8th International IFAC Symposium on
Dynamics and Control of Process Systems
Preprints Vol.2, June 6-8, 2007, Cancún, Mexico

ON-LINE ESTIMATION FOR PROCESS CONTROL AND OPTIMIZATION APPLICATIONS

Tor Steinar Schei

Cybernetica AS

Abstract: Design of Kalman filter type and moving horizon estimators for on-line estimation applications based on first principles models is reviewed. Important design issues are discussed, such as: model development; choice of process noise model and selection of model parameters for on-line estimation; use of asynchronous and delayed measurements; and off-line estimation of fixed but uncertain model parameters. The main conclusion, which is substantiated through application examples, is that robust and reliable estimation applications based on first principles models of considerable complexity, can be designed and implemented for use in an industrial environment.

Keywords: State estimation; Parameter estimation; Estimation algorithms; Kalman filters; Moving horizon estimation; On-line estimation; Industrial applications.

1. INTRODUCTION

In recent years the interest in optimization of process operations has increased significantly in most process industries. The use of stationary process models for steady-state real-time optimization (RTO) has been applied for optimization of continuous process operations for a long time in some process industries. However, optimal operation of processes which are characterized by continuous operations with frequent grade changes, semi-batch or batch operations often demand repetitive dynamic optimization, e.g. implemented using nonlinear model predictive control techniques. Such applications are based on fundamental process models derived from first principles, which require real time estimation of process states and uncertain process parameters.

This paper focuses on the development and implementation of on-line estimation applications in the process industries. Even though the state and parameter estimation problem occurs in a much wider context than as part of a feedback control or on-line optimization system, the focus will be on the latter.

Process models are assumed to be formulated on a time-discrete nonlinear stochastic state-space form:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_{k-1}, \mathbf{v}_{k-1}) \quad (1)$$

$$\mathbf{y}_k = \mathbf{g}(\mathbf{x}_k, \boldsymbol{\theta}, \mathbf{u}_{k-1}) + \mathbf{w}_k \quad (2)$$

where

- $\mathbf{x}_k$ is the n_x - vector of states at time t_k ;
- $\boldsymbol{\theta}$ is the n_θ - vector of model parameters;
- $\mathbf{u}_{k-1}$ is the n_u - vector of measured process inputs, which are assumed constant over the time interval $[t_{k-1}, t_k]$;
- $\mathbf{y}_k$ is the n_y - vector of output measurements at time t_k ;
- $\mathbf{v}_{k-1}$ and $\mathbf{w}_k$ are sequences of independent random process and measurement noise variables (white noise sequences);
- $\mathbf{f}(\mathbf{x}_{k-1}, \boldsymbol{\theta}, \mathbf{u}_{k-1}, \mathbf{v}_{k-1})$ is the nonlinear process model. $\mathbf{f}$ is generally the solution to a system of differential-algebraic equations (DAE) or ordinary differential equations (ODE) between the sampling instants t_{k-1} and t_k ;
- $\mathbf{g}(\mathbf{x}_k, \boldsymbol{\theta}, \mathbf{u}_{k-1})$ is the nonlinear measurement model;
- the basic sample interval $[t_{k-1}, t_k]$ is assumed to be constant.

The state estimation problem is to determine an estimate of the state $\hat{\mathbf{x}}(t_k) = \hat{\mathbf{x}}_k$ given the chosen model structure, an a priori initial state estimate $\bar{\mathbf{x}}(t_0) = \bar{\mathbf{x}}_0$, and a sequence of noisy measurements $Y(t_k) := \{\mathbf{y}(t_0), \ldots, \mathbf{y}(t_k)\}$. Rawlings and Bakshi (2006) give a recent overview of currently available methods for state estimation in linear, constrained and nonlinear systems. In the stochastic setting chosen here, the conditional density of the state given the measurements is the natural statistical distribution of interest. The complete conditional density is difficult to calculate exactly, however, except for well-known simple systems, such as when $\mathbf{f}$ and $\mathbf{g}$ are linear, and $\mathbf{v}$, $\mathbf{w}$ and $\bar{\mathbf{x}}_0$ are normally distributed. In this case the conditional density is also Gaussian with mean and covariance provided by the well-known Kalman filter. When $\mathbf{f}$ and $\mathbf{g}$ are nonlinear, however, the conditional density is not Gaussian, and obtaining a complete solution is generally impractical. Moreover, when state estimation is used as part of a feedback control system, the state estimator must meet other requirements. The estimate must be found during the available sample interval of the system as each measurement becomes available. The on-line requirements provide further limitations on what is achievable in state estimation.

This paper discusses two types of estimators for use in industrial process applications: *Kalman Filter* type of nonlinear filters and *Moving Horizon Estimators* (MHE). The discussion is based on the author's experience from implementing these estimation techniques in industrial applications.

2. KALMAN FILTER TYPE OF NONLINEAR FILTERS

2.1 The Extended Kalman Filter

The filtering problem is to determine an estimate of the state $\hat{\mathbf{x}}(t_{k|k}) = \hat{\mathbf{x}}_k$ at time t_k. The basic Kalman filter algorithm, which is an optimal minimum variance estimator, is derived for linear systems. The theory is extended to nonlinear systems with well-defined first-order derivatives in the *Extended Kalman Filter* (EKF) algorithm, which is based on first order Taylor approximations of state transition and observation equations around the estimated state trajectory. The EKF algorithm consists of a *prediction* part where the *a priori* state estimate $\hat{\mathbf{x}}(t_{k|k-1}) = \bar{\mathbf{x}}_k$ is determined from (1) by propagating the state estimate $\hat{\mathbf{x}}(t_{k-1|k-1}) = \hat{\mathbf{x}}_{k-1}$ at time t_{k-1} and the *mean* process noise $\bar{\mathbf{v}}_{k-1}$ through the nonlinear model $\mathbf{f}$. The *a priori* predicted measurement $\bar{\mathbf{y}}_k$ is then calculated from (2) based on the a priori state estimate $\bar{\mathbf{x}}_k$ and the mean of the measurement noise $\bar{\mathbf{w}}_k$. In the measurement *correction* part of the EKF algorithm the *a posteriori* state estimate is calculated as

$$\hat{\mathbf{x}}_k = \bar{\mathbf{x}}_k + \mathbf{K}(k)(\mathbf{y}_k - \bar{\mathbf{y}}_k) \tag{3}$$

where $\mathbf{y}_k$ is the measurement vector at time t_k and $\mathbf{K}(k)$ is the Kalman filter gain matrix. $\mathbf{K}(k)$ is calculated from the process noise covariance $\mathbf{V}_{k-1}$, the measurement noise covariance $\mathbf{W}_k$, the a priori state covariance estimate $\bar{\mathbf{X}}_{k-1}$, and from the partial derivatives of $\mathbf{f}$ and $\mathbf{g}$ in (1) and (2) with respect to the stochastic variables. Many variations on the same theme have been proposed such as the iterated EKF and the second-order EKF (Gelb, 1974).

Of the nonlinear filtering methods, the EKF method appears to be the most widely applied method in various industrial applications. The reason for the popularity of this method is assumed to be due to its relative simplicity and demonstrated effectiveness in handling many nonlinear systems with reasonable use of computational resources. However, there are also some problems with the EKF, such as the inability to accurately incorporate physical state constraints. Such constraints generally have to be enforced through simple projections of the a posteriori state estimate in (3). Other problems are due to the use of differentiation to determine the linear system matrices used for calculating state covariances and the Kalman filter gain. The EKF can easily fail due to non-existence of the partial derivatives in certain singular points of the state space (Schei, 1997), and even if the derivatives exist they may lead to poor approximations in terms of mapping the mean and the covariance of the state probability density distribution through the nonlinear functions $\mathbf{f}$ and $\mathbf{g}$.

2.2 Sigma Point Kalman Filters

Recent developments in dynamic filtering are the *Sigma Point Kalman Filters* (SPKFs). Like the basic Kalman filter, the SPKFs seek to determine a state estimate that minimizes the l_2-norm of the residuals. The SPKF technique differs from the basic Kalman filter in the estimate and covariance propagation prior to determining the a posteriori state estimate at the time of the measurement update. The differences lie in that the SPKFs do not linearize the dynamic system for the propagation, but instead propagate a cluster of points centered around the current estimate in order to form improved approximations of the conditional mean and covariance. The sigma points are chosen deterministically, for example as points on a selected covariance contour ellipse. A particular advantage with the SPKF approach is that these filters do not require knowledge or existence of the partial derivatives of the system dynamics and measurement equations. SPKFs have the additional advantage over the basic Kalman filter in that they can easily be extended to determine second-order solutions to the minimum l_2-norm measurement update, which increases the estimation accuracy when the system and measurement equations are nonlinear.

The First-Order (DD1) and Second-Order (DD2) *Divided Difference Filters* (Nørgaard, 2000) are examples of SPKF-class estimators; other examples can be found in Julier *et al.* (2000, 2004) and Ito *et al.* (2000). The Divided Difference Filters are generalizations of the filter introduced by Schei (1995, 1997).

3. MOVING HORIZON ESTIMATION

3.1 Full Information Estimator

The *full information estimator* is derived by maximizing the joint probability distribution for a trajectory of state values $\{\mathbf{x}_0,\ldots,\mathbf{x}_T\}$, given a sequence of measurements $\{\mathbf{y}_0,\ldots,\mathbf{y}_T\}$. It can be shown (Haseltine and Rawlings, 2003) that if the a priori state estimate $\bar{\mathbf{x}}_0$ and the process and measurement noise sequences $\{\mathbf{v}_0,\ldots,\mathbf{v}_{T-1}\}$, $\{\mathbf{w}_0,\ldots,\mathbf{w}_T\}$ are all normally distributed and independent stochastic variables, the full information estimator for the system in (1), (2) can be formulated as a least-squares optimization:

$$\Phi_T = \min_{\mathbf{x}_0,\ldots,\mathbf{x}_T} \Gamma(\mathbf{x}_0) + \sum_{k=0}^{T-1} \mathbf{v}_k^T \mathbf{V}_k^{-1} \mathbf{v}_k + \sum_{k=0}^{T} \mathbf{w}_k^T \mathbf{W}_k^{-1} \mathbf{w}_k \quad (4)$$

$$\text{where} \quad \Gamma(\mathbf{x}_0) = (\mathbf{x}_0 - \bar{\mathbf{x}}_0)^T \bar{\mathbf{X}}_0^{-1} (\mathbf{x}_0 - \bar{\mathbf{x}}_0) \quad (5)$$

$\mathbf{V}_k$ and $\mathbf{W}_k$ are covariance matrices for the process and measurement noise variables $\mathbf{v}_k$ and $\mathbf{w}_k$, and $\bar{\mathbf{X}}_0$ is the covariance matrix for the a priori state estimate $\bar{\mathbf{x}}_0$. The size of the problem in (4)-(5) grows as new measurements become available.

3.2 Arrival Cost and Moving Horizon Estimator

To overcome the computational limitations of the full information estimator, the problem is reformulated over a fixed-size estimation window of the *N* last measurements:

$$\begin{aligned} \Phi_T = & \min_{\mathbf{x}_{T-N},\ldots,\mathbf{x}_T} \tilde{\Phi}_{T-N}(\mathbf{x}_{T-N}) \\ & + \sum_{k=T-N}^{T-1} \mathbf{v}_k^T \mathbf{V}_k^{-1} \mathbf{v}_k + \sum_{k=T-N+1}^{T} \mathbf{w}_k^T \mathbf{W}_k^{-1} \mathbf{w}_k \end{aligned} \quad (6)$$

In the *Moving Horizon Estimator* (MHE) (Rao *et al.*, 2003) the problem in (6) is solved approximately and repeatedly at each sampling time subject to the model in (1)-(2), constraints in the state estimates, and possibly constraints on the process and measurement noise. In order to ensure feasibility of the optimization problem, constraints on the measurement noise should, however, generally be avoided. The *arrival cost*, $\tilde{\Phi}_{T-N}(\mathbf{x}_{T-N})$, summarizes the past information up to the start of the estimation window. The arrival cost can only be determined approximately, and the construction of this approximation is the key to preserving stability and performance of the MHE. Estimator divergence may result if the initial penalty biases old data by too strongly weighting the past estimates, while performance may suffer if the initial penalty neglects the old data by not sufficiently weighting them. Arrival cost is a fundamental concept in MHE because it allows the recasting of the full information problem as an approximate fixed horizon problem. An exact algebraic expression for the arrival cost can only be derived for linear, unconstrained systems under Gaussian assumptions (Rao *et al.*, 2001).

The statistically correct choice for the arrival cost is based on the conditional density of the state estimate $\hat{\mathbf{x}}_{T-N|T-N}$ at sample $T-N$, given data up to sample $T-N$. One possibility is to use the *a posteriori* state covariance determined by the EKF or one of the SPKF estimators as the arrival cost and to weight deviations from $\hat{\mathbf{x}}_{T-N|T-N}$. This is a *filtering scheme* since the penalty in the arrival cost is based on past measurements only. A problem with this scheme is that it tends to introduce oscillations in the state estimates. A *smoothing scheme*, where the arrival cost is based on the conditional density of the smoothed estimate $\hat{\mathbf{x}}_{T-N|T-1}$, given data up to sample $T-1$, will solve this problem but it will introduce other problems us discussed below.

3.3 Reducing the Size of the MHE Optimization Problem

If it is assumed that the optimization problem in (6) is solved using a *sequential* approach, i.e. the integration of the process model equations is performed independently of the optimization in (6), the MHE optimization problem has $n_x + N \cdot n_v$ independent variables. This size of the optimization problem might be prohibitive for many applications. For models with a large number of states it is impractical to estimate the entire state vector $\mathbf{x}_{T-N}$ in (6), and most or all of the individual states will have to be fixed to $\hat{\mathbf{x}}_{T-N|T-N}$ or $\hat{\mathbf{x}}_{T-N|T-1}$. This might lead to instability as pointed out above. However, stability problems can usually be prevented if the data window is sufficiently long. If the arrival cost is omitted from (6), the MHE optimization has $N \cdot n_v$ independent variables. As is discussed below, in order to obtain zero steady-state deviations in all predicted output measurements, n_v should at least be equal to n_y, but for high estimator performance, n_v should generally be $\sim 2n_y$. It is usually also desirable to choose a relatively short sampling interval and a long data window, particularly if the arrival cost is omitted. A feasible approach to reduce the number of independent process noise variables in (6) is to depart from the white noise assumption and parameterize the process noise sequence with a smaller number of independent parameters, e.g. the process noise is assumed to be constant for a number of consecutive sampling intervals. However, this will

deteriorate the performance of the estimator. Hence, designing an MHE application is generally a compromise between performance and computational requirements.

3.4 Solving the MHE Optimization Problem

The optimization problem in (6) can be solved by applying a sequential quadratic programming (SQP) algorithm and exploiting the structure to attain efficiency. The SQP algorithm is an iterative technique in which each step is obtained by solving an approximation to (6), in which the objective is replaced by a quadratic approximation and the constraints by linear approximations. Since (6) is solved repeatedly at each sample time, the initial estimates will normally be close to the optimal solution, and only a few SQP iterations are usually required at each sample time. By limiting the number of SQP iterations, the maximum computational time can be well defined and within the required time-limit for the particular application.

4. DEVELOPMENT OF ON-LINE ESTIMATION APPLICATIONS

This section will review some issues related to the development of on-line estimation applications for control and optimization purposes.

4.1 Model Development

The starting point is model development, which should be directed by the particular application needs. The model is assumed to be based on first principles, and should incorporate the main process nonlinearities. It should, however, be carefully considered how far the first principles modelling should be followed. Often complex mechanisms, which are difficult to model with high degree of confidence, should be replaced by unknown parameters which are estimated from plant data. Also, the numerical properties of the model should be carefully considered. The model must be sufficiently smooth with respect to perturbations of the independent variables if MHE optimization calculations are based on sensitivities obtained from numerical differentiation. Also, if analytical derivatives are used, they should exist and be continuous for all possible states. Since the estimator is part of a real time system with limited time between samples, the maximum computational time for each sample should be well defined. Hence, it is advantageous if the underlying ODE or DAE system in (1) can be solved with a fixed integration time step. This time step is determined during application development.

4.2 Process Noise Model

A critical design issue is the choice of the process noise model in (1). This choice depends on which model parameters are chosen for on-line estimation. A standard approach to modelling of uncertain parameters is to augment the state vector $\mathbf{x}$ with the uncertain parameters $\boldsymbol{\theta}$, and to model the time variation of these parameters as integrated white noise sequences:

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \boldsymbol{\upsilon}_{k-1} \tag{7}$$

The state $\mathbf{x}_k$ and the process noise $\mathbf{v}_{k-1}$ in (1) are augmented with $\boldsymbol{\theta}_k$ and $\boldsymbol{\upsilon}_{k-1}$. The choice of which parameters to estimate should be guided by an identifiability analysis. Usually it can not be assumed that the process excitations fulfil certain persistency requirements in order to ensure convergence of parameter estimates. Hence, the parameter vector $\boldsymbol{\theta}$ should normally be a set of parameters which is identifiable from stationary data, and which do not require any particular excitations in order to obtain convergence. Typical choices are heat and mass transfer coefficients, kinetic parameters, etc. By carefully selecting a number of parameters n_θ equal to the number of output measurements n_y, zero steady-state deviations in all predicted output measurements can normally be achieved.

The process noise model in (7) is usually not sufficient to obtain small measurement prediction errors, and additional process noise variables should be introduced. In the choice of process noise it is, however, essential that the basic balance equations are not violated. Process noise variables can typically be added to heat transfer rates, mass transfer rates, reaction rates, etc., such that mass, component and energy balances are fulfilled. This is important in many applications, e.g. to maintain a critical stoichiometric ratio between components or if the energy balance is used to estimate the conversion in a batch reactor. A simple example is the modelling of heat transfer from the reaction mixture to the cooling system in an exothermal reactor:

$$q = h_{k-1} A (T_r - T_{cw}) + v_{q,k-1} \tag{8}$$

$$h_k = h_{k-1} + \upsilon_{h,k-1} \tag{9}$$

In (8) and (9) q is the heat transfer from the reaction mixture, T_r and T_{cw} are the reactor and cooling water temperatures, and h is an overall heat transfer coefficient which depends on several factors such as mixing, viscosity of the reaction mixture, fouling, flow rate of cooling water, etc. In this example the heat transfer coefficient h is modelled as integrated white noise. In addition, white process noise is added directly on the heat transfer rate q. It is assumed that the reactor temperature and the outlet water temperature from the cooling system are measured. With this choice of noise model, the estimator can be tuned such that the measurement prediction errors are small without introducing aggressive corrections of the estimated heat transfer coefficient. This is important in order to maintain good prediction capabilities when the model is used in an NMPC or on-line optimization application. The same principle

can be used for modelling uncertain reaction rates. A kinetic parameter can be modelled as integrated white noise. In addition, white process noise should be added to the calculated reaction rate.

In applications where it is important to not violate basic balance equations, the MHE estimator has an advantage over Kalman filter type of estimators, since the balance equations can be fulfilled exactly, provided the stochastic model in (1) do not violate these balances and provided that the arrival cost is omitted such and $\mathbf{x}_{T-N}$ in (6) is fixed to $\hat{\mathbf{x}}_{T-N|T-N}$ or $\hat{\mathbf{x}}_{T-N|T-1}$ as discussed above. This is not the situation when Kalman filter type of estimators are applied, due to the linearization approximations involved in the calculation of model state updates from the measurement prediction deviations in (3). However, analysis of some applications have shown that also Kalman filter type of estimators can be designed to fulfil mass and energy balances with relatively high accuracy.

4.3 Process Output Measurements

When developing estimation applications, the characteristics of process measurements which are available for use as output measurements $\mathbf{y}$ in (2) are important. Available measurements are usually of the following types: continuous measurements which can be sampled with high sampling frequency; delayed measurements from various analytical instruments which are normally available at a fixed sampling rate; and manual laboratory measurements which are available at unequal intervals and with varying delays. The on-line estimation application is usually developed with a fixed basic sampling frequency. The continuous measurements will normally be available at each sample time while measurements from analytical instruments and in particular laboratory analyses are available at a lower and unequal rate. Both the Kalman filter type of estimators and the MHE estimator can easily be implemented to handle asynchronous measurements with arbitrary sampling intervals. In such an implementation, it must be determined which of the available measurements are valid at each basic sample time. This functionality is required also for the continuous measurements, because a system for on-line estimation should include validation of all process input and output measurements $\mathbf{u}$ and $\mathbf{y}$ in (1) in order to prevent outliers and faulty sensors from deteriorating the estimates.

Handling of measurement delays is more involved, particularly in Kalman filter type of estimators. If the measurement delays are not more than a few sampling intervals, this delay can be incorporated in the model by augmenting the state vector with delayed measurements. Another possibility is to store the old input and output data, and re-run the Kalman filter from the time when the measurement sample was drawn from the process. However, for processes where manual laboratory measurements must be used to correct the on-line estimator, the MHE estimator is generally preferred. In the MHE estimator delayed measurements are included by entering the measurement value at the correct place in the data window, in accordance with the time the measurement sample was drawn from the process. This approach is, however, not without problems. As discussed by Rawlings and Bakshi (2006), it is generally preferred to determine the arrival cost by using a *smoothing scheme* instead of the *filtering scheme*. However, when using the smoothing scheme in combination with delayed measurements, the total weight of the delayed measurements will depend on when the measurements were entered into the data window. Another problem is that when results from laboratory analyses are entered manually by the operator, it should be possible to correct an erroneous input after it has been entered into the data window. Both of these problems are solved by using the filtering scheme. However, as explained by Rawlings and Bakshi (2006), the filtering scheme tends to introduce oscillations in the state estimates. This is a problem which is still not satisfactorily solved.

4.4 Off-line Estimation of Constant Model Parameters

In (1)-(2) the parameter vector $\boldsymbol{\theta}$ is assumed to consist of uncertain and time varying parameters which are selected for on-line estimation. In addition there are usually a number of constant model parameters which are difficult to calculate exactly from first principles, and which should be estimated from logged process data during the application development. A feasible approach is to estimate a vector of uncertain but constant parameters $\boldsymbol{\eta}$, such that the measurement prediction errors obtained from one of the recursive estimators are minimized, using an SQP optimization approach. The optimization problem will generally also include estimation of uncertain initial states $\mathbf{x}_0$, corresponding to the start of the logged time series data used for the off-line estimation. These time series data should generally be obtained from excitation experiments which are designed to obtain sufficient identifiability of the parameter vector $\boldsymbol{\eta}$.

5. APPLICATION EXAMPLES

A few application examples where on-line estimation techniques have been implemented on various industrial processes are briefly reviewed below.

5.1 Suspension PVC Polymerization Process

Vinyl chloride monomer (VCM) is polymerized in a 140 m^3 batch autoclave with cooling jacket and overhead reflux condenser to maximize cooling capacity. A simple sketch of the autoclave is shown in Figure 1.

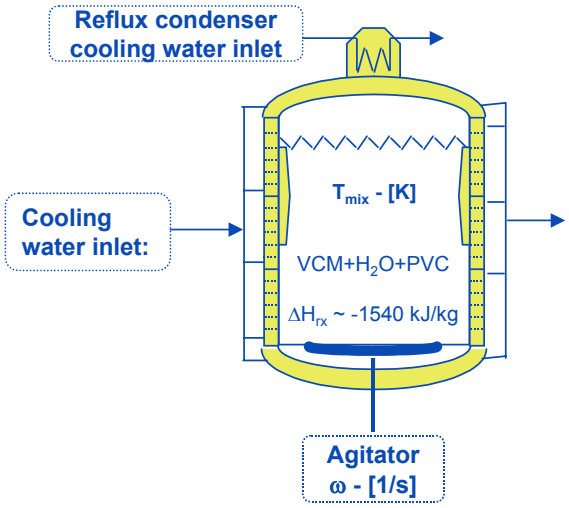

Fig. 1. Autoclave for batch polymerization of PVC. VCM is polymerized by dispersing liquid monomer droplets and initiators in water under intense agitation. The heat of reaction is removed by the use of a cooling jacket and an overhead reflux condenser.

A comprehensive model is developed which includes the reaction kinetics for this free-radical polymerization process, phase equilibria for the four-phase suspension process, material balances and energy balances for the suspension, the cooling jacket and the reflux condenser. In order to optimize the operation of this process, in particular the amount of initiators and the temperature profile during polymerization, it is essential that the model is able to accurately predict the product quality, which depends on the polymerization history, as well as the heat of reaction and the available cooling capacity throughout the entire polymerization stage. In order to accurately predict the energy balance, three model parameters are assumed to vary during the batch. A multiplicative correction factor for the polymerization rate was introduced to account for inaccuracies in the kinetic model. Similar correction factors were used for the heat transfer from the suspension to the cooling jacket and for the heat transfer in the reflux condenser.

A SPKF Kalman filter with 15 s sampling interval was designed for this application. Three output measurements are used for the estimation: the suspension temperature and the outlet cooling water temperatures from the jacket and the condenser. The on-line estimation of the three model parameters $\boldsymbol{\theta}$ ensure integral action in the predicted temperature measurements. In addition a number of other parameters $\boldsymbol{\eta}$ where estimated to fixed values during the application development. The on-line estimation is part of a nonlinear MPC (NMPC) application. The model is also used in a "run-to-run" optimization system (Schei *et al.*, 2001), where the three time-varying parameters are parameterized as functions of conversion.

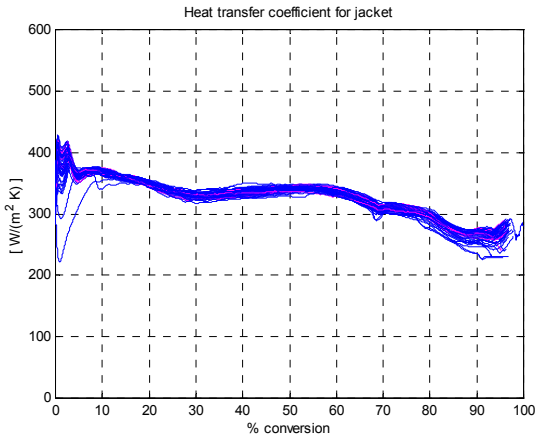

Fig. 2. Estimates of the overall heat transfer coefficient for the cooling jacket. Estimates from 50 consecutive batches are shown as functions of conversion.

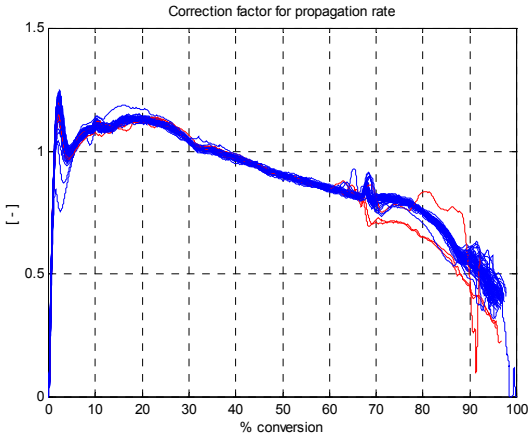

Fig. 3. Estimates of the multiplicative correction factor for the polymerization rate. Estimates from 50 consecutive batches are shown as functions of conversion.

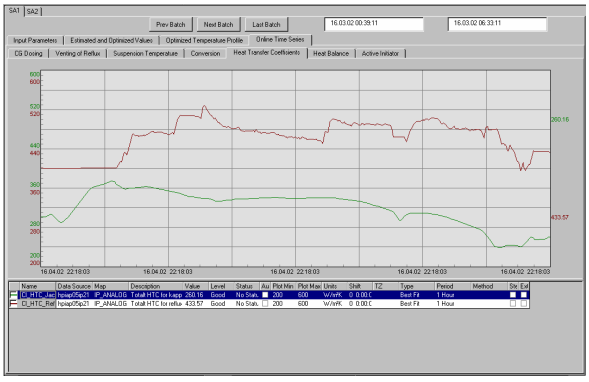

Fig. 4. Trend plot for estimated overall heat transfer coefficients for the reflux condenser (upper curve) and for the cooling jacket (lower curve).

Fig. 5. Trend plot for estimated reaction heat (upper curve) and estimated cooling from condenser and jacket (lower curves). The reflux condenser is not used during the initial part of the batch.

5.2 Phenol-formaldehyde Semi-batch Polymerization Process

Phenol-formaldehyde resole resins are produced in a 68 m^3 batch polymerization reactor with an internal coil system for reactor cooling and steam heating, and a reflux condenser for additional cooling. The resole resins are prepared by the reaction of phenol and formaldehyde, in the presence of sodium hydroxide and potassium hydroxide as the catalysts. Phenol and formaldehyde reacts through several possible reaction steps to form various methylolated phenols. The methylol phenols react further through condensation polymerization reactions to form resins with higher molecular weights. A model has been developed for this condensation polymerization process. The model includes: reaction kinetics for the various methylolation and condensation polymerization reactions; formaldehyde equilibria for aqueous solutions, hemiformal equilibria, and equilibria between the reactive ionized phenolic species and the corresponding unreactive neutral molecules; population balances for functional groups; and energy balances for the reactor and the cooling and heating systems. The resins are produced in semi-batch operations where formaldehyde and sodium hydroxide are charged to the reactor at several stages during a batch run. The reaction heat is high during parts of the batch, particularly during charging of formaldehyde and sodium hydroxide, and a powerful cooling system is required for this process.

A SPKF Kalman filter with 2 s sampling interval was designed for this application. An MHE estimator has also been tested with good results. The output measurements used for the estimation includes: reactor temperature, flow rate of water in the cooling coil, temperature of outlet water from the cooling coil, and the reactor weight scale reading. A number of parameters are estimated on-line: two multiplicative correction parameters for the methylolation and the condensation reactions; overall heat transfer coefficients for the cooling coil and for the steam heating coil; charging rate parameters for some of the raw materials; and two parameters related to the flow characteristic for the cooling coil with control valve. This on-line estimation application is part of an NMPC application where the position of the cooling water control valve is controlled directly. The estimated charging rate parameters are used to predict future charging rates for critical raw materials. The various parameters are only estimated during periods with sufficient identifiability of the individual parameters $\boldsymbol{\theta}$. E.g., the overall heat transfer coefficient for the cooling coil is not estimated during periods when the cooling water valve is closed.

5.3 Alkylation Batch Process

The production of X-ray contrast medium involves many batch operations. One of the last steps in the primary production is alkylation of substance Y to form the final basic substance P. A reactor vessel is loaded with a solvent and sodium hydroxide (NaOH) in which the solid Y is dissolved. The alkylation reaction is initiated when the alkylation agent A is added. The reaction is exothermic and a temperature control system maintains the temperature at the desired setpoint.

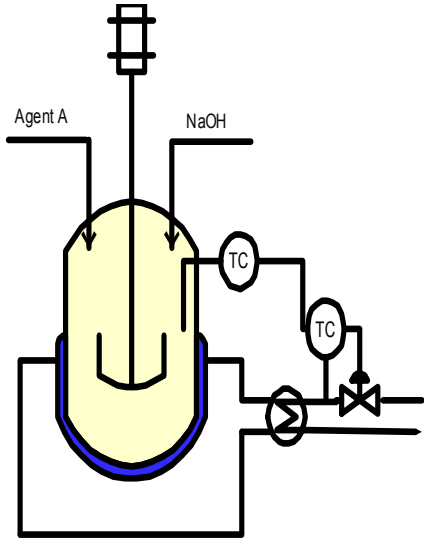

Fig. 6. Reactor vessel for alkylation batch process.

For this particular process it is important that the reaction rate follows a nominal trajectory. If the reaction proceeds too fast, too much by-product (W) will be formed, while if the reaction is too slow, the batch schedule is violated. The reaction rate is adjusted by adjusting the pH. The pH and the reaction rate are increased by adding sodium hydroxide, while pH and reaction rate are decreased by adding the alkylation agent A. These additions are the manipulated variables and are small compared to the initial amounts. The pH is also affected by the initial amount of Y and other components. A dynamic model including mass balances for the various chemical components is developed, and the kinetic parameters are fitted to data from more than 50 batches. The energy balance is not modelled for this application, and the reaction temperature is a measured input to the kinetic model.

During the course of the reaction, samples are taken at two different instances and the concentrations of Y and W are analyzed, in addition to the pH. Only when the results of these analyses are available, corrective actions are implemented on the process. In addition, an extra analysis is taken at the end of the batch. An NMPC application is designed to control the concentrations of Y and W at their set-points and within constraints at the final batch time. Reducing the batch time is not an issue for this process.

A *full information estimator* is designed for this application. Hence, the length of the data window, T in (4)-(5), increases throughout the batch, from zero up to the total batch time. Uncertainties are ascribed to the initial conditions $\mathbf{x}_0$. The initial amounts of Y, NaOH and A are all measured, but still there are uncertainties in the actual amount due to measurement errors and due to small contaminations of the agents. There is no process noise $\mathbf{v}_k$ in this application, except some minor uncertainties related to the corrective actions taking place at two time instances during the batch. The output measurements include the analyzed concentrations of Y, W and pH, which is done twice during the batch. At these two time instances the uncertainties in the Y and W measurements are different. It is difficult to measure pH in this system so little weight is put this measurement. As the laboratory measurements become available they are inserted into the data window according to the time-stamp when the samples were taken from the process. In addition, an alternative technique for analyzing Y and W with much higher frequency based on NIR is also applied. However, the laboratory measurements are still required for calibration of the NIR measurements. These measurements were readily added to the *full information estimator*. This estimator is well suited for batch processes with scarce quality measurements, which are delayed and asynchronous, and where it is essential to achieve maximum use of the information content in the measurements.

The figures below show filtered and smoothed estimates of the concentrations of Y and W as well as the pH. The laboratory measurements are indicated with small circles.

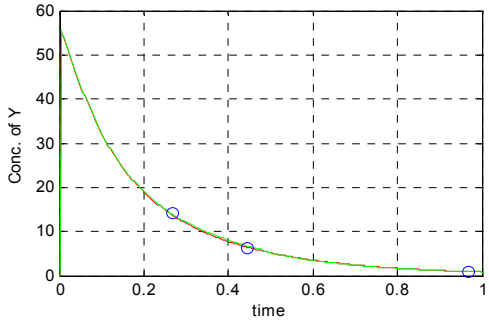

Fig. 7. Filtered and smoothed estimate of the concentration of Y. The two curves are almost indistinguishable.

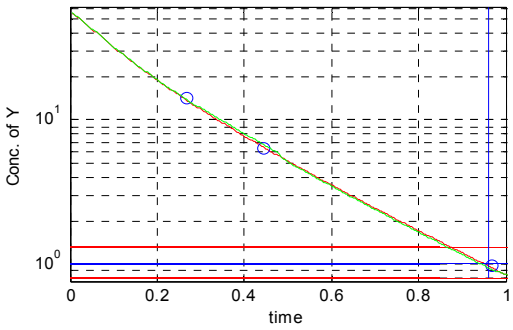

Fig. 8. Logarithmic scale for estimated concentration of Y. Setpoint and minimum and maximum constraints for the final concentration of Y are indicated in the plot.

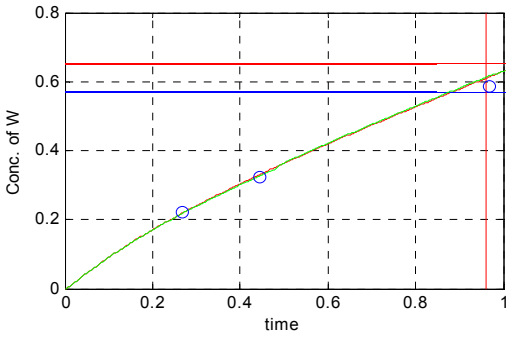

Fig. 9. Filtered and smoothed estimate of the concentration of W. The two curves are almost indistinguishable. Setpoint and maximum constraint for the final concentration of W are indicated in the plot.

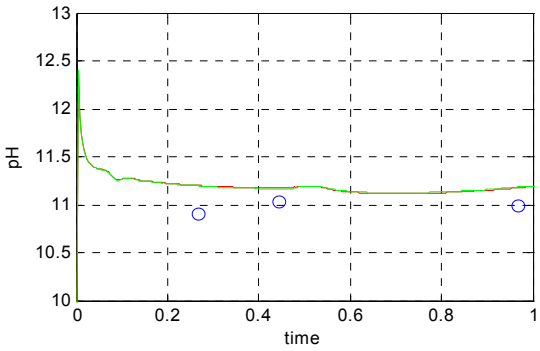

Fig. 10. Filtered and smoothed estimate of the pH. The two curves are almost indistinguishable.

6. CONCLUSIONS

This paper discusses two types of algorithms for use in estimation applications based on *first principles* models: *Kalman Filter* type of nonlinear filters and *Moving Horizon Estimators* (MHE). The discussion is based on the author's experience from implementing these estimation techniques in industrial process applications. After an introductory review of basic principles, some important design issues are discussed, such as: model development;

choice of process noise model and selection of model parameters for on-line estimation; use of asynchronous and delayed measurements; and off-line estimation of fixed but uncertain model parameters. The design of an estimation application is typically done in an iterative procedure including model refinement and validation, choice of noise model, tuning of process and measurement noise covariances, estimation of constant model parameters as well as other design issues related to the specific choice of estimation algorithm. The paper ends with a review of some industrial applications.

The main conclusion of this paper is that robust and reliable estimation applications based on first principles models of considerable complexity, and based on the current available theory, can be designed and implemented for use in an industrial environment. There are, however, still some issues which should be the focus of further research. One such specific issue is the problem related to the *arrival cost* in MHE estimation and the problems encountered when the arrival cost is based a *filtered* state estimate or a *smoothed* state estimate. A much wider area of further research is how to best solve the combined problem of the MHE optimization problem in (6) in connection with the integration of the underlying system of differential-algebraic equations.

REFERENCES

Gelb, A. (Ed.). (1974). *Applied optimal estimation.* The M.I.T. Press. Cambridge, Massachusetts.

Haseltine, E.L. and J.B. Rawlings (2002). *A critical evaluation of extended Kalman filtering and moving horizon estimation.* TWMCC Tech. Rep. 2002-03. Department of Chemical Engineering, University of Wisconsin-Madison.

Ito, K. and K. Xiong. (2000). Gaussian Filters for Nonlinear Filtering Problems. *IEEE Transactions on Automatic Control*, **45**, 5, 910-927.

Julier, S., J. Uhlmann and H.F. Durrant-Whyte (2000). A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators. *IEEE Transactions on Automatic Control*, **45**, 477-482.

Julier, S.J. and J.K. Uhlmann (2004). Unscented Filtering and Nonlinear Estimation. *Proceedings of the IEEE*, **92**, 3, 401-422.

Nørgaard, M., N.K. Paulsen and O. Ravn (2000). New developments in state estimation for nonlinear systems. *Automatica*, **36**, 1627-1638.

Rawlings, J.B. and B.R. Bakshi (2006). Particle filtering and moving horizon estimation. *Computers and chemical engineering*, **30**, 1529-1541.

Schei, T.S, H. Ludvigsen, P. Singstad and B. Foss (2001). Operator Support System for Optimization of Suspension PVC Batch Polymerization Reactors. *Proc. European Control Conference*, Porto, Portugal.

Schei, T.S. (1995). A finite-difference approach to linearization in nonlinear estimation algorithms. *Proc. American Control Conference*, Seattle, Washington, **1**, 114-118.

Schei, T.S. (1997). A finite-difference method for linearization in nonlinear estimation algorithms. *Automatica*, **33**, 2051-2058.

Rao, C.V., J.B. Rawlings and J.H. Lee (2001). Constrained linear state estimation – a moving horizon approach. *Automatica*, **37**(10), 1619–1628.

Rao, C.V., J.B. Rawlings and D.Q. Mayne (2003). Constrained state estimation for nonlinear discrete-time systems: Stability and moving horizon approximations. *IEEE Transactions on Automatic Control*, **48**(2), 246–258.