

# IDENTIFICATION OF TIME-VARYING MODAL MODELS

J. P. Norton

*Dept. of Mathematics and Integrated Catchment Assessment & Management Centre,  
Australian National University, Canberra, Australia.; [john.norton@anu.edu.au](mailto:john.norton@anu.edu.au)  
and Dept. of Electronic, Electrical & Computer Eng., University of Birmingham, UK*

Abstract: Modal, decoupled models of linear systems have easily interpreted parameters: gains and poles for independent rate processes, hence time constants and damping ratios. Time-varying modal models can fit a wide range of time-invariant, non-linear systems, and the time variation shows the effects of non-linearity. Modal parameters can be estimated via an ARMAX model, but the calculation is ill conditioned, so direct estimation of time-varying modal parameters is examined. The modal components of the output cannot be observed so must be estimated along with the parameters, a non-linear estimation problem. It is treated as state estimation, using the extended Kalman filter (EKF) and optimal smoothing. The capabilities and limitations of the EKF are investigated by simulation. With care, good results can be obtained even in difficult cases, e.g. with abrupt change in an input or output offset. Copyright © 2005 IFAC

Keywords: Parameter estimation; linear models; modal models; time-varying parameters; optimal smoothing

## 1. INTRODUCTION

In the identification of non-linear systems, linear but time-varying (LTV) models can act as a half-way step to non-linear, time-invariant models. The idea is to examine the time variation of the LTV model parameters to see how the model structure must be modified to produce a time-invariant model. For example, a gain found to vary in sympathy with the input indicates an input non-linearity. There exist long-established recursive algorithms to estimate the parameters of LTV models (Mayne, 1963; Lee, 1964; Norton, 1975, 1976; Young, 1984; Norton, 1986; Niedzwiecki, 2000).

Time-varying parameters may be handled efficiently in identification by representing them as random walks. The extent of variation of individual parameters can then be controlled selectively through the variances of their increments, with flexibility not afforded by forgetting factors or deterministic trend models. A wide range of behaviour can be accommodated by a simple random walk (SRW) (Norton, 1975) or integrated random walk (IRW) (Norton, 1976) model for each time-varying parameter. Each requires only a single number to be specified. For an SRW, the variance of its zero-mean, wide-sense-stationary increments controls the extent of its variation. For an IRW, the variance of the second differences controls the smoothness. The range of variation of an IRW is insensitive to its variance, so a parameter can be estimated initially, without fine tuning, as an IRW to find its range of variation, which is then used to give the variance of an

SRW model (Norton, 1976). Alternatively the variances can be treated as meta-parameters, found by maximum-likelihood estimation (Young *et al.*, 1991).

Parameters modelled as random walks are effectively state variables. Along with sampled earlier inputs and outputs, they determine the model output through an observation equation. Formulating LTV parameter estimation as state estimation has the advantage of allowing the parameter estimates to be improved by fixed-interval optimal smoothing (Bryson and Ho, 1975; Norton, 1975). The sequence of estimates from a “filtering” run through the input-output records is refined by reverse-time “smoothing” which exploits the information about the parameters present in later output samples. This is essential with short records, to distinguish genuine parameter variation from the convergence transient. Optimal smoothing also removes the lag between parameter variations and their estimates incurred by filtering.

LTV identification with optimal smoothing is long established and has been found effective in a range of applications (Norton, 1975, 1976; Young *et al.*, 1991; Evans *et al.*, 2001; Chanut and Norton, 2003; Norton and Chanut, 2005). This paper examines modifications to extend its scope. Specifically, the technique is extended to cover decoupled modal models, moving-average noise models independent of the plant model, and records with constant or time-varying offsets in input or output.

Section 2 compares autoregressive-moving average-

exogenous (ARMAX) and modal model structures and finds several reasons for replacing the linear pseudo-regressions often employed for parameter estimation by models containing products of unknowns. Section 3 recalls an algorithm for identifying LTV ARMAX models by extended least squares and optimal smoothing, and suggests an alternative to handle products of unknowns while retaining enough linearity to allow optimal smoothing. In Section 4 the procedure is used on very demanding examples to test the performance obtainable. Section 5 draws conclusions.

## 2. ARMAX AND MODAL MODELS

Approximate minimum-variance, linear, unbiased recursive parameter estimation often employs the pseudo-regression ARMAX model

$$\begin{aligned}
y_t &= -a_1 y_{t-1} - \dots - a_n y_{t-n} + b_1 u_{t-1} + \dots \\
&+ b_n u_{t-n} + c_1 e_{t-1} + \dots + c_n e_{t-n} + e_t \\
&\equiv A(q^{-1})y_t + B(q^{-1})u_t + (1 + C(q^{-1}))e_t \\
&\equiv \mathbf{h}_t^T \mathbf{x}_t + e_t
\end{aligned} \tag{1}$$

where the subscripts  $t-i$  indicate discrete time,  $u$  and  $y$  are observed input and output, the regressor and unknown parameter vectors are

$$\begin{aligned}
\mathbf{h}_t &\equiv [-y_{t-1} \dots -y_{t-n} \quad u_{t-1} \dots u_{t-n} \quad e_{t-1} \dots e_{t-n}]^T \\
\mathbf{x}_t &\equiv [a_1 \dots a_n \quad b_1 \dots b_n \quad c_1 \dots c_n]^T
\end{aligned} \tag{2}$$

and  $\{e\}$  is zero-mean, white noise. Operator  $q^{-1}$  is a one-sample delay and  $A(q^{-1})$ ,  $B(q^{-1})$ , and  $C(q^{-1})$  are polynomials. Including  $C(q^{-1})e_t$  in  $\mathbf{h}_t^T \mathbf{x}_t$  makes the “noise” term  $e_t$  in (1) uncorrelated with  $\mathbf{h}_t$ , avoiding estimation bias in  $\mathbf{x}_t$  but requiring approximation of the unknown  $e$ 's in  $\mathbf{h}_t$  by estimates from earlier updates. The variances of the white, zero-mean increments in the random-walk model

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{w}_{t-1} \tag{3}$$

(user-supplied, zero for constant parameters) define the diagonal covariance matrix  $\mathbf{Q}_{t-1}$  of  $\mathbf{w}_{t-1}$ .

As (1) is linear in  $\mathbf{x}_t$ , “optimal” filtering and

smoothing can be applied, as described in Section 3.

Statistical efficiency of the estimates is reduced by use of estimated  $e$ 's in  $\mathbf{h}_t$ ; minimum covariance and zero bias are not generally achieved. This may not be significant, as ideal statistical properties are unachievable since (1) and its associated assumptions idealise the behaviour of the records, neglecting unmodelled high-order dynamics, non-linearity and non-stationarity. Only if the errors in the  $e$ 's dominate such systematic modelling errors are they of concern. However, (1) raises other practical issues.

First, writing (1) in output-error form

$$y_t = \frac{B(q^{-1})}{1 - A(q^{-1})} u_t + \frac{1 + C(q^{-1})}{1 - A(q^{-1})} e_t \tag{4}$$

the noise submodel is not parsimonious, as  $1+C$  must cancel  $1-A$  unless all noise is affected by all the plant poles. Conversely, if the parsimonious model

$$y_t = \frac{B(q^{-1})}{1 - A(q^{-1})} u_t + (1 + C(q^{-1}))e_t \tag{5}$$

is rewritten as a pseudo-regression, the noise model  $(1+C)e$  is multiplied by  $1-A$ . The products of parameters in  $CA$  then spoil the linearity needed for the optimal filtering and smoothing algorithm.

Second, a linear model such as (1) or (4) commonly describes small-signal behaviour, about given input and output values, of a system with some non-linearity between total input and total output. Common practice is to refer the input and output to their sample means, subtracting the mean from each record before parameter estimation. If the total input-total output relation is  $y = f(u)$  and  $u' \equiv u - u^*$ ,  $y' \equiv y - y^*$ , then for small changes about  $(u^*, y^*)$

$$\begin{aligned}
y' &= f(u^* + u') - y^* \\
&\equiv f(u^*) + \left. \frac{\partial f}{\partial u} \right|_{u^*} u' - y^* \equiv \left. \frac{\partial f}{\partial u} \right|_{\bar{u}} u' \equiv g(u^*)u'
\end{aligned} \tag{6}$$

which is valid only if  $y^* = f(u^*)$ , i.e. about *mutually compatible* input and output values. Taking a mean does not commute with a non-linear relation  $f(\cdot)$ , so the sample means  $(\bar{u}, \bar{y})$  are generally incompatible. If  $\bar{y} = f(\bar{u}) + \tilde{y}$ , the small-signal relation about  $(\bar{u}, \bar{y})$  is  $y' = y - \bar{y} \equiv g(\bar{u})u' - \tilde{y}$ , not  $y' \equiv g(\bar{u})u'$ . As  $f$  is unknown, so is bias  $\tilde{y}$ . The solution is to write the model explicitly for deviations from nominal:

$$y_t = y_t^* + \frac{B(q^{-1})}{1-A(q^{-1})}(u_t - u_t^*) + \text{noise} \quad (7)$$

It may be desirable to allow the nominal values to vary with time, to allow for drift and abrupt changes:

$$y_t = y_t^* - a_1(y_{t-1} - y_{t-1}^*) \dots - a_n(y_{t-n} - y_{t-n}^*) + b_1(u_{t-1} - u_{t-1}^*) + b_n(u_{t-n} - u_{t-n}^*) + \text{noise} \quad (8)$$

The  $u^*$ 's and  $y^*$ 's cannot be supplied from earlier updating steps like the unknown  $e$ 's in (1), so (8) is bilinear in its unknowns.

The third practical issue is the uninformative nature of ARMAX parameters. Their credibility and the implications of their time variation are best assessed by transforming them into poles  $p_i$  and gains  $g_i$  in the modal model

$$y_t = y_t^* + \sum_{i=1}^n \frac{g_i q^{-1}}{1 - p_i q^{-1}} (u_t - u_t^*) + c_1 e_{t-1} + \dots + c_n e_{t-n} + e_t \quad (9)$$

Each  $|p_i|$  and  $|g_i|$  relates directly to the time constant and gain coefficient of an underlying rate process. For complex modes  $\angle p_i$  and  $\angle g_i$  relate to the frequency and initial phase of an oscillatory process. The snag is that the transformation from (8) to modal model (9) is ill conditioned. For instance, for  $n=2$  the normalised sensitivities of  $g_1$  to  $a_1 = -p_1 - p_2$ ,  $a_2 = p_1 p_2$ ,  $b_1 = g_1 + g_2$  and  $b_2 = -g_1 p_2 - g_2 p_1$  are

$$\begin{aligned} \frac{a_1}{g_1} \frac{\partial g_1}{\partial a_1} &= \frac{(p_1 + p_2)(p_1 g_2 - p_2 g_1)}{g_1 (p_1 - p_2)^2} \\ \frac{a_2}{g_1} \frac{\partial g_1}{\partial a_2} &= \frac{p_1 p_2 (g_1 - g_2)}{g_1 (p_1 - p_2)^2} \\ \frac{b_1}{g_1} \frac{\partial g_1}{\partial b_1} &= \frac{p_1 (g_1 + g_2)}{g_1 (p_1 - p_2)} \end{aligned}$$

$$\frac{b_2}{g_1} \frac{\partial g_1}{\partial b_2} = \frac{g_1 p_2 + g_2 p_1}{g_1 (p_2 - p_1)} \quad (10)$$

Some are large if the poles are close and the gains comparable in magnitude. The alternative of directly identifying the modal model entails writing (9) as

$$y_{i,t} = p_{i,t} y_{i,t-1} + g_{i,t-1} (u_{t-1} - u_{t-1}^*), \quad i = 1, 2, \dots, n \quad (11)$$

$$y_t = y_t^* + \sum_{i=1}^n y_{i,t} + c_{1,t} e_{t-1} + \dots + c_{n,t} e_{t-n} + e_t$$

which has unknowns

$$\mathbf{x}_t \equiv [p_{1,t} \dots p_{n,t} \quad g_{1,t} \dots g_{n,t} \quad y_{1,t} \dots y_{n,t} \quad c_{1,t} \dots c_{n,t} \quad e_{t-1} \dots e_{t-n} \quad y_t^* \quad u_{t-1}^*]^T \quad (12)$$

Again products of unknowns occur in the pseudo-linear

$$y_t = \mathbf{h}_t^T \mathbf{x}_t + e_t \quad \text{resulting from (11) and (12).}$$

All three practical considerations thus give rise to models containing products of unknowns. The remainder of the paper discusses application to model (11) of an optimal-smoothing-based estimator which covers all three situations.

### 3. PARAMETER ESTIMATION ALGORITHMS FOR LTV MODELS

The filtering pass producing  $\hat{\mathbf{x}}_t, t = 1, 2, \dots, N$  is

$$\begin{aligned} \hat{\mathbf{x}}'_t &= \mathbf{F}_{t-1} \hat{\mathbf{x}}_{t-1}; \mathbf{P}'_t = \mathbf{F}_{t-1} \mathbf{P}_{t-1} \mathbf{F}_{t-1}^T + \mathbf{Q}_{t-1} \\ \mathbf{g}_t &= \mathbf{P}'_t \mathbf{h}_t; m_t = 1 + \mathbf{g}_t^T \mathbf{h}_t; \\ \mathbf{P}_t &= \mathbf{P}'_t - \mathbf{g}_t \mathbf{g}_t^T / m_t \\ v_t &= y_t - \mathbf{h}_t^T \hat{\mathbf{x}}'_t; \hat{\mathbf{x}}_t = \hat{\mathbf{x}}'_t + \mathbf{g}_t v_t / m_t \end{aligned} \quad (13)$$

where  $\mathbf{P}_t$  is  $\text{cov} \hat{\mathbf{x}}_t / \sigma_{et}^2$  and  $\sigma_{et}^2 = E[e_t^2]$ . In  $\mathbf{h}_t, e_{t-1}$  to  $e_{t-n_c}$  are approximated by earlier residuals  $\hat{e}_{t-i} = y_{t-i} - \mathbf{h}_{t-i}^T \hat{\mathbf{x}}_{t-i}$ . The smoothing pass over  $t=N, N-1, \dots, 1$  is

$$\begin{aligned} \boldsymbol{\rho}_t &= \mathbf{F}_t^T \boldsymbol{\mu}_t; \hat{\mathbf{x}}_{t|N} = \hat{\mathbf{x}}_t - \mathbf{P}_t \boldsymbol{\rho}_t \\ \boldsymbol{\mu}_{t-1} &= \boldsymbol{\rho}_t - \mathbf{h}_t (v_t + \mathbf{h}_t^T \boldsymbol{\rho}_t) / m_t \end{aligned} \quad (14)$$

This algorithm is computationally cheap and stable, unlike some algebraic equivalents (Norton, 1975), but requires storage of all filtered  $\mathbf{P}$ 's. For an ARMAX model with SRW models of time-varying parameters,  $\mathbf{F}_t = \mathbf{I}$  and a simpler version requiring inversion of  $\mathbf{F}_t$  can be used (Norton, 1975).

The non-zero diagonal elements of  $\mathbf{Q}$  are tuned to keep the ratio  $\rho = \text{m.s. } v_{t|N} / \text{m.s. } \hat{e}_{t|N}$  not far above 1 with m.s.  $\hat{e}_{t|N}$  small and credible parameter variations. Too much or too little parameter variation causes larger m.s. prediction error, raising  $\rho$  well above 1. This tuning technique requires no assumptions about distribution or stationarity.

With SRW models for all time-varying parameters, model (11) can still employ optimal filtering and smoothing if linearised about the latest state estimate:

$$\mathbf{h}_t \equiv \left[ \mathbf{0}^T \quad \mathbf{1}^T \quad e_{t-1} \dots e_{t-n_c} \quad c_1 \dots c_{n_c} \quad 1 \quad 0 \right]^T \quad (15)$$

and transition matrix  $\mathbf{F}_t$  is identity except for a unit-downward-shift matrix for the  $e$ 's as partition (4,4);  $\text{diag}(y_{1,t}, \dots, y_{n,t})$  and  $\text{diag}(p_{1,t}, \dots, p_{n,t})$  as partitions (3,1) and (3,3); and  $-\mathbf{h}_t^T$  as row  $2n+n_c+1$ . The modal gains can alternatively be assigned to the output, simplifying  $\mathbf{F}_t$  but complicating  $\mathbf{h}_t$ . As in any extended Kalman filter (EKF), linearisation errors may dictate performance and even prevent convergence from poor initial estimates.

#### 4. RESULTS

Records of 2560 input-output pairs were generated with  $n = 2, p_1 = 0.8, p_2 = 0.6, g_1 = 4, g_2 = -3,$   
 $n_c = 2, n_c = 2, c_1 = 0.8, c_2 = 0.3$  in (11). In some runs  $g_1$  varied sinusoidally about 4 with peak amplitude 2 and, and/or offsets  $\bar{u} = -0.2, \bar{y} = 2$  reversed signs instantaneously for certain periods. The input record and noise-generating sequence  $\{e\}$  were white and u.d. over  $(-0.5, 0.5)$ . The m.s. output SNR was 10. To get insight into the behaviour of time-varying estimates, a large number of single realisations were examined. A small selection, illustrating generic conclusions, are discussed below.

Identifiability analysis *via* state observability would give data-dependent results, as the linear operators relating state at any instant to succeeding observations are the results of linearisation about current estimates. The simulation runs reported focus on identifiability, the effectiveness of smoothing, the influence of initial

estimates and variances, and tuning of  $\mathbf{Q}$ . For brevity results are by default for initial estimates 0.9, 0.5 for the poles with normalised variances 0.4 (to avoid risk of instability); 1, -1 for the modal gains (unequal to impose a definite order and avoid repeated poles, the unequal initial pole values being insufficient to ensure this); 0 for the modal outputs with variances normally  $10^6$ ; 0 for the  $c$ 's and  $e$ 's, with variances 10 and 1 (the latter by definition, as (13), (14) are normalised by the variance of  $e$ ); 0 for  $\bar{u}$  and  $\bar{y}$  with variance  $10^6$ .

As expected, constant  $\bar{u}, \bar{y}$  are found not to be separately identifiable, varying greatly with the specified error variances of their initial estimates but influencing the m.s. value  $s_v$  of the output 1-step prediction error by only a few percent. Fixing either at its correct value allows the other to be identified, mostly to well within 1%, over a range of initial variances from 100 to  $10^6$  for the  $g$ 's, modal outputs, and whichever of  $\bar{u}$  and  $\bar{y}$  is free. Estimates of  $p_1$  range from 0.816 to 0.842,  $p_2$  0.459 to 0.547,  $g_1$  2.46 to 3.19 and  $g_2$  -1.68 to -2.17. Even so,  $s_v$  varies only by 8% and the steady-state gain  $G$  by 4%, with all but one  $G$  within 1% of the actual value 12.5. The EKF is thus reasonably robust and captures the main features of the dynamics, but with accuracy which might be inadequate.

Figure 1 shows estimates obtained after crude tuning of  $\mathbf{Q}$  when  $\bar{y}$  reverses sign from  $t=450$  to  $t=850$  while  $g_1$  varies sinusoidally between 2 and 6.

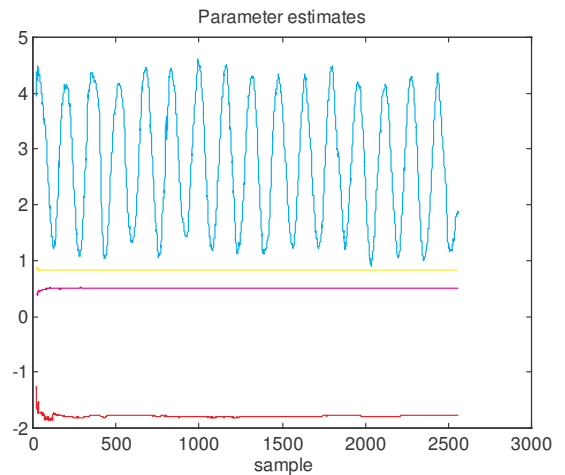
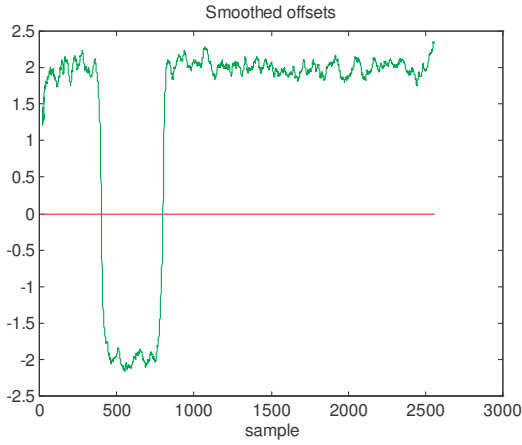


Figure 1(a) Smoothed estimates: top to bottom

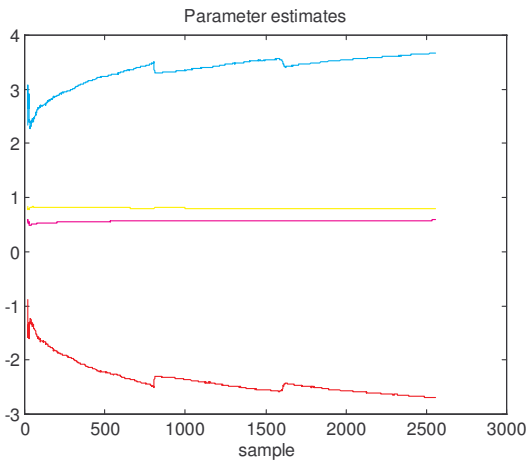
$$g_1, p_1, p_2, g_2$$



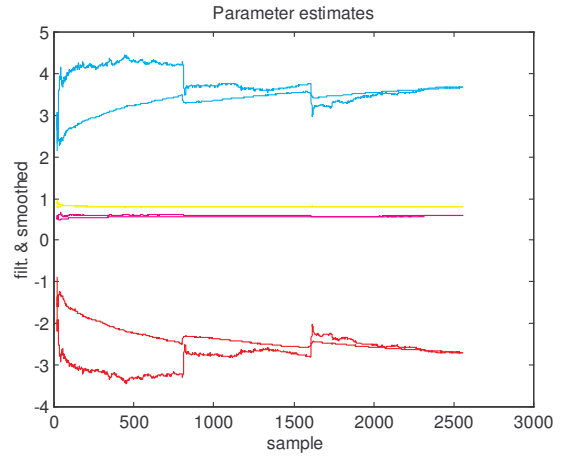
**Figure 1(b)** Smoothed estimates:  $\bar{y}$  (varying),  $\bar{u}$

Similar mean errors to the previous case are evident, and the variation of  $g_1$  is underestimated (even with its increment variance larger), but the variation in  $\bar{y}$  is captured well and at the same time the timing and nature of that in  $g_1$  are well caught.

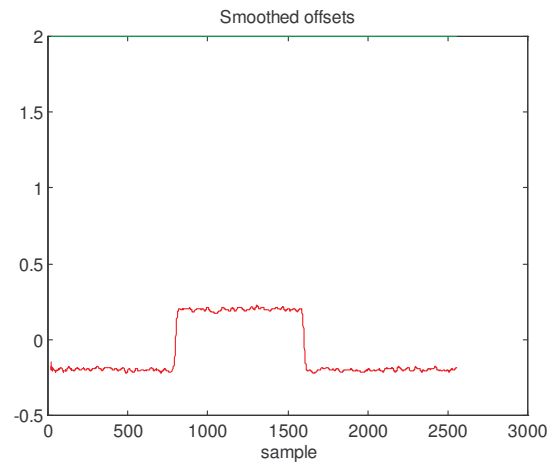
In the next example  $\bar{u}$  reverses sign from  $t=850$  to  $t=1650$ , a severe test as the algorithm has to distinguish the effects of these changes on the output from those of a change in one or both modal gains. The first graph of Figure 2(a) shows that, with the variances of the increments in  $g_1$  and  $g_2$  specified as zero, optimal smoothing cannot quite remove the spurious variation in  $g_1$  and  $g_2$  (because of linearisation error), but Figure 2(b) confirms that it has reduced it very considerably. Figure 2(c) shows excellent estimation of  $\bar{u}$ .



**Figure 2(a)** Smoothed estimates; order as in Fig. 1(a)



**Figure 2(b)** Filtered and smoothed estimates



**Figure 2(c)** Smoothed estimates of  $\bar{u}$

Other broad conclusions from the cases investigated so far are that the biases cannot be removed by tuning of  $\mathbf{Q}$ , the initial estimates or their specified variances, or by elaborating the noise model or allowing extra process noise terms to account for linearisation errors; the m.s. 1-step output prediction error is a reliable guide to model quality; the estimated covariance of the parameter errors is not a reliable guide; the specified variances of the initial estimates of the noise-model ( $c$ ) parameters must not be too large (as is reasonable, as the noise model should not be allowed the freedom to interact with the time-varying model of the dynamics); convergence is slow compared with that for the equivalent ARMAX model; the values; optimal smoothing diverges in some cases (with relatively large initial errors and  $\mathbf{Q}$ ) where filtering does not; and the relation between  $\mathbf{Q}$  and the variation of the parameter estimates is straightforward and consistent enough to make tuning  $\mathbf{Q}$  fairly easy.

## 5. CONCLUSIONS

A combination of a modal model, modelling of time-

varying parameters as random walks, extended Kalman filtering and optimal smoothing can give useful results in demanding cases including abruptly changing input or output offset (but not both) and/or strongly time-varying modal gain. In contrast to ARMAX models treated by extended least squares and optimal smoothing, the noise model is decoupled from the plant dynamics. However, there remains doubt over how much of the bias found in the poles and modal gains is reducible in principle by an improved algorithm, and how much is intrinsic to the form of the model. Insensitivity of the prediction performance to these biases suggests that they are largely intrinsic.

Norton, J. P. and J. G. Chanut (2005). Linear time-varying models to investigate complex distributed dynamics: a rainfall-runoff example, *Mathematics & Computers in Simulation*, to appear.

Young, P. C. (1984). *Recursive Estimation and Time Series Analysis*, Springer-Verlag, Berlin.

Young, P. C., C. Ng, K. Lane and D. Parker (1991). Recursive forecasting, smoothing and seasonal adjustment of non-stationary environmental data, *Journal of Forecasting* **10**, 57-89.

## REFERENCES

- Bryson, A. E. and Y-C. Ho (1975). *Applied Optimal Control: Optimization, Estimation and Control*, John Wiley and Sons, New York.
- Chanat, J. G. and J.P. Norton (2003). Interpreting hydrologic response using transfer function models with time-varying parameters: an example from the Virginia Blue Ridge, *Proc. MODSIM 2003 Int. Conf. on Modelling & Simulation, Townsville, Australia*, 933-938.
- Evans, C., P. J. Fleming, D. C. Hill, J. P. Norton, I. Pratt, D. Rees and K. Rodriguez-Vazquez (2001). Application of system identification techniques to aircraft gas turbine engines, *Control Engineering Practice* **9**, 135-148.
- Lee, R. C. K. (1964). *Optimal Identification, Estimation and Control* MIT Press, Cambridge, Mass.
- Mayne, D. Q. (1963). Optimal nonstationary estimation of the parameters of a linear system with Gaussian inputs, *J. Electron. Control*, **14**, 101-112.
- Niedzwiecki, M. (2000). *Identification of Time-Varying Processes*, John Wiley & Sons, Chichester, UK.
- Norton, J. P. (1975). Optimal smoothing in the identification of linear time-varying systems, *Proc.IEE* **122**, 663-668.
- Norton, J. P. (1976). Identification by optimal smoothing using integrated random walks, *Proc.IEE* **123**, 451-452.
- Norton, J. P. (1986). *An Introduction to Identification*, Academic Press, London and New York.