. A simple example is the estimation of a heat transfer coefficient (HTC) in a heat exchanger. Because of fouling occurring during process operation, the value of this coefficient will generally be different for each data set. The HTC values in the various data sets can be estimated from corresponding exchanger outlet temperature measurements, simultaneously with all other (time-invariant) model parameters. For more complicated online estimation problem formulations such as state estimation, this strategy needs to be considered and generalized, which is a topic for more research.

Model Validation

Validation of FPMs often needs to be more extensive for control applications than for other OMBAs. First, validation of the steady state predictions are done for test data sets not included in the parameter estimation data sets. The process gains between all measured inputs and the controlled variables are computed for the model and mapped out over all of the polymer products (grades).

These gains are reviewed with process engineers for validation. If the gains are not correct, more data analysis is required and additional data may be collected. In some cases some simple step testing may have to be employed to validate the gains. In other instances, data sets have been "created" to provide the missing information on the proper sensitivity. In other cases, lab tests have been available for product molecular property analysis, but these are often expensive and difficult to obtain. This process of estimation/validation is iterative until the model passes the established prediction requirements, including proper sensitivity. Admittedly, in practice this process has not always included the best available technology, such as use of parameter confidence limits and parameter observability information.

Since the model is based on first principles, its dynamic behavior is mostly governed by vessel volumes, certain kinetic parameters and regulatory control tuning factors. There has been no dynamic parameter estimation problems used for the applications in this case study, only comparison of the dynamic predictions of the model with the process data (historical data replay task). Some filter factors or other dynamic tuning factors can be manually adjusted after viewing the dynamic prediction results to get better agreement of the model predictions with the process. This aspect of the model parameter identification can be improved by a better dynamic estimation problem formulation and is an area deserving further analysis.

Online Model Calibration

The polymer process models used in the applications for this case study incorporate feedback through online disturbance estimation. Most disturbance estimation was performed using the dynamic "square" nonlinear updating (observer-like) approach; some disturbance estimation was performed using the MHE approach. The disturbance estimation problem has feedback tuning parameters that need to be established. The application engineers use their experience to protect against update strategies that incorrectly change model sensitivities due to bad measurements or improper model structure. The lack of recognition of this issue can result in poor predictions, and in fact this was the root cause of performance problems in a number of RTO applications in the 1990s. This aspect can be tested with real process data before controller commissioning begins. This is an area of development where better support tools and technical approaches could remove the requirement for the experience factor.

Reliability, Robustness and Computation Time

One concern mentioned for the deployment of FPMs in control applications is their solution reliability (convergence of the nonlinear model equations and/or optimization problem). Linear and quadratic programming methods are now almost 100% reliable to solve the MPC control problems. Nonlinear FPMs embedded in the controller require a nonlinear programming (NLP) solver. Control applications that use the FPM in an external way to provide gains also have the challenge of converging in the face of large process data changes. Part of the application development process is to do extensive robustness testing in a controller simulation environment. In addition to testing normal control scenarios such as configured CV prioritization and constraint handling, the nonlinear models must be tested for solution reliability in the face of large input changes, very ill-conditioned constrained scenarios, among others. Polymer process models are notoriously difficult to solve due to extreme nonlinearity and bad scaling (kinetic constants for some processes around 10^{+35}, moment variables around 10^{-10}). Over time, good learning experiences with these applications has resulted in the development of many approaches to addressing this concern. The following factors combine to provide a highly reliable FPM application for polymer processes:

- Rigorous, automated data validation and preparation
- Reliable nonlinear programming solver
- Use of proven online calibration strategies/designs
- Good equation-oriented model formulation, including proper scaling
- Control problem formulation that accepts feasible, suboptimal NLP solutions
- Model variable validity limits used during the solution (keeps model away from problems due to nonconvex regions).

With these factors addressed, the number of failures per year can be reduced to a very small number as a percentage of the total runs. Since the application is starting from a previous solution, the NLP optimal solution is usually obtained each cycle. It is rare to see the NLP feasibility mode active, where it reaches maximum iterations and obtains a feasible suboptimal improved solution.

For the EO approach to solving the dynamic control/optimization problem, the simultaneous method in Renfro et al (1987) has been used successfully for many years. This allows a dynamic polymer process flowsheet model to be solved and simultaneously optimized over a time horizon in the required cycle times for the level of FPM rigor described at the beginning of this section. Again, the NLP feasibility mode provides protection against excessive cycle times if the iteration limit is set to a low value. However, to date the largest NMPC application has been around 10MVs and 35 CVs, with 10-20 future prediction intervals. For larger NMPC applications, other formulations and options may be needed to solve reliably within the required cycle times. Control engineers need options during the application design phase to reduce model complexity, improve solution times with special algorithms, or have more flexible control problem formulation options (e.g. suboptimal but fast) to continue increasing the scope and size of these types of control applications to levels similar to those for linear MPC and still meet the computational constraints dictated by the required cycle time.

For an application running once every 1-3 minutes (cycle time), 24/7 about 3-4 failures per month on average occur, which is very good on a percentage basis. In most of these cases, the controller will pick up with a successful solution on the next cycle. These failures are usually related to some unforeseen missing data validation logic, where an invalid or inconsistent input gets into the model. Since OMBAs are running frequently, it is easy to compare information from two subsequent cycles to deduce problem root causes. Occasionally a failure occurs whose resolution is not obvious and requires a detailed model or solver analysis. This is a challenge issue for the future - to eliminate these failures or at least to devise a systematic work process for a process engineer to address this class of application failure issues.

FPM-Related Commissioning Issues

Poor performance of a nonlinear controller may be related to model predictions, and the root cause needs to be identified. For large sets of CV/MVs, it may be difficult to get all of the model gains correct and this fact is sometimes found out during commissioning. If this polymer process is different than other previous applications, the gains may not be known. Moreover, it is not always easy to correct the gains of an FPM since they are not available as explicit expressions that can be modified. One example of a control application where this was a difficulty was a High Impact Polystyrene (HIPS) process involving multiple CSTRs and multiple plug flow reactors. Because of weakly observable parameters related to back mixing in the plug flow reactors, some of the gains predicted by the FPM were not correct. Some identification testing was subsequently required to enable the gains to be estimated. Even then, it was still a difficult exercise to make use of the values of these gains in the context of parameter estimation or any other adjustment applied to the FPM. In contrast, empirical nonlinear models with explicit expressions for gains that can be constrained during the (offline) parameter estimation problem would have been easier to adjust. In linear MPC, control engineers regularly manually change gains in the linear step response models to correct them after the identification is done. This is an extremely useful and important feature of the modeling approach for control applications which is not readily available in FPMs. Consequently, in many cases a simple empirical model has been used to replace a FPM CV/MV pair in order to address this particular issue – this is one form of hybrid first-principles/empirical (or “gray box”) modeling, an area which may deserve more research attention in this context. Alternatively, it may be possible for gain and other engineering knowledge to be incorporated directly in offline parameter estimation formulations. In addition, utilities for analyzing observability and uncertainty of model parameters, comparable to those already available in state-of the-art offline modeling tools, would be highly desirable.

Model Maintenance

FPMs actually eliminate a number of maintenance concerns associated with linear MPC applications. Since regulatory control models are embedded in the process models, retuning of the PID controllers in the process is directly accounted for in the model. The tuning parameters and other control calculations are included in the process model. This eliminates the reduction in application performance and/or the need for re-identification that is required in most linear MPC applications under this scenario. Process equipment and configuration changes are handled differently in FPM-based control applications. Adding or removing process equipment can reduce linear MPC performance or entirely invalidate many submodels. For modeling tools with flowsheeting capability, the model can be updated and the control application can be put back into service in a reasonable amount of time.

Also, when new polymer grades are added for an existing catalyst, in many cases the controller can meet performance requirements with no additional work. This is due to the good extrapolation/interpolation capability for FPMs. On the other hand, if a new catalyst is introduced, then an additional offline calibration step may be needed to update the model. This task is not as easy as it needs to be for a process control engineer as the current linear identification process. Better nonlinear identification tools will help in this part of FPM maintenance.

In the past, because of a lack of visual modeling tools for dynamic FPMs (of the quality of current steady-state simulation and design software), maintenance of these models could be a challenge for the novice. The architecture and functionality of first principles based modeling systems targeted at operations needs to be improved to the point where process engineers can maintain FPM based applications in a manner similar to that in standard process simulation packages. Also of critical importance is the ability to deploy a model with an equation oriented solution capability to meet the needs of model calibration, control and optimization calculations. Sequential modular (SM) architectures are not well suited for these tasks but they can be a great help with model initialization in a combined EO/SM architecture as mentioned previously.

In summary, FPMs can be used in process control applications to achieve benefits and have reasonable maintenance costs. The project execution work process and tools need to be improved so that the applications can be done in a time frame that is competitive with simpler approaches that will achieve lower but acceptable benefits. The maintenance of these applications needs to be improved as well to achieve a high sustainability in a more cost effective way, using engineers with reasonable control knowledge, but not necessarily modeling expertise.

Conclusions and Future Challenges

Online applications based on FPMs have been supporting process operations and control for more than 40 years. There is still much untapped value in using such models more extensively. There are also a number of barriers that can be addressed through improvements in technology, software and general work flows and procedures.

The expanded online use of FPMs in process operations faces a number of challenges. From the commercial viewpoint, the customer's end objectives for a particular online application can usually be met in multiple ways. Companies that deliver such applications need to do so in a timely fashion in a profitable way within their own required margins. FPMs have the potential to deliver more value, but can be more complex – hence riskier – to develop and deploy, which reduces its attractiveness in comparison with more empirical modeling approaches. Moreover, beyond the initial development and commissioning of an OMBA, it is not clear how to quantify the sustainability issues, in particular regarding the tradeoffs between increased maintenance requirements against the increased benefits of brought by FPMs. This is currently done by experience and development of a knowledge base but could perhaps be done more systematically.

On the technical side, the paper has discussed the advantages of the EO process modeling approach in the context of OMBAs. For the wider application of dynamic FPMs in online application, it is essential for the EO modeling tools used for their construction to become as easy and straightforward to use as existing steady-state flowsheet simulators. Fortunately, current state-of-the-art modeling technology is already moving in this direction, largely driven by the demands of offline applications and their requirement for detailed FPMs within increasingly wider process envelopes. A similar trend is seen in offline model calibration, e.g. in the use of multiple data sets from dynamic experiments, for parameter estimation, the assessment of the accuracy of model parameters derived from such data, and the model-based design of experiments aimed at producing optimal information content. Such techniques are now routinely applied to data from specially designed lab- and pilot-scale experiments. However, the situation is less straightforward when model calibration needs to rely mostly or entirely on historical plant data. In particular, easier tools and methods need to be made available to analyze the "richness" of such data, and tools for nonlinear identification (essentially nonlinear parameter estimation from dynamic plant data) need to be made as powerful and reliable as those currently available for linear model identification. Moreover, as illustrated by the Case Study, it is often necessary to take account of additional "external" process knowledge (e.g. constraints on MV/CV gains and other information). Ways of incorporating this information within the FPM calibration procedure need to be developed.

Irrespective of the sophistication of the modeling tool, the cost of developing a new FPM from scratch may be considerable. This can partly be addressed by leveraging modeling investment from across the process lifecycle. However, it has to be recognized that models used for operator training simulators, process design, and plant control all have different needs and priorities, and it is a challenge for the modeling software to help reconcile these conflicting requirements to reduce model maintenance issues.

Another significant challenge for modeling tools is to support the type of mathematical calculations required by online applications. Some of these (e.g. dynamic simulation and optimization) are already supported by the more advanced process modeling tools, but the challenge is to achieve the degree of robustness and efficiency required for online use. As we have seen, progress in this direction will need both further algorithmic improvements and exploitation of new computing paradigms. Other mathematical calculations, such as algorithms for state estimation/online FPM calibration, have not yet been incorporated within commercially available modeling technology. Such a development would greatly facilitate the migration of FPMs from offline to online application.

We have also seen that the paradigm of a single, fixed-structure FPM may not be able to provide the flexibility required by more complex OMBAs, and that a degree of "on-the-fly" horizontal and vertical model configuration may be essential. The challenge is to develop software architectures that can support such complex operations in a general, reliable and efficient manner.

The authors are confident that addressing the issues highlighted above will result in a wider use of currently existing OMBAs, and will greatly facilitate the development and accelerate the deployment of a number of new and powerful OMBAs that will deliver additional value to process operations in the future.

Acknowledgements

The authors would like to acknowledge the following for their review and contributions to this topic: from Honeywell, Ravi Nath, Jeff Kelly, Grant Stephenson, and from PSE, Ying-Sheng Cheng, James Marriott and Pablo Rolandi.

References

Allgöwer F., Findeisen, R., Nagy, Z.K. (2004). Nonlinear Model Predictive Control: From Theory to Application. *J. Chin. Inst. Chem. Engrs.*, *35*, 299-315.

Allsford, K.V, Renfro, J.G., Goodman, J. (2008). Experiences in the Building of Dynamic Flowsheet Models for Embedded Nonlinear Control of Polymer Processes. *AIChE National Meeting*, November 2008.

Asprey, S. P., Macchietto, S. (2000). Statistical tools for optimal dynamic model building. *Comput. Chem. Eng*. *24*, 1261–1267.

P. Astrid, (2004). *Reduction of Process Simulation Models: a Proper Orthogonal Decomposition Approach*, Ph.D. Thesis, Technical University of Eindhoven, Netherlands.

Backx, T., Bosgra, O., Marquardt, W. (2006). Industrial Challenges in Modelling of Processes or Model Reduction. In *Proc. Intl. Symp. Adv. Contr. Chem. Proc. (ADCHEM 2006)*, International Federation of Automatic Control, 143-151.

Cox, R.K., Smith J.F., Dimitratos, Y. (2006). Can Simulation Technology Enable a Paradigm Shift in Process Control? Modeling for the Rest of us. *Comput. Chem. Eng*., *30*, 1542-1552.

Dünnebier, G., van Hessem, D., Kadam, J., Klatt, K.-U., Schlegel, M. (2005). Optimization and Control of Polymerization Processes. *Chem. Eng. Tech*. *28*, 575–580.

Engell, S. (2007). Feedback Control for Optimal Process Operations. *J. Proc. Cont*., *17*, 203-219.

Froisy, B., Hart C., Lingard, S., Papastratos, S., Dixon, D., McGrath, P. (1999). Industrial Application of On-line First Principle Dynamic Models Using State Estimation. AIChE Annual Meeting, Dallas, Texas, November 1999.

Garcia, C.E., Prett, D.M., Morari, M. (1989). Model Predictive Control: Theory and Practice – a Survey. *Automatica, 25*, 335-348.

Hartwich, A., Stockmann, K., Terboven, C., Feuerriegel, S., Marquardt, W. (2010), Parallel Sensitivity Analysis for Efficient Large-Scale Dynamic Optimization. *Optim. Eng*., DOI 10.1007/s11081-010-9104-4.

Haseltine, E. L., Rawlings, J.B. (2005). Critical Evaluation of Extended Kalman Filtering and Moving Horizon Estimation. *Ind. Eng. Chem. Res*., *44*, 2451-2460.

Hedengren J.D., Allsford, K.V., Ramlal J. (2007). Moving Horizon Estimation and Control for an Industrial Gas Phase Polymerization Reactor, In *Proc. of American Control Conference* (ACC), New York, NY.

Huang, R. (2010). *Nonlinear Model Predictive Control and Dynamic Real Time Optimization for Large Scale Processes*, PhD Thesis, Carnegie Mellon University, Pittsburgh, PA.

Keeping B.R., Pantelides C.C. (1998). A Distributed Memory Parallel Algorithm for the Efficient Computation of Sensitivities of Differential-Algebraic Systems. *Math Comput Simul 44*, 545–558

Lee, J. (2011). Model Predictive Control: Review of the Three Decades of Development. *Intl. J. Contr. Autom. Syst*., *9*, 415-424.

Martin Rodriguez, H., Cano A., Matzopoulos, M. (2010). Improve Engineering via Whole-Plant Design Optimization. *Hydrocarbon Processing* (December), 44-51.

Marquardt, W. (2001). Nonlinear model reduction for optimization based control of transient chemical processes. In *Proc. Chemical Process Control Conference VI*, Tucson, Arizona, 12–42.

Mesbah, A., Landlust, J.L., Versteeg, C., Huesman, A.E.M., Kramer, H.J.M., Ludlage, J.H.A., van den Hof, P.M.J., (2010). Model-based Optimal Control of Industrial Batch Crystallizers. In *Proc. 20th European Symposium on Computer Aided Process Engineering – ESCAPE20*, Elsevier B.V.

Nath, R., Libby, D.J., Duhon, H.J., (1986). Joint Optimization of Process Units and Utility System, *Chem. Eng, Prog*., *82*, 31-38.

Pistikopoulos, E.N. (2009). Perspectives in Multiparametric Programming and Explicit Model Predictive Control. *AIChE J., 55*, 1918-1925.

Qin, S.J., Badgwell, T.A. (2003). A Survey of Industrial Model Predictive Control Technology. *Control Engineering Practice*, *11*, 733–764.

Renfro, J.G, Morshedi A.M., Asbjornsen O.A. (1987). Simultaneous Optimization and Solution of Systems Described by Differential/Algebraic Equations, *Comput. Chem. Eng.*, *11*, 503-517.

Rolandi, P.A., Romagnoli, J.A. (2005). A Framework for Online Full Optimizing Control of Chemical Processes. In *Proc. ESCAPE 15*, Elsevier, pp. 1315–1320

Romijn, R., Ozkan, L., Weiland, S., Marquardt W., Ludlage, J. (2008). A Grey-Box Modeling Approach for the Reduction of Nonlinear Systems, *J. Proc. Contr., 18*, 906–914.

Sargent, R.W.H. (1958). Applications of an Electronic Digital Computer in the Design of Low Temperature Plant, *Trans. Instn. Chem. Engrs*., *36*, 201-214.

Sun, C., Hahn, J. (2005). Reduction of Stable Differential-Algebraic Equation Systems via Projections and System Identification, *J. Proc. Contr., 15*, 639–650.

Van Brempt, W., Van Overschee, P., Backx, T., Moen, Ø., Kiparissides, C., Chatzidoukas, C. (2003a). Plantwide Economical Dynamic Optimization: Application on a Borealis BORSTAR Process Model. In *Proc. ADCHEM 2003*, Hong Kong.

Van Brempt, W., Van Overschee, P., Backx, T., Ludlage, J., Hayot, P., Ootsvogels, L., Rahman, S. (2003b). Grade-Change Control using INCA Model Predictive Controller: Application on a DOW Polystyrene Process Model. In *Proc. Am. Control Conf. 2003*, 5411-5416.

Van den Berg, J. (2005). *Model Reduction for Dynamic Real-Time Optimization of Chemical Processes*, PhD thesis, Technical University of Delft, Netherlands.

Wattamwar, S.K., Weiland, S., Backx, T. (2010). Identification of Low-Order Parameter-Varying Models for Large-Scale Systems. *J. Proc. Contr., 20*, 158-172.

Würth, L., Hannemann, R., W. Marquardt, W. (2009). Neighboring-extremal Updates for Nonlinear Model-Predictive Control and Dynamic Real-time Optimization, *J. Proc. Contr. 19*, 1277–1288.

Yoon, S., MacGregor, J.F. (2000). Statistical and Causal Model-Based Approaches to Fault Detection and Isolation, *AIChE J*., *46*, 1813-1824.

Young, R.E., Bartusiak, R.D., Fontaine, R.W. (2001). Evolution of an Industrial Nonlinear Model Predictive Controller, Preprints on Chemical Process Control-CPC VI, CACHE, Tucson, AZ, 399-401.

Zavala, V. M., Biegler, L.T. (2009a). The Advanced-step NMPC Controller: Optimality, Stability and Robustness. *Automatica*, 45, 86-93.

Zavala, V. M., Biegler, L.T. (2009b). Optimization-Based Strategies for the Operation of Low-Density Polyethylene Tubular Reactors: Moving Horizon Estimation. *Comput. Chem. Eng*., *33*, 379-390.