

Towards Control-Relevant Forecasting in Supply Chain Management

Jay D. Schwartz*, Daniel E. Rivera*¹, and Karl G. Kempf†

*Control Systems Engineering Laboratory
Department of Chemical and Materials Engineering
Arizona State University, Tempe, Arizona 85287-6006

†Decision Technologies
Intel Corporation, 5000 W. Chandler Blvd., Chandler, Arizona 85226

Abstract The focus of this paper is understanding the effects of demand forecast error on a tactical decision policy for a single node of a manufacturing supply chain. The demand forecast is treated as an external measured disturbance in a multi-degree-of-freedom feedback-feedforward Internal Model Control (IMC) based inventory control system. Because forecast error will be multifrequency in nature, the effect of error in different frequency regimes is examined. A mathematical framework for evaluating the effect of forecast revisions in an IMC controller is developed. A Simultaneous Perturbation Stochastic Approximation (SPSA) optimization algorithm is implemented to develop an optimal tuning strategy under these conditions. For the IMC-based inventory controller presented it is concluded that the most desirable performance may be obtained by acting cautiously (e.g. implementing small changes to factory starts) to initial forecasts and gradually becoming more aggressive on starts until the actual demand change is realized.

I. INTRODUCTION

A supply chain is composed of the (interconnected) components necessary to transform raw materials into a salable product. Maintaining an efficient supply chain is necessary for a company's financial survival. In many instances, such as in the semiconductor manufacturing arena, supply chain performance will suffer from both long production times and demand forecast errors [4].

Control-oriented approaches have been recently proposed to deal with the problems inherent in manufacturing supply chains [2][4][8]. In these approaches, the demand is treated as an exogenous signal. This paper extends these approaches by attempting to gain a broader understanding of disturbance/demand modeling and the effects of forecast error on an Internal Model Control (IMC) [6] based tactical decision policy.

The presence of error in a demand forecast will adversely affect decision-making in a supply chain. Inaccurate market research, order changes, outdated demand models, and misforecasting of product/business cycles may have a negative effect on the profitability of a supply chain dependent

corporation. Eliminating error from a demand forecast may be impossible; however, it may be possible to mitigate its detrimental effects. Therefore, it is important to understand the effects of error on a supply chain decision policy.

The relationship between demand forecast error and deviation from an inventory setpoint is studied for a control-oriented tactical decision policy in a single node of the manufacturing process. An understanding of this relationship represents one step towards a fundamental understanding that will allow manufacturing planners to deal with inherently erroneous forecasts in an educated manner.

This paper begins with a discussion of the modeling of an inventory position using a fluid analogy. In Section II, a model-based inventory controller relying on Internal Model Control is presented. The multi-degree-of-freedom formulation allows the controller to be independently tuned for setpoint changes, forecasted demand changes, and unforecasted demand changes. In Section III, the closed-loop transfer functions describing forecast error are derived and the effect of erroneous forecasts is studied in both the time and frequency domains. A mathematical framework for the implementation of forecast signal revisions within the IMC decision policy framework is developed. Finally, a Simultaneous Perturbation Stochastic Approximation (SPSA) optimization algorithm is employed to determine optimal controller tuning parameters for a variety of systems.

II. REPRESENTATION OF THE SYSTEM AND CONTROLLER

A. Inventory/Level Control

A single node of a manufacturing supply chain can be modeled using a fluid analogy. The factory is represented as a pipe with a particular throughput time, θ . The inventory is represented as a tank containing fluid. Delivery to customer from inventory is modeled as a pipe with a transportation time of θ_d (Fig. 1). Note that the delivery time (θ_d) is considered to be zero for this and all other examples presented. The dynamics relating fluid level (net stock, $y(t)$) to inlet pipe flux (fab starts, $u(t)$) and outlet pipe flux (customer demand, $d(t)$), composed of the forecasted customer demand, $d_F(t - \theta_F)$,

¹To whom all correspondence should be addressed. phone: (480) 965-9476; fax: (480) 965-0037; e-mail: daniel.rivera@asu.edu

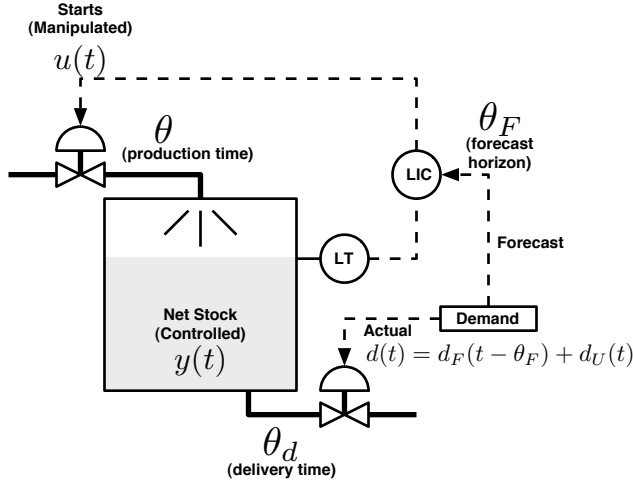


Fig. 1. Fluid analogy and combined feedback/feedforward control scheme.

plus unforecasted customer demand, $d_U(t)$ is represented in (1). Note that θ_F is the forecast horizon.

$$y(s) = p(s)u(s) - p_{d1}(s)p_{d2}(s)d_F(s) - p_{d2}(s)d_U(s) \quad (1)$$

$$y(s) = \frac{e^{-\theta s}}{s}u(s) - e^{-\theta_f s} \left(\frac{1}{s}\right) d_F(s) - \left(\frac{1}{s}\right) d_U(s)$$

It is desirable to meet customer demand and maintain the inventory level at a specified target. This can be accomplished by adjusting the factory starts. A combined feedback/feedforward control scheme will use both the inventory position and a forecast of future product demand to determine quantity of factory starts. This is shown schematically in Fig. 1.

B. Three-Degree-of-Freedom Internal Model Control

A multi-degree-of-freedom Internal Model Control (IMC) [6] structure is considered as the decision policy in this work. With this structure independently tuned controllers can be utilized for setpoint tracking (i.e., meeting an inventory target), measured disturbance rejection (i.e., meeting forecasted demand), and unmeasured disturbance rejection (i.e., satisfying unforecasted demand).

The plant and disturbance models ($p(s)$ and $p_d(s)$, respectively) are represented by integrators with a time delay (the production time θ for the plant model and the forecast horizon θ_F for the disturbance model). Note that $\tilde{p}(s)$ and $\tilde{p}_d(s)$ represent estimates of the plant and disturbance models, respectively. The components of the disturbance model are p_{d1} , the time delay, and p_{d2} , the integrator.

The aforementioned controllers correspond to q_r for setpoint tracking, q_F for measured disturbance rejection, and q_d for unmeasured disturbance rejection. Eqn. 2 is the general form of the closed-loop transfer functions relating net stock to setpoint changes $r(s)$, measured disturbances $d_F(s)$, and unmeasured disturbances $d_U(s)$.

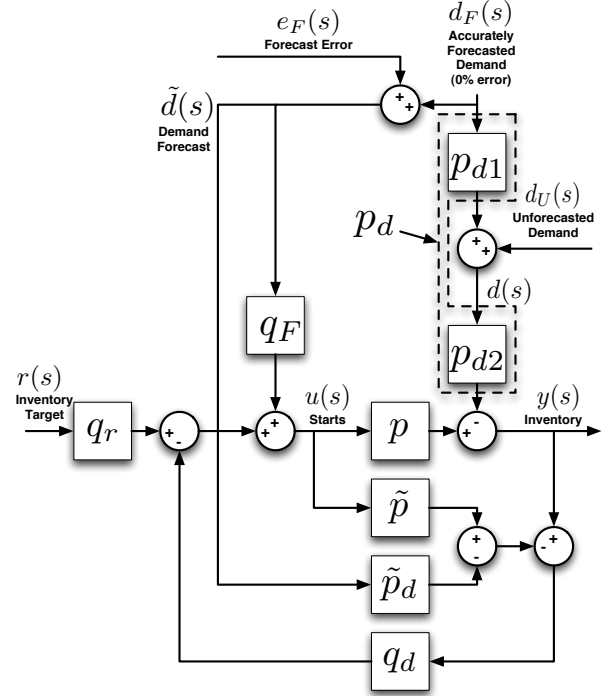


Fig. 2. 3 DoF IMC Structure with forecast error introduced.

$$y(s) = p(s)[q_r(s)r(s) + q_F(s)d_F(s) - q_d(s)(p(s)u(s) + \tilde{p}(s)u(s) - \tilde{p}_d(s)d_F(s))] - \frac{d_U(s)}{s} \quad (2)$$

The controllers shown in (3), (4), and (5) are designed for H_2 -optimality [5] and have been augmented with low-pass IMC filters [6]. The feedforward controller $q_F(s)$, shown in (5), is selected based on the relationship between production time and forecast horizon. The controller implemented if $\theta_F > \theta$ could accomplish perfect control given an exact demand forecast and a value of 0 for the user-adjustable parameter λ_F . The controller implemented if $\theta_F < \theta$ can not accomplish perfect control since the effect of any control action would occur after a demand change had already been realized. Note that for the measured disturbance IMC controller, a filter is not needed for physical realizability when $\theta_f \geq \theta$.

$$q_r(s) = \frac{s}{(\lambda_r s + 1)^{n_r}} \quad \lambda_r \geq 0 \quad n_r \geq 1 \quad (3)$$

$$q_d(s) = \left(s(\theta s + 1) \frac{(n_d \lambda_d s + 1)}{(\lambda_d s + 1)^{n_d}} \right) \quad \lambda_d \geq 0 \quad n_d \geq 3 \quad (4)$$

$$q_F(s) = \begin{cases} \frac{[e^{-(\theta_f - \theta)s}](n_f \lambda_F s + 1)}{(\lambda_F s + 1)^{n_f}} & \text{if } \theta_f \geq \theta \\ \frac{[(\theta - \theta_f)s + 1](n_f \lambda_F s + 1)}{(\lambda_F s + 1)^{n_f}} & \text{if } \theta_f < \theta \end{cases} \quad (5)$$

$$\lambda_F \geq 0 \quad n_f \geq 2$$

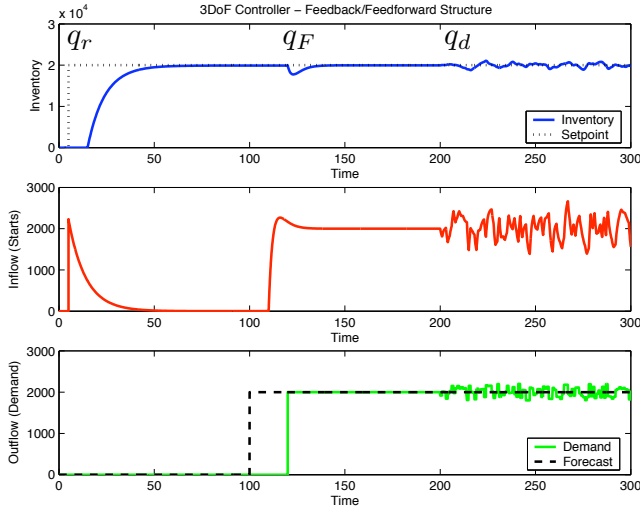


Fig. 3. Response to setpoint change, anticipated demand, and unforecasted demand for the 3 DoF IMC inventory control structure ($\lambda_r = 9, \lambda_F = 3, \lambda_d = 3$).

Fig. 3 shows the closed-loop response to a setpoint change, forecasted demand change, and unforecasted demand variability. A demand forecast is received at $t = 100$. The forecast shows that a positive demand change of 1000 units will occur at $t = 120$. In this example the forecast horizon ($\theta_F = 20$) is greater than the production time ($\theta = 10$). The first feedforward control action can be seen at $t = 110$. The controller compensates for the lost inventory and returns the inventory level to the steady-state target.

C. Time and Frequency Domain Response to Forecast Error

As noted previously, the focus of this paper is the effect of forecast error on the controlled and manipulated variables. An erroneous forecast will be modeled by adding a non-zero signal to the demand forecast signal (see Fig. 2). For the purposes of this example the demand signal will be considered to be constant and the forecast error will be fed directly to the feedforward controller.

Note that the transfer functions relating starts and inventory position to the forecast error are composed of two parts. The first part of each equation is contribution from the feedforward (forecast) controller. The second part of each equation is the contribution from the feedback (inventory excess/backlog) controller.

Eqns. 6 and 7 are the closed loop transfer functions that relate inventory position, $y(s)$, and starts, $u(s)$, to the forecast error, $e_F(s)$. The feedforward controller, $q_F(s)$, is chosen from (5) based on the relationship between the forecast horizon, θ_F , and the production time, θ .

$$\frac{y(s)}{e_F(s)} = q_F(s)p(s) - q_d(s)\tilde{p}_d(s)p(s) \quad (6)$$

$$\frac{u(s)}{e_F(s)} = q_F(s) - q_d(s)\tilde{p}_d(s) \quad (7)$$

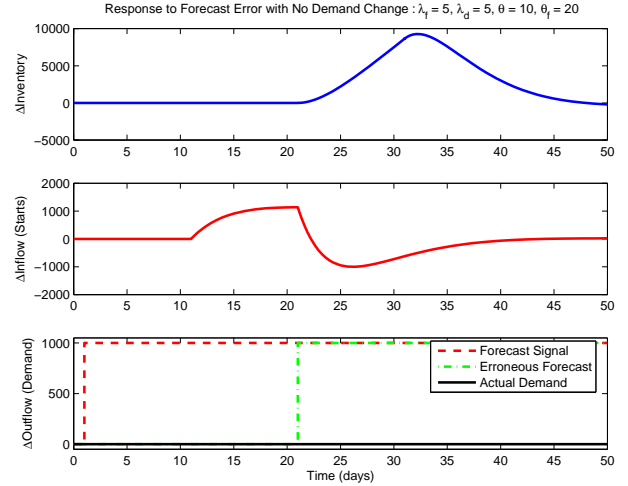


Fig. 4. IMC controller policy response to forecast error of +1000 units for $\lambda_f = 5, \lambda_d = 5$. Note that the forecast signal is received θ_F days before the demand change is supposed to occur ($\theta = 10, \theta_F = 20$).

Given the closed-loop transfer functions that relate inventory and starts to forecast error, the time-domain response to a forecast error signal can be determined. Fig. 4 shows the response to an erroneous forecast of +1000 units for a forecast time (θ_f) of twenty days and a production time (θ) of ten days. The forecast signal is received at $t = 1$. At $t = 11$ the feedforward action is visible, the controller is attempting to compensate for the anticipated demand change at $t = 21$. At $t = 21$ no demand change is realized and the feedback control action is implemented. The controller then decreases starts to return the inventory level to the specified target.

To more broadly understand the effect of tuning, the frequency response of the system will be examined. Fig. 5 shows frequency response for values of the user-adjustable parameters $\lambda_F = 5$ and $\lambda_d = 5$ (Case 1). The notched shape of the amplitude ratio suggests that the effect of forecast error at low and high frequencies will be attenuated, while the effect of forecast error within the notch bandwidth will be amplified. Note that this will affect both starts and inventory level responses.

Fig. 5 also shows the frequency response of the system given different values of the user-adjustable tuning parameters, (Case 2: $\lambda_F = 5$ and $\lambda_d = 1$). Decreasing the value of λ_d makes the feedback controller more aggressive. As a consequence, the feedback controller will have a greater tendency to act on noise or high-frequency components of the forecast error signal. The maximum amplitude ratio shifts to the right (toward increased high-frequency amplification) for both the inventory and starts. In addition, the amplitude ratio for the manipulated variable shifts upward (resulting in increased amplification at all frequencies).

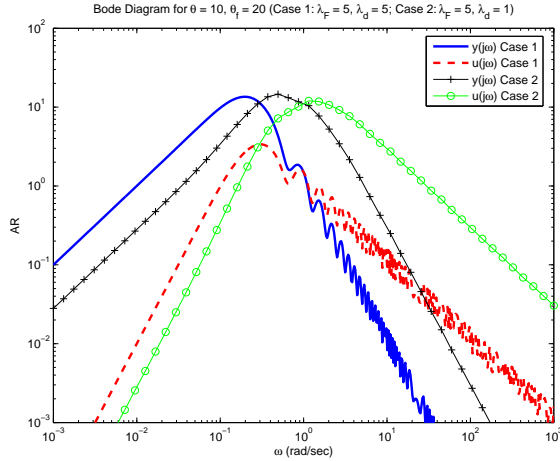


Fig. 5. Amplitude ratio describing the response to forecast error ($\frac{\tilde{d}}{e_F}$ and $\frac{u}{e_F}$) for different tunings. The feedforward controller is tuned identically in both cases ($\lambda_F = 5$). The feedback controller is tuned more aggressively in Case 2 ($\lambda_d = 5$ in Case 1 vs. $\lambda_d = 1$ in Case 2).

III. FORECAST ERROR ANALYSIS UNDER REVISION

A. Closed-Form Transfer Functions For a System Under Forecast Revision

In practice it is desirable to revise a forecast as more accurate estimates of a demand change become available. This will be implemented as an initial forecast followed by a series of revisions. Fig. 6 is a block diagram describing the forecast revision process under a multi-degree of freedom IMC feedback/feedforward structure. This section is concerned with the development of closed-loop transfer functions that relate the forecast error to changes in the starts and inventory.

The following is a list of assumptions and definitions used during the development of a model for implementing forecast revisions. An erroneous demand forecast is received, $\tilde{d} = d_F + e_F$, where d_F is the actual realized demand and e_F is the forecast error. The demand change will occur at day θ_F . The time period from day 0 to day θ_F is the *overall forecast window*. The first forecast ($\gamma_1 = 0$) will become available before the throughput time window is entered. Forecast revisions will be received at day γ_i (within the throughput time window). Each forecast signal will be modeled as $K_i e_F$ where $\sum_{i=1}^N K_i = 1$ given N forecast signals ($N - 1$ revisions). For a specific forecast revision, the expected demand change will occur at day $\Theta_{F,i} + \gamma_i$. The time period $\Theta_{F,i}$ is the *signal specific forecast window*. Each feedforward controller that acts within the throughput time window will use $\theta - \Theta_{F,i}$ in place of the parameter θ in (5) to represent the difference between the throughput time and the time at which a demand change is expected.

Given the assumptions listed above, the closed-loop starts response (8) will be composed of the feedforward (forecast) signals and the feedback (inventory excess/backlog) signal. The closed-form inventory response, (9), is obtained by

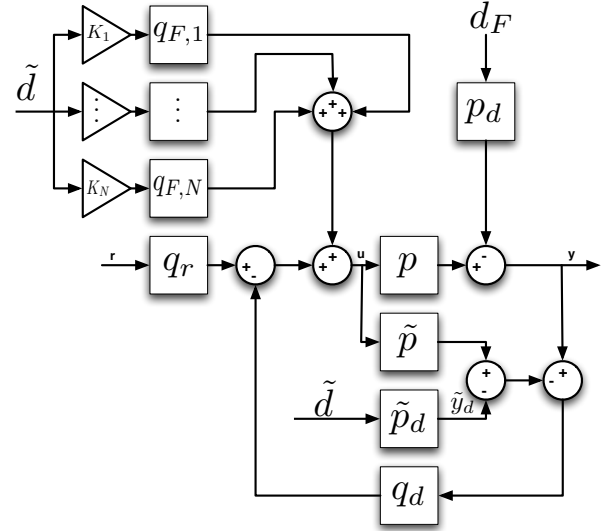


Fig. 6. Three DoF IMC Structure with N forecast signals. Each forecast revision signal is handled via a separate IMC controller. Each controller ($q_{F,i}$) utilizes a separate adjustable parameter (λ_i). This allows for a different response to each forecast revision signal. The initial forecast is received outside of the production window, revisions are received within the production window. The feedback control acts only on forecast error, $e_F = \tilde{d} - d_F$.

multiplying (8) by the plant model, $p(s)$. By implementing the controllers presented in (3) - (5), the transfer functions that relate inventory and starts to a forecast error signal can be developed.

$$\frac{u(s)}{e_F(s)} = \sum_{i=1}^N [K_i e^{-\gamma_i s} q_{F,i}(s)] - q_d(s) \tilde{p}_d(s) \quad (8)$$

$$\frac{y(s)}{e_F(s)} = \left[\sum_{i=1}^N [K_i e^{-\gamma_i s} q_{F,i}(s)] - q_d(s) \tilde{p}_d(s) \right] p(s) \quad (9)$$

B. Monte Carlo Simulation Analysis of Forecast Revisions

Fig. 7 shows the response of the IMC inventory control system to a series of forecast revisions leading to an actual demand change. An initial forecast is received at $t = 0$ for a single demand change of $\sim +1100$ units at $t = 50$, (the actual value of the demand change at $t = 50$ is $+1000$ units). The current forecast is shown by the dashed line in the Δ Demand Forecast plot. The error limits for the current forecast are shown by the dots in the Δ Demand Forecast plot, for a given simulation the forecast at any time will fall within the limits.

Since the first forecast is available outside of the throughput time window ($(\theta_F = 50) > (\theta = 40)$), the first feedforward controller waits $\theta_F - \theta$ days before implementing a change in starts (visible at $t = 10$ in the Δ Starts plot). The Tuning plot shows the value of the user-adjustable parameter λ for the controller that is currently acting on the forecast.

Forecast revisions are implemented at $t = 20, 30$, and 40 . The feedforward action is immediately visible in the

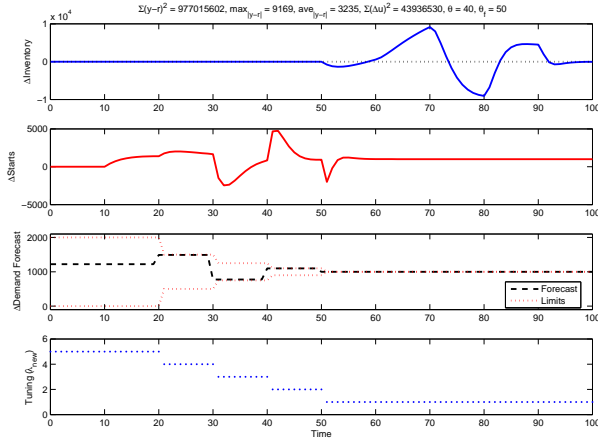


Fig. 7. Response to forecast revisions. Initial forecast received at $t = 0$. $N = 3$ revisions occurring at $t = 20, 30,$ and 40 . The actual demand change (+1000 units) is realized at $t = 50$. $\theta = 40, \theta_F = 50, \lambda_F = 5, \lambda_{F1} = 4, \lambda_{F2} = 3, \lambda_{Ff} = 2, \lambda_d = 1$.

$\Delta Starts$ plot since controllers acting within the throughput time window have no time delay. The demand change of +1000 units is realized at $t = 50$. The feedback controller then reduces starts to compensate for the fact that the previous forecast was an overestimate of the demand change.

Changes in inventory can be seen θ days after a controller move (see $\Delta Inventory$ plot). Note that value of each forecast revision signal can assume any value between the funnel-shaped limits shown in Fig. 7. While there is uncertainty in the forecast for all $t < 50$, the accuracy of the forecast signals increase with time.

IV. SIMULATION-BASED OPTIMIZATION

A general understanding of the relationship between the effect forecast error and control system tuning was presented in Section III. Given that a framework for implementing forecast revisions has been developed, it is desirable to understand how such a control system can be tuned to mitigate the effects of forecast error/uncertainty. A Simultaneous Perturbation Stochastic Approximation (SPSA) search provides a mechanism for this [7].

SPSA is an appealing method for the stochastic optimization of the proposed control system; since every additional forecast revision implemented in the proposed framework adds another dimension to the search space in which an optimal set of controller tunings will be found. SPSA has shown exhibited superior performance when used to optimize systems with a high number of parameters [7]. SPSA has also shown greater accuracy than comparable stochastic approximation methods [3].

As noted previously, deviation from setpoint is minimized by tuning the controllers more aggressively (lower values of the adjustable parameter, λ). However, changes in starts become more aggressive as well. Eqn. 10 was used as an

objective function to measure the performance of the tactical decision policy discussed in this study. Using this equation, the SPSA [7] algorithm was used to conduct simulation based optimization of the independent controller tunings.

$$J = \Gamma \sum_{t=50}^{100} (y - r)^2 + \Psi \sum_{t=1}^{100} (\Delta u)^2 \quad (10)$$

The SPSA algorithm consists of the following steps.

- 1) *Initialization*: An initial guess of the optimal λ values is made ($\hat{\theta}_k$). The coefficients of the gain sequences (a_k, c_k) are selected using guidelines provided by Spall [7].
- 2) *Perturbation Vector Generation*: A random perturbation vector (Δ_k) is generated. Each element of the vector is independently generated using a Bernoulli ± 1 distribution with a probability of $\frac{1}{2}$ for each possible outcome.
- 3) *Objective Function Evaluation*: Two measurements of the objective function are obtained: $f(\hat{\theta}_k + c_k \Delta_k)$ and $f(\hat{\theta}_k - c_k \Delta_k)$.
- 4) *Approximate the Gradient*: The simultaneous perturbation approximation of the gradient, $\hat{g}(\hat{\theta}_k)$, is determined using (11). Note that the common numerator in all components of $\hat{g}(\hat{\theta}_k)$ reflects the simultaneous perturbation of all the components in $\hat{\theta}_k$.

$$\hat{g}(\hat{\theta}_k) = \frac{f(\hat{\theta}_k + c_k \Delta_k) - f(\hat{\theta}_k - c_k \Delta_k)}{2c_k \Delta_k} \quad (11)$$

- 5) *Update the Estimate*: The standard stochastic approximation form (12) is used to update $\hat{\theta}_k$ to $\hat{\theta}_{k+1}$.

$$\hat{\theta}_{k+1} = \hat{\theta}_k - a_k \hat{g}(\hat{\theta}_k) \quad (12)$$

Fig. 8 shows the results of using the SPSA algorithm to optimize controller tunings for the 2-dimensional search space depicted in Fig. 4. Also depicted in Fig. 8 is the optimization path for the Finite Difference Stochastic Approximation (FDSA) algorithm [1]. The FDSA algorithm closely (but not exactly) follows the gradient descent path towards the optimum.

For the case shown in Fig. 8, both the SPSA and FDSA algorithms converge to the optimum value in approximately 25 iterations. SPSA requires two simulations (one gradient measurement) per iteration while FDSA requires four simulations per iteration (two gradient measurements, one for each parameter). This is consistent with the theory that for a p -dimensional optimization problem, FDSA typically requires p times as many measurements (simulations) as the SPSA algorithm [7].

Fig. 9 shows the results of using the SPSA algorithm to conduct a simultaneous search for the optimal band of control system tuning parameters for the system depicted in Fig. 7. Note that optimal performance, for the chosen objective function, is obtained by implementing detuned control initially and eventually becoming more aggressive on the manipulated variable as the forecast signals increase in accuracy.

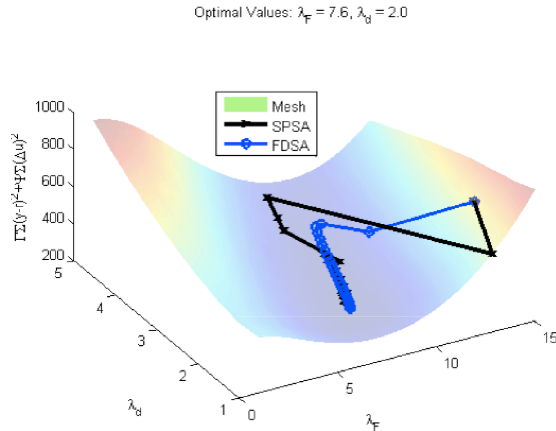


Fig. 8. SPSA and FDSA optimization algorithm paths for a 2-dimensional search space. A single feedforward controller is used (tuning is shown by λ_F). The feedback controller is tuned by manipulating λ_d . Note that increasing values of lambda denote increasing controller detuning.

The adjustable parameter for the initial feedforward controller converges to $\lambda_F = 9$. This controller will be more detuned than the other controllers for a given forecast signal. The three feedforward controllers that act within the production window (q_{F1} , q_{F2} , and q_{Ff}) become more aggressive (λ values become smaller) as the system progresses towards the realization of the demand change. Finally, the tuning parameter for the feedback controller converges to $\lambda_d = 4$. This value is smaller than the adjustable parameter values for the other controllers and denotes more aggressive controller behavior.

V. CONCLUSIONS AND FUTURE WORK

The effect of forecast error on a multi-degree-of-freedom IMC single node inventory control system has been studied. The magnitude and shape of the amplitude ratio for the forecast error transfer function was determined to be a function of the user-adjustable tuning parameter λ . It was shown that the use of a combined feedback/feedforward controller caused high and low frequency forecast error signals to be attenuated, while forecast error signals of intermediate frequency are amplified. It may be worthwhile to characterize the frequency spectrum of a forecast signal and reduce the error within this intermediate bandwidth.

The use of aggressively tuned controllers may yield the least deviation from inventory setpoints for the controller considered in this study. However, based on the nature of the forecast error considered, it was found that the use of detuned feedforward controllers in conjunction with an aggressive feedback controller may provide the most desirable overall response (in terms of trading off inventory deviation vs. starts changes).

There are three significant conclusions that should be given attention as a result of this work. If the decision policy described in this work is subjected to raw demand forecast changes, it will be most responsive to signals containing error with power in the intermediate frequencies.

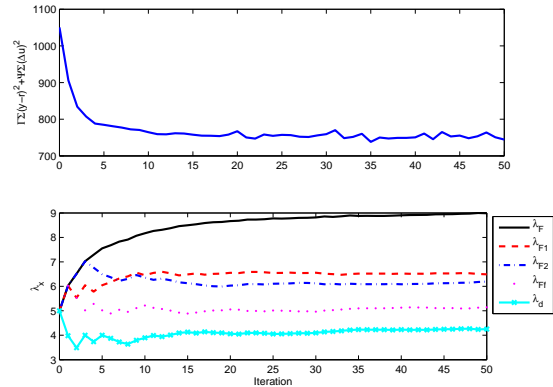


Fig. 9. SPSA optimization algorithm path for a 5-dimensional search space. A single feedforward controller is used (tuning is shown by λ_F) outside the throughput time. The three feedforward controllers acting within the throughput time window are represented by λ_{F1} , λ_{F2} , and λ_{Ff} . The feedback controller is tuned by manipulating λ_d . Note that increasing values of lambda denote increasing controller detuning.

If decisions need to be made on how to best model forecasts, one should focus on reducing forecast error over the intermediate bandwidth where sensitivity to error is greatest. Finally, the tactical decision policy presented can be tuned to attenuate the effects of forecast error.

Future work consists of an equivalent analysis for Model Predictive Control based decision policies [8]. The SPSA technique will be used to determine the optimum move suppression for a series of forecast revisions. Eventually this framework will be extended to larger and more diverse network topologies and Multi-Input-Multi-Output problems.

VI. ACKNOWLEDGEMENTS

The authors would like to acknowledge support from the Intel Research Council and the National Science Foundation (DMI-0432439).

REFERENCES

- [1] J. R. Blum. Multidimensional stochastic approximation methods. *Annals of Mathematical Statistics*, 25:737-744, 1954.
- [2] M. W. Braun, D. E. Rivera, M. E. Flores, W. M. Carlyle, and K. G. Kempf. A Model Predictive Control framework for robust management of multi-product, multi-echelon demand networks. *Annual Reviews in Control*, 27:229-245, 2003.
- [3] D. C. Chin. Comparative study of stochastic algorithms for system optimization based on gradient approximations. *IEEE Transactions on Systems, Man, and Cybernetics - Part B*, 27:244-249, 1997.
- [4] K. G. Kempf. Control-oriented approaches to supply chain management in semiconductor manufacturing. In *Proceedings of the 2004 American Control Conference*, pages 4563-4576, Boston, MA, 2004.
- [5] D. R. Lewin and C. Scali. Feedforward control in the presence of uncertainty. *Industrial Engineering Chemistry Research*, 27:2323-2331, 1988.
- [6] M. Morari and E. Zafriou. *Robust Process Control*. Prentice-Hall, Englewood Cliffs, New Jersey, 1988.
- [7] J. C. Spall. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. John Wiley and Sons, Inc., Hoboken, New Jersey, 2003.
- [8] W. Wang, D. E. Rivera, and K. G. Kempf. Centralized Model Predictive Control strategies for inventory management in semiconductor manufacturing supply chains. In *Proceedings of the American Control Conference*, pages 585-590, Denver, CO, 2003.