16th European Symposium on Computer Aided Process Engineering
and 9th International Symposium on Process Systems Engineering
W. Marquardt, C. Pantelides (Editors)

Simultaneous Dynamic Validation/Identification of Mechanistic Process Models and Reconciliation of Industrial Process Data

Pablo A Rolandi[a], José A Romagnoli[b]

[a]*Process Systems Enterprise Ltd., 107a Hammersmith Bridge Road, London W6, UK*
[b]*Department of Chemical Engineering, Louisiana State University, Baton Rouge, LA, 70803 USA*

Abstract

Process models are subject to parametric uncertainty and raw process-instrumentation data are corrupted by systematic and random errors. In this work, we present a framework for dynamic parameter estimation and data reconciliation aiming at integrating model-centric support tools and industrial process operations. A realistic case-study for the rectification of the mass balance of an industrial continuous pulping system is presented. The incentive for gross-error estimation during model-based production accounting and inventory analysis is demonstrated.

1. Introduction

An assumption implicit in the execution of any hybrid data-driven/model-based activity is that both the mechanistic model and experimental data describe the behaviour of the process system accurately. In the case of industrial manufacturing systems, these conditions are rarely met. Physical, chemical, and biochemical process phenomena are complex and, therefore, difficult to model conceptually and mathematically. Thermodynamic and transport properties and reaction rates are difficult to characterise experimentally and, hence, subject to parametric uncertainty. Modelling industrial process systems aggravates these problems since site-specific operating conditions render theoretical and/or empirical modelling of some natural phenomena virtually intractable. Concurrently, plant data is abundant and readily available in industrial process systems. However, raw process-instrumentation data are corrupted by systematic and random errors undermining the solution performance of any hybrid data-driven/model-based activity making use of experimental data pools. *Joint parameter estimation and data reconciliation* techniques provide a framework for simultaneous dynamic validation/identification of mechanistic process models and reconciliation of industrial process data.

2. Problem definition

In general terms, the parameter-estimation problem can be stated as finding the optimal estimate of the vector of parametric variables θ, subject to the constraints imposed by the fundamental principles of conservation (i.e. the mathematical model of the process). Conventionally, optimality implies maximising the likelihood of predicting the experimental set of measurements or, alternatively, minimising a measure of the distance between experimental measurements and predicted values. Similarly, simultaneous data reconciliation and gross-error estimation can be stated as finding an optimal vector of random measurement errors $\boldsymbol{\varepsilon}$ and systematic errors β so that the

corrected measurements satisfy the fundamental principles of conservation imposed by the mechanistic process model. Within the scope of data reconciliation, optimality is usually associated to minimising a measure of the error between experimental measurements and predicted values.
In this work, we shall adopt a notation that eases the declaration of a given process-engineering problem in terms of the conventions of state-of-the-art high-level symbolic languages. Thus, we define a dynamic estimation problem (DEP) as follows:

$$\begin{gathered}
\min_{\theta,\beta,\omega,\gamma} \varphi(\tilde{z}(t), z(t), \sigma(t)) \\
F(\dot{x}(t), x(t), y(t), u(t), p, \theta, \beta) = 0,\ t \in [0, t_f] \\
I(\dot{x}(0), x(0), y(0), u(0), p, \theta, \beta) = 0 \\
\sigma(t) = \sigma(\tilde{z}(t), z(t), \omega, \gamma),\ t \in [0, t_f] \\
\theta^{\min} \le \theta \le \theta^{\max} \\
\beta^{\min} \le \beta \le \beta^{\max} \\
\omega^{\min} \le \omega \le \omega^{\max} \\
\gamma^{\min} \le \gamma \le \gamma^{\max}
\end{gathered} \tag{1}$$

For each measuring device, the reconciled (corrected) measurement $\mathbf{z}$, the raw measurement $\tilde{\mathbf{z}}$, the measurement error $\boldsymbol{\varepsilon}$ and the measurement bias β are given by:

$$\mathbf{z} = \tilde{\mathbf{z}} + \boldsymbol{\varepsilon} + \beta \tag{2}$$

In the *error-in-variables measured* (EVM) method for parameter estimation/data reconciliation, all measured process variables are assumed to contain systematic and random errors. In order to engage in the solution of this type of DEP using commercial general-purpose process-engineering software tools, a problem formulation consistent with the characteristics of Eq. (1) is needed. In this work, we shall assume that $\left\|\boldsymbol{\varepsilon}^{ip}\right\|_\infty << \beta^{ip}$, where *ip* indicates input process variables; according to this, Eq. (2) becomes:

$$\mathbf{z}^{ip} = \tilde{\mathbf{z}}^{ip} + \beta^{ip} \tag{3}$$

Here, $\mathbf{z}^{ip}$ can be uniquely determined from $\tilde{\mathbf{z}}^{ip}$ since, structurally, the number of unknowns in Eq. (3) is one. Consequently, the overall number of decision variables to be estimated is N (*i.e.* the dimension of the vector β). Because in plant-wide industrial applications the number of measuring devices to be reconciled is generally in the order of a few hundreds, this approach renders large-scale EVM DEP solvable with available advanced process modelling (APM) tools. From a physical perspective, $\left\|\boldsymbol{\varepsilon}^{ip}\right\|_\infty << \beta^{ip}$ denotes sensors with *good precision and poor calibration* (this situation is the most common in industrial manufacturing plants).
Defining a dynamic estimation problem requires selecting a subset of measurements from the experimental data set that is consistent with the purpose of the particular estimation experiment. In the case of industrial process systems where raw plant data is

abundant, this process is rarely a trivial task. For example, process measurements are available at sampling periods which are orders of magnitude smaller than the characteristic time constant of the process system, leading to a phenomenon of *data over-sampling*. As a consequence, raw plant data is not adequate for populating this experimental data subset, and it is advisable to perform some data pre-processing and conditioning in order to improve the solution performance of the dynamic estimation experiment. In this work, the methodology for *reconstruction of process trajectories* (RPT) proposed by Rolandi & Romagnoli (2006) has been used to reduce the number of observations in the experimental data subset and simultaneously smooth high-frequency temporal fluctuations in process variables. Even though the details of this technique are out of the scope of this contribution, RPT has improved the accuracy, efficiency and robustness of industrial DEP.

3. Methodology

A successful and meaningful definition of parameter estimation/data reconciliation problems requires a painless integration of empirical data and large-scale dynamic models. State-of-the-art commercial process-engineering tools lack support mechanisms for manipulating plant data seamlessly and incorporating this information in the formulation of hybrid data-driven/model-based problems. This has precluded the routine validation of plant-wide mechanistic models, as well as the widespread use of advanced model-centric technologies (MCTs) such as joint parameter estimation and data reconciliation techniques.

In a companion paper (Romagnoli & Rolandi, 2006), the authors proposed a novel architecture for process-engineering software development and introduced the notion of the so-called *Problem Definition Component* (PDC). This software object supports a novel methodology for definition of parameter estimation/data reconciliation problems which is based on the refinement of instances of *process-engineering data models*. In this paradigm, *Data Model Templates* (DMTs) determine what information (predominantly the model's structure and control system's objects) is available to the user and how this information can be manipulated by the end-user. On the other hand, *Data Model Definitions* (DMDs) represent valid model-based activities and associated experimental process data. DMDs are generated by the user as a series of refinements of the original DMTs according to particularities of the conceptual definition of a given process-engineering problem. This definition process is regulated entirely by the nominated PDC. Due to space constraints, the *Problem Definition Environment* (PDE) of the System for Support of Process Operations (SYSS-PRO) software prototype will not be shown in this work.

Two data models are needed in order to fully describe the mathematical definition of a dynamic estimation problem. These structures are the so-called *Process Data Object* (PDO) and the *Dynamic Estimation Problem* data models (DEP). In brief:

- **PDO model**: it contains data representing raw experimental process data in a form suitable for combined discrete/continuous process modelling; not only does this structure support data pre-processing, conditioning and reconstruction techniques, but it also maps process instrumentation from which data was retrieved to the corresponding input/output variables of the process and model variables.
- **DEP model**: it contains data determining structural and numerical information of general dynamic estimation problems; this structure is given by a series of control (input), measured (output) and parametric (decision) process variables which maps into the corresponding model variables and process-instrumentation objects

(devices); this allows, for instance, a better characterisation of the objective function, selection of process operating parameters and/or measurement biases to estimate, determination of forcing input conditions, etc.; it also keeps information on upper and lower bounds and initial guesses of decision variables.

4. Case-study

With the exception of the contribution by Özyurt & Pike (2004), dynamic parameter estimation and data reconciliation of industrial process systems represented by large-scale mathematical process models are rare in the open literature, and the solution of this kind of problems still poses several challenges to the research community.

In this work, a large-scale mechanistic model of the continuous pulping system of a world-class industrial pulp and paper mill is used to illustrate viability of the proposed framework. Overall, the implementation of the resulting large-scale mathematical model gives rise to approximately $1.5 \cdot 10^4$ differential-algebraic equations; among these, there are $1.4 \cdot 10^4$ algebraic equations, $9.7 \cdot 10^2$ ordinary differential equation, and $3.2 \cdot 10^2$ structural degrees-of-freedom. Concurrently, there are approximately $3.6 \cdot 10^2$ statuses within the state transition network. gPROMS was used as the modelling and solution engine (MSE).

The goal of this case-study is to reconcile historian process data focussing on the closure of the general mass balance of the continuous pulping system. We will also aim at demonstrating that the abundance of plant data in today's industrial manufacturing systems can be readily exploited to solve realistic model-based parameter estimation/data reconciliation problems.

4.1. Problem specification

In this case-study, process-instrumentation data obtained from the historian throughout 24hr of operation is used. A set of 26 input process variables is reconstructed in order to force the behaviour of the continuous process system according to experimental process conditions. Among these, 21 are controlled variables and 5 are disturbances. A combined implicit/explicit *state initialisation procedure* is used to determine the initial state of the process; the details of this technique are outside the realms of this manuscript.

Two parametric process variables are subject to *estimation*. The *wood chip impregnation factor* is a measure of the flowrate of steam condensate bounded to the interstitial space between wood chips before entering to the chip meter in the feed line. Changes in wood handling operations and operating conditions of the chip bin affect the impregnation of wood, changing the free-liquor pattern flow and affecting the extent of the pulping reactions and closure of the overall mass balance. The *pre-multiplier of the fundamental kinetic model* also determines the extent of the pulping reactions, accommodating for seasonal wood-composition fluctuations and inadequate wood handling operations. In this case-study, three flow-measurement devices are *rectified*: the *overall white liquor addition*; the *wash filtrate addition* to the digester's bottom; and the *black liquor extraction* from the upper screens of the digester. Data from eight sensors are used for the purpose of estimation; three of them are output measured process variables and five are input measured process variables.

A *weighted least-squares* objective function minimising the difference between the model predictions and experimental observations is used in this case-study. The *weights* of each sensor are proportional to the expected value of the measured time series; hence, relative deviations contribute equally to the magnitude of the objective function irrespectively of the nature of the measured process variable.

The *potential* for model-based joint parameter estimation/data reconciliation of a large-scale complex industrial process system is demonstrated in this case-study. The problem results in the estimation of *five* parametric process variables (*three* of them are measurement biases) from an experimental data pool of *eight* measured variables and *twenty-six* control variables. The *challenge* of this joint parameter estimation/data reconciliation case-study lies on the combination of large amounts of process data and a large-scale mechanistic process model to solve an involved process-engineering problem of interest to mill personnel.

4.2. Analysis of results

Figures 1 to 3 show the confidence regions for the wood chip impregnation factor and the kinetic pre-multiplier, the white liquor and wash filtrate addition biases, and the upper extraction flow bias and kinetic pre-multiplier, respectively.

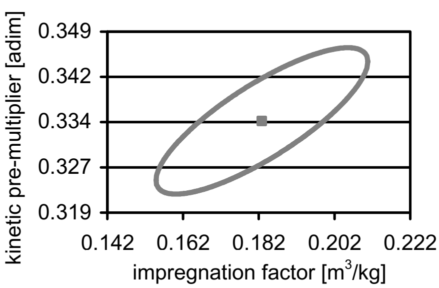

Figure 1. Confidence region.

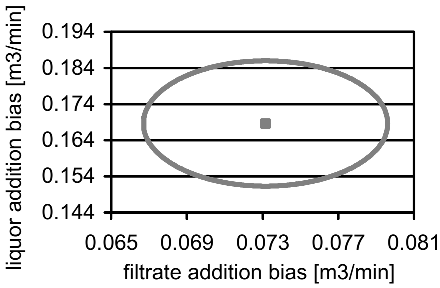

Figure 2. Confidence region.

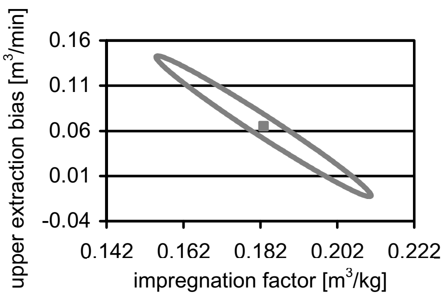

Figure 3. Confidence region.

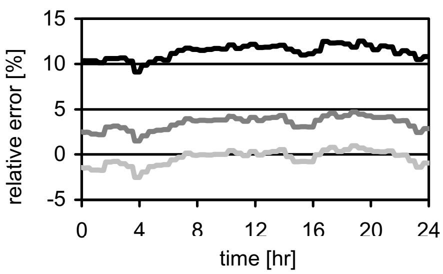

Figure 4. Mass balance closure.

Figure 4 shows three different trajectories corresponding to the closure of the general mass balance according to *different degrees of awareness* on the status of the process system. For instance, the conservation of mass in the continuous pulping area on the basis of the measured volumetric flowrates is inaccurate by approximately 11.3%; this discrepancy is due to the trivial fact that mass flowrates and not volumetric flowrates should be used in this calculation; unfortunately, this information is rarely available from industrial process instrumentation. Hence, *accurate inventory analysis is virtually impossible without the aid of model-based software tools*. In light of these facts, the mechanistic model of the continuous pulping area is used to examine the fulfilment of the principles of conservation. In effect, when the calculated mass flowrates are used, the closure of the general mass balance can be verified, on average, by a reasonable 3.5%. However, joint parameter estimation/data reconciliation enables us to approach this problem from a different perspective. Indeed, the mechanistic process model could be used to attain a more accurate compliance of experimental plant data with the fundamental laws of conservation provided that the plant/model mismatch was attributable not only to the mechanistic process model but also to experimental plant data. When gross errors are corrected, the overall conservation of mass is verified within a 0.7% error. These results substantiate the idea that *gross-error detection and estimation is more critical to advanced industrial process data management systems than the conventional reconciliation of random errors.*

Since production accounting and inventory analysis are based on the information arising from process instrumentation, systematic errors introduce a bias in these calculations. From a practical viewpoint, it would be reasonable to estimate those biases which have a strong impact on inventory analysis, or whose quantification is vital for other operational purposes (*e.g.* inferential soft-sensing). In the case of an industrial continuous pulping system, a cost analysis reveals that the most significant sources of revenue and expenses are likely to be the production of pulp, the cost of chip consumption and the cost of evaporation of weak black liquor.

The cost of evaporation of weak black liquor can be partially reconciled from the estimate of the bias of the upper-screen extraction flow meter. The estimated 6.4% error in this process measurement is associated to a material stream which accounts for nearly 32% of the overall weak black-liquor extraction flow from the continuous cooking digester at this nominal production level (~3.1m^3/min). Additionally, the treatment of the black liquor in the evaporation area comprises approximately 56% of the variable costs of operation of the continuous pulping area (~ 88US$/min). Hence, a 6.4% measurement error on such a critical process stream is equivalent to a production miscalculation of approximately 0.50 million US$ per year, or an inventory error of roughly 32 thousands cubic meters per year.

5. Conclusions

The ability to manipulate plant data seamlessly and to define dynamic estimation problems in the industrial workplace is *critical* to the success of advanced MCTs. In this paper we described a novel software architecture aiming at this goal, and we presented two process-engineering data models enabling this paradigm shift. A prototype estimation/reconciliation environment was built to ease the manipulation of these data models while defining joint parameter estimation/data reconciliation problems of industrial relevance. A large-scale process model of an industrial continuous pulping system was used. The accuracy of the process model was improved and process-instrumentation data was rectified by joint parameter estimation/data reconciliation techniques. Also, the closure of mass balances was improved drastically, and gross-errors estimation was found to be critical for accurate production accounting, inventory analysis and soft-sensing of industrial process systems. This provided an economic incentive for applying the proposed framework for joint parameter estimation/data reconciliation supporting the advanced operation of industrial process systems.

References

Özyurt, D.B., Pike, R.W. (2004). Theory and practice of simultaneous data reconciliation and gross error detection for chemical processes. Computers & Chemical Engineering, 28, 381-402.

Rolandi, P.A. and Romagnoli, J.A. (2006). Integrated model-centric framework for support of manufacturing operations. Part ii: The simulation environment. Computers and Chemical Engineering, submitted for publication.

Romagnoli, J.A., Rolandi, P.A (2006). Model-centric technologies for support of manufacturing operations, PSE 2006/ESCAPE 16.