

Moving Horizon Parameter Estimation using Qualitative Knowledge and Multiple Linear Models

Edward P.Gatzke, Francis J. Doyle III

Department of Chemical Engineering
University of Delaware
fdoyle@udel.edu

Keywords: Parameter Estimation, Multiple Linear Models, Fault Diagnosis, Moving Horizon, Mixed Integer Linear Programming

Prepared for Presentation at the November 1998 National Meeting, Miami
Copyright 1998, University of Delaware
Unpublished

AIChE shall not be responsible for statements or opinions contained in papers or in its publications.

Abstract

Details and results of a combined moving-horizon fault detection/isolation and parameter estimation technique will be presented. Parameters are treated as disturbances that affect the plant model. The estimation technique performs a 1-norm minimization of plant model measurement mismatch over a finite horizon by adjusting parameter values over the same horizon. To implement process knowledge, this minimization is constrained such that only S different faults (parameters) may change during a specific horizon window. Model estimates and parameter values for the entire horizon window are calculated at each iteration. Multiple linear models are used to capture nonlinear process characteristics such as asymmetric response, changing dynamics and varied gains.

1 Introduction

Modern advanced control and optimization methods increasingly require detailed information about the target system. Parameter estimation techniques can provide a portion of this needed information. However, linear estimation techniques can become inadequate in cases involving nonlinear systems. On the other hand, nonlinear estimation techniques can quickly become intractable for problems of moderate size. Realistic systems may have many unmeasured parameter variations (unmeasured disturbances) which can and should be estimated for use in control strategies. Such unmeasured disturbances can include system faults. A fault may not be a catastrophic event, but rather a change in the system that adversely limits or modifies the system properties. A method for detecting faults and estimating parameter values is desirable.

This paper presents a method for obtaining parameter estimates, given process measurements and adequate system models. This new method uses a multiple linear models and Mixed Integer Linear Programming (MILP) to formulate and solve a Moving Horizon (MH) optimization problem. A Continuous Stirred Tank Reactor (CSTR) case study is presented, along with results.

1.1 Related Approaches

There are many different methods for attempting fault diagnosis, parameter estimation, and nonlinear modeling. Presented is a partial listing of some relevant work.

Fault detection and isolation has become an important topic for chemical engineers attempting to minimize process down time and industrial accidents. Rule-based diagnosis and root cause analysis has been presented in numerous sources [5]. A method for diagnosis based on digraph representations has also had some success [11][7]. Many fault detection and isolation methods are adequate for dealing with faults on a qualitative level, but in many cases one desires a more useful estimate of the current system state.

Parameter estimation can be treated as a state estimation problem. Kalman filtering has been used extensively to solve linear estimation problems. Nonlinear estimation techniques have also been applied. The Extended Kalman Filter (EKF) has met with much success [12][15]. Moving horizon methods can produce results similar to those of Kalman Filtering. An overview of nonlinear estimation is found in [10]. One should also consider the similarity of moving horizon estimation methods to moving horizon optimal control methods (such as MPC Control and QDMC).

Multiple model approaches hold great interest and promise because of the ability to represent a complex nonlinear system using established and tractable techniques. Typically, the complex nonlinear system is partitioned into regions where a local model is assumed to accurately represent the system. Many different methods for partitioning model spaces and switching between models have been proposed. See [9] for general discussion of multiple models. Fuzzy logic model selection [13] and Gaussian model selection [8] have both been used as multiple model methodologies. A multiple model moving horizon approach was proposed in [2]. A multiple model approximation has been applied to fault diagnosis in [1].

Fault diagnosis usually infers the use of qualitative rules. Parameter estimation can be treated as a moving horizon optimization problem. The combination of estimation and diagnosis results in a problem that can be formulated and solved using Mixed Integer (MI) optimization methods. A Mixed Integer Quadratic Programming (MIQP) formulation has been proposed in [14] and [3]. This type of formulation expresses qualitative rules about a system as constraints involving integer variables. Previous work has demonstrated the usefulness of the method on small scale linear systems.

1.2 Proposed Approach Advantages and Limitations

The proposed method uses a Mixed Integer Linear Programming (MILP) formulation. This formulation minimizes either the total or maximum absolute error between the system measurements and the model predictions. Many estimation techniques use a Quadratic Problem formulation which minimizes the 2-norm of the measurement-model error. The LP formulation allows for real-time solution of realistic problems, while still giving accurate results. The multiple model formulation empowers the estimation method with the ability to represent typical system nonlinearities, such as asymmetric response, changing system gains, and variable system dynamics. This results in improved parameter estimates over linear methods. A principle drawback of the multiple model approach is that it requires accurate system models and increases the size of the optimization problem.

2 Methodology

This section will describe how to set up and solve the horizon based estimation problem.

2.1 Formulation

The current formulation solves the problem:

$$\min_{i = k - h \dots k} ||m_1(y(i) - \hat{y}(i))||_1 + ||m_2\Delta\Theta(i)||_1$$

where k is the current time, y_i is the process measurement vector at time i , $\hat{y}(i)$ is the vector of process model estimates at time i , m_1 is a scaling vector for weighting or normalizing the measurements, m_2 is a scaling vector for weighting and normalizing the change in the parameter estimate, h is the length of the moving horizon, Θ_i is the vector of parameter estimate at time i , and $\Delta\Theta_i$ is given by:

$$\Delta\Theta(i) = \Theta(i) - \Theta(i - 1)$$

Θ_{i-h} is the value of the parameter estimate θ_{i-h+1} from the previous horizon window. An impulse response formulation is used to calculate the response of the process model estimate, $\hat{y}(i)$ to changes in the system parameters, Θ . An individual system parameter (or fault) may be described at time i as $\Theta_j(i)$. This individual parameter estimate is the sum of the contributions from each model for that parameter:

$$\Theta_j(i) = \theta_{j,1}(i) + \dots + \theta_{j,n}(i)$$

where n models are used for a certain parameter variation. Mixing coefficients are not used to select models based on operating regime. The models are selected to best fit the available data, given the existing constraints. The impulse response formulation for $\hat{y}(i)$ is:

$$\hat{y}(i) = \sum_{j=1}^F M_{j,1}\theta_{j,1}(i) + \dots + M_{j,n}\theta_{j,n}(i)$$

where $M_{j,n}$ are the impulse response coefficients for model $M_{j,n}$. F is the total number of disturbances modeled in the formulation. The index j represents the number of possible faults (or disturbances). The index n represents the number of models used for a single fault or disturbance.

In order to handle asymmetric response to a parameter change, at least two models are used. The values for all $\theta_{j,n}(i)$ (and therefore all $\Theta_j(i)$) are constrained to be positive values,

≥ 0 . This means that for negative changes of a parameter, the formulation actually uses the negative of the impulse response matrix. This type of formulation assume that large positive and negative parameter changes do not occur in a parameter over a single window length.

The formulation to this point only includes continuous variables and few constraints. Solving this formulation without additional constraints typically yields an underspecified problem that matches the measurement values with the estimated measurement values exactly. At this point, one may make the assumption that only a limited number of disturbances can affect the system during a single horizon. This leads one to use binary decision variables to represent whether or not a fault has occurred in the current horizon window. The following constraints are added:

$$\theta_{j,1}(i) + \dots + \theta_{j,n}(i) \leq M f_j \quad \forall i, \forall j$$

$$\sum_{j=1}^F f_j \leq S$$

for all fault parameters j , fault models $1..n$, and all horizon indices (times) i . The value M is a large number that ensures whenever $\theta_{j,n}(i)$ is nonzero, f_j switches from 0 to 1. S is the total number of faults that can occur in a horizon window. In the presented results, response to positive and negative changes in a parameter are treated as separate fault events.

The size of the basic problem can now be calculated. Assuming that there are N models for each fault, the horizon length is h , and there are O system outputs, in the one-norm case, the total number of variables is:

$$F + hFN + hO + hF$$

where F is the number of binary fault variables, hFN are the number of variables for the parameter values $\theta_{j,n}$, hO are the number of variables used to calculate the one-norm of the measurement-model error, and hF is then number of variable used to calculate the one-norm of the change in parameters over the horizon.

2.2 Solution Method

Solving a large scale MILP problem can be computationally intensive. The use of multiple models only makes the problem larger. For a moderately sized problem, there may be thousands of variables to solve. Even with increased computational power, the task seems impossible in real time. Making modifications to the MILP solution strategy allows one to make dramatic speed increases in the solution time.

In general, when $S = 1$ (the one fault case), the solution for this formulation is trivial. There are only F possible combinations of integer variables available. One may directly solve the large LP formulation using an MILP, but having the knowledge that there are only F possible cases allows one to find a more direct solution. In each possible case, only the variable for the given fault are nonzero; the variable for all the other faults are constrained to zero. Setting up and solving F different LP problems for a single fault is computationally trivial. The LP problem with the minimal objective function will be the overall MILP solution.

In the multiple fault case this becomes less trivial. There are many combinations of faults to explore and enumeration quickly becomes impossible. A standard MILP solution method will evaluate a large LP problem at each node in the branch and bound tree. Intelligently using branch and bound can serve to speed the computational process.

The proposed solution method is outlined as follows-

- Solve small single fault problem for each fault
- Order the best objective function fits from the F possible solutions
- Evaluate MILP by modified branch and bound:
 - reduce problem when possible
 - skip nodes if applicable
 - enumerate cases if useful

The MILP can start at a node other than the root node, root node being the total LP relaxation. For example, in a 2 fault case ($S = 2$), assume that solving the single fault problem for each possible fault results in the f_1 having the best fit, f_2 having the next best, and so on to f_F . This requires F solutions to small (size 1) problems. Now add the constraints that $f_1 = 1$ and $f_2 = 1$. If $f_1 = 1$ and $f_2 = 1$ all other faults are constrained to 0 in the $\sum f_i \leq S = 2$ example. Therefore one may solve this node as a moderately small problem (size 2). The objective function of this solution becomes the new best integer solution for the problem, which is also an upper bound on the overall MILP problem. The entire enumeration of possible integer solutions with $f_1 = 1$ numbers only F . These cases can all be quickly evaluated as moderately small problems. The best solution of these cases now becomes the best integer solution. Using the branch and bound technique, the node corresponding to $f_1 = 0$, all other f_i unconstrained must be evaluated. If the objective function of this node is larger than the integer bound, the solution procedure can stop with the current best integer solution as optimal. The solution of this node is a large LP problem,

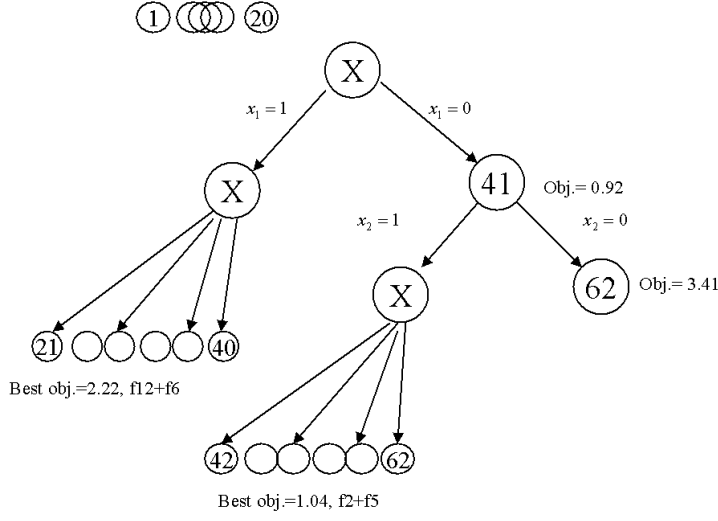


Figure 1: Modified branch and bound example for $S=20$

taking much longer than the smaller problems. The solution LP solution at this node can be warm started with a feasible starting point using the results from the original F small optimization problems of size 1. If the value of the objective function at this node is less than the optimal integer solution, the true optimal solution must lie in the $f_1 = 0$ branch. There are now $F - 1$ possible integer solution size 2 problems with the constraints $f_1 = 0$ and $f_2 = 1$. These moderately small problems can be quickly enumerated. A new best integer solution may be found. The node with constraints $f_1 = 0$, $f_2 = 0$, all other f_i unconstrained must be solved. Again this node is a large LP problem, taking much longer than the smaller problems. The solution LP solution at this node can also be warm started with a feasible starting point using the results from the original F small optimization problems of size 1. If the value of the objective function at this node is greater than the optimal integer solution, the true optimal integer solution has already been found.

The procedure can be repeated as need until a valid optimal integer solution is found. Ultimately, this may in some cases require that all possible fault combination be evaluated. It is expected that these cases will be rare in practice. This modified solution method intelligently modifies the MILP solution engine such that large scale problems need not be evaluated. By solving smaller problems corresponding to single fault cases and exploiting the known constraints of the problem, a good integer solution can be found early in the

computation. By exploiting the structure of the problem, the LP nodes for computation can be decreased in problem size, greatly increasing computational efficiency.

2.3 Formulation Issues

For illustration purposes, consider the following simple example with the forcing function d representing an exogenous disturbance:

$$\frac{dx}{dt} = -x - (d - 1)^2, \quad x_o = -1$$

$$y = x + 1$$

It is desirable to develop models the system using step tests. The response to disturbance steps at time 0 from $d = 0$ to $d = 1$ and $d = 0$ to $d = 1.8$ is given by:

t	$y, \Delta d = 1$	$y, \Delta d = 1.8$
0	0	0
1	0.63	0.23
2	0.86	0.31
3	0.95	0.34
4	0.98	0.35
55	0.99	0.36
6	1	0.36

This system has input multiplicities. The steady state locus is given in Figure 2. The dynamic character of the response to step changes are identical, no matter the input level. The steady state gain for the system is higher for the lower level input than for the high level input. Using these two models in the current estimation formulation will result in biased parameter estimates. The scaled impulse response coefficients are:

t	M_1	M_2
0	0	0
1	0.63	0.13
2	0.23	0.05
3	0.09	0.02
4	0.03	0.01
55	0.01	0
6	0.00	0

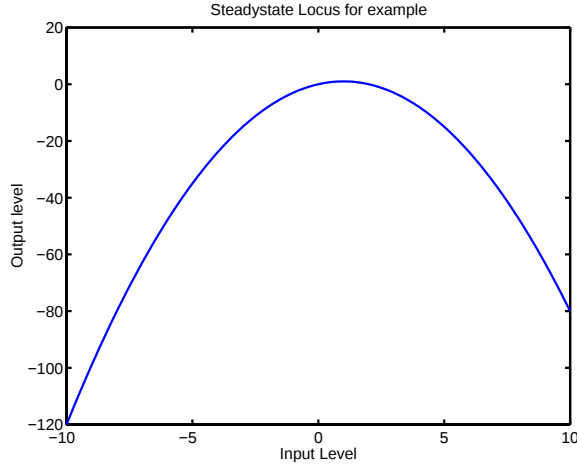


Figure 2: Steadystate locus for example

Now assume that a disturbance hits the real system. The response of the system to a step change from 0 to 1.9 is:

t	$y, \Delta d = 1.9$
0	0
1	0.32
2	0.44
3	0.48
4	0.50
55	0.51
6	0.51

The best fit for this measurement data (assuming that $\Theta(0) = 0$) would be to use only the first model to fit the data. The resulting parameter estimate would be 0.51 for the entire horizon length.

This example shows that biased estimates are possible in some cases. If the local models accurately capture dynamic differences, the parameter estimates can be much more accurate.

In another case, assume that only one model is used for a single fault. Then the impulse response for a single output system is:

$$y = M_1\Theta_1 + M_2\Theta_2 + \dots M_F\Theta_F$$

In some cases, the output response to a parameter change may also be formed from the combination of other faults. This means that a nonzero vector x can be found such that for

a given disturbance model M_i ,

$$M_i = [M_1 \quad \dots \quad M_j \quad \dots \quad M_F]x, \quad j \neq i$$

This can be true when the constraint for maximum number of faults is not used in the estimation formulation ($S = F$). In this case, when allowing output biases as faults, any disturbance response could be created by the linear combination of the output biases. In effect this system is unobservable. Depending on the weighting matrices m_1 and m_2 , either the actual disturbance or a linear combination of the output bias models or a combination of both will be given as the parameter estimate. Setting S to a value close to 1 rather than F allows the optimization to correctly identify the disturbance.

2.4 Degenerate Solution Test

In some cases, the measured response to a fault may resemble the response of another fault or combination of faults. It is useful to know which faults are distinguishable from other faults or fault sets. If one assumes that all measurements may potentially be biased, the measurement response to a fault may be represented as either a change in the fault parameter or a combination of changes in the measurement biases.

In the current formulation, parameters are assumed to remain constant over the horizon. Only S different parameters may change over an optimization window, and these parameters are constrained to be positive values. In this way, asymmetric response can be captured in the linear models. The traditional methods of checking the rank of the observability matrix fail in these conditions because the parameters are limited to positive values. A method for finding degenerate fault sets is desirable.

First, assume that M_i is the impulse response matrix model for fault i . In the entire system there may be F total faults, $M_i \in [M_1 \dots M_F]$. Given a set of N models, we wish to find out which models in the set $M_1 \dots M_F$ may be used to represent the subset N models, $M_i, i \in N$. For notational convenience, we define the subset in question as the first N models of the total set, $[M_1 \dots M_N]$ of $[M_1 \dots M_F]$. Defining an integer variable f_i which represents the boolean value of whether or not a model set is required to represent the subset, the MILP problem can be formulated as:

$$\min_{[x,y]} \sum_{i=1}^F f_i$$

subject to:

$$\begin{aligned}
[M_1 \dots M_N]x - [M_1 \dots M_N \dots M_F][y_1 \dots y_N \dots y_F]^T &= 0 \\
\sum y_i - V f_i &\leq 0, \forall i \\
\sum x &\geq B \\
\sum_{i=1}^N f_i &\leq N - 1 \\
0 &\leq x \leq U \\
0 &\leq y \leq U \\
f_i &\in [0, 1]
\end{aligned}$$

where the vectors x and y_i are used to find any set of parameters that will produce similar output values across the horizon. The first constraint is used to force x and y to use the models to find an equivalent output value across the horizon. The second constraint (with V any suitable large value) forces f_i to become 1 when any element of y_i is nonzero. The third constraint forces the elements of x to be nonzero. The fourth constraint forces the set of y_i found to represent the subset to not be the subset itself. The last three constraints ensure that the elements of x and y are constrained to be positive numbers and all f_i are binary.

The result of this problem is the minimal set of faults (set $i \in J$) that could possibly be used in an optimization problem to represent the given subset of models. The problem can be re-solved with the additional constraint:

$$\sum f_i \leq J - 1$$

The new problem finds the minimal set of faults to represent the subset of models that is not the original set or the new found set. This iterative procedure can be used to find all of the potential degenerate model sets for a set of models by solving the MILP and adding the new constraint, and then resolving. Eventually, too many constraints will be added and no new set of degenerate models can be found.

Due to the total number of fault constraints and weightings in the moving horizon formulation, this does not mean that the degenerate sets are likely to occur when used in the moving horizon optimization process.

3 Case Study

A description of the Van Der Vusse CSTR model may be found in [4] or [6]. This model simulates a first-order reaction for $A \Rightarrow B$ with two competing reactions $B \Rightarrow C$ and $2A \Rightarrow D$.

$$\begin{array}{ll}
k_1 o = 1.287 \cdot 10^{12} h^{-1} & \Delta H_{RBD} = -41.85 \frac{kJ}{mol A} \\
k_2 o = 1.287 \cdot 10^{12} h^{-1} & \rho = 0.9342 \cdot 10^{-4} \frac{kg}{l} \\
k_3 o = 1.287 \cdot 10^{12} h^{-1} & C_P = 3.01 \frac{kJ}{kg K} \\
E_1 = -9758.3 K & C_{PK} = 2.0 \frac{kJ}{kg K} \\
E_2 = -9758.3 K & k_w = 4032 \frac{kJ}{h m^2 K} \\
E_3 = -8560 K & A_R = 0.215 m^2 \\
\Delta H_{RAB} = 4.2 \frac{kJ}{mol A} & m_k = 5.0 kg \\
\Delta H_{RBC} = -11 \frac{kJ}{mol A} & V_R = 0.1 m^3
\end{array}$$

Table 1: Model parameters

$$\begin{aligned}
\frac{dC_A}{dt} &= \frac{\dot{V}}{V_R} (C_{AO} - C_A) - k_1(v) C_A - k_3(v) C_A^2 \\
\frac{dC_B}{dt} &= -\frac{\dot{V}}{V_R} C_B + k_1(v) C_A - k_2(v) C_B \\
\frac{dv}{dt} &= \frac{\dot{V}}{V_R} (v_O - v) - \frac{(k_1(v) C_A \Delta H_{RAB} + k_2(v) C_B \Delta H_{RBC} + k_3(v) C_A^2 \Delta H_{RAD})}{\rho C_p} + \frac{k_w A_R}{\rho C_p V_R} (v - v_K) \\
\frac{dv_K}{dt} &= \frac{1}{m_K C_{PK}} (\dot{Q}_K + k_w A_R (v - v_K)) \\
k_i &= k_{io} e^{\left(\frac{E_i}{v+273.15}\right)} \\
\dot{Q}_k &= F_K C_{PK} (v_{ko} - v_k)
\end{aligned}$$

All reaction models use temperature dependent Arrhenius reaction rates. The model has four states: concentration of A, concentration of B, reactor temperature, and cooling jacket temperature. With this process, it is desirable to produce as much product B as possible. This system exhibits many highly nonlinear characteristics which include: input multiplicity, gain sign change, asymmetric response, and minimum to nonminimum phase behavior. These nonlinear characteristics are very prevalent at the most desirable operating point. See Figure 3 for a comparison of dynamic response to a step change in the feed flow rate. From this figure, you can see that inverse response is observed for changing to a low flow rate and not observed for changes to higher flow rates. As seen from the process values at steady state, the process gain changes from positive to negative causing input multiplicity. The optimal operating point is at the convergence of changing gains and inverse to non-inverse response.

The process model equations, parameters, and operating conditions may be found in Appendix A. All simulation results were performed using MATLAB 5.1.

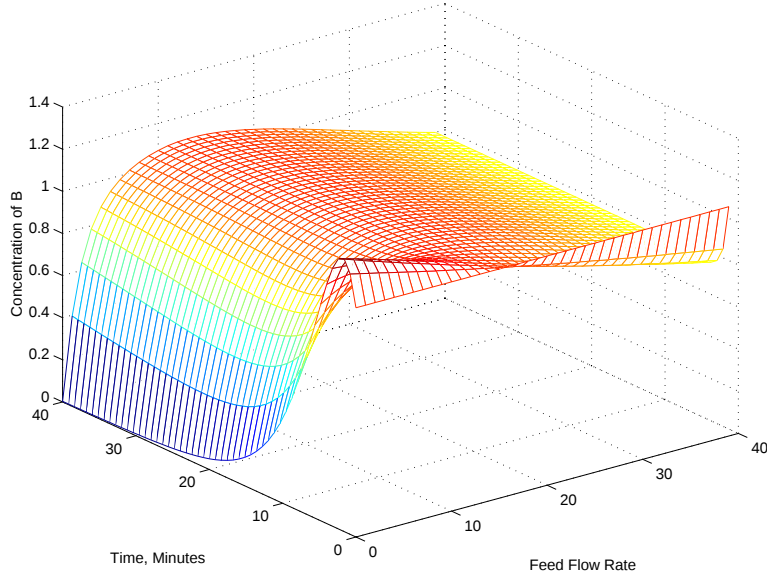


Figure 3: Response of model to step changes at time 0, initial feed flow = 14.9

4 Results

Results are presented using the Van Der Vusse case study. It is assumed that the four states of the CSTR are directly measurable. An initial steady state point was selected near the process optima. Step response models were developed for ten different parameter variations. The step response models were evaluated for $\pm 5\%$ and $\pm 25\%$ for a total of 20 different models. A moving horizon window of size 25 was chosen to account for system dynamics.

The systems accurately distinguishes between parameters in the single fault case ($S = 1$). All twenty faults can be distinguished and accurately estimated. Given a change in a parameter the, estimation method can estimate the parameter value very closely. For example take the set of measurements in Figure 8. These measurements are taken from the Van Der Vusse reactor model after a 10% increase in the feed flowrate at time = 5 minutes. In Figure 5 the results for estimating this 10% increase in the feed flow are shown. The 10% increase hit the system 20 minutes before this estimation snapshot was computed. The two models used to estimate the total parameter are shown. As one can see in Figure 5, both the 5% model and the 25% model are used when estimating a 10% disturbance. The

Table 2: Potential faults for CSTR case study

f_2	increase in feed flow rate	f_2	decrease in feed flow rate
f_3	increase in coolant flow rate	f_4	decrease in coolant flow rate
f_5	increase in feed temperature	f_6	decrease in feed temperature
f_7	increase in coolant temperature	f_8	decrease in coolant temperature
f_9	increase in feed concnetration	f_{10}	decrease in feed concentration
f_{11}	increase in jacket heat transfer coefficient	f_{12}	decrease in jacket heat transfer coefficient
f_{13}	increase in C_a measurement bias	f_{14}	decrease in C_a measurement bias
f_{15}	increase in C_b measurement bias	f_{16}	decrease in C_b measurement bias
f_{17}	increase in v measurement bias	f_{18}	decrease in v measurement bias
f_{19}	increase in v_k measurement bias	f_{20}	decrease in v_k measurement bias

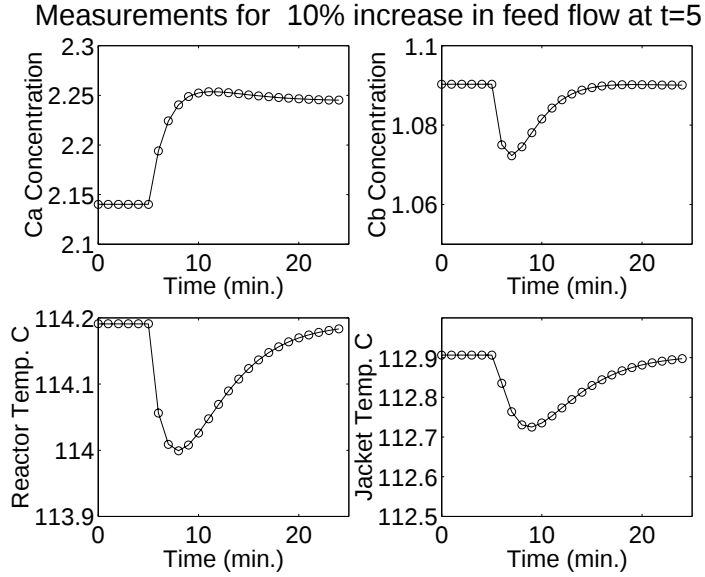


Figure 4: Measurements for 10% increase in feed flow at time = 5

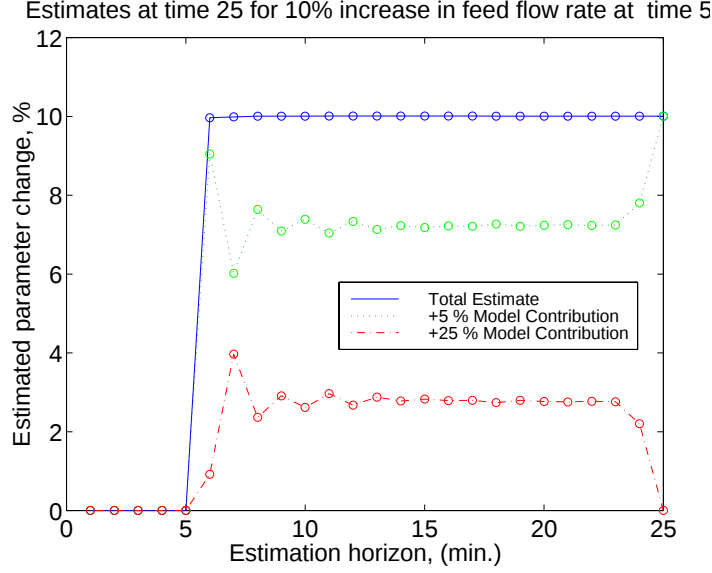


Figure 5: Estimates using multiple models

two model estimates can be summed directly because the models were normalized before the optimization took place. In Figure 6 the estimation results for using single models are shown. As can be seen, the estimates are not as accurate as the multiple model case. The correct fault is distinguished from the others, but the estimates are slightly off when not using the multiple linear models.

The evolution of a fault over time is also of interest. This estimation method estimates the value of the parameters over the entire horizon, not just the current estimate or the initial value estimate. In Figure 7 the actual input values over the estimation horizon for different times are shown. As time marches forward, new measurements are available and the estimation process is accomplished again.

In Figure 8 the estimates over the moving horizon for different times are displayed. In After the fault is fully evolved, the amount of error in the estimate becomes larger. At this point, the change in the measurements is becoming negligible, so any model error can lead to increased estimation error.

The estimation method works well in multiple fault scenarios as well. In the following example, the input feed flow rate is decreased by 20% at time $t-25$ and the feed concentration is decreased by 10% at time $t-15$. The output measurements for this parameter change are

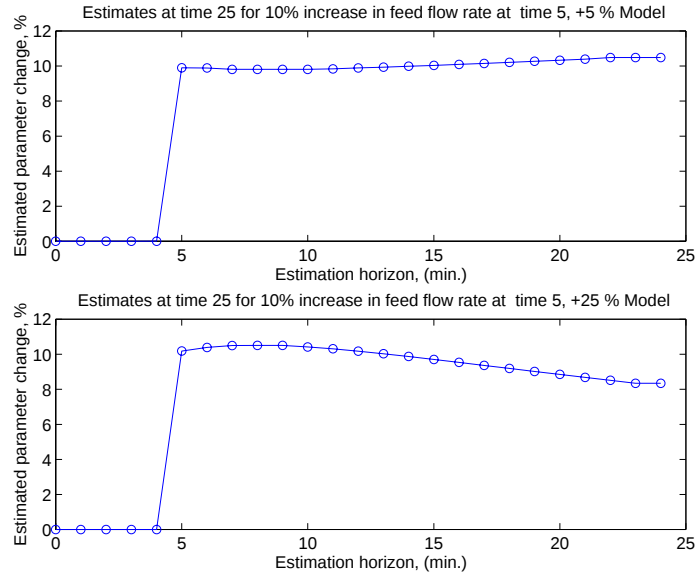


Figure 6: Estimates using single models

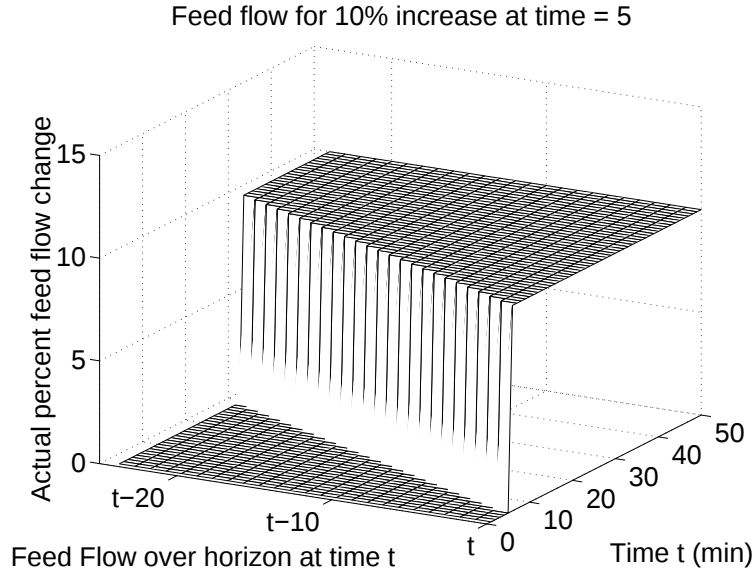


Figure 7: Actual parameter values over horizon for 10% increase in feed flow

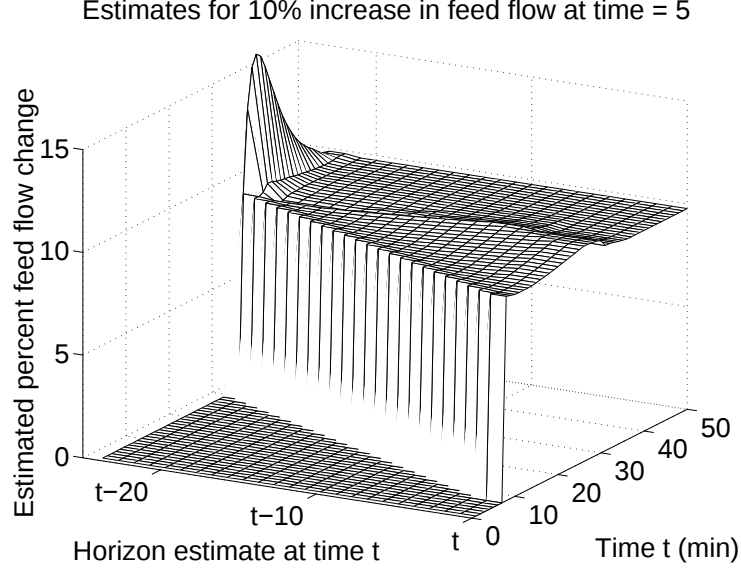


Figure 8: Horizon parameter estimates for 10% increase in feed flow

given in Figure 9. The overall parameter estimates for time $t = 25$ are shown in Figure 10. Here, the parameter estimates accurately picked the correct faults from the group of 20 and also accurately estimated the parameter values.

4.0.1 Potential other results:

1. S=F case, no int, uses all faults to perfectly fit data (can do now)
2. Accurate in detecting correctly all faults (can do now) Graph? 20 plots of estimates for all 20 faults?
3. Accurate in estimating ramped fault level (can do now)
4. Accurate in presence of significant noise (can do now)

5 Conclusions

In this paper, we have developed and used a moving horizon method for detecting and estimating parameter changes. Using a moving horizon allows the use of a finite amount

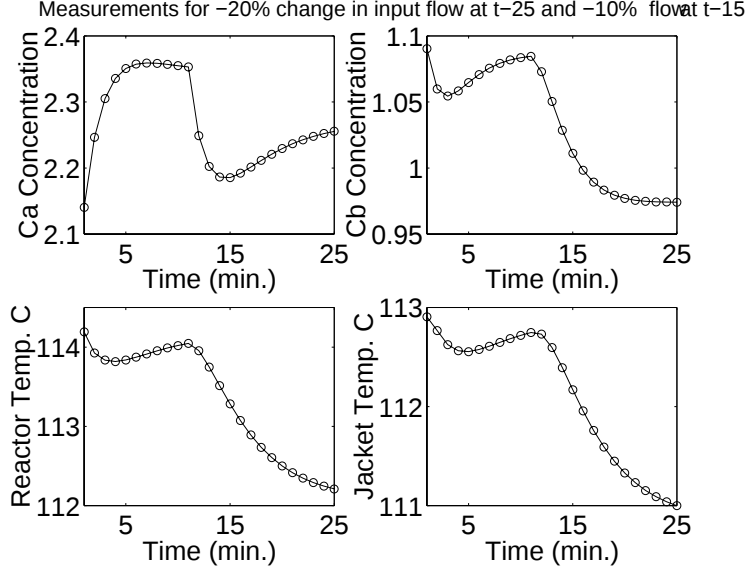


Figure 9: Measurements for dual fault case, -20% input feed flow step at t=25 and -10% feed concentration step at t=15

of data in the optimization problem. This method makes use of multiple linear models to capture the true behavior of the actual nonlinear system. The problem can be formulated as a MILP with an integer constraint on the total number of changing parameters over the horizon.

There are limitations to this method, some of which have been espoused in this article. Using multiple linear models can lead to biased estimates in some cases. This will not always occur, but can be seen in some mathematical examples. The results from one problem produce a potential fault set. This set may not be the only set of faults that can accurately portray the existing data. This means that multiple model output values may exist. Due to optimization weights and a single optimal solution, it should be rare but possible to find different sets of faults yielding the same or similar horizon based estimation results. Computationally, solving a large scale MILP can be tedious. For multiple fault cases, enumerating all possible cases is typically a poor solution. An improved MILP solution method is proposed that can exploit the problem structure and existing information.

In the future, even larger problems may be explored. In the proposed method, separate models for positive and negative parameter shifts as well as small and large parameter shifts were used. Use of traditional linear models can allow for solution of larger scale problems

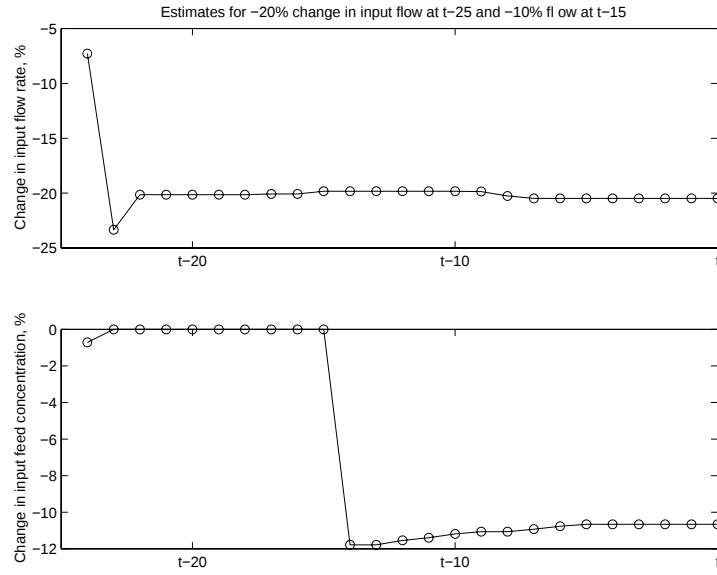


Figure 10: Estimates for dual fault case -20% input feed flow step at t-25 and -10% feed concentration step at t-15

(at least 4x), but the system estimates should be less accurate. This may be a valid tradeoff in systems where only linear models are available.

References

- [1] P. Balle, D. Fussel, and O. Hecker. Detection and Isolation of Sensor Faults on Nonlinear Processes Based on Local Linear Models. In *Proceedings of the American Control Conference*, pages 468–472, 1997.
- [2] A. B. Banerjee, Y. Arkun, B. Ogunnaike, and R. Pearson. Estimation of Nonlinear Systems Using Linear Multiple Models. *AIChE J.*, 43(5):1204–1226, 1997.
- [3] A. Bemporad and M. Morari. Control of systems integrating logic, dynamics, and constraints. Technical report, Institut für Automatik, ETH, 1997.
- [4] H. Chen, A. Kremling, and F. Allgöwer. Nonlinear Predictive Control of a Benchmark CSTR. In *Proc. of the European Control Conf.*, 1995.
- [5] L. W. Chen and M. Modarres. Hierarchical decision process for fault administration. *Computers in Chemical Engineering*, 16(5):425–448, 1992.
- [6] S. Engell and K.-U. Klatt. Nonlinear Control of a Non-Minimum-Phase CSTR. In *Proc. American Control Conf.*, pages 2941–2945, San Francisco, CA, 1993.
- [7] M. Iri, K. Aoki, E. O’Shima, and H. Matsuyama. An algorithm for diagnosis of system failures in the chemical process. *Computers and Chemical Engineering*, 3:489–493, 1979.
- [8] T. A. Johansen and B. A. Foss. Constructing NARMAX models using ARMAX models. *Int. J. Control*, 58(5):1125–1153, 1993.
- [9] R. Murray-Smith and T. A. Johansen, editors. *Multiple Model Approaches to Modelling and Control*. Taylor and Francis, 1997.
- [10] K.R. Muske and T.F. Edgar. *Nonlinear Process Control*, chapter 6. Nonlinear State Estimation, pages 311–370. Prentice Hall, 1997.
- [11] D. S. Nam, C. Han, C. Jeong, and E. S. Yoon. Automatic construction of extended symptom-fault associations from the signed digraph. *Computers in Chemical Engineering*, 20:5605–5610, 1996.
- [12] D. Robertson and J. H. Lee. Integrated state estimation, fault detection, and diagnosis for nonlinear systems. *Proc. American Control Conf.*, 1993.

- [13] A. Rueda. Approximation of Nonlinear Systems by Dynamic Selection of Linear Models. In *IEEE Canadian Conference on Electrical and Computer Engineering*, pages 270–273, 1996.
- [14] M. L. Tyler and M. Morari. Qualitative modeling using propositional logic. In *AIChE National Meeting*, 1996.
- [15] K. Watanabe and D. M. Himmelblaus. Incipient fault diagnosis of nonlinear processes with multiple causes of faults. *Chemical Engineering Science*, 39(3):491–508, 1984.