NOVEL PLANT TEST FOR MPC

Simone L. Kohare  * Jorge A. Mandler **

Fax: (610) 481-8803, Phone: (610) 481-6170.
** Same affiliation as first author. Phone: (610) 481-3413.

Abstract: The present work is a method of plant testing for the purpose of open-loop system identification and model-based control of chemical processes, in particular, model predictive control (MPC). It is a multivariable test technique using plant-friendly and optimal amplitude binary multi-frequency signals, in conjunction with multivariable parametric modeling techniques, to achieve significant time savings in plant testing by comparison with traditional step testing and even other multivariable testing methods in the literature.

Keywords: system identification, open-loop plant test, perturbation signal design, model predictive control, model-based control, optimization

1. INTRODUCTION

The widely accepted empirical way to characterize the dynamics of a system is to apply perturbation signals at the input channels and measure the response of the system to these signals. The input and output signals are then processed to give the required estimate of the dynamics of the system. This procedure is well known as system identification.

In much of the literature on system identification, little attention has been paid to the perturbation signal design itself, other than to the fact that the signals should be persistently exciting. In the case of linear systems this means effectively that the signal should adequately span the bandwidth of the system being identified. One of the main reasons for this lack of attention has been the emphasis in the literature on identification techniques for parametric models. The main focus has been on discrete transfer function models of single-input single-output (SISO), linear, time-invariant systems. Under these circumstances, there is not usually a great deal to choose among different perturbation signal designs. However, this is quite an idealized situation in most applications and, in practice, many questions on signal design issues do in fact arise. These questions are generally associated with how to establish a suitable compromise between persistent excitation and plant friendliness during test, how to reduce the testing time without compromising the information needed, e.g., for control design, process monitoring, etc.

On the issue of system identification for model predictive control (MPC), a more recent interest has emerged in the automatic control community. However, in practice, traditional step testing is still adopted by and large in open-loop data collection for MPC modeling.

Step testing assumes that only one input channel is moved at a time. Each step move is held for a relatively long length of time. In (Boyden, 1999) it is recommended that each input move should be held for an average of half the process settling time and a series of 15 – 20 moves should be executed...
for each input. In terms of their frequency characteristics, step signals tend to emphasize steady-state behavior and do not focus, therefore, on the closed-loop (faster) behavior. This means that a model with poor dynamic properties may be obtained. The plant testing time estimate for MPC, $t_{test}$, corresponding to this step testing technique can be expressed by the following formula:

$$t_{test} = (8 \ldots 10) \times IV \times T_{settling}$$  \hspace{1cm} (1)$$

where $IV$ is the number of independent variables, i.e., the sum of manipulated inputs and measurable disturbances and $T_{settling}$ is the process settling time. Naturally, $t_{test}$ can become prohibitively long for systems with long process settling time and/or large # of independent variables. For some applications in our experience, this number can easily vary between 1 to 2 months depending on the particular process characteristics. This excessive plant testing time generally translates into practical and economical infeasibility of MPC projects.

Besides step testing, single and multivariable PRBS testing have become quite popular in the more recent literature on system identification for process control. Some relevant work includes (Cott, 1995a; Cott, 1995b; Gaikwad and Rivera, 1994; Koung and MacGregor, 1994; Ijung, 1998; Rivera et al., 1990; Zhan, 1999; Zhu, 1998). In (Cott, 1995a) the author describes a benchmark problem proposed in the Process Identification Workshop at the 1992 Canadian Chemical Engineering Conference. Both industrial and academic participants in the workshop were challenged to identify models for the benchmark problem by using different perturbation signal designs and system identification techniques. Most of the participants utilized pseudo-random binary signals (PRBS) with different design characteristics. In (Zhu, 1998) the author proposes the use of PRBS as part of the so-called asymptotic method (ASYM) of system identification. In (Gaikwad and Rivera, 1994) and (Rivera et al., 1990) the authors make use of PRBS as perturbation signals for system identification. In (Koung and MacGregor, 1994) the authors rotate input signals in an attempt to better capture uncertainties in the process steady-state gains. The signals used as the basis for this design are PRBS. Thus, PRBS, which started as a periodic signal generated in the time-domain through shift register circuitry, received a frequency-domain interpretation and a multivariable design which made it attractive for system identification purposes. Nonetheless, PRBS displays several weaknesses that have been reported in the literature (Cott, 1995a; Cott, 1995b; Godfrey, 1994).

Other types of time-domain periodic signals have found little application so far. These are referred to as multi-level pseudo-random signals (also known as m-signals) (Docter, 1999; Zierler, 1959). Regarding frequency-domain identification, early work on signal design dates from the 1960s but most of the relevant papers are much more recent. However, they appear in journals ranging from Nuclear Science and Engineering to the International Journal of Control, from Industrial Engineering and Chemistry to International Shipbuilding Progress and in (at least) six different IEEE Transactions. As a result, this work is not as well known as it should be. The book edited by Keith Godfrey (Godfrey, 1994) is a relatively recent reference that collects some of this material.

In the present work, a thorough study of various perturbation signals for system identification, namely, pseudo-random binary signals, multi-level pseudo-random signals, sum-of-harmonics signals (van der Ouwerka et al., 1988; Rivera, 1999) and binary multi-frequency signals (Van den Bos, 1967; Buckner and Kerlin, 1972; Harris and Mellichamp, 1980; Paechlike and Rake, 1979; Van den Bos, 1970; Van den Bos and Krol, 1979), was carried out. The findings of this study led to the selection of binary multi-frequency (BMF) signals as perturbation signals for plant test. The factors that led to this selection will be discussed later.

2. MAIN RESULTS

The main goal of this work is to considerably shorten plant testing time without compromising, and actually improving, the obtained model quality. In order to achieve this goal, the following steps are used as part of a comprehensive plant test procedure:

- Determination of Control-Relevant Plant Information
- Perturbation Signal Design
- Multivariable Plant Test Design
- Plant-Friendliness Analysis of the Plant Test
- Optimal Input Signal Amplitude Selection
- Parametric Data Analysis

Figure 1 is a flowchart of the main elements of the system identification technique proposed in this work.

2.1 Control-Relevant Plant Information

Extracting plant information which is relevant for the purpose of process control has been a topic of discussion in the literature (see, e.g., (Rivera, 1999)). The approach has been suggested as a means of enhancing controller performance. In the present work, this approach has been identified
as one of the key components in reducing plant testing time.

Let us consider a first-order process with one input $u(t)$ and one output $y(t)$. If the input is excited with a unit step, the output response is given by $y(t) = K(1 - e^{-t/\tau})$, where $K$ is the process steady-state gain and $\tau$ is the time constant. By definition, the process is considered settled when $t \geq T_s \equiv 5\tau$, which implies that 99.33% of the response has been attained. However, if $t = 4\tau$, 98.17% of the response is still captured and if $t = 3\tau$, 95% of the response is obtained. Thus, for a first-order process, one can design the signal so that $\omega_{low}^\beta$ is then increased to $\frac{1}{3T_s}$, where $\beta \in [3, 5]$, and still obtain reasonable low frequency information on the process. Even if the process is not first order, most chemical processes' step responses near a given operating point resemble those of a linear system and similar arguments can be used to shorten testing time.

Now the question remains as to what should be the upper bound on the frequency range useful for control-relevant system identification. For any given linear system, a perturbation signal that spans the bandwidth of the system, i.e., $\omega \in \left[\frac{1}{T_s}, \frac{\alpha}{\tau_{dom}}\right]$ rad/unit of sampling time, where $\tau_{dom}$ is the estimated dominant time constant of the system, is adequate for the purpose of system identification if the system is to operate in closed-loop. However, if the purpose of the plant test is to identify a system suitable for operation in closed-loop, the closed-loop bandwidth must be considered instead. If the closed-loop is estimated to be $\alpha > 1$ times faster than the open-loop, then the bandwidth for the system in closed-loop is given by $\omega \in \left[\frac{1}{T_s}, \frac{\alpha}{\tau_{dom}}\right]$ rad/unit of sampling time. Therefore, this is the frequency range which the perturbation signal should span for control-relevant system identification.

Typically, $\alpha$ assumes values between 2 and 3 but it could be much higher if the controller is tuned more aggressively. Therefore, for closed-loop operation, $\omega_{high}^\beta = \frac{\alpha}{\tau_{dom}}$ rad/unit of sampling time.

This frequency window will dictate the minimum period of the signal used for system identification and the speed with which this signal should vary.

### 2.2 Binary Multi-Frequency Signal Design

An excellent compromise between flexibility in signal power distribution, small peak factor, short plant test and ability to obtain high signal-to-noise ratio can be achieved by using the so-called binary multi-frequency signals (BMF) (Van den Bos, 1967; Backner and Kerlin, 1972; Harris and Mellichamp, 1980; Paechl and Rake, 1979; Van den Bos, 1970; Van den Bos and Krol, 1979). Like the PRBS, these are binary, discrete-interval, periodic signals with period $P = NT_{mc}$ samples. Like the sinusoids, these signals display as much power as possible in certain harmonics specified by
the user. Nearly all designs are for $N$ a power of 2, which makes it easy to ensure no spectral leakage using FFT signal processing. However, this is not a necessity and the only hard restriction on $N$ is that it is an even number to guarantee that the signal is zero mean.

Because of the similar periodic auto- and cross-correlation properties of PRBS and BMF signals and the larger flexibility in the choice of $N$ for the BMF, the plant testing time with BMF can always be made equivalent or shorter than that with PRBS. Thus, it has been recognized in this work that one does not pay any penalty in testing time by having a zero mean signal with a user-defined power distribution.

The Frequency Domain Identification Toolbox in MATLAB has two routines used to generate and improve a single binary multi-frequency signal. In these routines, the user is allowed to choose the power associated with each harmonic in a specified frequency range. The computation is an optimization where the signal with the smallest peak factor and the largest percent of useful power is computed to match, as closely as possible, the user-defined power for each harmonic. The main drawback of these routines is the lack of reproducibility of results since the resulting BMF signal will change depending on the initial random seed generated by the software. These different signals will have correspondingly different peak factor and power distribution. Therefore, the software does not necessarily provide the signal with the best compromise between peak factor and power distribution and the inexperienced user may end up with a poor signal selection. In this work these issues are addressed by improving on the existing software.

The algorithm used in these MATLAB routines is discussed in (Van den Bos and Kool, 1979) with improvement via a search technique described in (Pachliike and Rake, 1979).

Because the BMF can be designed to concentrate power in a selected number of harmonics in the control-relevant bandwidth, it is generally possible to excite the system as much with a lower amplitude BMF as with a higher amplitude PRBS.

The first design requirement on the BMF signal will be dictated by the need to capture the process long-term behavior. The lowest frequency captured by a BMF signal of period $P = NT_{sw}$ is given by $\omega_{\text{low}} = \frac{2\pi}{NT_{sw}}$ rad/unit of sampling time. Therefore, the following inequality guarantees that the signal spans the process fundamental frequency:

$$\omega_{\text{low}}^{\text{input}} \leq \omega_{\text{low}}^{\text{process}} \implies \frac{2\pi}{NT_{sw}} \leq \frac{1}{T_{ss}}$$

$$\implies P = NT_{sw} \geq 2\pi T_{ss} \quad (2)$$

Further reduction in testing time can be achieved if the requirement on the low frequency information is relaxed as discussed in section 2.1, i.e.:

$$P = NT_{sw} \geq 2\pi\left(\frac{\beta}{\alpha}\right)T_{ss}, \quad \text{where } \beta \in [3, 5] \quad (3)$$

On the high-frequency side, the upper bound on the control-relevant frequency window, $\omega_{\text{high}}^{\text{process}}$, should never exceed the Nyquist frequency of the input signal which corresponds to $\omega_{\text{high}}^{\text{process}} \leq \omega_{\text{Nyquist}} \equiv \frac{\pi}{T_{sw}}$ rad/unit of sampling time. This condition generates the following upper bound on the switching time, $T_{sw}$:

$$\omega_{\text{high}}^{\text{process}} \leq \omega_{\text{Nyquist}} \implies \frac{\alpha}{\tau_{\text{dom}}} \leq \frac{\pi}{T_{sw}}$$

$$\implies T_{sw} \leq \frac{\pi T_{dom}}{\alpha} \quad (4)$$

Inequalities (3) and (4) together provide the general guidelines for the design of BMF signals for system identification. These requirements on $P$ and $T_{sw}$ are the same as for PRBS signals (Rivera, 1999). The fact that they extend to BMF signals is a new finding and it comes from the qualitative similarities between the periodic autocorrelation function (ACF) properties of BMF and PRBS signals.

An important additional degree of freedom in the BMF signal design is the signal power distribution over the harmonics that the signal spans. By properly designing the BMF signal, a signal with relatively small (large) $T_{sw}$ can still invest a reasonable amount of power at lower (higher) frequencies. This allows for a very tailored design of the perturbation signal for each individual application.

2.3 Multivariable Plant Test Design

The BMF signals described in section 2.2 were used to design a new multivariable plant test. The multivariable nature of the plant test based on BMF signals proposed here is a key contributor to plant testing time reduction. As discussed in section 1, multivariable designs for perturbation signals for system identification are known for PRBS and Schröder-phased sinusoids. The multivariable statistically uncorrelated designs for PRBS and Schröder-phased sinusoids are based on time and frequency domain arguments, respectively.

In the present work, time-domain characteristics of the BMF signal, namely, the periodic auto- and cross-correlation functions (ACF and CCF), dictate the multivariable design in spite of the frequency-domain origin of the signal. This is a novel approach to multivariable plant test design using a frequency-domain based signal and generating uncorrelated copies in time-domain.

The multiple BMF signals are designed so as to ensure the smallest possible periodic unbiased
CCF between input pairs during one process settling time $T_{ss}$. In order to achieve this goal, the guidelines imposed by equations (3) and (4) are still relevant but may be replaced by more conservative inequalities.

In order to obtain $m$ BMF signals with the smallest possible periodic CCFs, the initial (“mother”) BMF signal is generated and $m - 1$ delayed copies of this signal are created. Therefore, a new parameter appears in the MIMO design which is the delay between input channels.

The multivariable design, as described above, imposes the following requirements on the BMF signal period $P$:

$$P = NT_{sw} \geq 2\pi \left( \frac{\beta}{5} \right) T_{ss}, \quad \text{for } m \leq 6$$

$$P = NT_{sw} \geq m \left( \frac{\beta}{5} \right) T_{ss}, \quad \text{for } m > 6 \quad (5)$$

with the delay between input channels, $D$, equal to $D = \left( \frac{\beta}{5} \right) T_{ss}$.

The resulting $T_{sw}$ in the MIMO case still satisfies inequality (4).

Since the BMF signal period in the MIMO case is given by equations (5), the minimum plant testing time is expressed by:

$$t_{test} = (2\pi + 1) \left( \frac{\beta}{5} \right) T_{ss}, \quad \text{for } m \leq 6$$

$$t_{test} = (m + 1) \left( \frac{\beta}{5} \right) T_{ss}, \quad \text{for } m > 6 \quad (6)$$

Equation (6) is derived from the fact that one process settling time $T_{ss}$ worth of data (or the modified settling time $(\frac{\beta}{5})T_{ss}$ with $\beta \in [3, 5]$) needs to be discarded in the beginning of the test, since it contains information related to previous input moves and not to the planned test moves.

Equation (6) is a key result of the present work. By comparing equations (6) and (1) one can infer an average $85 - 90 \%$ reduction in plant testing time with the new technique when compared to step testing.

\subsection{Plant-Friendliness Analysis of the Plant Test}

Once the multivariable test has been designed to tailor a particular application, it is important to predict, ahead of time, the effect that the designed perturbation signals will have on the plant. This is especially important given the multivariable nature of the test and how the different input moves interact to affect the outputs.

In this work a novel method for evaluating plant-friendliness of the designed test has been devised. Through assumption of simple linear dynamics and what is believed to be reasonable size moves on each input variable, it is possible to predict the plant output behavior during test. Output constraint violations can also be anticipated. This method provides a means of checking when input move combinations might lead the outputs towards constraints and what effect the input move sizes will have on the output responses.

If a design is judged to be plant-unfriendly, the first step might be to re-order the perturbation signals with respect to the input variables and re-evaluate plant-friendliness. If this does not produce satisfactory results, the test design should be re-visited.

Figure 2 illustrates how the application of this technique enables screening of a more plant-friendly test design prior to the actual test. The two curves indicate predictions of a plant output behavior using the plant-friendliness analysis tool. The dashed curve corresponds to the initial design based on the available plant information. The solid curve was obtained by simply re-ordering the BMF signals with respect to the input variables. While both designs satisfy the output constraints indicated by the horizontal dotted lines, the test design corresponding to the solid curve is believed to be more plant-friendly because it promotes output excitation above and below the initial operating point for a proportional amount of time. In the designing corresponding to the dashed curve, the output tends towards the low limit constraint for the largest duration of the test. In this particular application, violation of the low limit constraint means off-spec product and, therefore, output excitation towards this constraint for a prolonged amount of time is not desirable. The same input amplitudes were considered in both cases.

\subsection{Optimal Input Signal Amplitude Selection}

In this work, once a plant test design is considered plant-friendly, optimal input amplitudes are
computed for maximum output excitation within desired constraints.

In order to solve the dynamic optimization problem, the same model used for the predictions in the plant-friendliness analysis is utilized. For simplicity, the technique will be discussed for a first order model

\[ G(t) = K \left(1 - e^{-\frac{t}{\tau}} \right), \]  

(7)

where \( K \) is the matrix of steady-state gain estimates and \( \tau = \frac{T_s}{3} \).

Since the input signals change at every \( T_sw \) samples and there are \( N \) moves in a signal period, the optimization is solved \( N \) times with \( t = T_sw \) in formula (7).

The vector of amplitudes \( \lambda_i > 0, i = 1, \ldots, m \), for the \( m \) inputs constitutes the set of decision variables. The optimization problem which is solved for \( k = 1, \ldots, N \), is given by:

\[
\text{OF:} \quad \min_{\lambda^1, \ldots, \lambda^m} \sum_{i=1}^{m} \ln \left( \frac{1}{\lambda_i^k} \right) \\
\text{IC:} \quad \Delta u_{\text{min}} \leq \lambda^k u^k \leq \Delta u_{\text{max}} \\
\text{OC:} \quad \Delta b_{\text{min}} \leq G(T_sw)\lambda^k u^k \leq \Delta b_{\text{max}} \\
\text{AC:} \quad \lambda^k > 0
\]  

(8)

where OF, IC, OC and AC stand for objective function, input constraints, output constraints and additional constraints, respectively. \( \lambda^k \) is an \( m \times m \) diagonal matrix whose elements are the input amplitudes \( \lambda_i^k \), and \( u^k \) is the \( m \times 1 \) combination of the \( m \) input moves at the \( k^{th} \) delay (this vector only contains elements equal to \(-1, 0, +1\)).

\( u_{\text{min}} \) and \( u_{\text{max}} \) are the minimum and maximum allowed input values. \( \Delta u_{\text{min}} \equiv u_{\text{min}} - u_{ss} \) and \( \Delta u_{\text{max}} \equiv u_{\text{max}} - u_{ss} \) are, respectively, the minimum and maximum input deviation constraints with respect to the desired setpoint input value, \( u_{ss} \). \( b_{\text{min}} \) and \( b_{\text{max}} \) are, respectively, the minimum and maximum output constraints which must be satisfied at all times. \( \Delta b_{\text{min}} \equiv b_{\text{min}} - y_{ss} \) and \( \Delta b_{\text{max}} \equiv b_{\text{max}} - y_{ss} \) are the minimum and maximum output deviation constraints with respect to the current output value \( y_{ss} \). In practice, \( u_{ss} \) and \( y_{ss} \) are the vectors of input and output values corresponding to the start of the plant test. Since the input deviations are always made around the initial input value \( u_{ss} \), this value is not updated. The output values, however, vary as the inputs are implemented and the constraints have to be measured against updated values (that is why \( y_{ss} \) is dependent on \( k \)).

If this procedure were implemented on-line, during the plant test, as an advisor for input amplitude selection at each \( T_sw \) samples, it should take feedback from the plant in order to correct the outputs (instead of the update based on the assumed linear dynamics). This feedback would lead to more realistic amplitudes since it is a way to correct for the approximate dynamics in (7).

Figure 3 shows how this technique worked in a real plant test case. The figure shows the predicted output response in time obtained with optimized input amplitudes and the actual output response during plant test obtained with slightly larger amplitudes for higher output excitation. The concern with this particular critical output is that it should not go below 99.9% since this is a product specification. The technique predicts that the test sequence will drive the output close to that limit at approximately 5250 min and again at 8300 min into the test. This indeed takes place in the actual test but the output at 5250 min actually goes below the desired limit due to the higher amplitudes used during the test. The model used for the output prediction and input amplitude optimization was a simple first order model with rough steady-state gain estimate.

2.6 Parametric Data Analysis

Because the main focus of this work is to shorten testing time, this also means that less data is available for model construction. Therefore, parametric modeling methods are the only alternative if reasonable quality parameter estimates are to be obtained.

In the present work, the types of models considered were MISO or MIMO state-space or ARX models. MIMO models were utilized for highly interacting sub-systems of outputs where the output interactions must be taken into account in the model. These are the models that best capture the complex multivariable interactions in the process. This is not an issue for linear processes but, since most chemical processes are highly non-linear, MISO and MIMO models may lead to very different responses. In this work it has been observed that MIMO identification for groups of
highly interacting outputs leads to better fit of the individual output responses than MISO identification.

3. CONCLUSION

In this manuscript a method of plant testing and system identification with emphasis on test time reduction has been discussed. The plant testing time reduction is achieved via selection of binary multi-frequency (BMF) signals as input perturbation signals designed to excite the system in a control-relevant frequency range. The plant testing is a multivariable test with each input being excited with a delayed copy of the “mother” BMF signal. On an average, the minimum testing time with multivariable BMF testing is 85–90% shorter than conventional step testing. Once the test has been designed, its plant friendliness can be evaluated ahead of the plant test execution via simulation of predictive models based on a priori knowledge of the system. The amplitudes of the perturbation signals used to excite the different input channels are optimized to maximize the output variability and the signal-to-noise ratio within pre-specified input and output constraints. The resulting data from plant testing is analyzed using parametric modeling techniques.

REFERENCES


