PUB and Data-Based Mechanistic Modelling:
the Importance of Parsimonious Continuous-Time Models

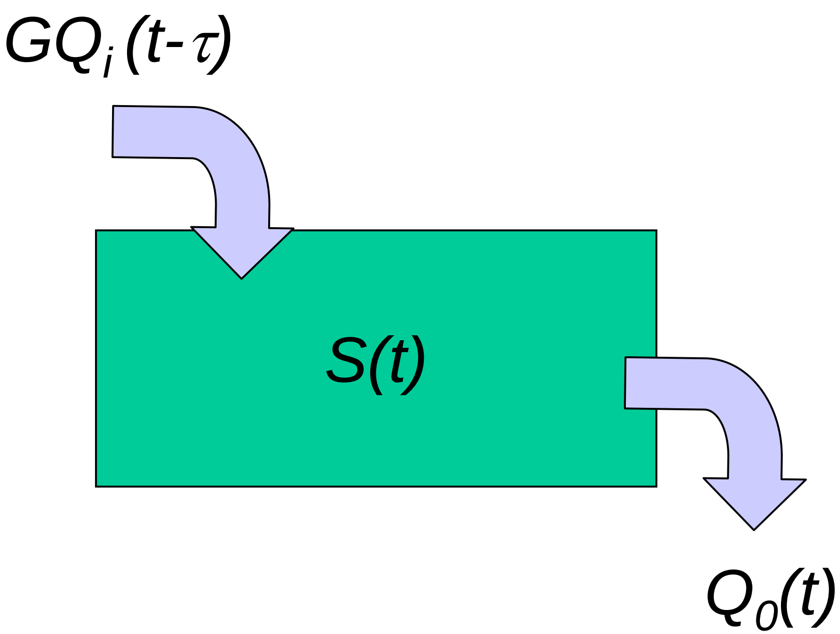
Peter C. Young and Renata J. Romanowicz
Centre for Research on Environmental Systems and Statistics,
Lancaster University, Lancaster, LA1 3EZ, UK

CRESearch
The Key to Success in Environmental and Biological Systems Analysis

LANCASTER UNIVERSITY

Continuous-Time (CT) Transfer Function Models

A first order example: *linear catchment storage*:



$$\frac{dS(t)}{dt} = GQ_i(t-\delta) - Q_o(t)$$

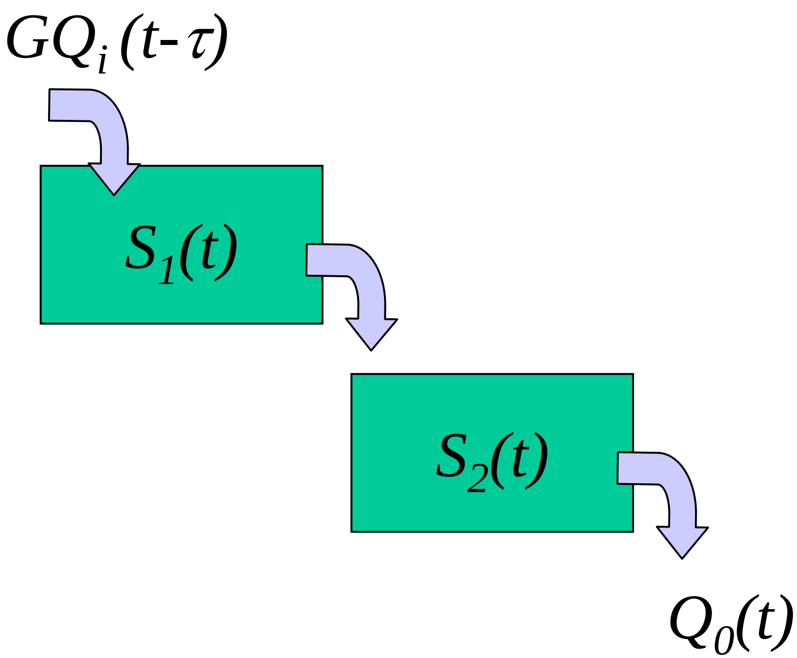
Assumption: $Q_o(t) = \rho S(t)$

$$Q_o(t) = \frac{b_0}{s+a_1} Q_i(t-\tau)$$

$$b_0 = \frac{G}{T} \;\; ; \;\; a_1 = \frac{1}{T}$$

Multi-Order CT Transfer Function Models

Example: Multi-Reach Lag-and-Route/ Nash Cascade (NC) Model



$$x(t) = \frac{B(s)}{A(s)} u(t-\tau)$$ (note: τ=0 for NC)

with additive stochastic noise:

$$y(t) = x(t) + \xi(t)$$

where,

$$A(s)=s^n+a_1\, s^1+\ldots+ a_n$$

$$B(s)=s^m+b_1\, s^1+\ldots+ b_m$$

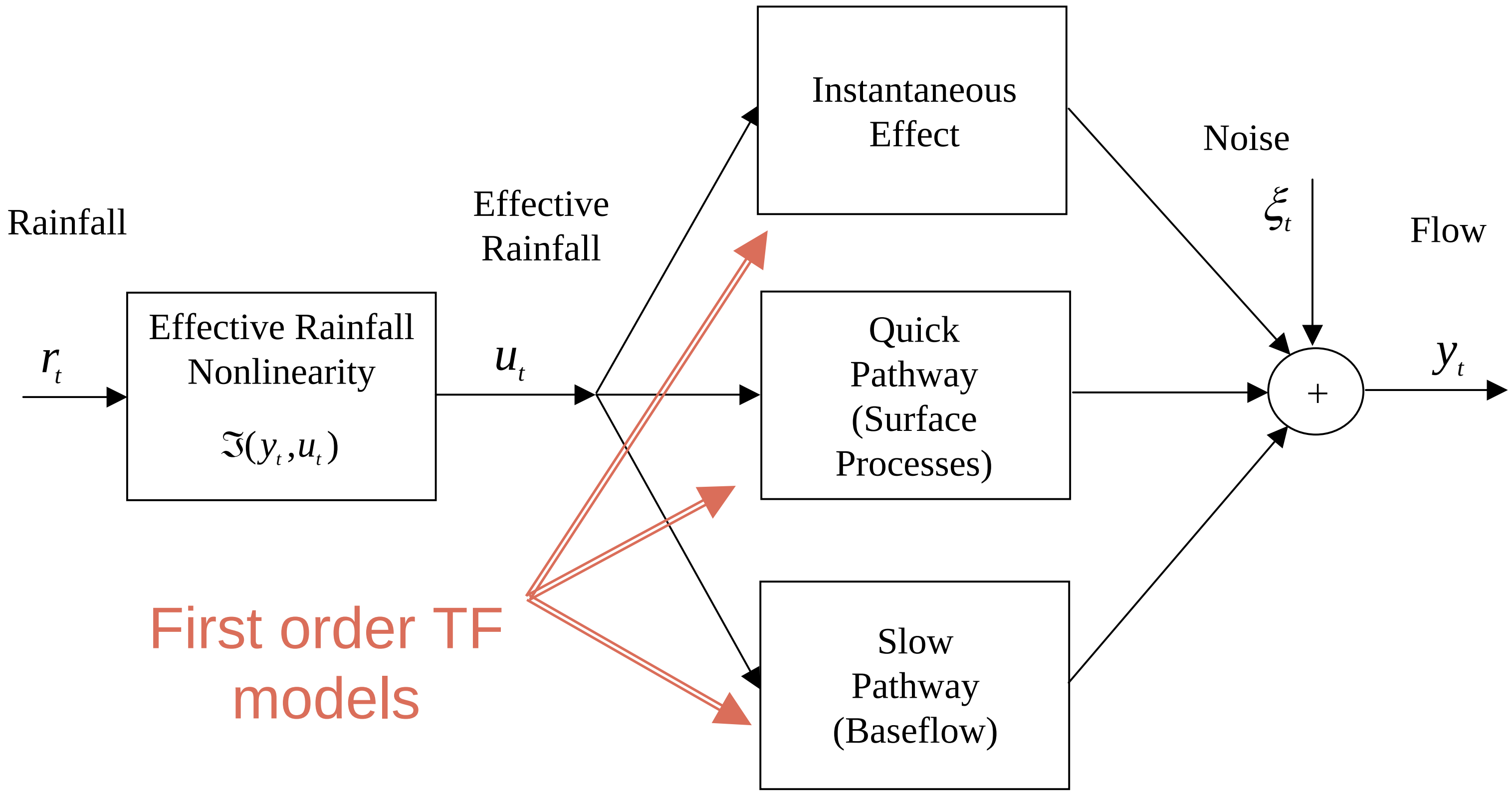
And s=d/dt is the derivative operator

Differential Equation Form of the CT Transfer Function Model

$$\frac{d^n y(t)}{dt^n} + a_1 \frac{d^{n-1} y(t)}{dt^{n-1}} + \ldots \; a_n y(t)$$
$$= \frac{d^m u(t-\tau)}{dt^m} + \ldots \; + a_m u(t-\tau) + \eta(t)$$

where $\eta(t) = A(s)\xi(t)$. The structure of this model, in differential equation or TF form, is defined by the triad $[n\, m\, \tau]$.

Decomposition of the TF Model: [2 3 0] Rainfall-Flow Example



Discrete Time (DT) Models

$$Q_{o,k} = \frac{\beta_0}{1+\alpha_1 z^{-1}} Q_{i,k-\delta}$$

where $Q_{0,k}$ is the flow measured at k sampling instant.

Relationship to continuous model parameters under zero-order hold (ZOH) assumption:

$$\alpha_1 = -\exp(-a_1 \Delta t)$$

$$\beta_1 = \frac{b_0}{a_1}\{1-\exp(-a_1 \Delta t)\}$$

Note: the parameters are a function of the sampling interval so, unlike CT models, the parameters change with the sampling interval.

Multi-order DT equivalent of the multi-order CT model:

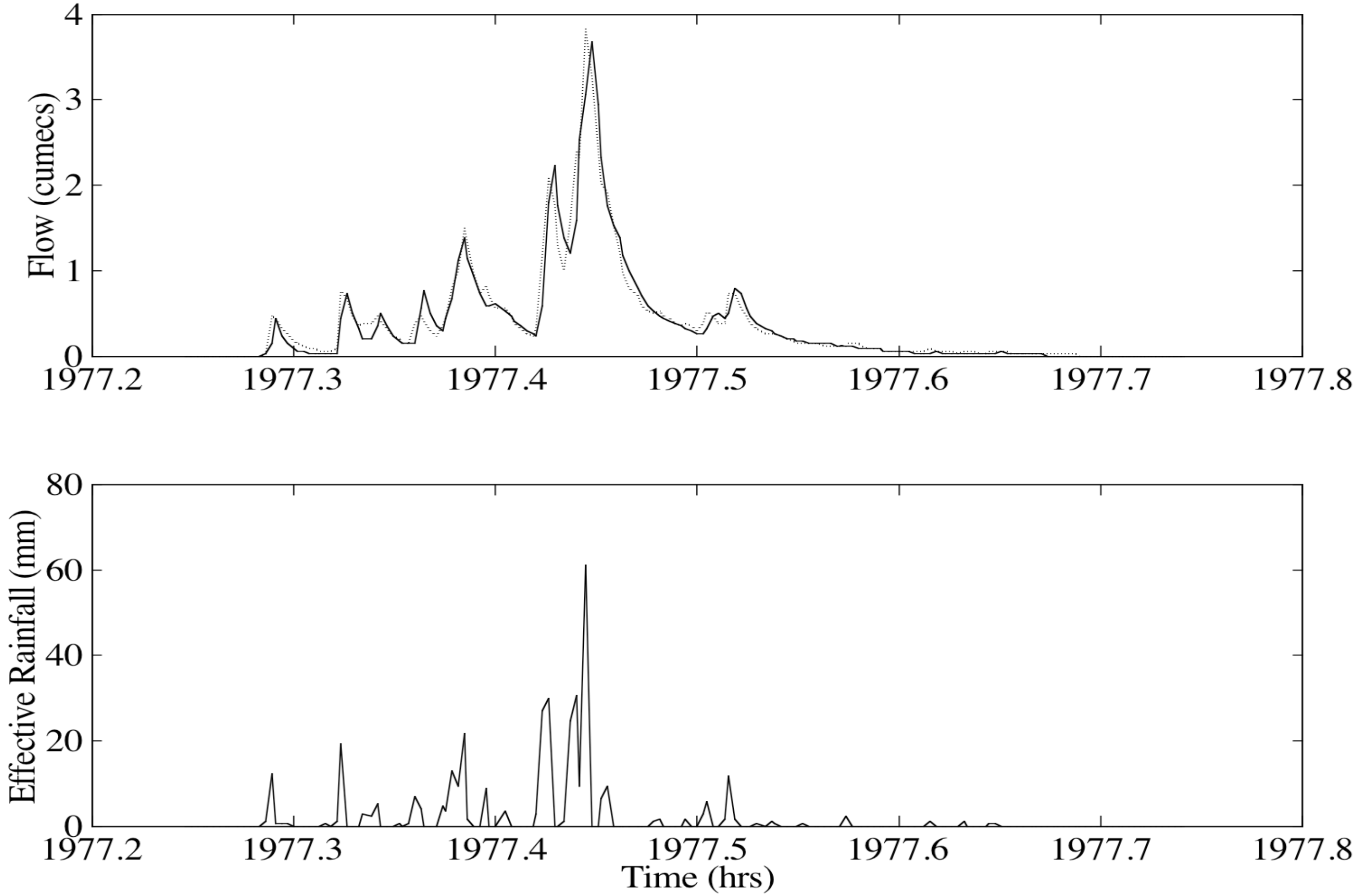
$$x_k = \frac{B(z^{-1})}{A(z^{-1})} u_{k-\delta}$$

$$y_k = x_k + \xi_k$$

$$A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_n z^{-p}$$

$$B(z^{-1}) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_m z^{-q}$$

Case Study: Rainfall-Flow Modelling



Daily effective rainfall and flow data for the ephemeral River Canning in Western Australia

Direct CT Identification and Estimation

(RIVC algorithm, CAPTAIN Matlab Toolbox)

Application of RIVC algorithm: [2 3 0] model

$$\hat{a}_1 = 0.457\,(0.032) \quad \hat{a}_2 = 0.0248\,(0.0045)$$

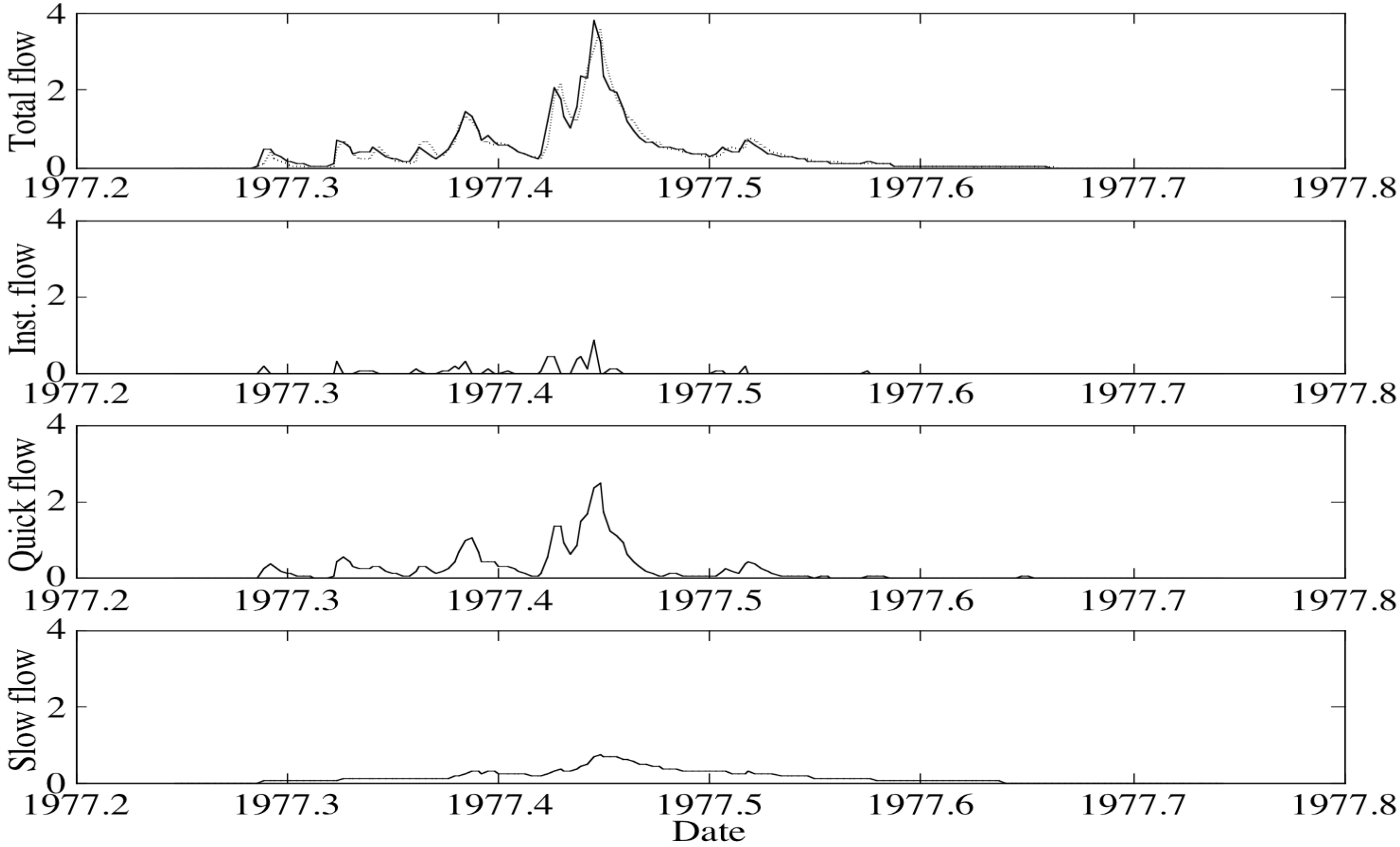
$$\hat{b}_0 = 0.0138\,(0.001) \quad \hat{b}_1 = 0.0505\,(0.002)$$

$$\hat{b}_2 = 0.0046\,(0.0008)$$

Model decomposes into parallel pathway form:

- instantaneous effect : 7.4%
- quick pathway : 54.1%
- slow pathway : 38.5%

Parallel Pathway Flows



RIVC model response compared with the gauged flow (top panel) and estimated instantaneous (within one day), quick and slow pathway flows below.

Indirect CT Identification and Estimation

(RIV algorithm, CAPTAIN Matlab Toolbox)

DT identification: model [2 3 0]

$$\hat{\alpha}_1 = -1.6034\,(0.008) \quad \hat{\alpha}_2 = 0.6244\,(0.007)$$

$$\hat{\beta}_0 = 0.0140\,(0.001) \quad \hat{\beta}_1 = 0.0151\,(0.002)$$

$$\hat{\beta}_2 = 0.0252\,(0.0013)$$

Discrete to continuous time conversion under ZOH assumption:

$$\hat{a}_1 = 0.4711 \quad \hat{a}_2 = 0.0264$$

$$\hat{b}_0 = 0.0140 \quad \hat{b}_1 = 0.0514 \quad \hat{b}_2 = 0.0049$$

Evaluation by Monte Carlo Simulation Analysis

This analysis was applied to a [2 3 0] CT simulation model, with data sampled at various rates from 5 mins. to 24 hours. The analysis reveals that:

- the direct RIVC algorithm rarely fails to converge: it has no failures for sampling intervals from 5 minutes to one hour; and a maximum of 2 failures in 50 at a sampling interval of 18 hours (a mean failure rate of only 0.3%);

- on the other hand, the indirect RIV (CAPTAIN Matlab Toolbox) and PEM-based (Matlab Identification Toolbox) indirect approaches perform poorly, particularly at high sampling rates, with mean failure rates of 7.4% and 17.4% respectively.

- as expected, the lowest number of failures for the RIV- based method occurs for longer sampling intervals of   hour, where the failure rate is only 3%. This is much superior to the PEM performance.

Conclusions

Why Use CT Models?

- The model parameters values are unique: unlike discrete-time models, they are not a function of the data sampling interval.

- Model parameter estimation is superior over a wider range of sampling intervals, particular at fast sampling intervals (e.g. 5-15 min. for rainfall-flow measurements).

- The model is in an ordinary differential equation form that relates directly to the formulation of physically meaningful models, such as those derived from mass, energy and momentum conservation - as required for *Data-Based Mechanistic* (DBM) modelling.

- These advantages should be particularly important in regionalization studies, where the need for well defined statistical estimates of the parameters is particularly important in defining any relationships between the parameter values and the catchment’s physical characteristics.