Design of Minimum Crest Factor Multisine Signals for "Model-on-Demand" System Identification and Model Predictive Control

Martin W. Braun, Raul Ortiz-Mojica, and Daniel E. Rivera

Control Systems Engineering Laboratory

Department of Chemical and Materials Engineering

Arizona State University, Tempe, Arizona 85287-6006

(602) 965-9476

email: daniel.rivera@asu.edu, martin.braun@asu.edu

Prepared for presentation at the Annual AIChE Meeting, Los Angeles, CA

November 12-17, 2000

Session: Modeling and Identification I, Paper 252c

September, 2000

UNPUBLISHED

Design of Minimum Crest Factor Multisine Signals for "Model-on-Demand" System Identification and Model Predictive Control

M. W. Braun, R. Ortiz-Mojica and D. E. Rivera *

Control Systems Engineering Laboratory

Department of Chemical, and Materials Engineering

Arizona State University, Tempe, Arizona 85287-6006

Abstract

The need for research in the area of input signal design for identification of nonlinear process systems has increased in recent years, in response to the plethora of work focused on nonlinear model selection. Minimum crest factor multisine signals are periodic, deterministic signals which can be designed to exactly meet desired frequency-domain requirements, and supply information content in the signal suitable for nonlinear identification. This paper proposes a structured design methodology for minimum crest factor multisine signals that allows process control engineers to maximize use of *a priori* information. By emphasizing use of prior knowledge, such as process constraints, desired control performance, and model structure, the methodology seeks to obtain an informative dataset for identification, while reducing the number of costly iterations in the experimental testing phase of system identification. Examples of the use of the methodology are provided through the nonlinear identification and control of a non-adiabatic CSTR simulation and a high-purity distillation simulation.

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

1 Introduction

Nonlinear modeling approaches such as neural networks, fuzzy modeling, semi-physical modeling, and block-structured techniques have become quite popular within the process control community (Seborg and Henson, 1997). The ability to estimate parameters for these model structures from normal operating data records has significant practical appeal. However, operational data often lacks suitable information content (e.g., persistence of excitation, identifiability requirements, etc.) and can have other undesirable properties (e.g., crosscorrelated inputs and disturbances) which make it unsuitable for system identification. The user is thus forced to consider the task of designing and executing identification experiments. Yet comparatively little attention has been paid to how one can generate informative data for use with nonlinear identification techniques. An even more difficult problem is how the user might introduce signals for plant excitation, while meeting safety, environmental, economic, or equipment constraints. In meeting these constraints, this excitation signal must still be able to get control-relevant information from the plant.

Previous work by the authors (Braun *et al.*, 1999; Braun and Rivera, 2000) has developed guidelines for the use of m-level pseudo-random sequences (m-level PRS) for identification of nonlinear plants. m-level PRS signals are conveniently generated by a deterministic algorithm. However, a disadvantage of the m-level PRS signal is that desired frequency limits may not exactly be met by the set of sequences available. This results in a signal that may be somewhat longer than necessary or requires that the user overdesign the high frequency content to bring the signal length within acceptable levels. In the case that the input amplitude is severely restricted, the m-level PRS signal may not provide significant variance in the output compared to other classes of signals with lower crest factors. m-level PRS signals will switch in one switching time between the limits of the input range. Because of actuator move-size constraints and other plant limitations, this is not a "plant-friendly" property.

Minimum crest factor multisine signals, on the other hand, are an attractive alternative to the m-level PRS signal. Multisine signals share with m-level PRS signals the properties of

being deterministic, periodic signals. However, a multisine signal can be designed to exactly meet a desired power spectrum profile, which often results in a shorter signal length than m-level PRS signals designed for the same bandwidth. Using a low crest factor input signal can improve the signal to noise ratio of the resulting plant output. A low crest factor indicates that most of the elements in the input sequence are distributed near the minimum and maximum values of the sequence. Multisine signals can be optimized to have a minimum crest factor and have no restrictions on the number of levels they can achieve within the input range. This provides better resolution of the input space, which is desirable for nonlinear process identification. Prior experience with the time domain realizations of these signals has shown that move sizes generally do not shift between the minimum and maximum values in one time step.

We present in this paper systematic guidelines for selecting the power spectrum/Fourier coefficients in a multisine input for nonlinear applications. To generate multisine signals with minimum crest factor, we apply the algorithm per (Guillaume *et al.*, 1991). The guidelines developed in our paper make use of available *a priori* information available to the user. For the MIMO identification problem, guidelines for designing a "zippered" set of minimum crest factor inputs are presented. The usefulness of both the design procedure and the minimum crest factor multisine signals is demonstrated in the context of a nonlinear 'Model-on-Demand" (MoD) estimation and control methodology.

"Model-on-Demand" (MoD) estimation (a.k.a. Just-in-Time (Stenman *et al.*, 1996)) is a data-driven technique similar to the Artificial Intelligence notion of Lazy Learning (Stenman, 1999). MoD determines several candidate neighborhoods of data (i.e. bandwidths) around the current operating point. Local models are constructed from the data in each of these candidate neighborhoods, and the models are evaluated using an automatic bandwidth selector such as the Localized Akaike's Final Prediction Error (FPE) or Localized Akaike Information Criteria (AIC). The MoD algorithm then uses the local model corresponding to the optimal bandwidth to estimate the next output. At the next time step, the algorithm is repeated. Since MoD implicitly uses the data to balance the bias/variance tradeoff, the quality of the

experimental data is critical to MoD performance. As a result, MoD represents a demanding nonlinear modeling technique from which we can evaluate the usefulness of proposed design procedure for minimum crest factor multisine signals. To evaluate the data in a control-relevant manner, MoD can be incorporated into a Generalized Predictive Control (GPC) framework for nonlinear control (aka. MoDMPC).

The subsequent section discusses the generation and design of minimum crest factor multisine signals for SISO systems. These ideas are then extended to the multi-input case. A brief treatment of MoD and MoDMPC is then made and the remaining sections are dedicated for meaningful case studies. A non-adiabatic CSTR simulation as derived by Denn (1986) provides a good test bed for the evaluation of user design decisions. A comparison of minimum crest factor multisine design decisions is made and these signals are compared against a well designed m-level PRS. User decisions are shown to affect the ability of these signals to provide data support for MoD estimation and control. The MIMO minimum crest factor multisine guidelines are demonstrated via closed-loop identification and control of the Weischedel-Mc Avoy high purity distillation column.

2 Minimum Crest Factor Multisines

Consider a multisine signal of the form

$$u(k) = \lambda \sum_{i=1}^{N_s/2} \hat{a}_u(i) cos(\omega_i kT + \phi_i). \tag{1}$$

λ is the scaling factor, T is the sampling time, n_s is the number of harmonics to reside within the frequency limits and $\hat{a}_u(i)$ is the relative amplitudes of those harmonics. Note $\hat{a}_u(i) = \sqrt{2\alpha_i}$. N_s is the length of the signal. Both N_s and n_s must satisfy

$$\omega_i = 2\pi i / N_s T \quad \text{and} \quad n_s \leq N_s/2. \tag{2}$$

The main issue in designing a multisine signal for system identification is the selection of the phases ϕ_i such that the time-domain realization of the signal does not contain a few very

large peaks or valleys that promote a poor distribution in the input range. There are many different quantitative measures of this phenomena discussed in the literature, most popular of which are the peak factor (pf)

$$pf = \frac{\max(u) - \min(u)}{2\sqrt{2}\text{rms}(u)}, \tag{3}$$

and the crest factor (cf)

$$cf_u = \frac{\ell_\infty(u)}{\ell_2(u)}. \tag{4}$$

Note rms(u), $\ell_2(u)$, $\ell_\infty(u)$ represent the root-mean-square, the 2-norm, and the Chebyshev norm of u, respectively. The Chebyshev norm is nondifferentiable with respect to k. The normalizing factor of $2\sqrt{2}$ is included in the definition of the peak factor such that the peak factor of a single sine wave is 1. It is recommended that the peak factor measure be used only on zero-mean multisine signals and other signals with symmetric input distributions (Godfrey, 1993). For the purposes of this paper, the crest factor is the chosen measure, since it has been observed to be a better indication of the compactness of a signal, and it is better at taking into account asymmetric peaks (Godfrey, 1993).

The first notable solution to this problem seeking to minimize peak factor was proposed by Schroeder (1970). This resulted in the closed form equation

$$\phi_i = 2\pi \sum_{j=1}^{i} j\alpha_j. \tag{5}$$

In Figure 1, the improvement this phase equation has in the time domain realization of a multisine is illustrated. The signal is better distributed over the input range and the severity of peaks in the signal is reduced.

2.1 Generation of Minimum Crest Factor Multisines

Guillaume and coworkers develop an optimization-based approach which seeks to approximate the minimization of the Chebyshev norm by sequentially minimizing the ℓ_p norm for $p = 4, 8, 16, \ldots$. Note the ℓ_2-norm remains invariant with respect to the phases ϕ_i. Under

certain assumptions, Guillaume and coworkers were able to show that the continuous ℓ_p norm could be equivalently represented in the discrete form

$$L_p(u(k)) = \frac{1}{N_s}\sum_{k=1}^{N_s}|u(k)|^p. \tag{6}$$

This expression was incorporated into a Gauss-Newton optimization algorithm with Levenburg-Marquardt Hessian approximation. This is accomplished by defining a cost function

$$\min_{\mathbf{p}} \frac{1}{N_s} e^T e, \tag{7}$$

where (8)

$$\begin{aligned} e &= [u_s(1)^{p/2} \quad u_s(2)^{p/2} \quad \dots \quad u_s(N_s)^{p/2}]^T \\ \mathbf{p} &= [\phi_2 \quad \phi_3 \quad \dots \quad \phi_{n_s}]^T \end{aligned} \tag{9}$$

Note that p must be even. The elements of the Jacobian $\mathbf{J}$ can be represented by

$$J_{ki} = -(p/2)u_s(k)^{(p/2-1)}\sqrt{(2\alpha_i)}\sin(\omega_i kT + \phi_i), \tag{10}$$

providing the iterative phase update equation

$$\mathbf{p}^{(i)} = \mathbf{p}^{(i-1)} - [\mathbf{J}^{(i-1)T}\mathbf{J}^{(i-1)} + \Lambda^{(i-1)}]^{-1}\mathbf{J}^{(i-1)T}e^{(i-1)}. \tag{11}$$

The reference does not detail the Hessian update, but study of the code provided by Guillaume and coworkers reveals that Λ is updated using a trust-region strategy. Guillaume and coworkers make use of a FFT based algorithm to compute $\mathbf{J}^{(i-1)T}\mathbf{J}^{(i-1)}$ and $\mathbf{J}^{(i-1)T}e^{i-1}$.

The sequential minimization problem begins with an initial guess of the phases ϕ_i. The phases determined at the end of the optimization for each value for p are used as the initial phases for the next value of p. The original work suggests that initial phases be determined by the Schroeder phase equation. Recently, additional work suggests a random phase initialization has some advantages for the novice user (Simon and Schoukens, 2000). Although a global solution cannot be guaranteed with this approach, most local minima are avoided and experimentally it performs better than other crest factor minimization techniques (Guillaume *et al.*, 1991).

2.2 Minimum Crest Factor Multisine SISO Design Guidelines

The user interested in designing minimum crest factor multisines for nonlinear process excitation is faced with the task of using available prior knowledge to design a useful signal. Often, *a priori* information can be estimated from existing operational data or through step testing. The user is also well served by spending some time thinking about the engineering logic or first-principles reasoning behind the process under study. The user typically has the following *a priori* information at their disposal.

1. Desired closed-loop speed of response
2. Desired low frequency information
3. Estimated range of dominant time constant
4. Order and structure of the model(s) to be fit
5. Acceptable signal length and amplitude

The guidelines of Rivera *et al.* (1997) can be directly applied to transform the above information into control relevant frequency requirements

$$\omega_* = \frac{1}{\beta \tau_{dom}^H} \leq \omega \leq \frac{\alpha}{\tau_{dom}^L} = \omega^*. \tag{12}$$

To meet these requirements N_s and n_s can be chosen such that

$$\omega_* \geq \frac{2\pi}{N_s T} \qquad \omega^* \leq \frac{2\pi n_s}{N_s T} \tag{13}$$

or equivalently,

$$N_s \geq \frac{2\pi\beta\tau_{dom}^H}{T} \qquad n_s \geq \frac{N_s T \alpha}{2\pi\tau_{dom}^L}. \tag{14}$$

τ_{dom}^H and τ_{dom}^L correspond to the high and low estimates of the dominant time constant of the system. α corresponds to the ratio of the desired closed-loop speed of response to the open-loop speed of response. Typical values for α are 2, or 3. β is used to incorporate the desired low frequency content. If the user would like low frequency information up to the 95%, 98%, or 99% settling time, the user may select a β of 3, 4, or 5. It should be noted that

the user's choice of β may come into conflict with the acceptable signal length requirement.

The user has the additional freedom to choose relative magnitudes of the harmonics α_i for harmonic suppression or for any shape of a desired power spectrum. This work chooses a notched spectrum design as shown in Figure 2. Since most process systems are highly over-sampled, the user has the option to specify the amplitudes of the high frequency sinusoids above ω^*. Thus, the user can maximize the power in the control relevant frequency range. The additional frequency components outside of the notch can also be used to provide better conditioning of a global parameter estimation procedure, when the number of sinusoids between ω_* and ω^* is small. High frequency content beyond ω^* dramatically improves the crest factor of the signal. This is often referred to as the "snow-effect". Additional low frequency sinusoids below ω_* are useful when the user intends to employ an iterative identification strategy. The user has the option to shift or expand the control relevant frequency range in subsequent signal designs, while maintaining the same fundamental frequency. The notch design results when amplitudes $\hat{a}_u(i)$ are chosen such that

$$\hat{a}_u(i) = \begin{cases} lf, & i = 1, \ldots, \gamma \\ 1, & i = 1+\gamma, \ldots, n_s+\gamma \\ hf, & i = n_s+\gamma+1, \ldots, N_s/2 \end{cases} \tag{15}$$

where

$$\gamma = floor\left(\frac{\omega_* N_s T}{2\pi} - 1\right) \tag{16}$$

Note that $floor(arg)$ is a MATLAB^{TM} argument that rounds arg to the nearest integer towards $-\infty$. A special case of the notch design is the low-pass design for which the user does not specify a value for lf, and $\gamma = 0$ in specifying the rest of the amplitudes in Equation 15. The user can also suppress multiples of harmonics as needed to independently identify linear and nonlinear contributions in a block structure model per the approaches mentioned in Barker and Zhuang (1997).

For linear model fitting, the number of unsuppressed harmonics up to n_s must meet the persistence of excitation requirement necessary for the model to be fit. In practical terms, this

requires the user to have more harmonics in the input design then the number of parameters in the model to be fit (Ljung, 1987). The number of harmonics can be increased by choosing higher values for α and β. Often minimum crest factor multisines will have more than enough levels to meet nonlinear identifiability requirements. The user should verify this for the model under study and consider stopping the crest factor minimization prematurely if this is not the case. The user may also choose a decomposition technique, and an evaluation of minimum crest factor multisines in this context can be found in Braun *et al.* (2000*b*). For nonlinear model identification, development of persistence of identification conditions is model dependent.

Probably the most difficult aspect of signal design for nonlinear system excitation is the choice of signal amplitude. Any steady-state information available from operational records or estimated through "back-of-the-envelope" calculations can reduce the number of iterations for this choice. For processes which are highly interactive, a closed-loop identification approach may be more useful to cover the operating regions of interest, while reducing the number of iterations.

For the user who requires a multisine signal with a nearly white autocorrelation function, the relative magnitudes supplied to the minimum crest factor optimization can be made the same as those for a well-designed multi-level pseudo-random sequence (m-level PRS). For guidelines on the design of m-level PRS, the user is referred to Braun *et al.* (1999) and Braun and Rivera (2000). When strict length requirements are also in effect, the number of sinusoids used from the m-level PRS spectrum can be truncated to n_s to satisfy the requirements in Equation 2. This often has a small effect on the off-peak autocorrelation coefficients.

There are a few drawbacks to the approach presented. The first being that maximum move size cannot be directly specified. If the process under study requires hard bounds on the maximum move size, the user must iteratively adjust the design parameters to try to meet the move size constraint. Additionally, the user must be considerate of the maximum number of levels that can be quantized by the plant's information and control system. If there is a

significant restriction on the number of levels for quantization, the actual power spectrum of the signal injected into the plant may be very different than that the user designed (Barker and Godfrey, 1999).

2.3 Minimum Crest Factor Multisine MIMO Design Guidelines

The idea of Schroeder-phased inputs designed with "zippered" power spectra to create independent excitation for MIMO systems per Rivera *et al.* (1997) can be applied to minimum crest factor multisine signals. For this we consider the perturbation signal $u_l(k)$ for the l-th channel

$$u_l(k) = \lambda_l \sum_{i=1}^{N_s/2} \hat{a}_u(li) cos(\omega_i kT + \phi_{li}), \tag{17}$$

where $l = 1, 2, \ldots, n_u$; λ_l denotes the l-th channel scaling parameter; $\hat{a}_u(li)$ represents the relative amplitude of harmonic i in the l-th channel, $\hat{a}_u(li) = \sqrt{2\alpha_{li}}$. ϕ_{li} represents the phase i in the l-th channel. An estimate must be made of the maximum τ_{dom}^{Hmax}, and minimum τ_{dom}^{Lmin} value of the dominant time constant between all input/output combinations. The user can calculate the control relevant frequency limits

$$\omega_* = \frac{1}{\beta \tau_{dom}^{Hmax}} \leq \omega \leq \frac{\alpha}{\tau_{dom}^{Lmin}} = \omega^* \tag{18}$$

The sampling time and signal length must satisfy

$$T \leq \min \left(\frac{\pi}{\omega^*}, \frac{\pi(1 - \frac{2n_u - 1}{n_u n_s})}{\omega^* - \omega_*} \right) \tag{19}$$

$$N_s \geq \max \left(2n_u n_s + \delta, \frac{2\pi(n_u + \delta)}{\omega_* T} \right) \tag{20}$$

$$\delta = floor \left(\frac{\omega_* N_s T}{2\pi} - n_u \right) \tag{21}$$

The user has the freedom to specify n_s to satisfy persistence of excitation requirements (in the linear model case), provided that

$$n_s \geq \frac{\frac{N_s T \alpha}{2\pi \tau_{dom}^{Lmin}} - \delta - 1}{n_u} + 1 \tag{22}$$

The relative magnitudes of the harmonics are specified such that

$$\hat{a}_u(i) = \begin{cases} lf, & i = 1, \ldots, \delta \\ 1, & i = 1 + \delta, \ldots, n_u n_s + \delta \\ hf, & i = n_u n_s + \delta + 1, \ldots, N_s/2 \end{cases} \tag{23}$$

$$\hat{a}_u(li) = \begin{cases} \hat{a}_u(i), & i = l, l + n_u, \ldots, l + n_u D \\ 0, & \text{otherwise} \end{cases} \tag{24}$$

$$D = \text{fix}\left(\frac{\frac{N_s}{2} - l}{n_u} \right) \tag{25}$$

Note fix(arg) is a MATLAB function that rounds the elements of arg to the nearest integer toward zero. By specifying the harmonic amplitudes as shown above, the input signals u_l ($l = 1, 2, \ldots, n_u$) are orthogonal to each other (ie. they are uncorrelated in a frequency-domain sense). For orthogonality to be satisfied, the non-zero Fourier coefficient in a specific frequency of one signal implies a zero-valued Fourier coefficient at the same frequency for the other signal for all frequencies. This idea is demonstrated in Figure 3.

One disadvantage of minimum crest factor multisine signals for multivariable excitation is that since the inputs are distributed to support the limits of the inputs, as the dimension of the multivariable problem increases, less and less support is provided for the interior of the regressor spaces. However, as noted before, the minimization problem can be terminated prematurely, allowing for signals with suboptimal crest factors, but which may better support the interior of the regressor spaces. Another symptom of the "curse-of dimensionality" is that as the number of inputs n_u increases, the frequencies contained in each individual multisine become spaced further and further apart.

2.4 Input Signal Design for Validation

To create a meaningful excitation signal for validation, many options are possible. The user may decide to simply repeat the estimation input signal, using one cycle for estimation and the other for validation. This would provide validation data which contains all the theoretical

requirements noted above. For nonlinear systems, this is not always advisable since this approach will not test the interpolative ability of the modeling technique being used, nor will it test the ability of estimation data to support the trajectories used for controlling the system.

An alternative technique for validation is to design inputs which have characteristics (e.g. shape, rates of change, trajectory) similar to the expected control action. These could be typical of inputs observed in operational data, or simply step inputs based on steady state information. The user can design these inputs to test the interpolative abilities of the modeling technique by choosing input levels other than those of the estimation input. This type of validation input design is intuitively appealing. A disadvantage may be that it does not excite the system over the control-relevant frequency range and does not take into account the theoretical requirements noted above.

One last possibility is to use another well designed minimum crest factor multisine signal that has been initialized with random phase initialization. If time limits for identification testing are flexible, the user may choose to combine the approaches just mentioned.

3 Model-on-Demand Estimation and Model Predictive Control

The MoD modeling formulation pursued in this paper follows from the approach of Stenman *et al.* (1996), and Stenman (1999). Consider a SISO process with nonlinear ARX structure, i.e.,

$$y(k) = m(\varphi(k)) + e(k), \qquad k = 1, ..., M, \tag{26}$$

where $m(\cdot)$ is an unknown nonlinear mapping and $e(k)$ is an error term modeled as i.i.d. random variables with zero mean and variance σ_k^2. The MoD predictor attempts to estimate a value $\hat{y}$ based on a local neighborhood of the regressor space $\varphi(t)$. The regressor vector is of the form

$$\varphi(t) = [y(t-1) \dots y(t-n_a) \quad u(t-n_k) \dots u(t-n_b-n_k)]^T \tag{27}$$

where n_a, n_b, and n_k denote the number of previous outputs, inputs, and delay in the model.

A local estimate $\hat{y}$ can be obtained from the solution of the weighted regression problem

$$\hat{\beta} = \arg \min_{\beta} \sum_{k=1}^{N} \ell(y(k) - m(\varphi(k), \beta)) \cdot W\left(\frac{\|\varphi(k) - \varphi(t)\|_M}{h}\right) \tag{28}$$

where $\ell(\cdot)$ is a scalar-valued and positive norm function, $\|u\|_M \stackrel{\triangle}{=} \sqrt{u^T M u}$ is a scaled distance function on the regressor space, h is a *bandwidth* parameter controlling the size of the local neighborhood N, and $W(\cdot)$ is a window function (usually referred to as the *kernel*) assigning weights to each remote data point according to its distance from $\varphi(t)$. The window is typically a bell-shaped function with bounded support. A tricube window function is chosen since it has a continuous derivative and smoothly descends to zero at the boundaries. This has the effect of reducing variance in the estimate (Stenman, 1999).

Assuming a local *linear* model structure,

$$m(\varphi(k), \beta) = \beta_0 + \beta_1^T(\varphi(k) - \varphi(t)), \tag{29}$$

a quadratic norm, $\ell(\varepsilon) = \frac{1}{2}\varepsilon^2$ is used and the model is linear in the unknown parameters. The estimate can be easily computed using least squares methods. If $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the minimizers of (28) using the model from (29), a one-step ahead prediction is given by

$$\hat{y}(t) = m(\varphi(t), \hat{\beta}) = \hat{\beta}_0. \tag{30}$$

Each local regression problem produces a single prediction $\hat{y}(t)$ corresponding to the current regression vector $\varphi(t)$. To obtain predictions at other locations in the regressor space, the weights change and new optimization problems have to be solved. This is in contrast to the global modeling approach where the model is fitted to data only once and then discarded. However, in a neighborhood around $\varphi(t)$, the local linear model provides an input-output linearization

$$A(q^{-1})y(t) = B(q^{-1})u(t - n_k) + \alpha, \tag{31}$$

where $A(q^{-1})$ and $B(q^{-1})$ are polynomials in the backward time-shift operator q^{-1} obtained from the components of $\hat{\beta}$, and

$$\alpha = \hat{\beta}_0 - \hat{\beta}_1^{\,T}\varphi(t) \tag{32}$$

is an offset term.

The bandwidth h controls the neighborhood size and has a critical impact on the resulting estimate since it governs a trade-off between the bias and variance errors of the estimate. Traditional bandwidth selectors produce a single global bandwidth; in MoD estimation, bandwidth is computed adaptively at each prediction. MoD selects candidate bandwidths using the following exponential update.

$$h_i = C_h \cdot h_{i-1} \tag{33}$$

where

$$C_h = 1 + \frac{0.3}{d}. \tag{34}$$

d denotes the dimension of the regressor space. h_0 is determined by a user defined parameter N_{min} which represents the least amount of data to be used for a local model estimate. MoD continually constructs local models, storing them along with the corresponding bandwidth h_i until the goodness-of-fit measure fails at a low level of significance, or N_{max} is exceeded. N_{max} is also a user-defined parameter. The optimal bandwidth h_{opt} is determined from the stored values, and the corresponding model is used for output prediction. Various goodness-of-fit measures are available for this purpose (Stenman, 1999). The default method used in this paper is the localized Akaike Information Criterion (AIC).

To incorporate the MoD estimator into an MPC controller, consider the MPC objective function given in Meadows and Rawlings (1997). Given the model description in Equation 31 and knowledge of the current system state, we seek a control that minimizes the objective function

$$J = \sum_{k=0}^{N-1} Q_e(k)(r(t+k+1) - \hat{y}(t+k+1))^2 + Q_u(k)u^2(t+k) + Q_{\Delta u}(k)\Delta u^2(t+k) \tag{35}$$

where $Q_e(k)$, $Q_u(k)$ and $Q_{\Delta u}(k)$ represent penalty weights on the control error, control signal, and control move size, respectively. Of the N_u future control actions that minimize J, only the first one is used. As new measurements become available, a new optimization problem is formulated whose solution provides the next control action. This is referred to as

the receding horizon principle. The main advantage of MPC is that hard or soft constraints can be specified by the user on move size, input magnitude, states or output.

Since optimization of Equation 35 can be computationally intensive for large prediction horizons N, often only 4 or 5 control increments are solved, with the implicit assumption that the control action will be held constant for the remaining moves. The practice of using smaller control horizons N_u has the effect of producing less aggressive controllers and providing stable control for non-minimum phase systems (Meadows and Rawlings, 1997).

To adopt the Generalized Predictive Control approach (Clarke *et al.*, 1987), this work makes use of a Control ARIMA model (CARIMA), shown in Equation 36, where $\Delta = 1 - q^{-1}$. This forces the controller to have integral action, since implicitly, the disturbance model is assumed to be nonstationary.

$$A(q^{-1})\Delta y(t) = B(q^{-1})\Delta u(t) \tag{36}$$

By making use of the Generalized Predictive Control formulation, the objective can be expressed in vector/matrix form as

$$\begin{aligned} J = &\| \boldsymbol{r} - \bar{\boldsymbol{y}} - \boldsymbol{S}\boldsymbol{\Delta u} \|^2_{Q_e} + \\ &\| \boldsymbol{T}\boldsymbol{\Delta u} + \bar{\boldsymbol{u}} \|^2_{Q_u} + \| \boldsymbol{\Delta u} \|^2_{Q_{\Delta u}} . \end{aligned} \tag{37}$$

$\mathbf{r}$ denotes the desired refe rence trajectory; $\bar{\mathbf{y}}$ denotes measured and known effects on the predicted output; $\mathbf{S}$ transforms the change in controls Δu into their effects on the predicted output; $\mathbf{T}$ relates the change in controls, such that the magnitude of the control can be penalized, if need be. For the *unconstrained* case, the minimizing control sequence is obtained explicitly by ordinary least squares theory. For the *constrained* case, the constraints can be re-formulated in matrix/vector form and the problem is solved efficiently using standard numerical optimization algorithms. More details on formulation of MoDMPC, and conditions to ensure stability, output constraint infeasibility handling, and additional feedforward ideas can be found in Stenman (1999), and Braun *et al.* (2000*a*).

4 Case Studies

Two case studies are presented to illustrate the benefits of the proposed methodology. Trade-offs that must be made in the design are examined based on analysis of support in the input/output spaces and performance of the MoDMPC controllers utilizing the input/output data.

4.1 CSTR SISO Identification and Control

The CSTR model is based on a first principles analysis of a hypothetical CSTR (Denn, 1986). The vessel is assumed to be perfectly mixed, and a single first-order exothermic, irreversible reaction, A $\Rightarrow$ B, takes place. A diagram showing the vessel and the surrounding cooling jacket is presented in Figure 4. Based on accounting and conservation principles, the lumped parameter equations describing the system are

$$\frac{dC_A}{dt} = \frac{F}{V}(C_{A_f} - C_A) - r \tag{38}$$

$$\frac{dT}{dt} = \frac{F}{V}(T_f - T) - \left(\frac{\Delta H}{c_p \rho}\right) r - \frac{UA}{c_p \rho V}(T - T_j) \tag{39}$$

$$r = k_o exp\left(\frac{-\Delta E}{RT}\right) C_A \tag{40}$$

This example is concerned with the Case 2 reactor parameter values presented by Bequette (1998) and shown in Table 1. The inlet stream is fed at a constant rate F, with constant concentration C_{Af} into the vessel. The final concentration of the reactant C_A is the controlled variable and the jacket temperature T_j is manipulated to keep the exit stream concentration C_A at setpoint. The exiting stream leaves at a rate F and since it is assumed the vessel is perfectly mixed, the exiting concentration and vessel concentration are assumed to be the same.

The input amplitude range chosen is $\pm 25\ K$. This range is particularly interesting because the diabatic CSTR exhibits an "ignition" as the reactor settles from a $+9\ K$ or greater increase in jacket temperature.

Table 1: Diabatic CSTR, Case 2 reactor parameters

Parameter	Units	Value
F/V	hr^{-1}	1
k_0	hr^{-1}	9703*3600
$(-\Delta H)$	kcal/kgmol	5960
E	kcal/kgmol	11843
ρc_p	kcal/($\text{m}^3{}^o\text{C}$)	500
T_f	^{o}C	25
C_{Af}	kgmol/m^3	10
UA/V	kcal/($\text{m}^3{}^o\text{Chr}$)	150
T_j	^{o}C	25

The goal is to develop a nonlinear Model-on-Demand Model Predictive Controller (MoD-MPC) to regulate production about the high concentration, low temperature steady-state. To achieve this end, 4 multisine signals were designed and 1 m-level PRS signal was designed for excitation of the CSTR. For a discussion on design of m-level PRS signals, the user is referred to Braun and Rivera (2000). Signal designs of length beyond one day's time were considered unacceptable. Previous attempts to formulate a linear PID controller revealed that a high bandwidth controller was needed to keep the reactor from "igniting." Therefore, an α of 5 and a β of 1 were selected for design of all 5 signals. A sampling time of 0.1 hrs was used for all signals and measurements. The validation input was designed as a combination of a swept multisine and random steps, such that the validation input power spectrum was well within the profiles of all of the estimation input power spectra. The amplitude and levels of the validation input were designed to test the interpolative ability of a MoD estimator using the estimation datasets.

For the minimum crest factor multisine signals, the combinations of shapes evaluated are presented in Table 2. The low frequency limit dictated that the multisine signals should have a length of 12.6 hrs. Figures 5 and 6 contain the time domain realizations of all 5 input

signals and the validation signal, and the power spectra of the input signals, respectively. From a "plant-friendliness" standpoint, the multisines are more attractive than the m-level PRS, mainly because the maximum step sizes are smaller than that for the m-level PRS. Another attractive feature is that the length of the multisine signal can be designed to precisely meet the low frequency limit.

Table 2: CSTR min cf multisine design configurations

Input Type	Harm. Sup'd h	Hi Freq. Amp. hf	Shape	Crest Factor cf	Max Move Size mms	Length (hrs)
multisine	none	0	flat	1.44	8.68	12.6
multisine	none	0.1	flat	1.24	30.48	12.6
multisine	even	0.1	flat	1.28	37.38	12.6
multisine	even	N/A	PRS	1.27	24.51	16.8
m-level PRS	even	N/A	PRS	1.39	50	16.8

By observing Figures 5 and 6, and taking into account the crest factor values listed in Table 2, again note that by allowing the multisine design some additional high frequency components, the crest factor can be significantly reduced. Harmonic suppression is observed to produce signals that have greater power in the unsuppressed frequencies. The validation signal excites the CSTR within the input range and provides power spectra and output well within the output limits of the 5 excitation signals. Figures 7 and 9 show the output and input/output space covered by the excitation signals, respectively. The legend of Figure 8 provides the root mean square (RMS) and maximum (MAX) errors of MoD estimation using a NARX [2 2 1] regressor structure, the AIC criteria with a 50 data point lower bound for local bandwidth selection, and a locally linear model structure. Figure 10 presents the controlled and manipulated variables for controlling the CSTR with a MoD-MPC controller with a constant configuration ($Q_e = 1$, $Q_u = 0$ and $Q_{\Delta u} = 0.001$) for all databases. The prediction horizon covered the 66% settling time and the control horizon included 3 moves. Note that the RMS and MAX setpoint errors are presented in the top plot of Figure 10 and

the maximum move size (MMS) and total move energy (TME) are included in the bottom plot of Figure 10. The control experiment was subject to two ramp setpoint changes and one step setpoint change at times 1, 5, and 15, respectively.

What can be concluded from the previous analysis is that in a control relevant sense, multisine signals can produce an input/output database which is just as informative as that from m-level PRS excitation and the database is capable of supporting the MoD estimator. While the signal length is 4.2 hrs shorter, validation using the even suppression minimum crest factor multisine provides better RMS error and similar maximum error compared to the 5-level signal. The main issues when choosing a multisine power spectrum are the uniformity and resolution the time-domain realization of that spectrum produces in the input/output space. For the multisines with harmonic suppression, better fits to the validation occur and less control energy is used for MoD-MPC because the support in the validation and control regressor spaces is more uniform. This is due to the inverse repeat property of signals with even harmonic suppression. This property naturally imposes a symmetric distribution in the input space.

4.2 High Purity Distillation MIMO Identification and Control

Distillation is one of the most important separation techniques in the chemical process industry (CPI). Distillation columns are considered in general, as a challenging control problem because of the inherent nonlinearities in the behavior of many process elements. In this section, a high-purity binary distillation column is studied. This system is considered as a challenging problem because of the following properties (Chien and Ogunnaike, 1992):

1. *Asymmetric nonlinearities.* Mirror image inputs produce corresponding responses that are fundamentally and significantly dissimilar in character.

2. *Interactions.* A change in any one input variable affects multiple outputs.

3. *Ill-conditioning.* The response characteristics such as steady state gain and time constants are strongly dependent on the direction and the magnitude of the manipulated

variable change.

This problem was first presented in the work of Weischedel and Mc Avoy (1980), and was further studied by Chien and Ogunnaike (1992) and Li and Lee (1994), among others. Gaikwad and Rivera (1997) studied the column using a MIMO PRBS input and one-input-at-a-time (i.e. SIMO) Schroeder-phased input. The work in Zong (1997) studied the column using a MIMO "zippered" Schroeder-phased signal. In this study, MIMO "zippered" minimum crest factor signals are developed with the techniques presented in Section 2 and used as the setpoint trajectory for a closed loop identification strategy. A closed-loop identification strategy is shown to cover the expected steady-state input and output spaces necessary for a nonlinear controller to transition the process between two operating conditions. With this data, a Model-on-Demand Model Predictive Controller is implemented to demonstrate the advantages of nonlinear control and compare the results to those previously reported.

The system under study is a methanol-ethanol distillation column with 27 trays and 56 states. The model of the system was developed using an overall and component mass balance for each tray in the column, pseudo steady-state energy balances, the Francis-Weir formula for liquid flow from each tray, and vapor-liquid equilibria relations. The controlled variables are the tray temperatures on tray 7 (bottom control point) and tray 21 (top control point). The reflux flow, L, and the vapor boil-up flow, V, are used as manipulated variables. Table 3 shows the operating information for the column.

The objective of control is to independently control the compositions of the distillate (tray 7) and bottom (tray 21) products. Two setpoint changes are traditionally reported in literature dealing with this problem. The first is a $+0.05$ temperature change at tray 7, while the temperature at tray 21 is held constant. The second is a $+0.05$ temperature change at tray 21, while the temperature at tray 21 is held constant. Control is based on the temperature instead of the compositions since it is reliable and easy to implement on-line.

A "zippered" minimum crest factor multisine signal was designed and generated for each input of the distillation column simulation. The dominant time constant for this system

Table 3: High purity distillation column operating information

Mixture	Methanol-Ethanol
Product Split	0.01–0.99
Number of Trays	27
Top Temperature	$608.0°R$
Bottom Temperature	$642.24°R$
Vapor Flow from Reboiler	$3.856\ mol/min$
Reflux Flow Rate	$3.384\ mol/min$
Feed Flow Rate	$1.50\ mol/min$
Feed Composition	0.5

ranges from 5 to 33 minutes. Previous attempts by other researchers to formulate a linear controller revealed that a high bandwidth controller was needed to achieve good control performance (Gaikwad and Rivera, 1997). Therefore, an $\alpha = 2$, $\beta = 1$, were selected to design the input signals. Thus, the input signals were purposely designed to capture information over a frequency band from 0.04 to 1.09 $radian/min$. The low frequency (lf) and high frequency (hf) relative magnitudes for the notch "zippered" power spectrum are 0.9 and 0.1, respectively. A sampling time of 1 $min.$, and a signal range of ± 0.1 was used for both signals. Using the multi-input guidelines resulted in a signal of length, $T \cdot N_s = 472\ min$, and $n_s = 40$ sinusoids. At first, a signal length of 472 may seem long, yet it is still shorter than most of the signals reported to test this system (Ravi Sriniwas *et al.*, 1995; Li and Lee, 1996).

Figure 11 shows the time domain realization of the excitation signals. Both signals display similar crest factor values. An initial investigation of open-loop MIMO excitation of the system did not provide sufficient coverage of the steady-state operating conditions associated with both setpoint changes. Many iterations with different selection of input amplitudes were performed and the space coverage shown in Figure 13, came the closest to covering both setpoints. Excitation in the weak directions to cover both setpoints was not possible with an open-loop identification strategy. Thus, these signals were used as setpoint trajec-

tories for the PID controllers developed by Chien and Ogunnaike (1992). The corresponding input and output data from the experiment are shown in Figure 12. Note that both the input and output spaces have coverage between the initial conditions and the final steady state values expected for the two setpoint changes as shown in Figure 13. The closed-loop identification experiment overcomes the limitations of the open-loop experiment by forcing the column in the weak directions and providing data coverage over both setpoints.

A MoDMPC controller, with a prediction horizon $N_p = 25$, control horizon $N_u = 10$, control error weighting of identity $Q_e = 1$ and move suppression weighting of $Q_{\Delta u} = 20$ for both inputs was implemented. The lower bounds on the neighborhood size $k_{min} = 75$, an AIC goodness-of-fit measure, along with a variance penalty $\alpha = 3$ was validated for the MoD algorithm. The control of the column in direction of setpoint # 1 for the approaches shown in the literature can be viewed in Figures 14 and 15 and contrasted against the performance of the MoDMPC controller. The control in the direction of the second setpoint is not included for the sake of brevity. The MoDMPC results demonstrate no overshoot and a settling time 20 *min* shorter than other controllers for the tray 21 temperature. The MoDMPC tray 7 temperature exhibits a maximum deviation half the size of the maximum deviation on the control variable allowed by the other controllers. The MoDMPC tray 7 temperature deviation also exists for half the time of the shortest deviation allowed by the other controllers. Both manipulated variable responses for the MoDMPC controller are much smoother in comparison to the performance of the other controllers.

5 Conclusions

Design guidelines for minimum crest factor multisines have been developed to translate the user's available *a priori* information into sensible design parameter selections for identification of nonlinear processes. These guidelines can be applied in both the single input and multiple input cases. Well designed minimum crest factor multisines can be used for linear identification, or they can support nonlinear controllers such that the resulting performance

is superior to linear control techniques. This was demonstrated using Model-on-Demand estimation and Model-on-Demand Model Predictive Control in the CSTR and Weischedel/Mc Avoy High Purity Column case studies.

6 Acknowledgments

The authors would like to thank Patrick Guillaume of Vrije Universiteit Brussel, Belgium for use of his minimum crest factor optimization code and Anders Stenman of Linköping University, Sweden for use of his MoDMPC code.

References

Barker, H. A. and K. R. Godfrey (1999). System identification with multi-level periodic perturbation signals. *Control Engineering Practice* **7**, 717–726.

Barker, H.A. and M. Zhuang (1997). Design of pseudo-random perturbation signals for frequency-domain identification of nonlinear systems. In: *Proceedings of the llth IFAC Symposium on System Identification*. Kitakyushu, Japan. pp. 1635–1640.

Bequette, B. W. (1998). *Process Dynamics: Modeling, Analysis, and Simulation*. Prentice Hall PTR. Upper Saddle River, NJ.

Braun, M. W. and D. E. Rivera (2000). Design of multi-level pseudo-random sequences for control-oriented identification of nonlinear process systems. *Submitted for Publication*.

Braun, M. W., D. E. Rivera and A. Stenman (2000*a*). A 'model-on-demand' identification and control methodology for nonlinear process systems. *Submitted for Publication*.

Braun, M.W., D.E. Rivera, A. Stenman, W. Foslien and C. Hrenya (1999). Multi-level pseudo-random signal design and "model-on-demand" estimation applied to nonlinear identification of a rtp wafer reactor. In: *Proc. of the American Control Conf., San Diego, CA, USA, 2-4 June 1999*. pp. 1573–1577.

Braun, M.W., R. Ortiz-Mojica and D.E. Rivera (2000*b*). Design of minimal crest factor multisinusoidal signals for "plant-friendly" identification of nonlinear process systems. In: *Proceedings of the l2th IFAC Symposium on System Identification, Santa Barbara, CA, USA, 2-4 June 2000.*

Chien, I-L. and B. A. Ogunnaike (1992). Modeling and control of high-purity distillation columns. In: *1992 AIChE Annual Meeting.* Miami Beach, FL. paper 2a.

Clarke, D. W., C. Mohtadi and P. S. Tuffs (1987). Generalized predictive control – part i. the basic algorithm. *Automatica* **23**(2), 137–148.

Denn, M. M. (1986). *Process Modeling.* Longman. New York.

Gaikwad, S. V. and D. E. Rivera (1997). Multivariable frequency-response curve fitting with application to control-relevant parameter estimation. *Automatica* **33**(6), 1169–1174.

Godfrey, K. (1993). *Perturbation Signals For System Identification.* Chap. Introduction to Perturbation Signals for Frequency-domain System Identification, pp. 60–125. Prentice Hall International. Hertfordshire, UK.

Guillaume, P., J. Schoukens, R. Pintelon and I. Kollár (1991). Crest-factor minimization using nonlinear chebyshev approximation methods. *IEEE Trans. on Inst. and Meas.*

Li, W. and J. H. Lee (1994). Closed-loop identification of ill-conditioned multi-variable systems. In: *AIChE Annual Meeting.* San Francisco, CA.

Li, W. and J. H. Lee (1996). Frequency-domain closed-loop identification of multivariable systems for feedback control. *AIChE Journal* **42**(10), 2813–2827.

Ljung, L. (1987). *System Identification: Theory for the User.* Prentice Hall PTR. Upper Saddle River, NJ.

Meadows, E. S. and J. B. Rawlings (1997). *Nonlinear Process Control.* Chap. Model Predictive Control, pp. 233–310. Prentice Hall PTR. Upper Saddle River, NJ.

Ravi Sriniwas, G., Y. Arkun, I-L. Chien and B. A. Ogunnaike (1995). Nonlinear identification and control of a high-purity distillation column: a case study. *Journal of Process Control* **5**(3), 149–162.

Rivera, D. E., S. Zong and W. Ling (1997). A control-relevant multivariable system idenfification methodology base on orthogonal multifrequency input perturbations. In: *Proceedings of the llth IFAC Symposium on System Identification*. Kitakyushu, Japan. pp. 595–600.

Schroeder, M. R. (1970). Synthesis of low-peak-factor signals and binary sequences with low autocorrelation. *IEEE Trans. Info. Theory*.

Seborg, D. E. and M. A. Henson (1997). *Nonlinear Process Control*. Chap. Introduction, pp. 1–8. Prentice Hall PTR. Upper Saddle River, NJ.

Simon, G. and J. Schoukens (2000). Robust broadband periodic excitation design. *IEEE Trans. on Instrumentation and Measurement* **49**(2), 270–274.

Stenman, A. (1999). Model on Demand: Algoritms, Analysis and Applications. PhD thesis. Linköping University. SE-581 83 Linköping, Sweden.

Stenman, A., F. Gustafsson and L. Ljung (1996). Just-in-time models for dynamical systems. In: *Proc. IEEE CDC*. Kobe, Japan. pp. 1115–1120.

Weischedel, K. and T. J. Mc Avoy (1980). Feasibility of decoupling in conventionally controlled distillation columns. *Ind. Eng. Chem. Fundam.* **19**, 379–384.

Zong, S. (1997). Minimal crest factor input design and frequency domain curve fitting for control relevant identification. Master's thesis. Arizona State University.

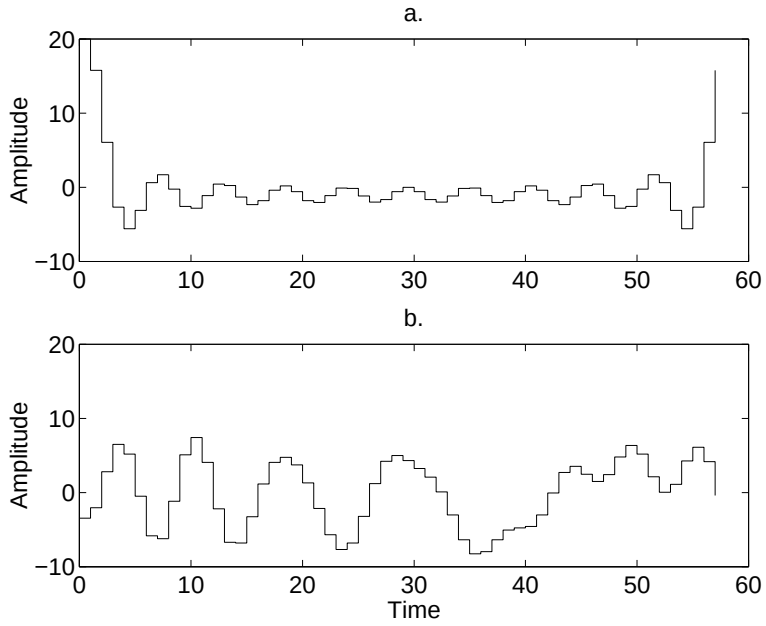

Figure 1: a.) multisine signal with $\phi_i = 0 \; \forall \; i$, b.) multisine signal with phases chosen per Equation 5

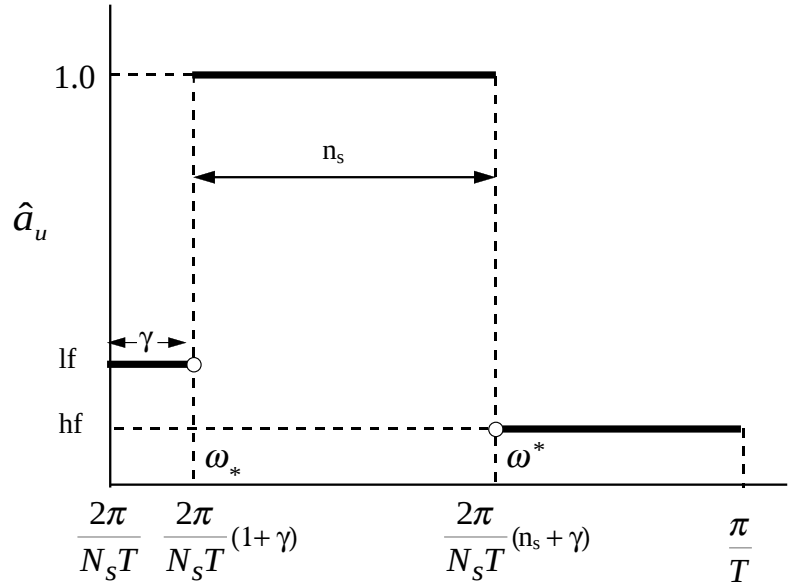

Figure 2: Power spectra design for single-input excitation

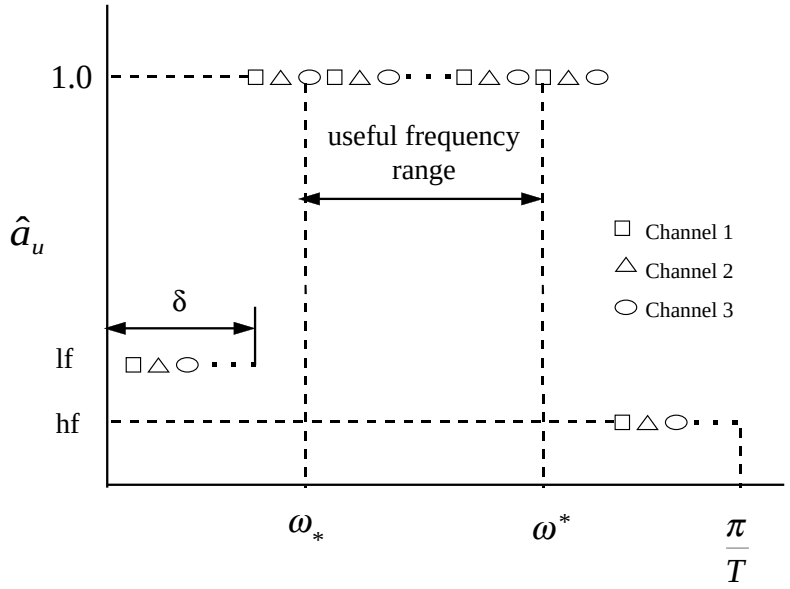

Figure 3: "Zippered" power spectra for multi-input excitation

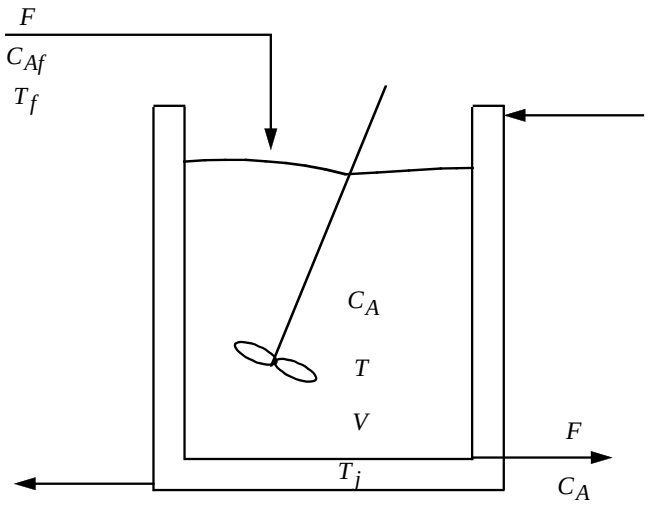

Figure 4: Schematic of the diabatic CSTR

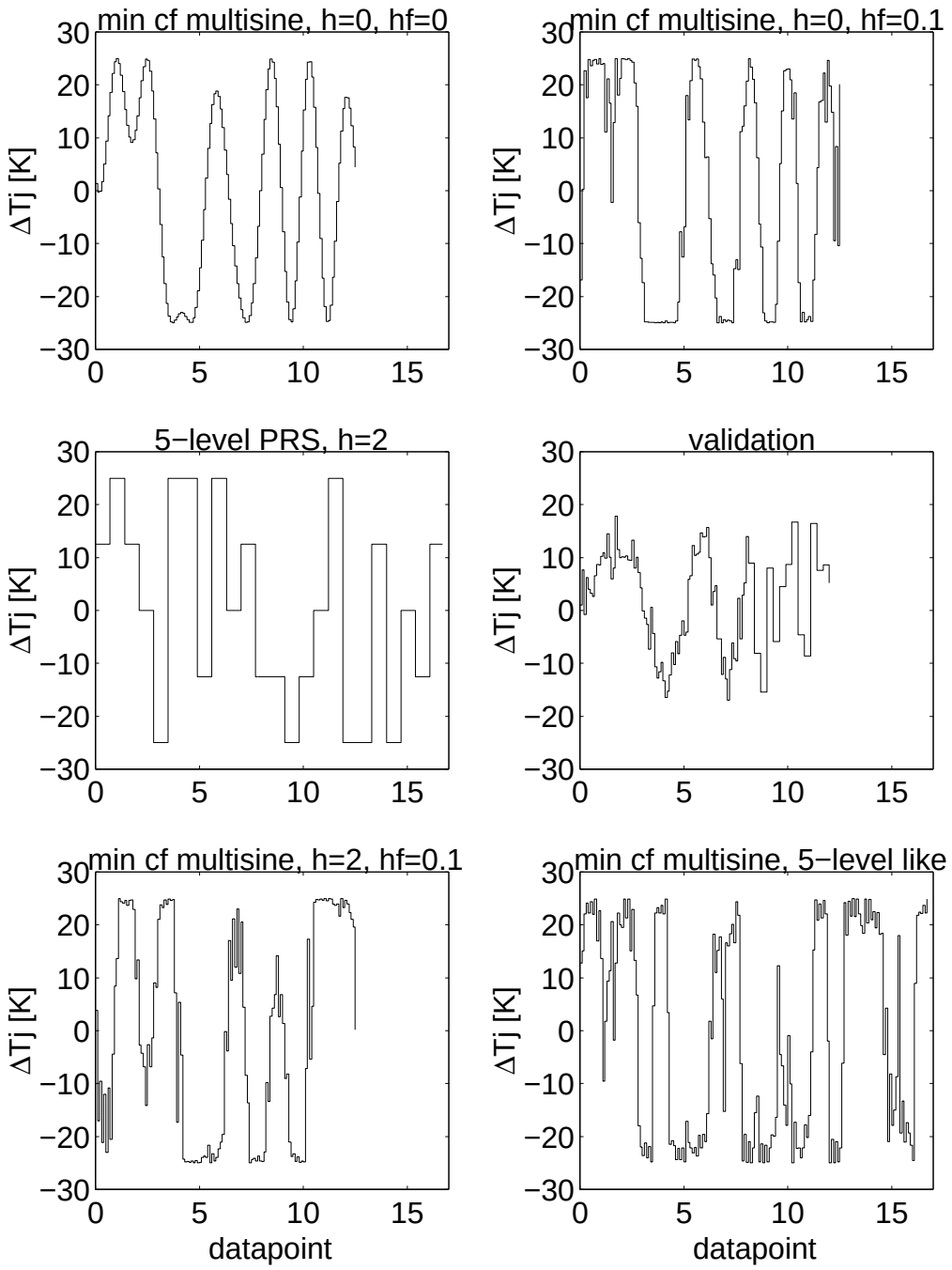

Figure 5: CSTR excitation and validation signals

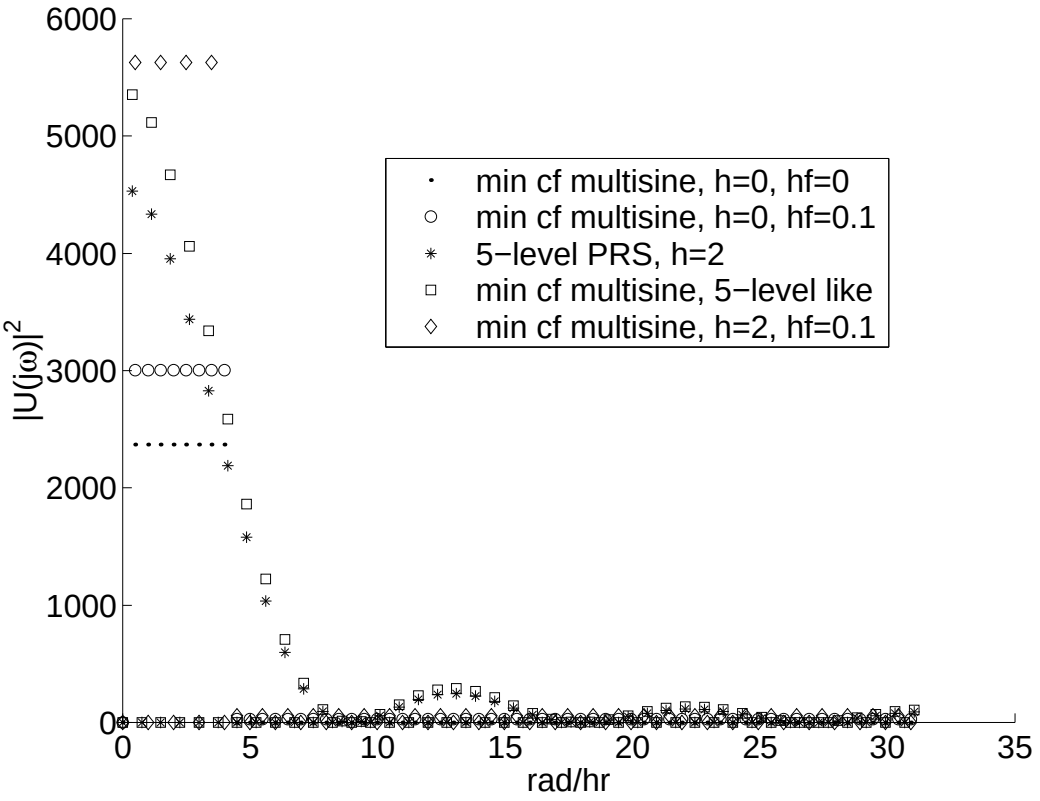

Figure 6: CSTR excitation power spectra

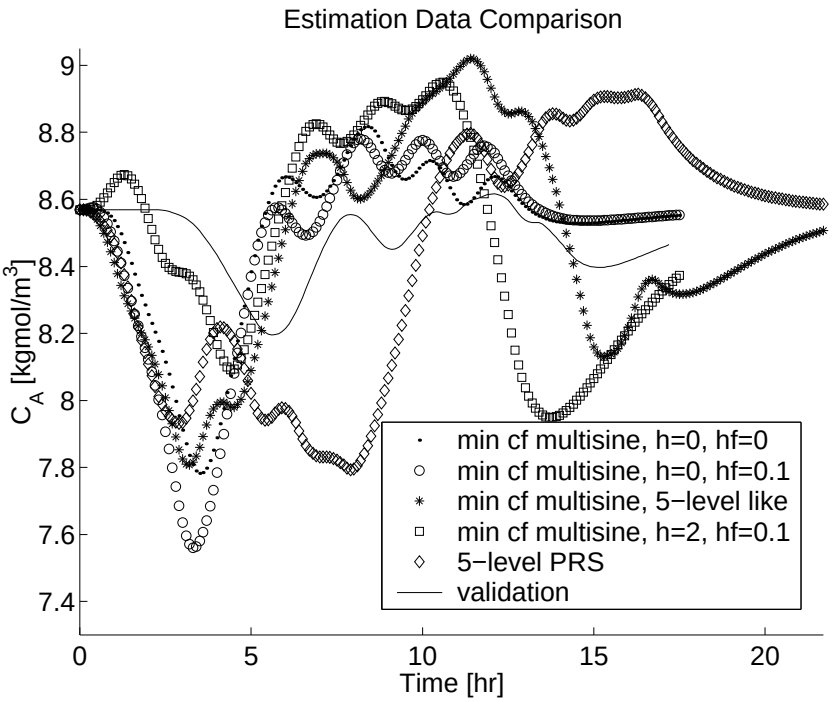

Figure 7: CSTR output data

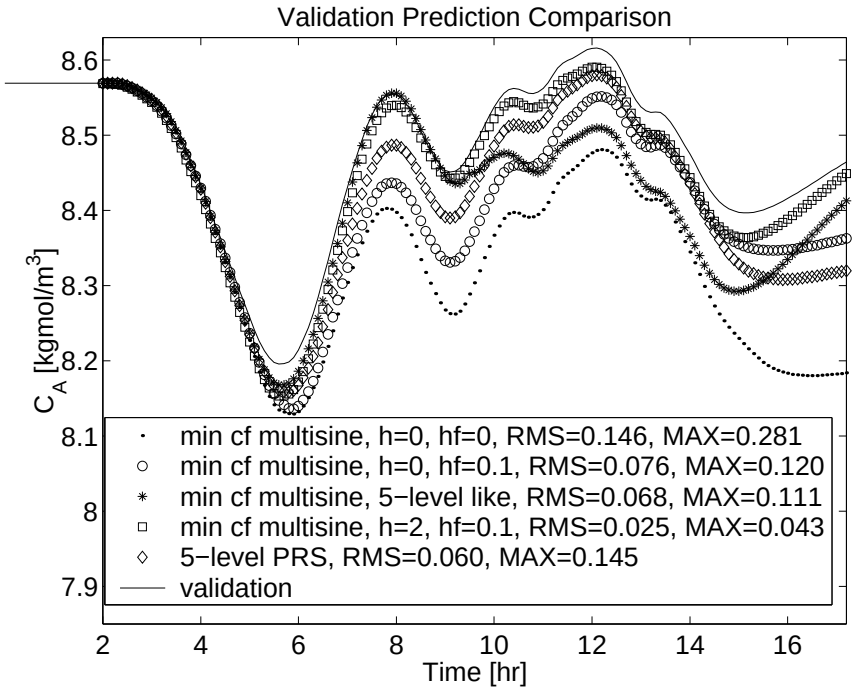

Figure 8: CSTR database validation

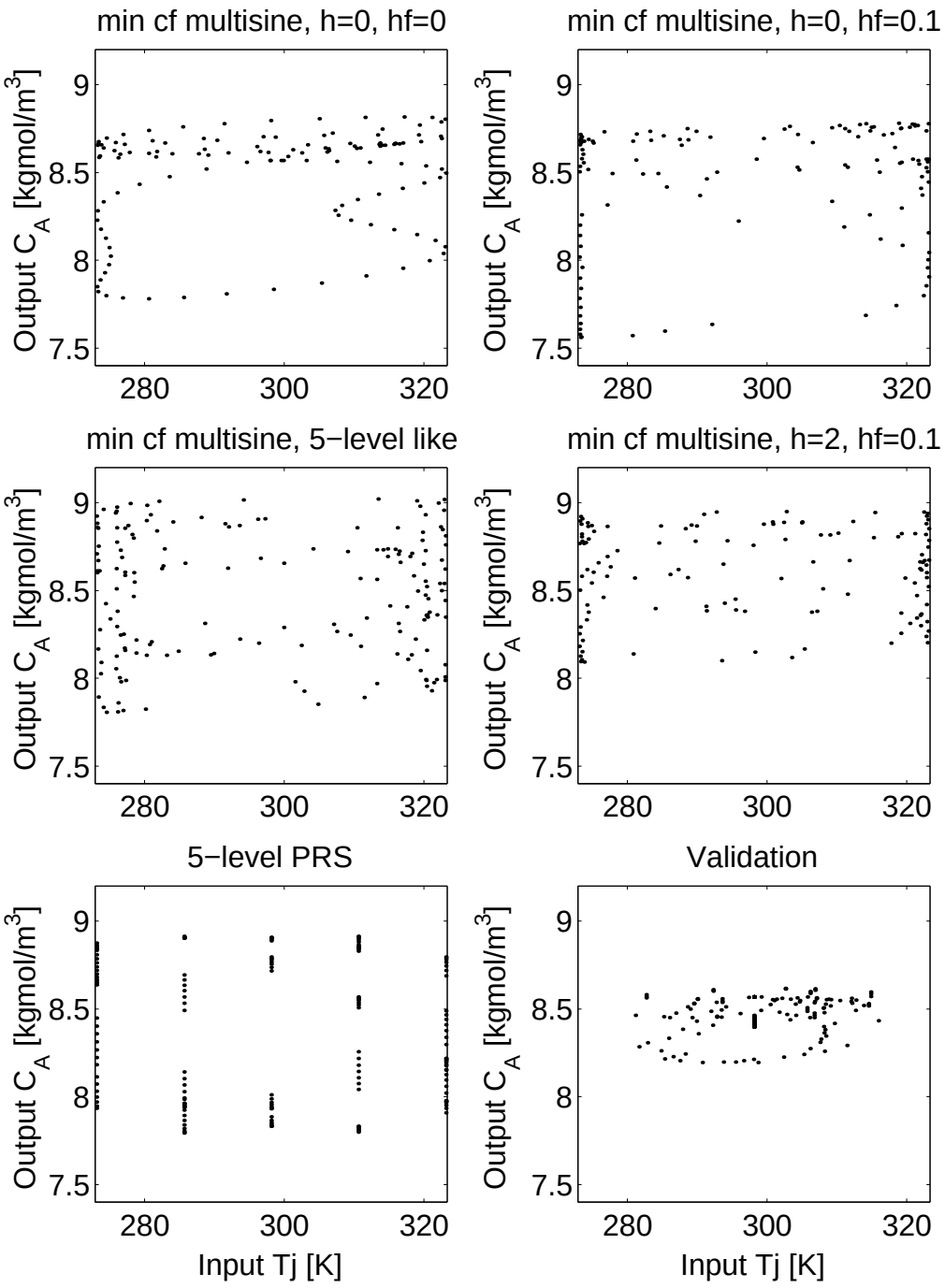

Figure 9: CSTR input/output spaces

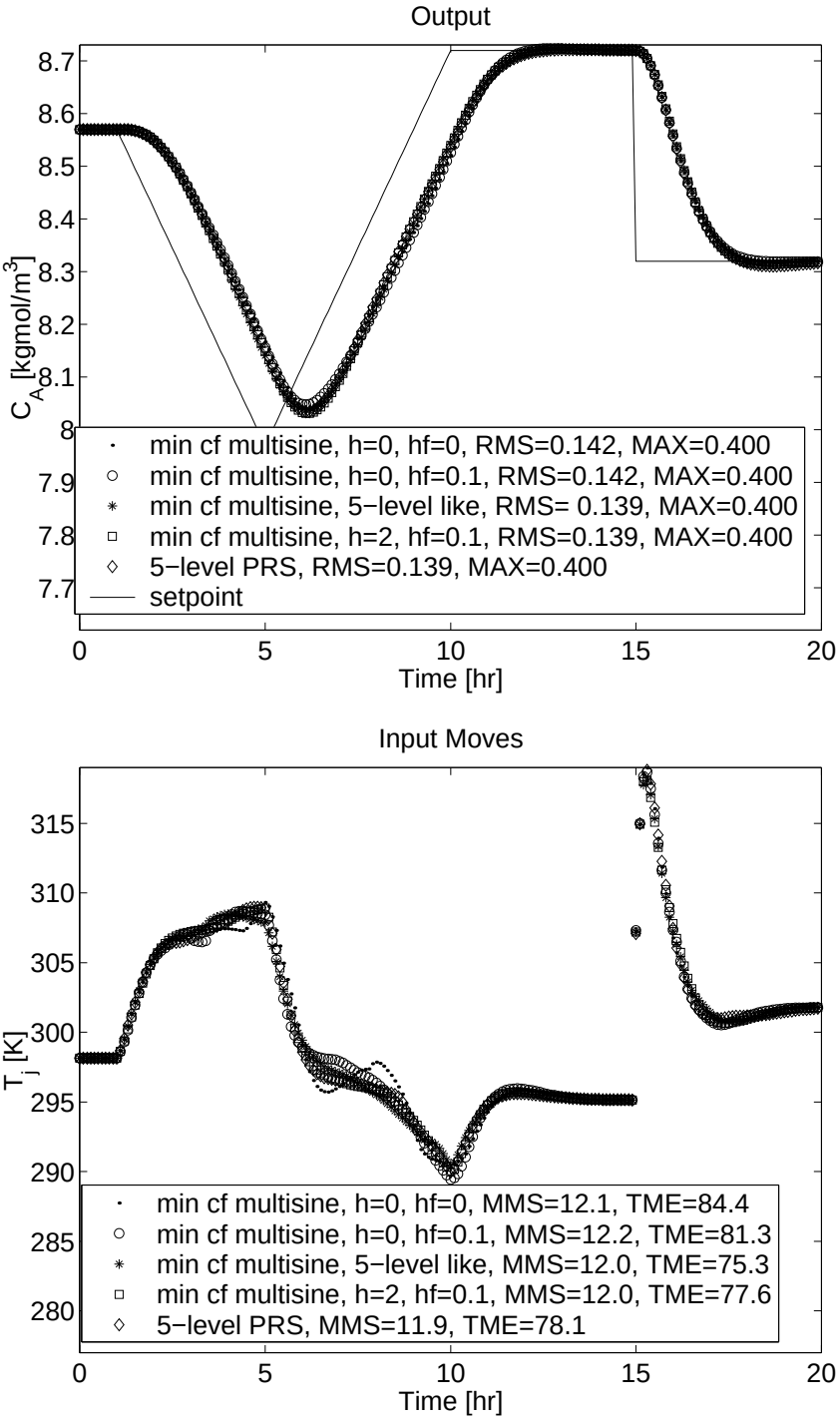

Figure 10: CSTR control (top) and manipulated (bottom) variable responses

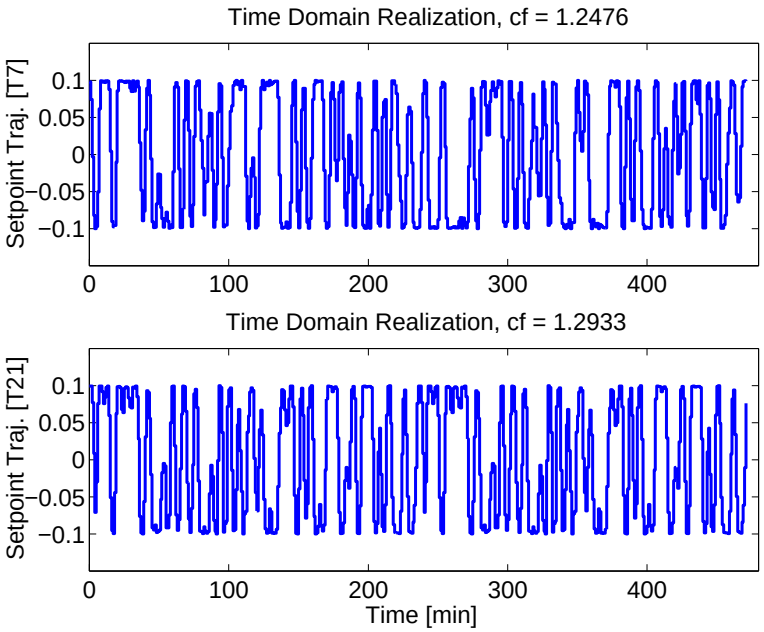

Figure 11: Time domain realization of the MIMO "zippered" minimum crest factor multisine signals

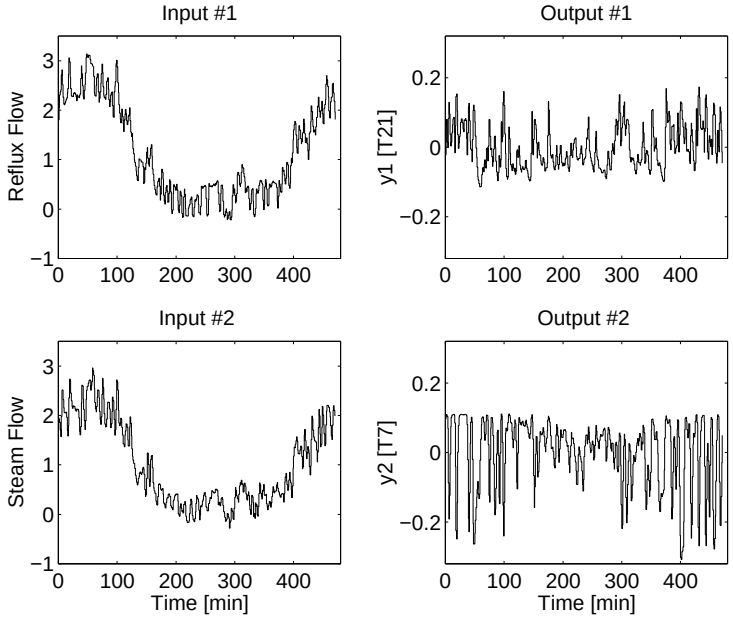

Figure 12: Closed-loop identification manipulated and control variable responses

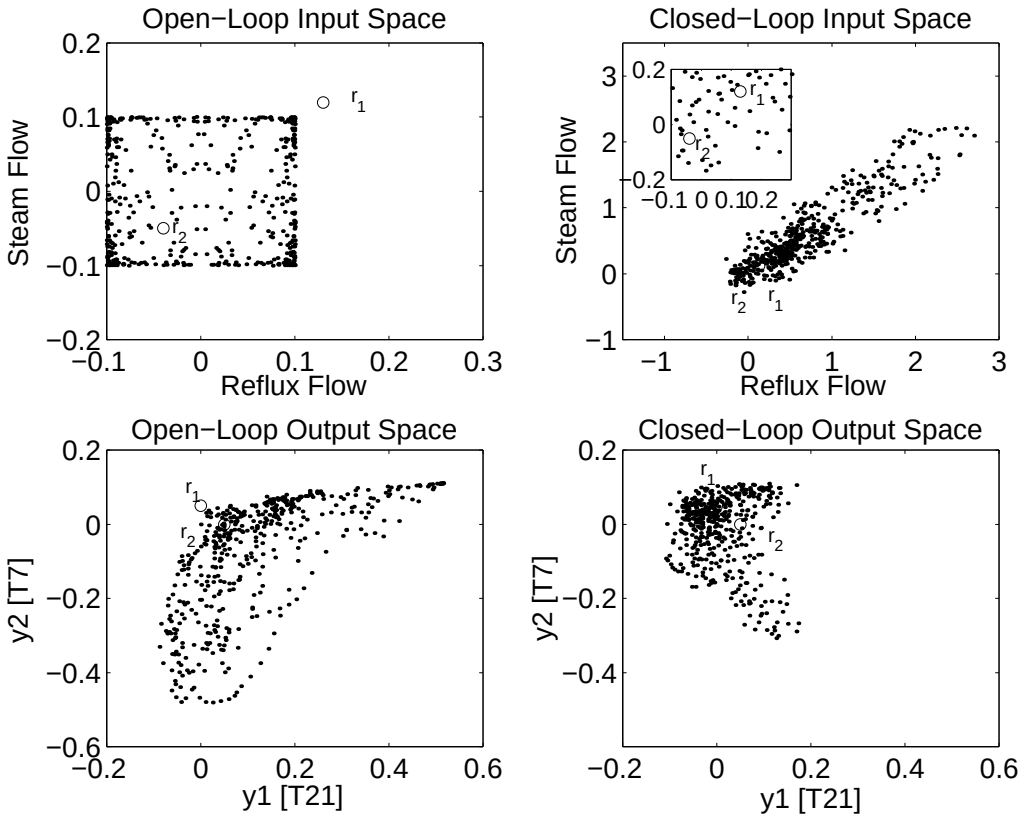

Figure 13: Input and output spaces covered by open and closed-loop identification, respectively; r_1-setpoint change for tray 7; r_2-setpoint change for tray 21

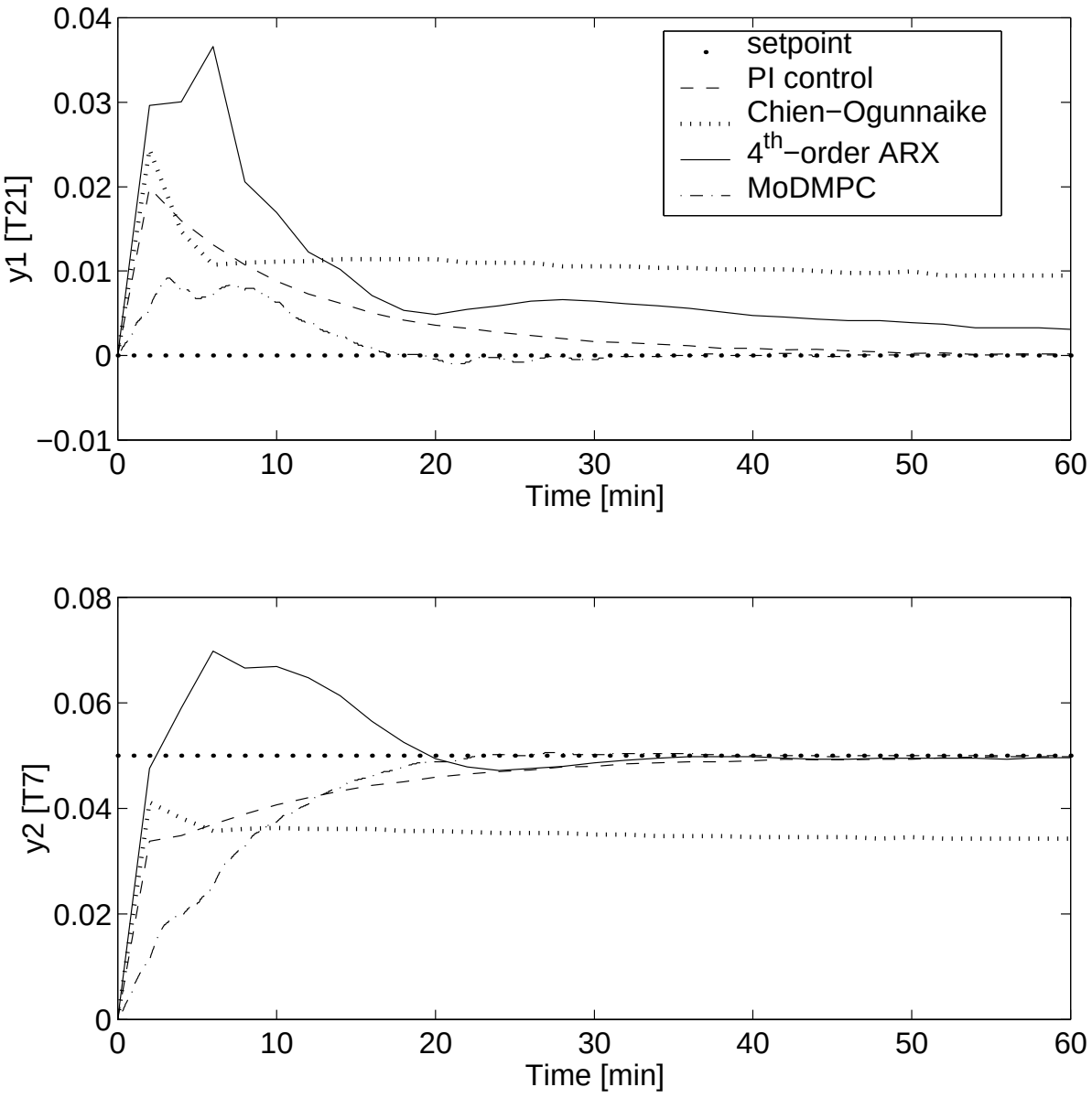

Figure 14: Control variable responses to setpoint change in tray temperature 7

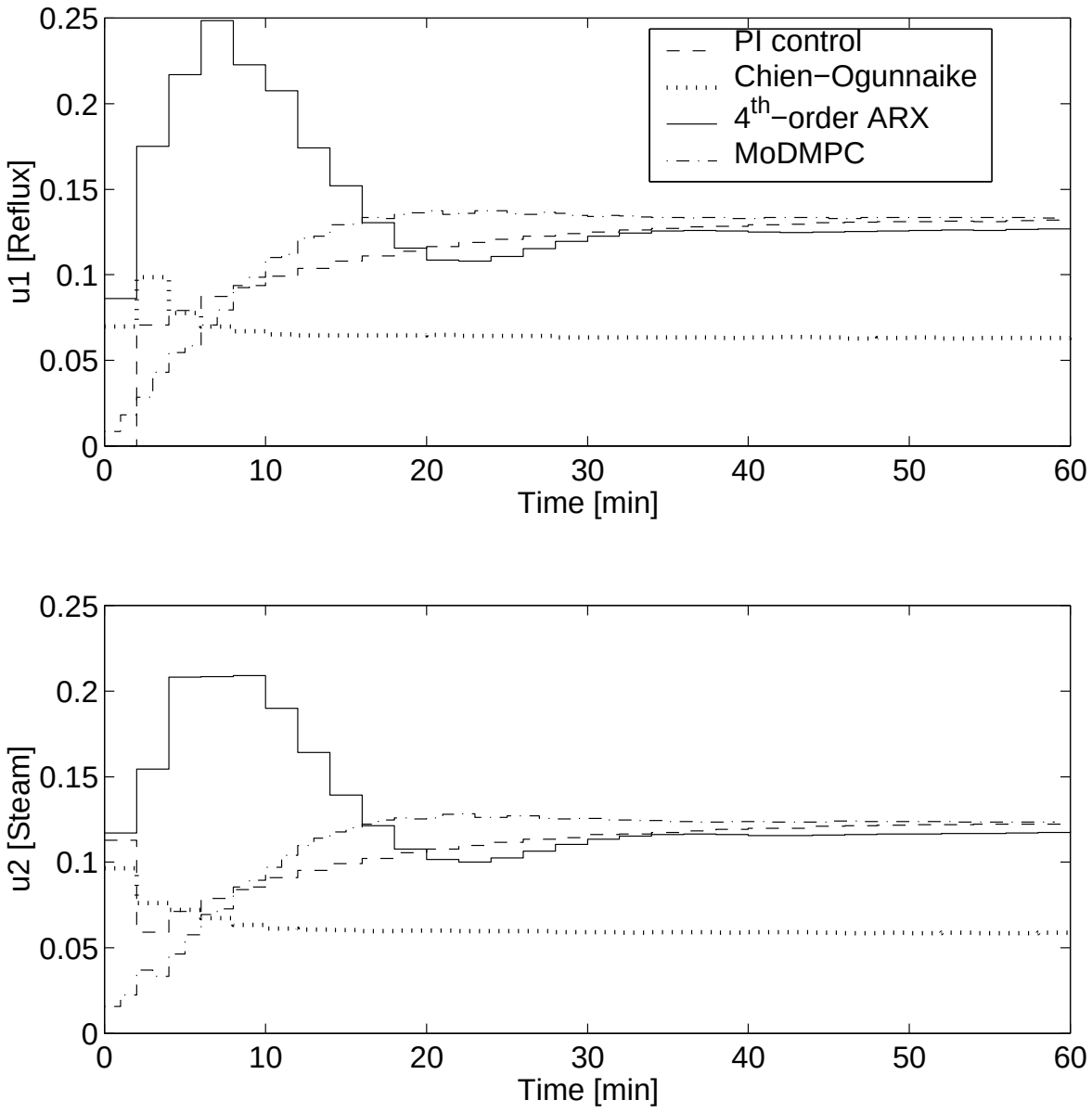

Figure 15: Manipulated variable responses to setpoint change in tray temperature 7