2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

Control-Relevant Demand Forecasting for Management of a Production-Inventory System

Jay D. Schwartz*, Manuel R. Arahal†, and Daniel E. Rivera*[1]

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

†Department of Systems Engineering and Automatic Control
University of Seville, Seville, Spain 41092

Abstract— **Forecasting highly uncertain demand signals is an important component for successfully managing inventory. We present a control-relevant approach to the problem that tailors a forecasting model to its end-use purpose, which is to provide forecast signals to a tactical inventory management policy based on Model Predictive Control (MPC). The success of the method hinges on a control-relevant prefiltering operation that emphasizes goodness-of-fit in the frequency band most important for achieving desired levels of closed-loop performance. A multi-objective formulation is presented that allows the supply chain planner to generate demand forecasts that minimize inventory deviation, starts change variance, or their weighted combination when incorporated in an MPC decision policy. The benefits obtained from this procedure are demonstrated on a case study where the estimated demand model is based on a AutoRegressive (AR) process.**

I. Introduction

Inventory management is a critical component of supply chain management. In recent years, a number of approaches inspired by chemical process control have been proposed to manage short-term "tactical" decision-making for inventory management problems associated with semiconductor manufacturing supply chains [1], [2], [3]. These approaches rely on Model Predictive Control [4], an optimization-based control technology that has experienced wide success in the process industries. In these approaches, demand is treated as an exogenous "disturbance" signal that must be properly "rejected" by a sensibly-designed control system. In this paper we highlight the interplay between forecast error and the decision policy, which relies on potentially erroneous demand forecasts. Specifically, understanding what properties of forecasts that have the most effect on the end-use decision policy can be used to develop contextualized algorithms that are able to simplify and increase the effectiveness of demand forecasting techniques. The control-relevant demand forecasting procedure presented in this paper addresses the issue of how demand forecasts are generated, and in doing so gives supply chain planners greater flexibility regarding how the demand "disturbance" is rejected. Such control-relevant forecasts will ultimately lead to improved production and inventory control.

The relationship between demand forecast error and changes in inventory and starts is examined for a control-oriented tactical decision policy in a single node of a supply chain. To accomplish this goal, ideas in control-relevant identification [5] and control-relevant demand modeling for production-inventory systems [6] are applied. A control-relevant understanding of the problem dictates that not all characteristics of the demand process need to be retained when the performance criteria of the end-use decision policy is considered, and consequently, placing an emphasis during modeling on the most important regions of time and frequency will have a beneficial effect. A sensibly designed prefilter applied to demand estimation data can be used for this purpose. A multi-objective formulation for the prefilter design problem is presented that allows the user to minimize inventory variance, starts variance, or a weighted combination. Results from the case study show that the approach improves performance in a flexible manner, reducing inventory variance, factory starts variance, or achieving a reasonable compromise between these objectives.

The paper is organized as follows. Section II provides background material on the modeling of a production/inventory system using a fluid analogy. Section III shows the the development of a tactical decision algorithm relying on Model Predictive Control. In Section IV, the closed-loop transfer functions describing forecast error are developed and the effect of erroneous forecasts is studied in both the time and frequency domains. This analysis serves as the basis for the procedure that performs control-relevant demand forecasting. Section V is a case study that applies the approach to manage a production/inventory system. The contrasting results obtained from the multi-objective formulation, and a comparison with a traditional "open-loop" approach that does not rely on a control-relevant understanding are presented. Section VI highlights the important conclusions that can be drawn from the analysis in this paper, as well as some areas for future research.

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

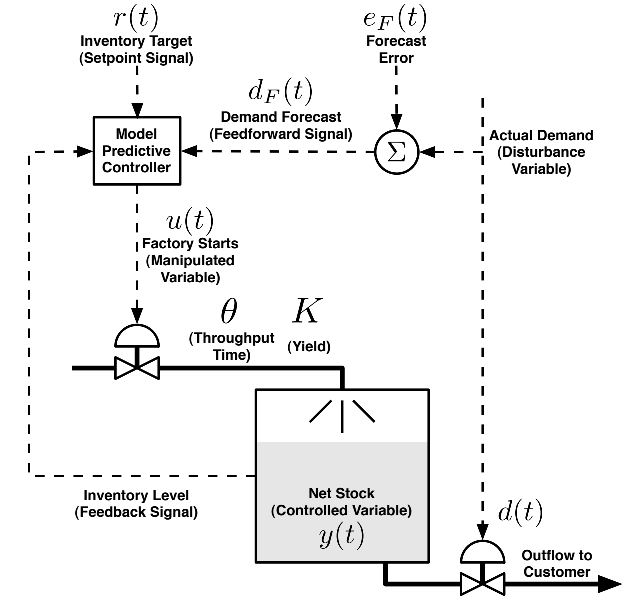

Fig. 1. Fluid analogy for a production/inventory system

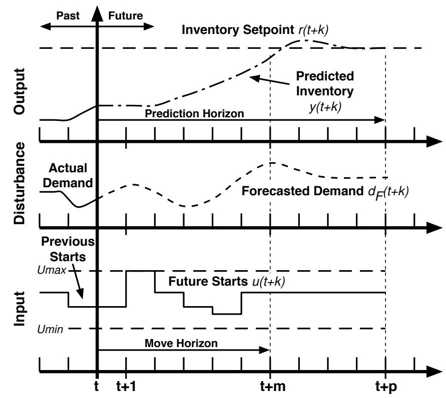

Fig. 2. Receding horizon representation of Model Predictive Control relying on demand forecasts, as applied to production/inventory control.

II. Modeling a Production/Inventory System using a Fluid Analogy

Developing a dynamic model is the first step towards producing a control system for inventory management. In most applications dynamic modeling is difficult whereas in manufacturing the most significant uncertainties are not in the process model but in the supply and demand signals. To begin we use a fluid analogy to describe a generic manufacturing process; this reduces the process to a collection of pipes and tanks. This abstraction not only makes the analysis more reasonable, it also gives it wider appeal. Such a fluid analogy is shown in Figure 1. The factory is modeled as a pipe with a particular throughput time θ and yield K. The inventory is modeled as material (fluid) in a tank. Applying the principle of conservation of mass to this system leads to a differential equation relating net stock (tank level, $y(t)$) to factory starts (pipe inflow, $u(t)$) and customer demand (tank outflow, $d(t)$). This differential equation is represented by the following z-domain discrete transfer function, where the demand is composed of the forecasted customer demand, $d_F(t-\theta_F)$, plus unforecasted customer demand, $d_U(t)$.

$$y(z) = \frac{Kz^{-\theta}}{z-1}u(z) - \frac{z^{-\theta_F}}{1-z^{-1}}d_F(z) - \frac{1}{1-z^{-1}}d_U(z) \quad (1)$$

Where θ_F is the forecast horizon and $d_F(t)$ represents an estimate of demand θ_F days into the future. Based on (1) it is possible to derive tactical decision policies that manipulate factory starts to maintain inventory level at a designated setpoint. If knowledge of future customer demand is available, it is advantageous to use feedforward compensation. A decision policy based on this nominal model (kept constant for purposes of this paper) that relies on a demand forecast is presented in the ensuing section.

III. Model Predictive Control

The application of Model Predictive Control [4], [7], [8] to the inventory management problems considered in this paper follows along the lines of the conceptual framework presented in [1]. In a computation interval the factory starts are optimized over a move horizon to minimize predicted deviations from inventory targets given an anticipated demand signal, as shown in Figure 2. The starts level corresponding to the first entry in the move horizon is implemented and the process is repeated. A meaningful MPC objective function formulation is as follows:

$$\min_{\Delta u(k|k)\ldots\Delta u(k+M-1|k)} J \quad (2)$$

where $\Delta u(k|k)\ldots\Delta u(k+M-1|k)$ represents the computed sequence of starts changes made at time k, and the individual terms of J correspond to:

$$J = \overbrace{\sum_{\ell=\theta}^{P} Q_e(\ell)(\hat{y}(k+\ell|k) - r(k+\ell))^2}^{\text{Keep Inventories at Inventory Planning Setpoints}} + \overbrace{\sum_{\ell=1}^{M} Q_{\Delta u}(\ell)(\Delta u(k+\ell-1|k))^2}^{\text{Penalize Changes in Starts}} \quad (3)$$

subject to constraints on inventory capacity ($0 \leq y(k) \leq y_{max}$), factory inflow capacity ($0 \leq u(k) \leq u_{max}$), and changes in the quantity of factory starts ($\Delta u_{min} \leq \Delta u_k \leq \Delta u_{max}$). Equation 3 is a multi-objective expression that addresses the main operational objectives in the supply chain. The first term is a setpoint tracking term intended to maintain inventory levels at user-specified targets over time. The second term is a move suppression term that penalizes changes in the factory starts. The emphasis given to each one of the sub-objectives in (3) (or to specific system variables within these objective terms) is achieved through the choice of weights ($Q_e(\ell)$ and $Q_{\Delta u}(\ell)$). These can potentially vary over the prediction and move calculation horizons (P and M, respectively).

IV. Achieving Control-Relevant Demand Forecasting

In this paper we are primarily concerned with the role that customer demand $d(t)$ plays in supply chain dynamics, since it is essential when designing an inventory control system that the demand characteristics be known [9]. Since this demand is unknown in advance it has to be predicted to be able to compute the best sequence of factory starts according to Equation 2. Historically, such a prediction has been based on a least-squares fit of the data to a certain model structure. Therefore, the family of models used to characterize the demand has played an important role. In this paper we also contend that the intended use of the model will also play a prominent role.

A. Time- and Frequency-Domain Analysis of Forecast Error

Fig. 3(a) shows the time-domain inventory and starts responses to a unit pulse in forecast error. The controller anticipates the future demand change and increases starts accordingly beginning at day 10. When no demand change is realized, starts are reduced to return the inventory level to the setpoint. For unconstrained MPC and using Equation 1 for the production/inventory system, the closed-loop system is linear [7] and its frequency responses can be adequately represented via nonparametric methods. These responses are represented by Finite Impulse Response models, from which frequency responses are generated. Fig. 3(b) shows that the response of the MPC decision policy to forecast error is characterized by notch filters, where high and low frequencies are attenuated, and only forecast error in an intermediate bandwidth is amplified. For this simulation the throughput time θ is 3 days and there is no yield loss ($K = 1$). MPC tuning parameters are as follows: inventory deviation penalty $Q_e = 1$, starts change penalty $Q_{\Delta u} = 5$, prediction horizon $P = 20$ days, and a move calculation horizon of $M = 10$ days.

Understanding Fig. 3(b) is critical towards developing a control-relevant demand modeling procedure. The MPC decision policy is relatively insensitive to "very fast" (i.e., high-frequency) and "very slow" (i.e., low-frequency) forecast errors; hence having an accurate demand model in these frequency bands is much less important than in the intermediate bandwidth where forecast error is amplified. The size of the bandwidth of importance will be determined by how the decision policy is tuned and the process throughput time. In addition, the outcome of interest will further influence the frequency band of importance to the problem; consider that the bandwidth associated with how starts changes are affected by forecast error lies at higher frequencies than for inventory.

B. Control-Relevant Modeling based on Historical Data

We now describe in this subsection how to achieve control-relevant demand modeling using prefiltered AR model estimation. Given access to a data set that contains past demand values it is possible to develop an auto-regressive model

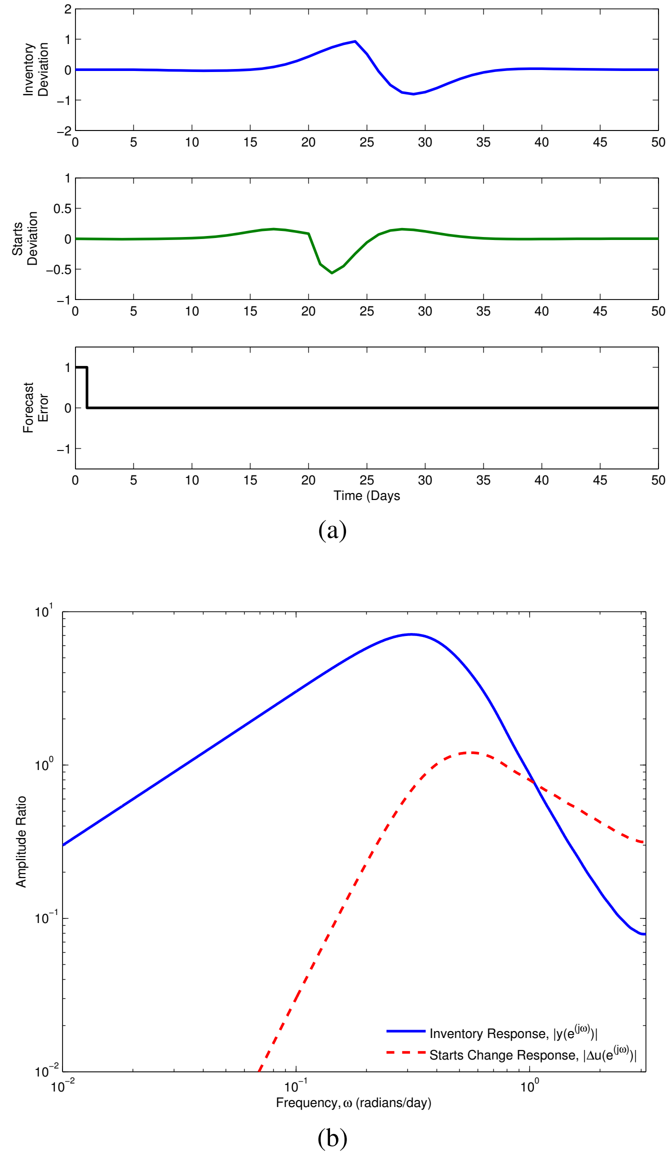

Fig. 3. Characteristic time-domain (a) and frequency-domain (b) responses of a production/inventory system to a forecast error pulse.

AR(n_a) [10]. This model uses past demand values $d(t)$ to generate the prediction $\hat{d}(t)$ as

$$\hat{d}(t+1) = a_1 d(t) + a_2 d(t-1) + \ldots + a_{n_a} d(t - n_a + 1) \quad (4)$$

and is amenable to control-relevant prefiltering and the identification of customer demand models. Other model families can be used, for instance the more general class of ARMAX models. In inventory control, however, the most common situation is that only past values of the demand are known with enough confidence to base the forecast on. The following analysis can be extended to the general ARMAX class of models.

Assume that the true demand model also has the AutoRegressive (AR) form, where a normalized demand signal $x(t)$ is obtained by subtracting the mean and obvious trends from $d(t)$.

$$A(z)x(t) = \xi(t) \quad (5)$$

$\xi(t)$ is a suitable perturbation signal (white noise with zero mean). The polynomial $A(z)$ has the form

$$A(z) = 1 + a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a} \quad (6)$$

Given an estimate of polynomial A, denoted as $\hat{A}$, the predictor model expressed as

$$\hat{x}(t \mid t-1, \hat{A}) = \left(1 - \hat{A}(z)\right) x(t) \quad (7)$$

computes the one-step-ahead prediction resulting in the one-step-ahead prediction error

$$e^x(t \mid t-1, \hat{A}) = x(t) - \hat{x}(t \mid t-1, \hat{A}) \quad (8)$$

In order to relate the performance of the predictor with the power spectrum characteristics of the signals it is necessary to write a formula for the prediction error. It is easy to show that

$$e^x(t \mid t-1, \hat{A}) = \frac{\hat{A}(z)}{A(z)} \xi(t) \quad (9)$$

is the way in which the one-step-ahead prediction error is related to the probabilistic model and to the polynomial estimate $\hat{A}$. For ease of notation we will refer to $e^x(t \mid t-1, \hat{A})$ simply as $e(t)$ in the following paragraphs. If a prefilter $L(z)$ is applied to the data, the filtered prediction error is obtained as $L(z)e(t)$. The control relevant estimation problem should determine values for $\hat{A}$ such that

$$V = \frac{1}{N} \sum_{t=1}^{N} \left(L(z)e(t)\right)^2 \quad (10)$$

is minimized. This minimization should put more emphasis at the frequencies relevant for the control problem; to this end the prefilter $L(z)$ must model the frequency responses for the system response to forecast error described in Section IV-A. Using Parseval's theorem one obtains

$$\lim_{N\to\infty} V = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left|L(e^{j\omega})\right|^2 \left|\frac{\hat{A}(z)}{A(z)}\right|^2_{z=e^{j\omega}} \Phi_\xi(\omega) d\omega \quad (11)$$

The stochastic variable ξ is not known but it is expected that the power spectrum Φ_ξ would have a broad support, in particular it is expected to have a non negligible tail at high frequencies. A constant distribution over ω can be a good approximation (in fact it is exact in the case when ξ is white noise).

A supply chain planner may choose to reduce either inventory deviation, stockout, or changing a factory setup depending on their associated costs. It is useful then to use the following metric to measure the performance of the control system.

$$J_e = \sum_{t=0}^{\infty} (1-\gamma) e_c^2(t) + \sum_{t=0}^{\infty} \gamma \Delta u^2(t) \quad (12)$$

where e_c is the control error (inventory deviation from setpoint) and γ is used as a weight to emphasize reduction on either inventory deviation from setpoint ($\gamma = 0$) or factory starts variance ($\gamma = 1$). The control-relevant prefilter corresponding to the objective (12) is

$$|L(e^{j\omega})|^2 = (1-\gamma)|L_{e_c}(e^{j\omega})|^2 + \gamma |L_{\Delta u}(e^{j\omega})|^2 \quad (13)$$

where the prefilters $L_y(e^{(j\omega)})$ and $L_{\Delta u}(e^{(j\omega)})$ are obtained from the unconstrained MPC impulse response to forecast error analysis as described in Section IV-A.

C. Algorithm for Developing Control-Relevant Prefilters

This section summarizes the algorithm for arriving at the prefilters that accomplish control-relevant demand modeling. The control-relevant filter is developed from the underlying process dynamics, controller tuning parameters, and the value of the user-adjustable parameter γ. The procedure is easy to implement using MATLAB® software. It consists of the following steps:

1) *Open-Loop Model Specification*: Identify an open-loop manufacturing process model to obtain values for the nominal throughput time θ and the nominal process yield K.
2) *MPC Initialization*: Choose a move calculation horizon M greater than the nominal throughput time. Correspondingly, choose a prediction horizon P greater than the move calculation horizon. Then choose weights for penalizing starts changes and inventory deviation from setpoint ($Q_{\Delta u}$ and Q_e, respectively). Enable Type-II output disturbance rejection; the system incorporates an integrating process and Type-I control action is insufficient to reject ramp disturbances.
3) *Closed-Loop Simulation*: In closed-loop simulation introduce a forecast error pulse with a magnitude of one unit and duration of one time interval. Measure the corresponding change in inventory deviation and starts change (which may be obtained by differencing the starts signal).
4) *Choosing Emphasis*: The closed-loop simulation yields FIR filters that act to minimize inventory deviation and starts change variance, respectively. Calculate the amplitude ratio of each filter and normalize it by the supremum. Then choose a value for the parameter γ to meet a supply chain objective or performance requirement. Finally, combine the amplitude ratios according to Equation 13.
5) *Filter Generation* A curve fitting procedure is then used to obtain an Infinite Impulse Response filter that matches the amplitude ratio of the control-relevant prefilter. A standard curve fitting algorithm for rational discrete-time transfer functions can be used for this purpose, such as the output-error minimizing algorithm as implemented in the MATLAB® function `invfreqz`. The resulting filter will be applied to the historical demand signal to develop a control-relevant AR forecasting model.

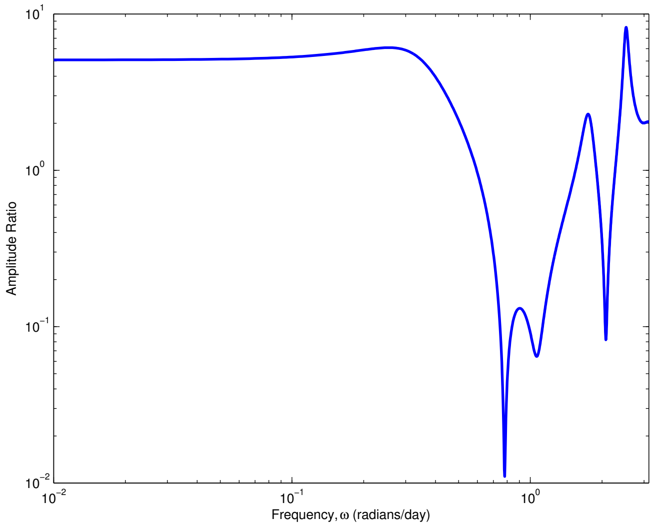

Fig. 4. Amplitude ratio of the ARMA(10,10,5) demand model used for the case study.

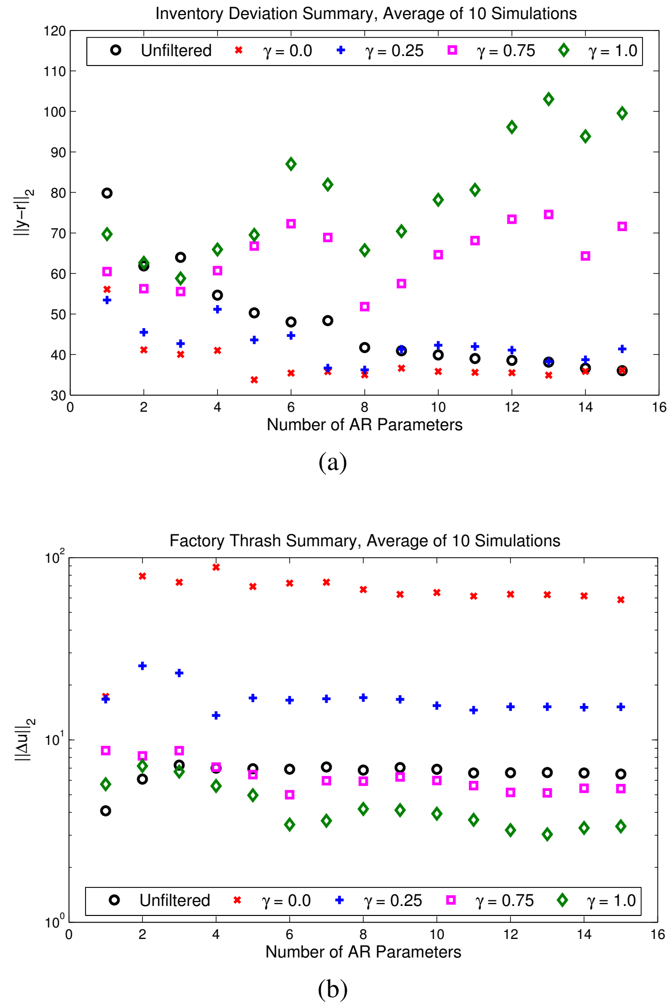

Fig. 5. Summary of closed-loop results for a demand model based on the ARMA (10,10,5) process whose amplitude ratio is shown in Figure 4. In (a) inventory deviation from setpoint is shown as a function of model order and γ; (b) summarizes factory thrash similarly.

V. Case Study

For this case study we use an MPC-based tactical decision policy to manipulate factory starts to keep the inventory level at a target subject to uncertain demand. The demand used for the experiments is generated by an AutoRegressive Moving Average (ARMA) process whose amplitude ratio is shown in Figure 4. The manufacturing process is modeled as a pipe with a throughput time of 3 days and no yield loss. The MPC prediction horizon (P) is 20 days, the MPC move optimization horizon M is 10 days, and MPC weights for penalizing starts changes and inventory deviation from setpoint ($Q_{\Delta u}$ and Q_e) are 5 and 1, respectively. At each control interval we execute the prefilter design procedure outlined in Section IV-C. Therefore, the AR forecasting model is updated at each time interval to reflect new demand information. In a control interval the demand modeling procedure must gather historical demand data, filter the data according to a control-relevant objective, identify an AR model from the filtered data, develop a prediction from the resulting AR model, and use the prediction in the optimal control move calculation.

Figure 5 displays inventory control performance as a function of the user-adjustable parameter γ for an aggregation of ten simulation results. A planner can choose to minimize inventory deviation from setpoint by selecting values of γ close to zero. Factory thrash is minimized by selecting γ close to one. When filtering is used to minimize inventory deviation ($\gamma = 0.0$) we obtain the desired result. However, this also resulted in the highest level of factory thrash. The converse is true for the factory thrash minimizing filter ($\gamma = 1.0$). The use of intermediate values of γ results in a trade-off between factory thrash and inventory deviation. The control-relevent modeling procedure yields the most benefit when applied to low order models. As the model order increases the benefit with respect to minimizing inventory deviation decreases.

Figure 6 displays performance as a function of the user-adjustable parameter γ for a single simulation. During the first 1000 days of the simulation the controller is run in feedback-only mode; no demand forecast is used. As expected, improved performance can be obtained by including a demand prediction in the MPC move calculation. From days 1000 to 2000 the demand prediction is generated without the benefit of the control-relevant filtering procedure shown in this paper. The rest of the simulation shows the effect of applying a control-relevant forecast, in terms of increasing γ, on inventory deviation $||y - r||_2$ and factory thrash $||\Delta u||_2$. Setting $\gamma = 0$ results in a 36% reduction in inventory deviation relative to the unfiltered case, but causes significant fluctuations in factory starts. Using a value of $\gamma = 1$ yields a 26% reduction in factory thrash relative to the unfiltered case. Choosing intermediate values of γ gives substantial flexibility to a supply chain planner.

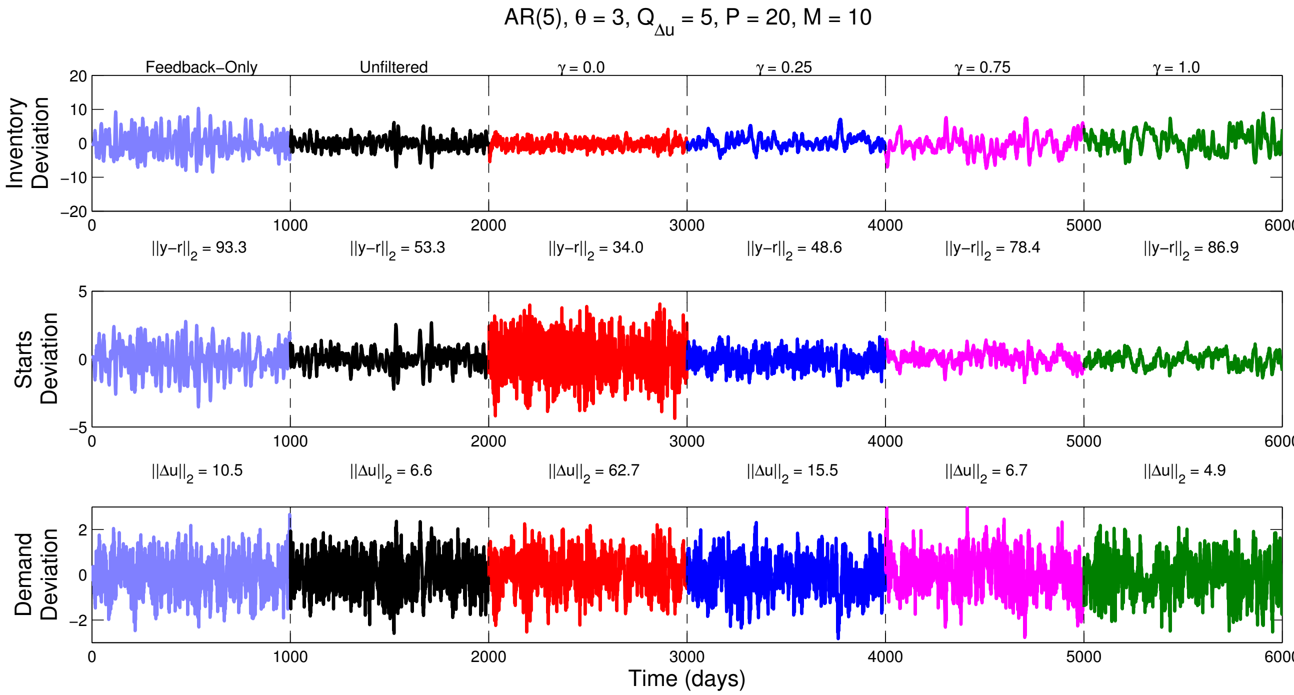

Fig. 6. This closed-loop simulation begins in feedback-only mode; the controller manipulates factory starts to keep the inventory level at a setpoint subject to uncertain demand. From days 1000 to 2000 a demand forecast is available to the controller, but the prediction is generated from unfiltered data. The time series from day 2000 on shows the benefit of the control-relevant forecasting procedure and the effect of the parameter γ on inventory deviation and factory thrash.

VI. Conclusions

We have developed a control-relevant "end-use" inspired framework for demand forecasting meaningful for supply chain management. The success of the method hinges on a control-relevant prefiltering operation applied to demand estimation data that emphasizes a goodness-of-fit in regions of time and frequency most important for achieving desired levels of closed-loop performance. This framework allows a supply chain planner to minimize inventory deviation from setpoint, factory thrash, or a weighted combination of the two objectives. In practice, applying control-relevant demand forecasting as presented in this paper will enhance a planner's ability to meet management objectives, such as keeping inventory within desired limits while simultaneously minimizing costly changes to the factory schedule. Extensions of the demand modeling procedure to multivariable systems where the MPC policy manages inventory in a multi-echelon supply chain are currently under investigation.

VII. Acknowledgments

This work was supported by grants from the National Science Foundation (CMMI-0432429) and the Intel Research Council.

References

[1] 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-echolon demand networks," *Annual Reviews in Control*, vol. 27, pp. 229–245, 2003.

[2] J. D. Schwartz, W. Wang, and D. E. Rivera, "Simulation-based optimization of model predictive control policies for inventory management in supply chains," *Automatica*, vol. 42, no. 8, pp. 1311–1320, 2006.

[3] W. Wang, D. E. Rivera, and K. G. Kempf, "Model Predictive Control strategies for supply chain management in semiconductor manufacturing," *International Journal of Production Economics*, vol. 107, no. 1, pp. 56–77, 2007.

[4] C. E. García, D. M. Prett, and M. Morari, "Model predictive control: theory and practice-a survey," *Automatica*, vol. 25, no. 3, pp. 335–348, 1989.

[5] D. E. Rivera, J. F. Pollard, and C. E. García, "Control-relevant prefiltering: A systematic design approach and case study," *IEEE Transactions on Automatic Control*, vol. 37, no. 7, pp. 964–974, 1992.

[6] J. D. Schwartz and D. E. Rivera, "Control-relevant demand modeling for supply chain management," in *Proceedings of the 14th Annual IFAC Symposium on System Identification*, Newcastle, Australia, 2006, pp. 267–272.

[7] E. F. Camacho and C. Bordons, *Model Predictive Control*. London: Springer-Verlag, 1999.

[8] S. J. Qin and T. A. Badgwell, "A survey of industrial model predictive control technology," *Control Engineering Practice*, vol. 11, pp. 733–764, 2003.

[9] C. D. Lewis, *Demand forecasting and inventory control*. Woodhead Publishing Limited, 1997.

[10] A. Pankratz, *Forecasting with Dynamic Regression Models*. John Wiley and Sons, Inc., 1991.