Direct Estimation of Continuous-Time Models From Sampled Data

Peter Young

Lancaster Environment Centre, U.K.

Australian National University, Canberra

p.young@lancaster.ac.uk

Prepared in honour of Maurice Priesley

Comemoration Day 18th December, 2013

Models of Stochastic Dynamic Systems

- *Discrete-Time* (DT) models: at first sight, these seem ideally suited to identification and estimation from sampled time series data that predominates in this digital age. But, as we shall see, they do not suit all applications.

- *Continuous-Time* (CT) models: are theoretically problematic and more difficult to estimate. Statistical interest has focussed on *Stochastic Differential Equation* (SDE) models (e.g. Nielsen et al, 2000) for application in areas such as finance, digital communications and ecology.

- *Hybrid CT* (HCT) models: In many scientific and engineering applications, the interest is much more in the relationship between measured input stimuli and noisy outputs than in the 'noise' on the data. The HCT approach that I discuss here is most appropriate for such applications.

Advantages of Continuous–Time Models in Science and Engineering

- Provide *good physical insight* into the system properties: most *conceptual models are based on the conservation laws formulated in CT*, so that the model parameters have an immediate physical interpretation, with recognizable units
- The parameters are *not a function of any sampling interval* Δt, provided this is selected sensibly (in contrast to a DT model).
- *Ideally suited for 'stiff' dynamic systems* where the *eigenvalues have widely spread modal frequencies* (often encountered in environmental science).
- Can be estimated much better from *rapidly sampled data*.

- Preserve a priori knowledge and lead to *more parsimonious models*.
- Can cope with *non-uniformly sampled data* and *fractional time delays*.
- Includes, in the HCT algorithm, inherent, *data filtering* that both *ensures statistical efficiency* and allows for the *generation of the optimally filtered derivatives that are required for optimal estimation*.
- Can be estimated from relatively *simple impulsive/step inputs* and can *handle initial conditions easily*.
- Can be *converted to a similar performance, discrete-time model at any sampling interval*; and this DT model will always have the same dynamic properties as the CT model.
- Ideally suited for use with the *'hybrid' (continuous-discrete) Kalman Filter* for use in *forecasting and data assimilation* with *irregular sampling*.

Differential Equations and Transfer Functions

$$\frac{d^n x(t)}{dt^n} + a_1 \frac{d^{n-1} x(t)}{dt^{n-1}} + \cdots + a_n x(t) = b_0 \frac{d^m u(t-\tau)}{dt^m} + \cdots + b_m u(t-\tau)$$

Introducing the differential operator $s^r = d^r/dt^r$, this equation can be rearranged to yield the *Transfer Function* form of the model:

$$x(t) = \frac{b_0 s^m + b_1 s^{m-1} + \cdots + b_m}{s^n + a_1 s^{n-1} + \cdots + a_n} u(t-\tau) = \frac{B(s)}{A(s)} u(t-\tau)$$

or, when the system is affected by additive noise $\xi(t)$,

$$y(t) = \frac{B(s)}{A(s)} u(t-\tau) + \xi(t)$$

The Hybrid Box-Jenkins Model

$$x(t) = \frac{B(s)}{A(s)} u(t_k - \tau) \quad : \text{ CT System Model}$$

$$\xi(t_k) = \frac{D(L)}{C(L)} \epsilon(t_k) \quad : \text{ DT Noise Model}$$

$$\epsilon(t_k) = \mathcal{N}(0, \sigma^2) \quad : \text{ White Noise Input}$$

$$y(t_k) = x(t_k) + \xi(t_k) \quad : \text{ Measured Output}$$

where L is the backward-shift or 'lag' operator (sometimes B, z^{-1} or q^{-1}).
Or, informally:

$$y(t_k) = \frac{B(s)}{A(s)} u(t_k - \tau) + \frac{D(L)}{C(L)} \epsilon(t_k)$$

This is clearly nonlinear in the parameters, so iterative *Pseudo-Linear Regression* (PLR) regression is used to estimate the parameters.

PLR and Refined Instrumental Variable Estimation

The **R***efined* **I***nstrumental* **V***ariable* algorithm for **C***ontinuous-time* systems (RIVC) is an *en bloc* and *recursive* method for estimating the parameters of the hybrid BJ model by Maximum Likelihood using iterative PLR optimization. The statistical assumptions about $\epsilon(t_k)$ suggest that a suitable error function is obtained by reference to:

$$\epsilon(t_k) = \frac{C(L)}{D(L)} \left[y(t_k) - \frac{B(s)}{A(s)} u(t_k - \tau) \right]$$

which can be written as,

$$\epsilon(t_k) = \frac{C(L)}{D(L)A(s)} \left[A(s)y(t_k) - B(s)u(t_k - \tau) \right]$$

Minimization of a least squares criterion function in $\epsilon(k)$, measured at the sampling instants, provides the basis for optimal stochastic estimation. However, since the polynomial operators commute in this linear case, the *hybrid prefilter*:

$$f(s,L) \triangleq \frac{C(L)}{A(s)D(L)} \quad \left(\text{note decomposed}: \quad f_c(s) = \frac{1}{A(s)};\ f_d(L) = \frac{C(L)}{D(L)}\right)$$

can be taken inside the square brackets to yield[1]:

$$\epsilon(t_k) = A(s)y_f(t_k) - B(s)u_f(t_k - \tau)$$

where the subscript f denotes prefiltering by the hybrid prefilter $f(s,L)$. This expression is linear in the parameters but it involves prefiltered variables that require knowledge of the parameters in $A(s)$, $C(L)$ and $D(L)$.

[1] This filter can be related directly to the optimal Kalman filter (Young, 1979)

The RIVC Algorithm: PLR Model

The RIVC algorithm (Young & Jakeman, 1980; Young, 2011) is the continuous-time equivalent of the earlier RIV algorithm (Young, 1976) for discrete-time transfer function model estimation. Here, the TF system estimation model at the k^{th} sampling instant is written in the following *Pseudo-Linear Regression* form:

$$y_f^{(n)}(t_k) = \boldsymbol{\phi}^T(t_k)\boldsymbol{\rho} + \epsilon(t_k)$$

where,

$$\boldsymbol{\phi}^T(t_k) = [-y_f^{(n-1)}(t_k) \cdots - y_f^{(0)}(t_k) \; u_f^{(m)}(t_k - \tau) \cdots u_f^{(0)}(t_k - \tau)]$$
$$\boldsymbol{\rho} = [a_1 \; \ldots \; a_n \; b_0 \; \ldots \; b_m]^T$$

Note that all of the variables in this PLR have to be prefiltered, but the hybrid prefilter is straightforward to implement in Matlab

The RIVC Algorithm: Iterative Optimization

- The RIVC estimation algorithm involves ML optimization of parameter vector $\boldsymbol{\theta} = [\boldsymbol{\rho} \ \ \boldsymbol{\eta}]^T$, where $\boldsymbol{\eta}$ is the vector of ARMA model parameters, i.e.,

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \ \mathcal{L}(\boldsymbol{\theta}\,,\,\mathbf{y}\,,\,\mathbf{u})$$

  in which $\mathbf{y}$ and $\mathbf{u}$ are the input and output vectors. As usual, $\mathcal{L}(\boldsymbol{\theta}\,,\,\mathbf{y}\,,\,\mathbf{u})$ is the likelihood based on the sum of squares of $\epsilon(t_k)$,

- The optimization involves iterative solution of the optimal IV normal equations for the system parameter vector $\boldsymbol{\rho}$ based on the PLR relationship. At the jth iteration, with N samples:

$$\hat{\boldsymbol{\rho}}_N^j = \left[(\hat{\boldsymbol{\Phi}}_N^j)^T \boldsymbol{\Phi}_N^j\right]^{-1} (\hat{\boldsymbol{\Phi}}_N^j)^T \mathbf{y}_N^j \quad \text{OR} \quad \boldsymbol{\rho}_N^j = \boldsymbol{\rho}_N^{j-1} + \left[(\hat{\boldsymbol{\Phi}}_N^j)^T \boldsymbol{\Phi}_N^j\right]^{-1} (\hat{\boldsymbol{\Phi}}_N^j)^T \mathbf{e}_N^j$$

  where $\hat{\boldsymbol{\Phi}}(t_k)$ is the optimal IV matrix. In order to implement the hybrid prefiltering, this solution involves concurrent update of the ARMA model parameter estimate $\hat{\boldsymbol{\eta}}_j$, using the IVARMA algorithm (Young, 2011), and exploits the asymptotic independence of $\boldsymbol{\rho}$ and $\boldsymbol{\eta}$ in the ML estimation of Box-Jenkins models (see D. A. Pierce, 1972).

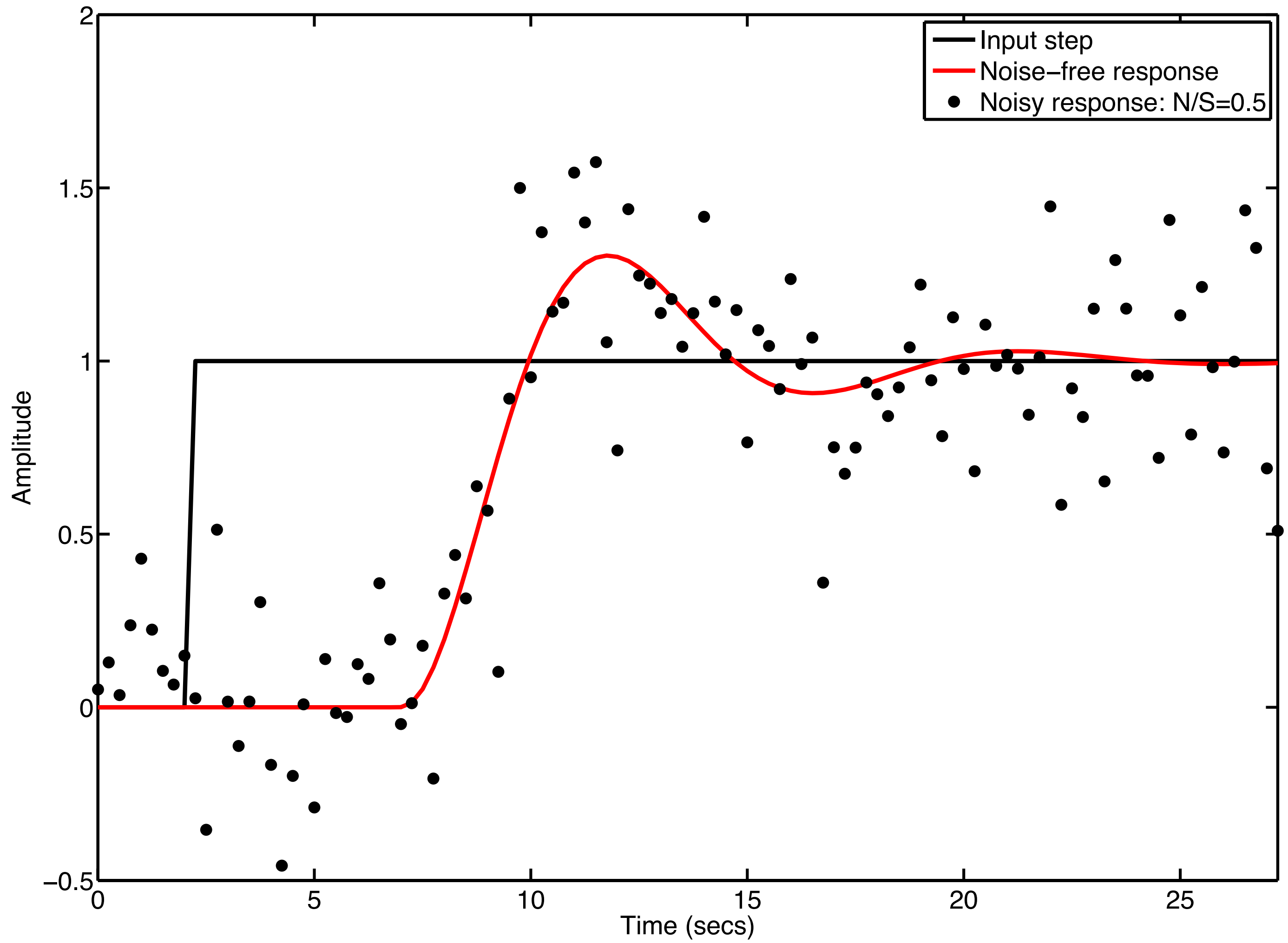
Simulation Example

$$x(t) = \frac{0.5}{s^2 + 0.5s + 0.5} u(t-5)$$
$$y(t_k) = x(t_k) + \epsilon(t_k)$$

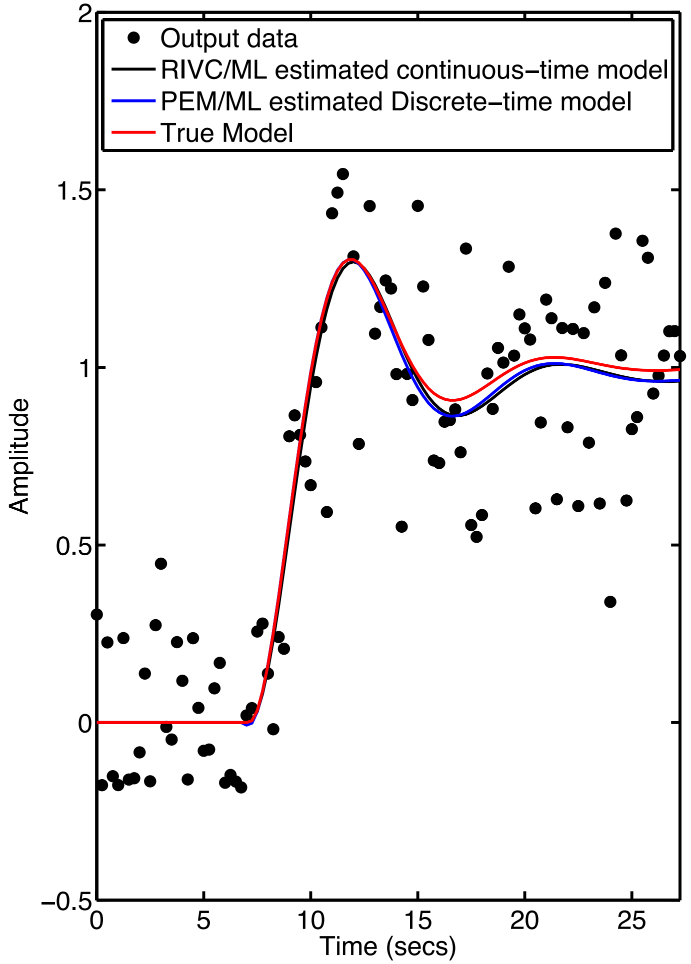
The equivalent discrete-time model is obtained using the 'zero-order hold' assumption for the intersample behaviour of the input (exactly true for step inputs). At a sampling interval of $\Delta t = 0.25$, the model takes the form:

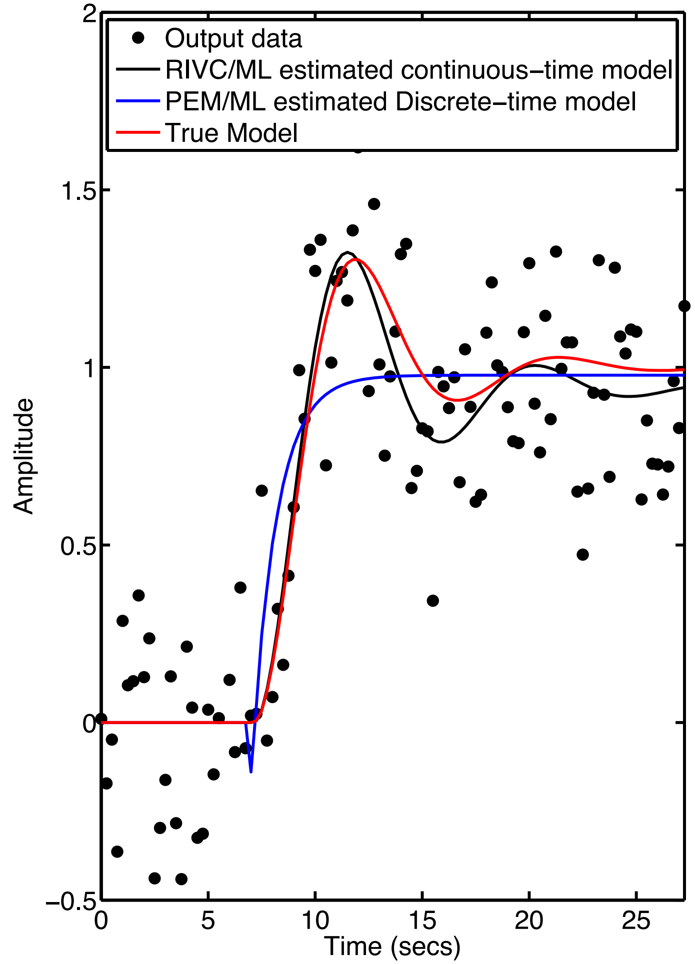
$$y(t_k) = \frac{0.01496 + 0.01434L}{1 - 1.853L + 0.8825L^2} u(t_k - 20) + \epsilon(t_k); \quad \epsilon(t_k) = \mathcal{N}(0, \sigma^2)$$

with σ^2 set to give a N/S ratio of 0.5 by standard deviation. Note how this involves one additional parameter. The step response is shown in the next slide.

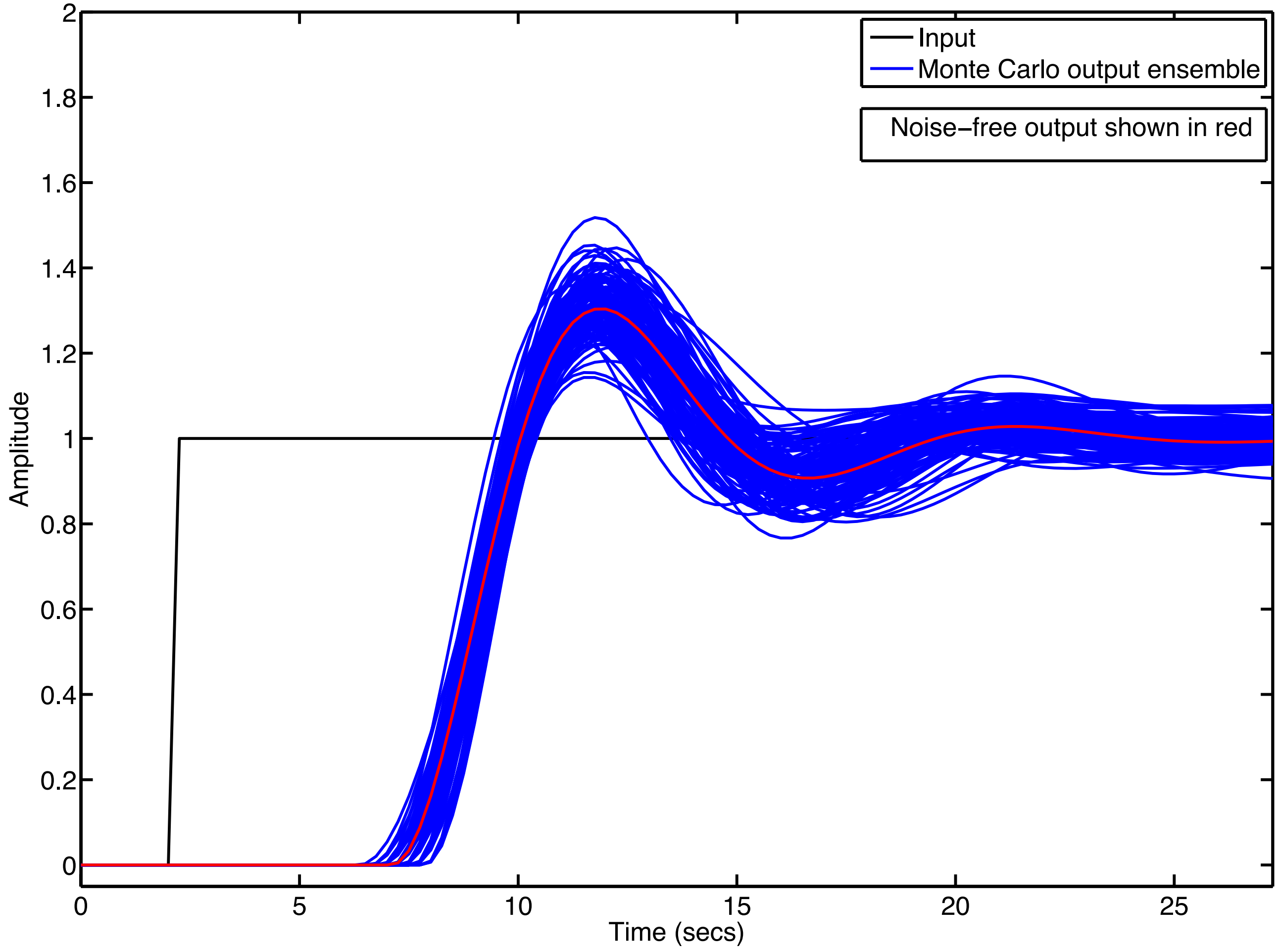


Noise-free and noisy step responses of the example system.

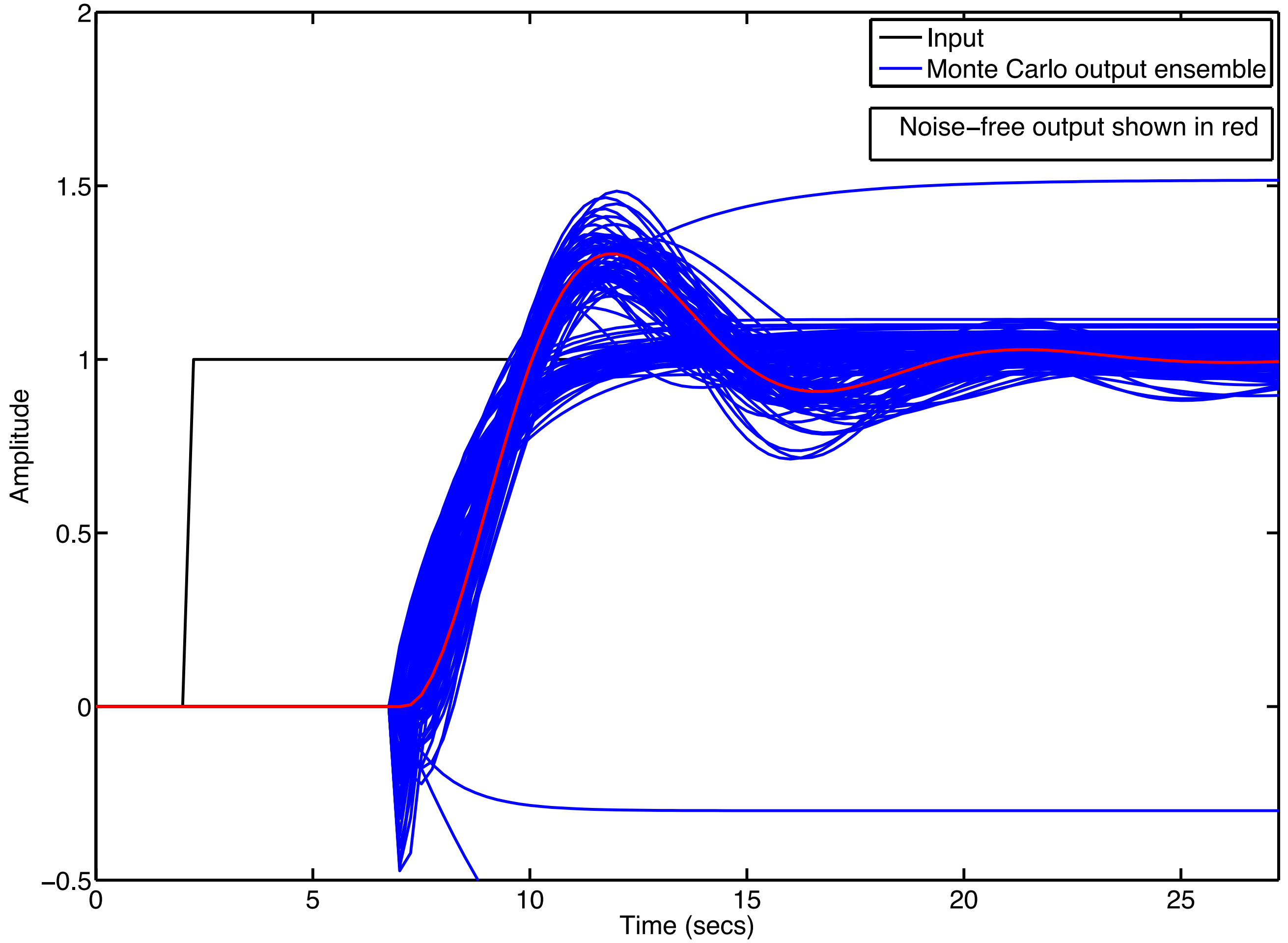




Typical estimated CT and converted DT model time responses.



Monte Carlo ensemble: direct from RIVC estimation.



Monte Carlo ensemble: indirect, converted from DT estimation.

MCS Comparison: Direct and Indirect Estimation of CT System

Parameter	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_0$	% Failure
True Values	0.5	0.5	0.5	
Direct RIVC Estimates	0.503	0.502	0.502	No Failures
MCS standard deviation	0.039	0.025	0.027	
Single Run estimated standard error	0.034	0.021	0.023	
Indirect from DT Estimation	0.502	0.501	0.501	12% Failures
MCS standard deviation	0.033	0.029	0.029	

MCS results based on 100 realizations, using using a square wave input signal with 1020 input-output samples, sampling interval $\Delta t = 0.25$ secs.; and additive white noise. Failures are caused by an estimated DT model with a negative real root, probably due to the short sample size and high noise level.

MCS Results for RIVC: Highly Coloured Additive Noise

Parameter	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_0$		
True Values	0.5	0.5	0.5	-0.98	0.20
Direct RIVC Estimates	0.500	0.500	0.505	-0.979	0.196
MCS standard deviation	0.043	0.020	0.029	0.004	0.020
Single run estimated standard error	0.039	0.018	0.029	0.004	0.020
Indirect from DT Estimation	0.493	0.497	0.501	-0.978	0.195
MCS standard deviation	0.045	0.024	0.032	0.005	0.022

MCS results based on 100 realizations, using a square wave input signal with 2420 input-output samples, sampling interval $\Delta t = 0.25$ secs.; but highly coloured additive noise generated by an ARMA(1,1) process, introducing significant autocorrelation up to lag 60.

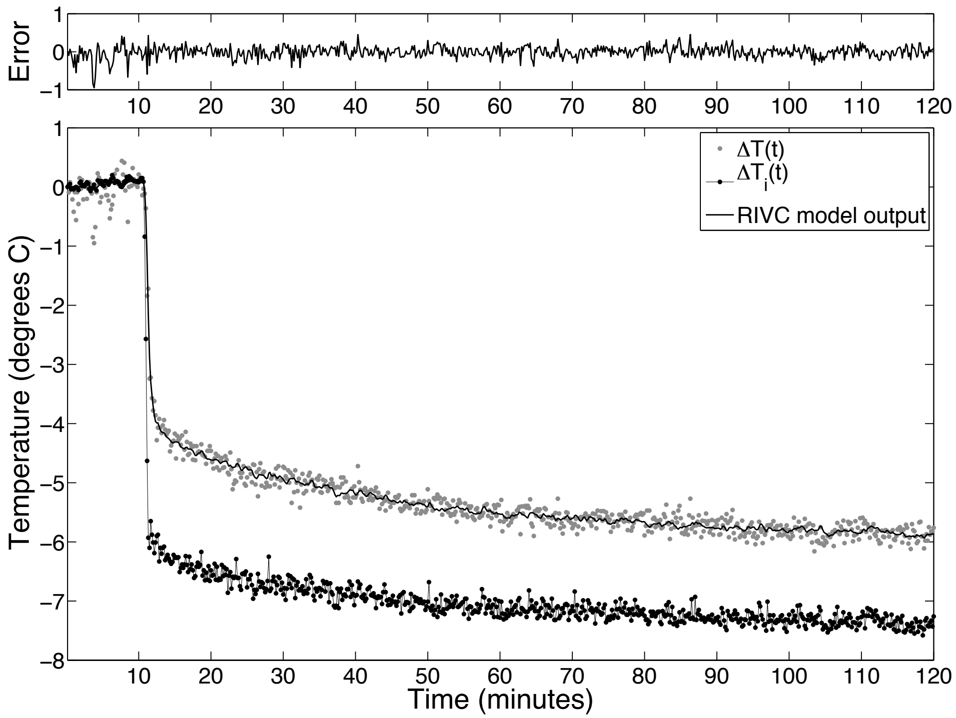
Direct and Indirect Estimation from Rapidly Sampled Data

Parameter		$\hat{a}_1$	$\hat{a}_2$	$\hat{b}_0$	% Failure
True Values		0.5	0.5	0.5	
Direct RIVC Estimates	$\hat{\boldsymbol{\rho}}$	0.500	0.501	0.501	No Failures
MCS standard deviation	SD	0.025	0.017	0.018	
Single run estimated standard error	SE	0.026	0.017	0.017	
Indirect from DT Estimation	$\hat{\boldsymbol{\rho}}$	0.497	0.491	0.491	54% Failures
MCS standard deviation	SD	0.035	0.046	0.046	

MCS results based on 100 realizations, using a square wave input signal with 4000 input-output samples, sampling interval $\Delta t = 0.025$ secs. (10 times faster sampling rate). Failures are caused by either: (i) roots of DT model too close to the unit circle; or (ii) an estimated DT model with a negative real root.

Data-Based Mechanistic Modelling Using RIVC

Practical Example I: Ventilated Chamber Dynamics



Data from experiment in a force-ventilated chamber.

The RIVC algorithm identifies a second order CT system with the following estimated TF model:

$$\Delta T(t) = \frac{1.9519s + 0.08828}{s^2 + 2.9487s + 0.11045}\Delta T_i(t) = \frac{1.9519s + 0.08828}{(s + 2.911)(s + 0.0379)}\Delta T_i(t)$$

- This can be decomposed by partial fraction expansion into a negative feedback connection of two first order processes with steady-state gains of $\{0.669,\ 0.130\}$ and widely spaced time constants of $\{0.343,\ 26.35\}$ minutes. The reason for the presence of the longer time constant in this model is that the installation is not effectively insulated by the outer chamber which, instead, acts as a temperature buffer zone with its own heat transfer characteristics.

- Using the inductive Data-Based Mechanistic (DBM) modelling approach, it is straightforward to relate the model to the differential equation model obtained by invoking the classical theory of heat transfer.

Practical Example II: Rainfall-Flow Modelling

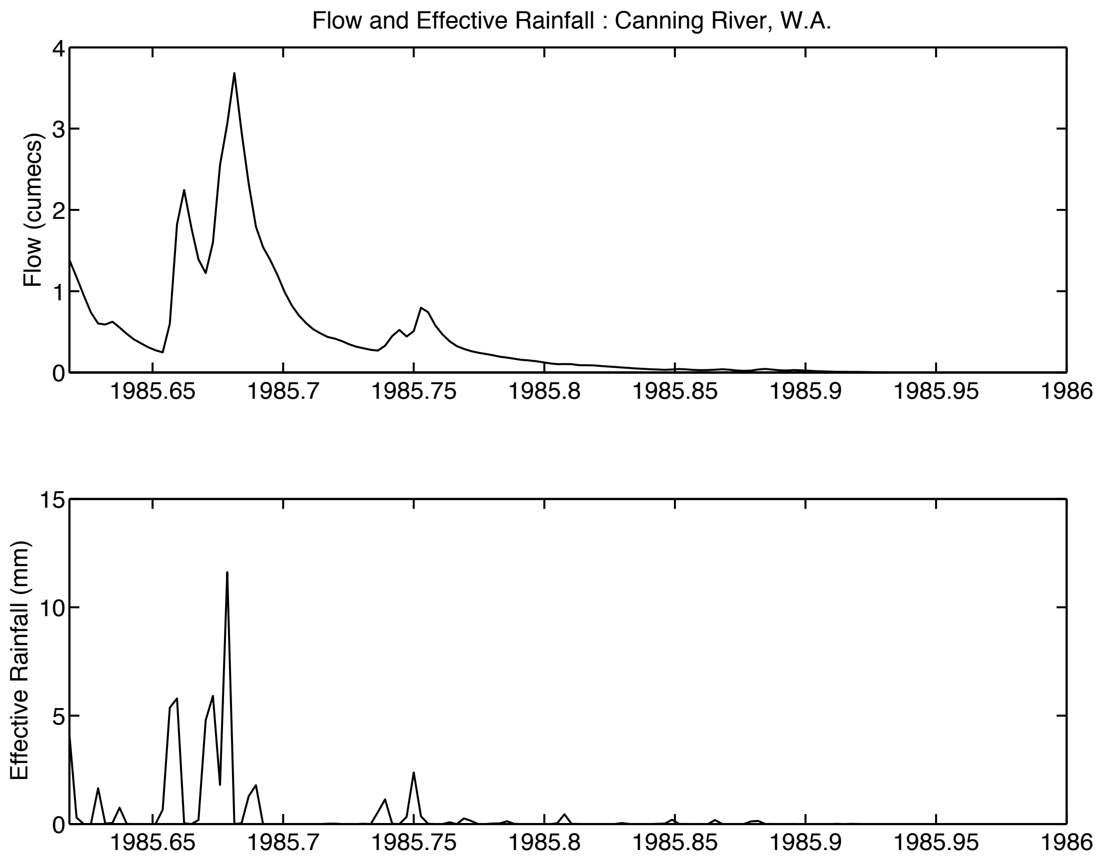


Figure 6: Rainfall and flow data for the Canning River, W.A.

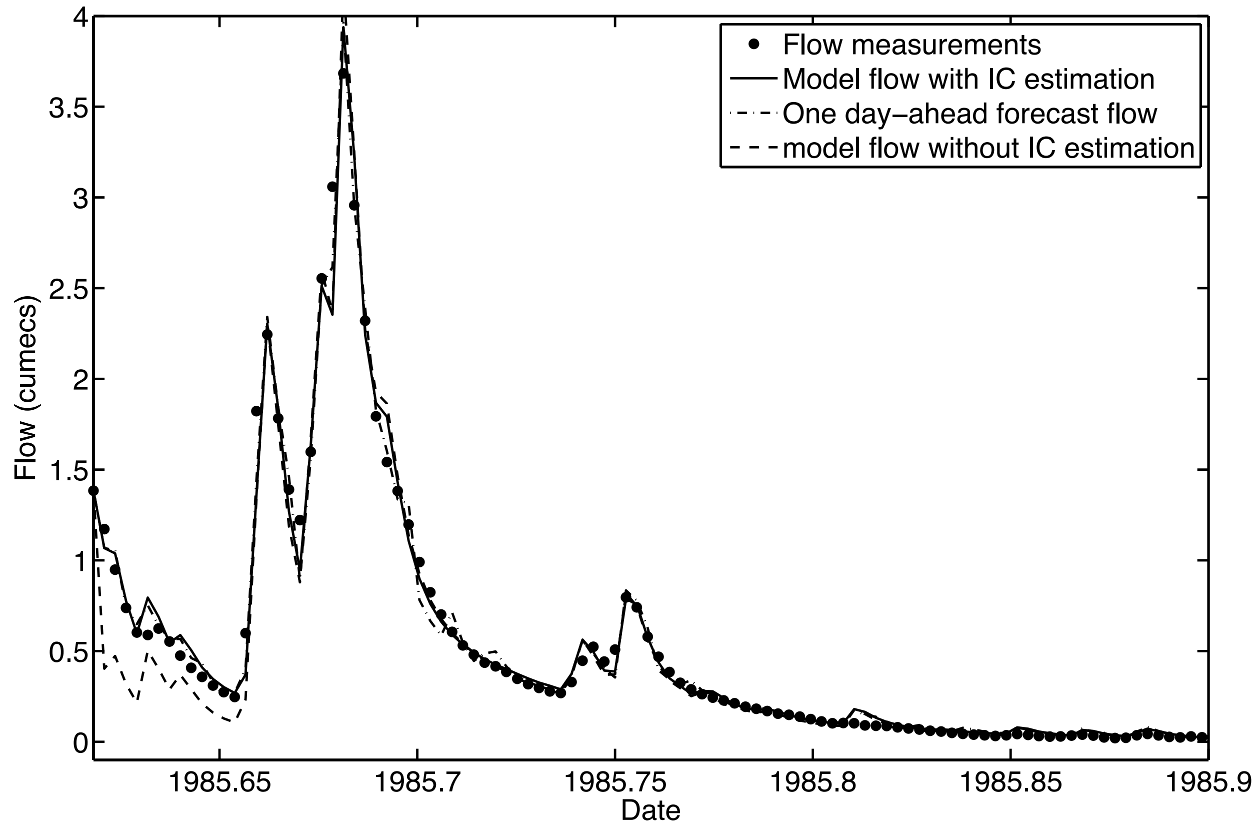
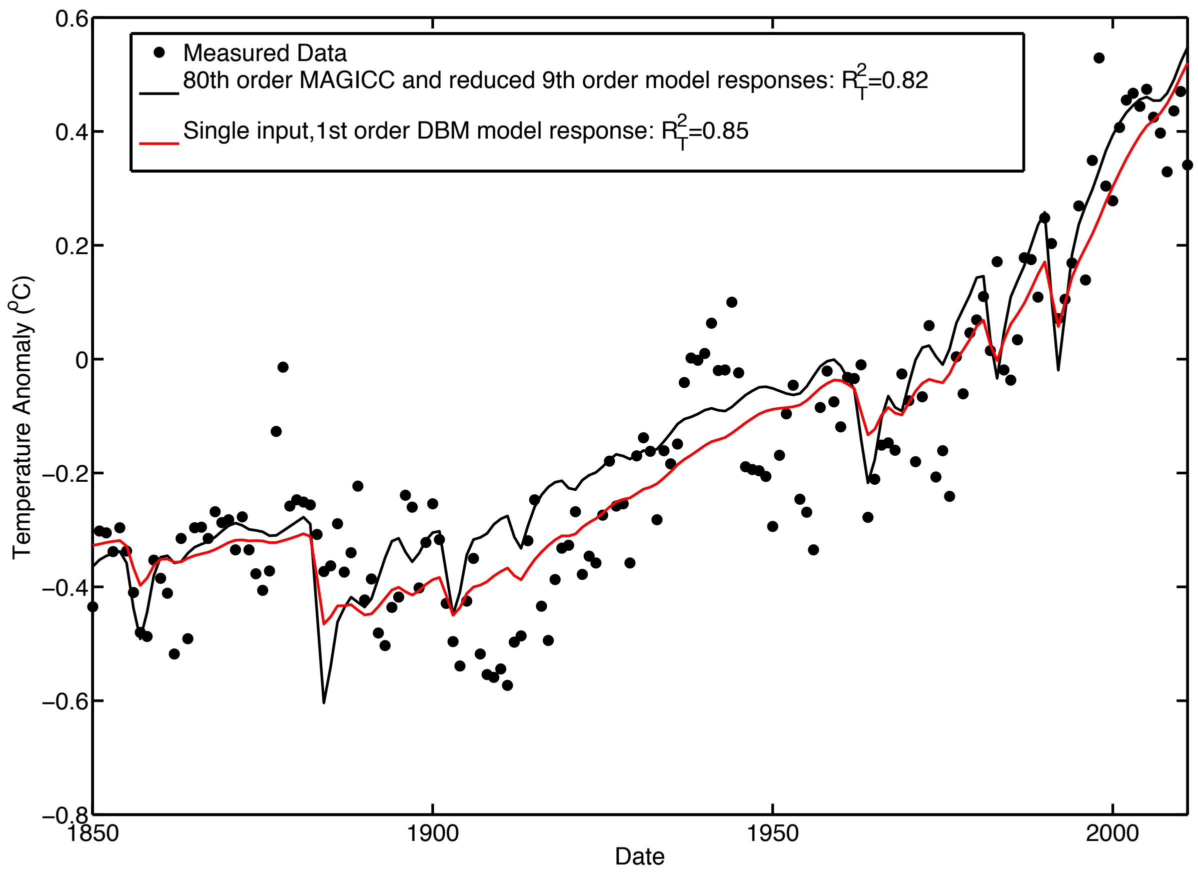


Figure 7: Comparison of SRIVC modelling results for the Canning River, with and without initial condition estimation.

Practical Example III: Global Climate Modelling



MAGICC, reduced order MAGICC and DBM model comparison.

Conclusions

1. The hybrid continuous-time BJ model, coupled with the optimal RIVC method for its identification and estimation, provides an additional approach to time series modelling that is not well known in the statistical/time series community.

2. When considering the stochastic modelling and forecasting of physical systems, there are advantages in utilizing continuous time (CT) models because they are more easily interpretable in physical terms than discrete-time, black box models.

3. Other advantages of CT modelling include the uniquely defined parameter values that are independent of the sampling interval, as well as the ability to handle: non-uniformly and rapidly sampled data; 'stiff' dynamic systems with wide ranging eigenvalues; and data with significant initial conditions.

4. The DBM approach using CT models can add a degree of mechanistic credibility that acts as an additional validation of the resultant models.