SYSTEM IDENTIFICATION IN A NOISE FREE ENVIRONMENT

L. Ljung

Department of Electrical Engineering, Linköping University, Linköping, Sweden

Abstract. In this contribution we consider the situation where no disturbances act on the system to be identified, but where - as always - it is not possible to find a model structure that exactly accommodates the true systems. The transient and asymptotic properties of the obtained estimates in this situation are considered. Special attention is paid to the practical problem of obtaining good adaptation properties by appropriate choices of forgetting factors and prefilters.

1 INTRODUCTION

Things take time. This fact of life manifests itself in System Identification as the fact that it takes time to construct a reliable model from data. According to the standard statistical paradigm the reason for this is the presence of disturbances whose effects must be averaged out over long measurement records. In fact, identification theory is usually primarily concerned with the asymptotic (as the data record tends to infinity) properties of these "noise" effects.

The implied conclusion would appear to be that in the absence of noise there would be no identification problem - Reliable models could be build more or less instantaneously. In fact, the Cramér-Rao inequality would suggest just that. Of course, this is not so in reality. Things still will take time, but now for other reasons than averaging out noise effects. Since the most important design issues in adaptation deal with trade-offs regarding the time-horizon in the model building, it is important to understand these other reasons.

The underlying reason why noise-free data do not lead to quick models is the lack of knowledge of a model structure where these data fit exactly. It is thus the model mismatch we must consider. Take any model $\mathcal{M}$, that produces a prediction $\hat{y}_{\mathcal{M}}(t)$ of the system's output at time t. The error between the actual, measured output $y(t)$ and $\hat{y}_{\mathcal{M}}(t)$ we denote

$$\varepsilon(t) = y(t) - \hat{y}_{\mathcal{M}}(t) \tag{1}$$

It is customary to single out two components of the error $\varepsilon(t)$

$$\varepsilon(t) = e(t) + \eta_{\mathcal{M}}(t) \tag{2}$$

Here $e(t)$ is an unpredictable disturbance ("*innovation*") that does not originate from the inputs applied to the process. The term $\eta_{\mathcal{M}}(t)$ is a *model error*, that stems from the fact that the model M does not give a 100 % correct description of the input-output properties of the system. This term could theoretically be made arbitrarily small by selecting sophisticated model structures (supported by an ever increasing data set). The term $e(t)$ can however never be made smaller. It is typically modelled as a white noise sequence.

Depending on our trade me may have different objectives for the model error. In control and adaptive control we typically use rather simple models, which may give model errors $\eta_{\mathcal{M}}(t)$ that are substantially larger than the innovation contribution $e(t)$ to $\varepsilon(t)$. To display and enhance the time-averaging aspects in such situation we shall in this paper assume that $e(t) = 0$, and thus concentrate on the case with noise-free data. The questions we address are: *What are the transient and asymptotic properties of models in case we have exact data? How should we choose design variables (Like forgetting factors and prefilters) to obtain good tracking properties in our adaptation algorithms?*

The roles of prefilters, forgetting factors and unmodelled dynamics have been extensively discussed in the literature on adaptive control. See, e.g. Wittenmark and Åström (1984), Rohrs et al (1984), and Kosut (1986, 1987, 1988).

2 A PREVIEW EXAMPLE

We shall consider a data set that contains quite little noise. It is obtained from a heated-air process, Feedback Process Trainer PT 326, and is described in Ljung (1987a), p 440 f (See Fig 17.12). It is also the data set DRYER5 of the System Identification Toolbox, Ljung (1987b). A portion of the data is shown in Figure 1.

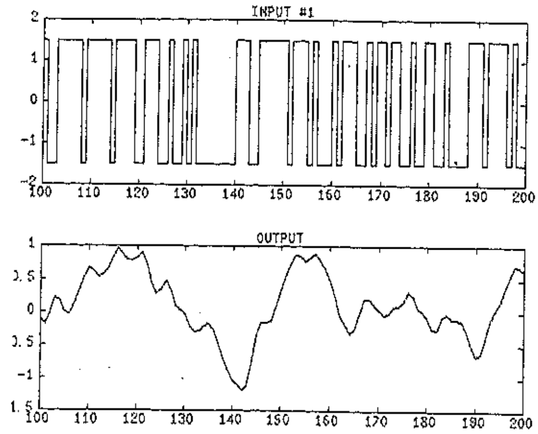

Figure 1: Input-output data from the heated-air process

Suppose we are interested in the transfer function of the system around frequency 12.5 rad/s. For this purpose we build models of the kind

$$y(t) + a_1 y(t-T) + \dots + a_{n_a} y(t - naT) =$$

$$= b_1 u(t - nkT) + \dots + b_{n_b} u(t - (nk + nb - 1)T) \quad (3)$$

relating the input sequence $\{u(t)\}$ (being the applied voltage to the heater) to the output sequence $\{y(t)\}$ (being the resulting temperature of the air). T is the sampling interval (0.08 sec). From such models we then compute

$$G(e^{i\omega T}) = \frac{b_1 e^{-ink\omega T} + \dots + b_{nb} e^{-i(nk-1+nb)\omega T}}{1 + a_1 e^{-i\omega T} + \dots + a_{n_a} e^{-ina\omega T}} \quad (4)$$

and evaluate if for $\omega = 12.5$.

We apply the recursive least squares procedure to the data with forgetting function $\lambda = 0.5$ (A very small value!) using the structure (3) with $nu = nb = 2$, $nk = 3$ (this is a good structure according to the analysis in Ljung (1982a)). In Figure 2 $|G(e^{i12.5 \cdot 0.02})|$ is shown as a function of time. The straight line is the value obtained for the whole data record. Despite the very low value of the forgetting factor the estimates do not fluctuate very much.

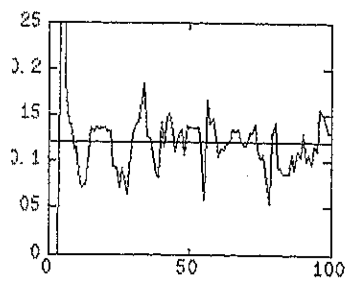

Figure 2: The amplitude of the transfer function at frequency 12.5 rad/s, for a model recursively identified with $na = nb = 2$, $nk = 3$ and forgetting factor $\lambda = 0.5$.

In Figure 3 the corresponding result is shown using the model structure $na = nb = nk = 1$.

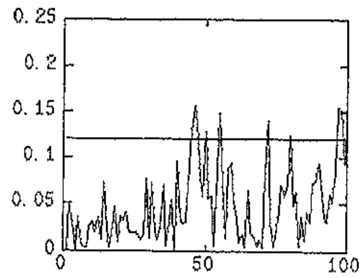

Figure 3: As Figure 2, but $na = nb = nk = 1$.

Figure 4 shows the result for this model structure but for $\lambda = 0.95$.

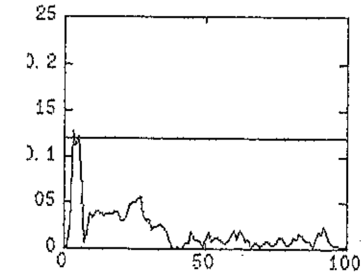

Figure 4: As Figure 3, but $\lambda = 0.95$.

In Figure 5 we have first filtered the input-output data through a 10th order Butterworth bandpass filter with cut off frequencies 10% above and below the desired frequency 12.5 rad/s. (This was obtained by the MATLAB command BUTTER(5,[0.9/π, 1.1/π), cf Moler et al (1986)).

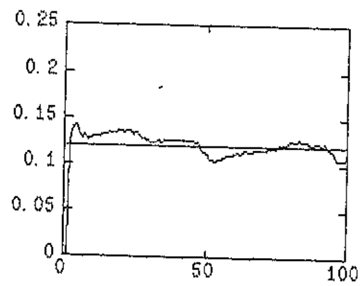

Figure 5: As Figure 4, but data first band pass filtered through a 10th order BP-Butterworth filter (See the text).

In Figure 6 we have computed the empirical transfer function estimate (cf Ljung (1987a), ch 6) consisting of the ratio of the output DFT to the input DFT.

$$Y_N^t(\omega) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} y(t-kT) e^{-i\omega kT} \quad (5)$$

$$U_N^t(\omega) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} u(t-kT) e^{-i\omega kT} \quad (6)$$

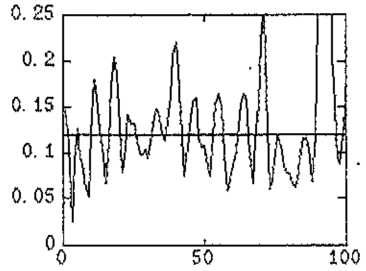

Figure 6: The empirical transfer function estimate (amplitude) at frequency 12.5 rad/s based on 50 past samples. Data as in Figure 1.

$$\hat{\hat{G}}_N^t(e^{i\omega T}) == Y_N^t(\omega)/U_N^t(\omega) \quad (7)$$

The figure is given for $\omega = 12.5$ and $N = 50$. Figure 7 shows the corresponding result for the filtered data series.

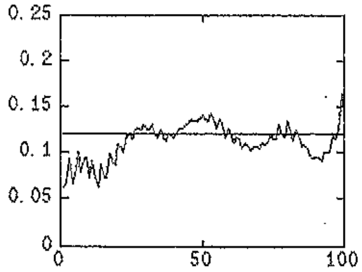

Figure 7: As Figure 6, but band-pass filtered data used.

The impulse response of the filter is shown in Figure 8.

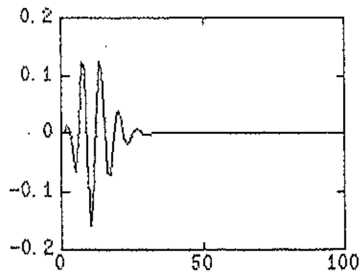

Figure 8: The impulse response of the band pass filter used in Figures 5 and 7.

3 SOME BASIC RESULTS

We shall consider the following set-up. The true system will assumed to be linear but not necessarily finite-dimensional and noise-free. The input-output relation can thus be written

$$y(t) = G_0(q)u(t) = (\sum_{k=1}^{\infty} g_k^0 q^{-k})u(t) =$$

$$= \sum_{k=1}^{\infty} g_k^0 u(t-k) \quad (8)$$

Here q^{-1} is the delay operator $q^{-1}u(t) = u(t-1)$. (We assume unit sampling interval for simplicity). The transfer function of the system is thus $G_0(e^{i\omega})$.

When we use parametric models we shall write them as

$$y(t) = G(q,\theta)u(t) + H(q,\theta)e(t) \quad (9)$$

Note that the model (9) may very well include a noise description $H(q,\theta)e(t)$ even if we believe the true system to be noise-free. It will still have a substantial effect on the results. In (9) θ is a finite dimensional parameter vector that parameterizes the transfer functions. The most common parameterization is the one corresponding to (3) for which

$$G(q,\theta) = \frac{B(q)}{A(q)} \qquad H(q,\theta) = \frac{1}{A(q)} \quad (10)$$

$$B(q) = b_1 q^{-nk} + \dots + b_{nb} q^{-nk-nb+1},$$

$$A(q) = 1 + a_1 q^{-1} + \dots + a_{na} q^{-na}$$

$$\theta = (a_1 a_2 \dots a_{n_a} \;\; b_1 \dots b_{n_b})^T \quad (11)$$

We shall sometimes also consider the further special case of a FIR (finite impulse response) word, corresponding to $na = 0$.

The basic way to estimate the parameter vector θ in (9) is to form the predictor

$$\hat{y}(t|\theta) = H^{-1}(q,\theta)G(q,\theta)u(t) + [1 - H^{-1}(q,\theta)]y(t) \quad (12)$$

compute the error

$$\varepsilon(t,\theta) = y(t) - \hat{y}(t|\theta) = H^{-1}(q,\theta)(y(t) - G(q,\theta)u(t)) \quad (13)$$

filter it through filter $L(q)$

$$\varepsilon_F(t,\theta) = L(q)\varepsilon(t,\theta) \quad (14)$$

and minimize the filtered errors:

$$\hat{\theta}_N = \arg\min_\theta \sum_{t=1}^{N} \varepsilon_F^2(t,\theta) \quad (15)$$

See, for example, Ljung (1987a) for more details. The estimate (15) will have the following properties:

$$\hat{\theta}_N \to \theta^* \quad as \quad N \to \infty \tag{16}$$

$$\theta^* = \arg\min_\theta \mathcal{L}\varepsilon_p^2(t,\theta) \tag{17}$$

where

$$\mathcal{L}f(t) \triangleq \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N} f(t) \tag{18}$$

Furthermore

$$\mathcal{L}\varepsilon_F^2(t,\theta) =$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |G_0(e^{i\omega}) - - G(e^{i\omega},\theta)|^2 \frac{\Phi_u(\omega)\cdot|L(e^{i\omega})|^2}{|H(e^{i\omega},\theta)|^2} d\omega \tag{19}$$

Here $\Phi_u(\omega)$ is the input spectrum, defined as

$$\Phi_u(\omega) = \sum_{\tau=-\infty}^{\infty} R_u(\tau)e^{-i\omega\tau} \tag{20}$$

$$R_u(\tau) = \mathcal{L}u(t)u(t-\tau) \tag{21}$$

All this follows in a rather straightforward fashion, see, e.g. Ljung (1987a), Ljung (1985).

We can also derive an expression for how quickly $\hat{\theta}_N$ approaches θ^*. For simplicity we give it only for the special use (10). This corresponds to the least squares method which gives us

$$\hat{\theta}_N = (\sum_{t=1}^{N} \varphi_F(t)(\varphi_F^T(t))^{-1} \sum_{t=1}^{N} \varphi_F(t)y_F(t) \tag{22}$$

$$\varphi(t) = [y(t-1)...y(t-na)\ u(t-nk)...u(t-nb-nk+1)]^T \tag{23}$$

$$\varphi_F(t) = L(q)\varphi(t)$$
$$y_F(t) = L(q)y(t) \tag{24}$$

Assume that $\{u(t)\}$ can be viewed as a stochastic process with suitable mixing properties. Then simple calculations give that

$$\sqrt{N}(\hat{\theta}_N - \theta*)\ \overrightarrow{dist}\ N(0,P) \tag{25}$$

i.e. the random variable on the left is asymptotically normal with zero mean and a covariance matrix given by

$$P = R^{-1}\ Q\ R^{-1} \tag{26}$$

$$R = \mathcal{L}\varphi_F(t)\varphi_F^T(t) \tag{27}$$

$$Q = \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N} E\ \varphi_F(t)\varphi_F^T(s)\varepsilon_F(t,\theta*)\varepsilon_F(s,\theta*) \tag{28}$$

$$\varepsilon_F(t,\theta*) = L(q)A_*(q)[G_0(q) - G_*(q)]u(t) \tag{29}$$

A_* and G_* here correspond to the limiting model. We notice, among other things, that even in the noise-free case the estimate will converge to its limit no faster than $1/\sqrt{N}$, unless $G_0(q) \equiv G^*(q)$.

Let us also comment on the recursive least squares algorithm with forgetting factor λ. It minimizes

$$V_t(\theta) = \sum_{k=1}^{t} \lambda^{t-k}\varepsilon_F^2(k,\theta)$$

recursively is t, giving the estimate $\hat{\theta}(t)$. Direct calculations (See Appendix A) show that

$$E(\hat{\theta}(t) - \theta*)(\hat{\theta}(t) - \theta*)^T \sim$$

$$\sim \frac{(1-\lambda)^2}{1-\lambda^2}P \approx \frac{1-\lambda}{2}P \tag{30}$$

for λ close to 1, where P is given by (26) - (29).

Empirical transfer functions

Consider now the estimate of the transfer function defined by (5) - (7). According to, for example, Theorem 2.1 of Ljung (1987a) we have

$$G_0(e^{i\omega}) - \hat{\hat{G}}_N^t(e^{i\omega}) = \frac{R_N(\omega)}{U_N^t(\omega)} \tag{31}$$

where

$$|R_N(\omega)| \le 2\cdot \max_t |u(t)| \cdot C_G/\sqrt{N} \tag{32}$$

where

$$C_G = \sum_{k=1}^{\infty} k|g_k^0| \tag{33}$$

If the method (5) - (7) is applied to filtered data, the only difference is that $u(t)$ is to be replaced by $u_F(t)$ in (32), and $U_N^t(\omega)$ in (31) should be the DFT of $\{u_F(t)\}$.

Again we see that the convergence to the limit transfer function is of the rate $1/\sqrt{N}$.

Several other techniques for estimating models, directly in the frequency domain have been discussed. See, e.g., LaMaire et al (1987), Parker (1988) and Parker and Bitmead (1987).

4 VARIABILITY DUE TO INPUT

In this section we shall discuss the variance (or variations) in the model that stem from the input.

4.1 A stochastic framework for the input

If we describe the input sequence as a realization of a stochastic process, we have a natural framework for assessing the variability of the model. We already in (25) - (29) established a result of this character. Let us comment on some aspects of this result.

1. The convergence rate is invariant under input scaling.
2. The rate is always $1/\sqrt{N}$ but the matrix P is proportional to the square of the error ε_F. In other words, the better the model, the quicker the convergence.
3. The convergence rate can be affected by the filter $L(q)$. More about that in Section 5.

Let us specialize to a Gaussian input. Then $\varphi_F(t)$, $\varphi_F(s)$, $\varepsilon_F(t,\theta^*)$ and $\varepsilon_F(s,\theta^*)$ will be jointly Gaussian and we may write in (28)

$$E\ \varphi_F(t)\varphi_F^T(s)\varepsilon_F(t,\theta^*)\varepsilon_F(s,\theta^*) =$$

$$= E\ \varphi_F(t)\varphi_F^T(s)\ E\ \varepsilon_F(t,\theta^*)\varepsilon_F(s,\theta^*) +$$

$$+ E\ \varphi_F(t)\varepsilon_F(s,\theta^*)\ E\ \varphi_F^T(s)\varepsilon_F(t,\theta^*) +$$

$$+ E\ \varphi_F(t)\varepsilon_F(t,\theta^*)\ E\ \varphi_F^T(s)\varepsilon_F(s,\theta^*) \qquad (34)$$

The first of these terms is the "standard" term that is obtained in the case with exact modelling but noise present. The last term is zero, since $E\varphi_F(t)\varepsilon_F(t,\theta^*) = 0$ is the defining relationship for θ^*.

A situation when the second term in (34) is zero is the following one:

$$\text{FIR model, } L(q) = 1,\ \{u(t)\}\text{white noise} \qquad (35)$$

To see this, we note that $G^*(q)$ will be

$$G^*(q) = \sum_{k=1}^{nb} g_k^0 q^{-k} \qquad (36)$$

and hence

$$\varepsilon_F(t,\theta^*) = \sum_{k=nb+1}^{\infty} g_k^0 u(t-k) \qquad (37)$$

Since $\varphi_F(t)$ contains only $u(t-1)...u(t-nb)$ we see that $E\varphi_F(t)\varepsilon_F(s,\theta^*) = 0$ whenever $t \geq s$. Hence

$$Q = \lim_{N\to\infty} \frac{1}{N}\sum_{t=1}^{N}\sum_{s=1}^{N} E\varphi_F(t)\varphi_F(s)\cdot E\varepsilon_F(t,\theta^*)\varepsilon_F(s,\theta^*)$$

We can now apply Theorem 4.1 of Ljung and Yuan (1985) to conclude that the variance of $\hat{G}_N(e^{i\omega}) - G_*(e^{i\omega})$ will (asymptotically for large nb) behave as

$$\frac{nb}{N}\cdot\frac{\Phi_{\varepsilon_F}(\omega)}{\Phi_u(\omega)}$$

Here $\Phi_{\varepsilon_F}(\omega)$ is the spectrum of $\{\varepsilon_F(t,\theta^*)\}$, i.e.

$$\Phi_{\varepsilon_F}(\omega) = |G_0(e^{i\omega}) - G_*(e^{i\omega})|^2\Phi_u(\omega) \qquad (38)$$

so that

$$Var(\hat{G}_N(e^{i\omega}) - G_*(e^{i\omega})) \sim \frac{nb}{N}\cdot|G_0(e^{i\omega}) - G_*(e^{i\omega})|^2 \qquad (39)$$

In the RLS-case (30) we get

$$Var(\hat{G}_t(e^{i\omega}) - G_*(e^{i\omega})) \sim \frac{nb}{2}(1-\lambda)|G_0(e^{i\omega}) - G_*(e^{i\omega})|^2 \qquad (40)$$

Even though the result here has been derived for a rather special case (white Gaussian input in an FIR model), the conceptual message is important: *The convergence rate is proportional to the bias error.* In applications of recursive identification the result is even more striking: The variance in (40) will reflect the variability of the estimate $\hat{G}_t(e^{i\omega})$ over time around the average value $G_*(e^{i\omega})$. This value is a biased estimate of the true transfer function $G_0(e^{i\omega})$ (G_* being the best approximation of G_0 that the model structure can offer). The surprising thing is that *the measured variability of $\hat{G}_t$ around G_* (which both are known to us) gives a simple and direct measure of the size of the bias error $|G_* - G_0|$!* Certainly G_0 is not known to us, and may be very far from G_*, so the easily estimated bound for the model error is very useful to us.

4.2 Fitting models to very short data records

Another way of studying the model variability due to the input is to consider what happens when models are fitted to minimal length data sequences.

Consider the first order model

$$y(t) + ay(t-1) = bu(t-1) \qquad (41)$$

Clearly, a and b can be uniquely determined from two consecutive equation (41) ($t = t_1, t = t_1+1$). This involves three values of $y(t)$: $y(t_1-1), y(t_1), y(t_1+1)$ and two values of $u(t)$: $u(t_1-1), u(t_1)$. Now, how can we characterize the model that is obtained in that way?

The answer is as follows: Let ω_*, A and φ be the unique values that solve the equations

$$y(t) = A\ \cos(\omega_* t + \varphi) \quad t = t_1 - 1,\ t_1,\ t_1 + 1 \qquad (42)$$

(since this is a system of non-linear equations, some regularity conditions on $y(t)$ $t = t_1 - 1,\ t_1, +1$ have to be required. We also constrain ω_* to be below the Nyquist frequency.) Note that ω_* will be a function of t_1:

$$\omega_*(t_1) \qquad (43)$$

The model that is obtained by fitting (41) to the data in question will then be given by

$$\frac{be^{-i\omega_* t}}{1 + ae^{-i\omega_* t}} = G_0(e^{i\omega_*}) \qquad (44)$$

Here $G_0(e^{i\omega})$ is the true transfer function of the system. *In other words, the obtained model is the first order approximation that gives exact fit to the true transfer function of the frequency $\omega_*(t_1)$, which is the apparent "Instantaneous frequency" represented by the three consecutive outputs aroundtime t_1.* To realize this, assume that the input-output sequences really are pure sinusoids of frequency ω_*. Nothing in the used data record, $y(t_1-1), y(t_1), y(t_1+1), u(t_1-1), u(t_1)$, can disprove that. For pure sinusoids, it is clear that the obtained model would be the one given in (44).

Suppose now that we evaluate the model transfer function at a pre-determined frequency ω_0. This value, $\hat{G}_t(e^{i\omega_0})$ will then be constant if either of the following two conditions are satisfied:

1) $\omega_*(t_1)$ does not depend on t_1. This happens if the input is a pure sinusoid and all transients in the output have died out.

2) The true system is of first order. It then does not matter at which frequency the fit is made.

The variability of $\hat{G}_t(e^{i\omega_0})$ when neither of these conditions is satisfied will then depend on the degree to which they are violated. A broad band output signal for a system that cannot be well approximated by a first order system will thus create considerable variation in the model. (Compare Figure 9 below).

When the model fit is extended over longer data records, the effect is that the model is smoothed (averaged) over the corresponding time span (with data dependent weighting coefficients, just as done in least squares solutions).

5 THE ROLE OF PRE-FILTERING

In this section we shall consider the role of prefiltering the input-output data before subjecting them to the identification procedure.

5.1 Parametric methods

The prefilter $L(q)$ in (13) - (15) plays two important, but different, roles.

First, it will affect the limiting result $\theta*$. This is evidenced by (17) - (19). The frequency bands where the prefilter has its support will be enhanced, according to (19) so that the model fit in those bands will be improved, at the expense of the fit in other bands. Chapter 13 of Ljung (1987) contains a detailed discussion of these aspects.

Second, the prefilter will also affect the variability of the model. This can be seen from expressions (25) or (30) together with (26) - (29).

Conceptually, we may write for (26)

$$P \sim (\text{energy in } \varepsilon_F)/(\text{energy in } \varphi_F) \qquad (45)$$

Here, in the noise free case

$$\text{energy in } \varepsilon_F \sim$$

$$\int_{-\pi}^{\pi} |G_*(e^{i\omega}) - G_0(e^{i\omega})|^2 \frac{|L(e^{i\omega})|^2 \Phi_u(\omega)}{|H_*(e^{i\omega})|^2} d\omega \qquad (46)$$

and

$$\text{energy in } \varphi_F \sim \int_{-\pi}^{\pi} |L(e^{i\omega})|^2 [\Phi_u(\omega) +$$

$$+ |G_0(e^{i\omega})|^2 \Phi_u(\omega)] d\omega \qquad (47)$$

G_*, H_* is the limiting model. By prefiltering the energy contents in ε_F and φ_F (see (24), (27), and (28)) will be decreased. If the relative energy reduction in ε_F and φ_F would be the same, the net result in P (see (26)) would be balanced by the contributions from R and Q, with no resulting decrease in variance. This is the situation that holds in the "standard case" when $\varepsilon(t)$ is dominated by a stochastic unpredictable error contribution $e(t)$. When the model error contribution $\eta(t)$ (see (2)) dominates, the situation is different, though. Then, since the fit is concentrated to a smaller frequency band by the prefiltering, the fit in this band can be substantially improved, and a reduction of energy of $\varepsilon_F(t)$ results, in addition to the reduction that the filter itself gives. Thus Q in (28) will decrease more than $R^{-1} \cdot R^{-1}$ increases, and the net result is a decrease in P. This is what we saw in Figures 4 and

5: Not only did the average level of the model change by the prefilter - its variability also decreased.

To be a bit more specific, assume that the prefilters that we apply are Butterworth Band-Pass filters. We may use one or the sum of several, to emphasize a certain, or several frequency bands. The locations and widths of these depend on the identification application. The choices we are faced with are thus

1. Width of BP-filter
2. Order of Butterworth filter.

Both these choices will affect the covariance matrix P in (26) as well as the length of the impulse response of the filter. The relationship between time response and frequency response, choice of order and the like is a standard topic in filter theory. See, e.g. Siebert (1985) or Parks and Burrus (1987). For our applications we may note the following:

Influence of width: Assume for simplicity that we use only one filter and a first order model. Let the width of the BP filter be W centered around frequency ω_0. For an ideal BP filter (46) gives

$$\text{energy in } \varepsilon_F \sim$$

$$\sim 2 \int_{\omega_0-W}^{\omega_0+W} |G_*(e^{i\omega}) - G_0(e^{i\omega})|^2$$

$$\frac{\Phi_u(\omega)}{|H_*(e^{i\omega})|^2} d\omega \approx [\text{small } W] \approx \ const$$

$$\int_{\omega_0-W}^{\omega_0+W} |G_*(e^{i\omega}) - G_0(e^{i\omega_0}) -$$

$$- G_0'(e^{i\xi})(\omega - \omega_0)|^2 d\omega \sim C \cdot W^3 \tag{48}$$

Here the second step follows by Taylor's expansion of G_0 around ω_0 and the third step from the fact that the first order model G_* can by best fit pick up the term $G_0(e^{i\omega_0})$. Similarly

$$\text{energy in } \varphi_F \sim C \cdot W \tag{49}$$

Hence, we find from (45) that the variance term will decay like

$$P \sim C\, W^2 \tag{50}$$

From the well known uncertainty relation for filters (e.g. Siebert (1985)) we also know that the response time of the filter will obey

$$T \gtrsim \sqrt{2\pi}/W \tag{51}$$

Influence of filter order. A Butterworth filter of order n behaves like

$$\frac{1}{1 + (\frac{\omega}{W})^{2n}} \tag{52}$$

outside the assigned BP domain. Taking those non-ideal effects into account in (48) gives a contribution

$$\text{energy in } \varepsilon_F \sim C_1 \cdot W^3 + \frac{C_2}{n} \cdot W \tag{53}$$

which means that

$$P \sim C_1 W^2 + \frac{C_2}{n} \tag{54}$$

The response time of an n:th order discrete time Butterworth filter is

$$T \approx C_3 \cdot n \quad (C_3 \approx 2) \tag{55}$$

5.2 Empirical transfer function estimate

Consider (5) - (7). The basic expression to study is (31). The upper limit (32) - (33) can be affected only by obtaining a small amplitude for $u(t)$, at the same time as $U_N^t(\omega)$ is large. By filtering the input and output through a BP filter around frequency ω_0 we do not decrease the amplitude of $|U_N^t(\omega_0)|$, but may substantially decrease the input amplitude. This accounts for the decreased variability in Figure 7, compared to Figure 6. Notice that for the empirical transfer function estimate (5) - (7), prefiltering does not affect the limiting model. Also this fact follows from (31) - (32).

6 PERFORMANCE LIMITATIONS

A major reason for using identification and recursive identification techniques is that the true system dynamics vary with time. It is then the task of the identification algorithm to track the variations. This leads to a classical trade-off problem between the tracking ability (high degree of alertness) and noise sensitivity. In the noiseless environment of this article, what are then the trade-offs? As we have seen, the recursively computed estimates, may exhibit some rather wild variations even in the noise free case. It is then still a problem to distinguish between those changes in the estimates that stem from actual variations in the system and those that correspond to "natural estimate variability".

In this section we shall discuss the tracking problem in a noise free environment. We start by an example.

Example 1. Consider the system

$$y(t) - 2.14y(t-1) + 1.553y(t-2) - 0.4387y(t-3) +$$

$$+0.042y(t-4) = 0.001(1.0u(t-2) +$$

$$+ 7.4u(t-3) + 0.924u(t-4) + 0.1764u(t-5) \quad (56)$$

(some as in Example 8.5 in Ljung (1987a)).

At sample 100 the gain of the system is halved.

The input is a white noise process. We are only interested in the transfer function at frequency 1 rad/s.

In Figure 9 we see the model transfer function at frequency 1, when the used model structure is given by (3) with $na = nb = 2$, $nk = 1$, and the forgetting factor is $\lambda = 0.95$.

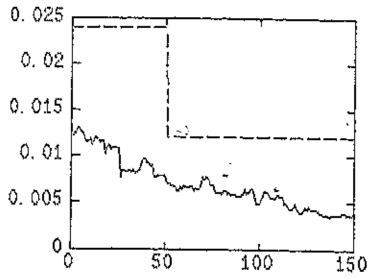

Figure 9: The amplitude of the model's transfer function at frequency 1 rad/s. Dashed line: True value. Model structure (3) with $na = nb = 2$, $nk = 1$ and forgetting factor $\lambda = 0.95$.

Figure 10 shows the same curve when the recursive least squares method has been applied to data filtered through a 5th order BP Butterworth filter with passband [0.9 1.1] rad/s, while a 3rd order filter has been used in Fig 11. Figures 12 and 13 show the corresponding results for the forgetting factor $\lambda = 0.8$

□

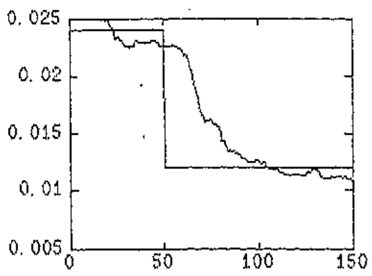

Figure 10: As Figure 9, but a 5th order band pass Butterworth filter first applied to the data. $\lambda = 0.95$.

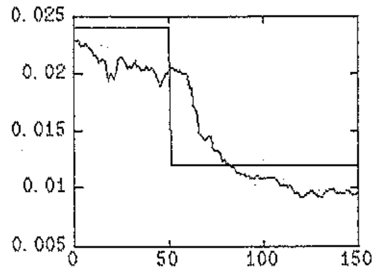

Figure 11: As Figure 10, but a 3rd order Butterworth filter used instead, $\lambda = 0.95$.

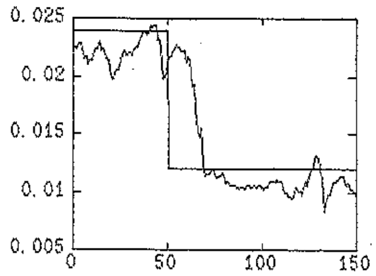

Figure 12: As Figure 10, but $\lambda = 0.8$.

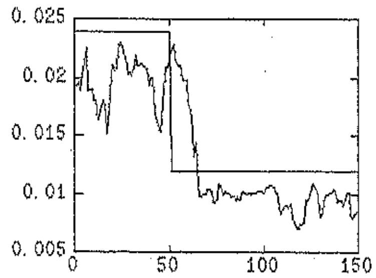

Figure 13: As Figure 11, but $\lambda = 0.8$.

Inspired by the example, let us discuss what it takes to detect a sudden system change quickly. We have two tools to play with: The forgetting factor λ and the prefilter $L(q)$. We are striving for low variability in the estimate, so that a definite change can be quickly and easily detected. How do we obtain low variability? According to (30) and (54) we have

$$Var\ \hat{\Theta}(t) \sim (1-\lambda)[C_1 W^2 + \frac{C_2}{n}] \quad (57)$$

Calm estimates are thus obtained by little forgetting (λ close to 1) and/or by narrow band, high order BP filters.

But these same features, which both involve averaging over old measurement will also delay the effect of a change. The forgetting factor recursive least squares algorithm, can be seen as a first order filter with time constant $1/(1-\lambda)$ (see, e.g. Ljung (1987) Ch. 11), and the BP filter will, according to (51) and (55) have a time constant of

$$T \gtrsim \frac{c}{W} + 2n$$

The delay due to forgetting and BP-filtering will thus behave like a time constant of

$$T_{delay} \approx \frac{1}{1-\lambda} + \frac{\sqrt{2\pi}}{W} + 2n \qquad (58)$$

The trade-off is thus expressed by (57) and (58): We would like both to be small. The situation resembles that is the classical tracking ability (58) - noise sensibility (57) balance. We have here though, the added feature of the BP-filter which gives useful extra flexibility in the trade-off.

The two expressions (57) and (58) also set theoretical limits as to how quickly a change can be detected. The practical choice will of course depend on the constants C_1 and C_2. These in turn depend on the model structure and on the true system. When a perfect fit is possible (the system belongs to the model set) C_1 and C_2 are both zero, and we may use a small value of λ and no filtering to yield detection within a few samples.

7 CHOICES OF FORGETTING FACTOR AND PREFILTER

We shall in this section summarize the user aspects we have discussed earlier. With a given model structure and a given input sequence, the two main tools to affect the model estimate are

- The prefilter $L(q)$ (According to (13) - (14), this can alternatively be seen as a noise model).
- The forgetting factor λ (other means to affect the adaptation rate can also be used, see e.g. Ljung (1987a), Ch 11).

Our goal is to achieve a good estimate of the system transfer function $G(e^{i\omega})$ in the frequency range $\omega \in \Omega$, at all times. We then have to consider the following items:

1. Make a preliminary selection of L to match the selected frequency range Ω.
2. Link the choice of λ to the choice of L (both order and width) so that the impulse response length of L is $\approx 1/(1-\lambda)$.
3. Choose λ as small as possible (with corresponding consequences for L according to 2.) with the constraint that the variability of the estimate is acceptably small.

If the input also can be selected by the user, it is of course desirable to let it be narrow band within Ω. The need for BP filtering is then decreased. However, if the system changes, these changes will excite the natural modes of the system. Fast modes pose no problems. Slow modes, below Ω will delay the change detection but can be eliminated by a prefilter. Note that a forgetting factor does nothing for this problem. Natural modes within Ω are tricky (and it is likely that we will have natural modes there, since our focus of interest in the frequency range typically is where "things happen"). These modes will beat with the forced output and cannot be eliminated neither by prefiltering nor by forgetting.

8 CONCLUSIONS

We have in this contribution considered a noise-free identification environment, where the true system cannot be exactly modelled in the chosen structure. Although idealized, this environment is probably closer to the truth in many applications than the conventional noise-corrupted-but-exact-model-matching setup.

The unmodelled dynamics will then take up and play the part of disturbances. Although many conventional results fail or change, the bottom line is that most basic features of the noise influence will carry over to the situation considered here. The most important difference is that the "signal-to-noise" ratio can be improved by prefiltering. This gives several new aspects to the choice of forgetting factors and prefilters for best tracking properties.

REFERENCES

Kosut, R. L. (1986). Adaptive Calibration: An approach to uncertainty modelling and on-line robust control design. *Proc. 25th IEEE CDC*, Athens, Greece.

Kosut, R. L. (1987). Adaptive Uncertainty Modelling: On-line robust control design. *Proc. 1987 ACC*, Minneapolis, MN.

Kosut, R. L. (1988). Adaptive Robust Control via Transfer Function Uncertainty Estimation. *Proc. 1988 ACC*, Atlanta, GA.

LaMaire et al (1987). A frequency domain estimator for use in adaptive control systems. *Proc. 1987 ACC*, pp 238-244, Minn., MN.

Ljung, L. (1985). A nonprobabilistic framework for signal spectra. *Proc. 24th IEEE Conf. on Decision and Control*, Fort Lauderdale, Fla, pp 1056-1060.

Ljung, L. (1987a). *System Identification - Theory for the user.* Prentice-Hall, 1987.

Ljung, L. (1987b). *The System Identification Toolbox.* The MathWorks Inc.

Ljung, L. and Zh-D. Yuan (1985). Asymptotic properties of black-box identification of transfer functions. *IEEE Trans. Autom. Control*, vol. AC-30, pp. 514-530.

Parker, P. J. (1988). Frequeny domain descriptions of lienar systems. Ph.D. Thesis, Dept of Systems Engineering, The Australian National University.

Parker, P. J. and R. R. Bitmead (1987). Adaptive frequency response identification. *Proc. 26th IEEE Conf. on Decision and Control*, Los Angeles, vol I. pp 348-353.

Parks, T. W. and C. S. Burrus (1987). *Digital Filter Design*, Wiley, New York.

Rohrs, C. E. et al (1984). Some design guidelines for discrete-time adaptive controllers. *Automatica*, vol. 20, pp 653-660.

Siebert, W. M. C. (1985). *Circuits, Signals and Systems*, MIT Press, Cambridge, MA.

Wittenmark, B. and K. J. Åstrãm (1984). Practical issues in the implementation of self-tuning control, *Automatica*, Vol. 20, no. 5, pp 595-605.