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

Peter C. Young

Renata J. Romanowicz

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 4YQ, U.K.
p.young@lancaster.ac.uk

Abstract: The problem of *Prediction in Ungauged Basins* (PUB) is intimately linked with the concept of regionalisation; namely the transfer of information from one catchment that is gauged to another that is not. But such regionalisation exercises can be dangerous and should be attempted only with great care. The present paper addresses what the authors believe to be one essential aspect of regionalisation: namely, the importance of considering only 'top-down' models that are parametrically efficient (parsimonious) and fully 'identifiable' from the available catchment data. We argue further that many mechanistic model parameters are more naturally defined in the context of continuous-time, differential equation models (normally derived by the application of natural 'laws', such as mass and energy conservation). As a result, there are advantages if such models are identified and estimated directly in this continuous-time, differential equation form, rather than being formulated and estimated as discrete-time models. The arguments presented in the paper are illustrated by an example in which the top-down, *Data-Based Mechanistic* (DBM) approach to modelling is applied to a set of precipitation-flow data. This involves the application of an advanced method of continuous-time transfer function identification and estimation; and the interpretation of this estimated model in physically meaningful terms, as required by the DBM modelling approach.

Keywords:Continuous-time transfer functions; data-based mechanistic; rainfall-flow; regionalisation.

1. INTRODUCTION

From a conceptual standpoint, most mathematical models of hydrological systems are formulated on the basis of natural laws, such as dynamic conservation equations, often expressed in terms of continuous-time (CT), linear or nonlinear differential equations. Paradoxically, *Transfer Function* (TF) models, which have been growing in popularity over the last few years because of their ability to characterise hydrological data in efficient parametric (or 'parsimonious') terms, are almost always presented in the alternative, discrete-time (DT) terms. One reason for this paradox is that hydrological data are normally sampled at regular intervals over time, so forming discrete time series that are in a most appropriate form for DT modelling. Another is that most of the technical literature on the statistical identification and estimation (calibration) of TF models deals with these DT models. Closer review of this literature, however, reveals apparently less well known publications [e.g. Young and Jakeman, 1980] dealing with estimation methods that allow for the *direct* identification of CT (differential equation) models from discrete-time, sampled data.

This paper first discusses the formulation, identification and estimation of CT models. It then considers a practical example in which the *Data-Based Mechanistic* (DBM) approach to modelling [e.g. Young, 2001 and the prior references therein] is applied to a typical set of effective rainfall-flow data from the ephemeral Canning River of Western Australia.

2. CONTINUOUS-TIME TF MODELS

Let us consider first a conceptual catchment storage equation in the form of a continuous-time, linear storage (tank or reservoir) model: see, for example, the review papers by O'Donnell, Dooge and Young in Kraijenhoff and Moll [1986]; or the recent book by Beven [2001]. Here, the rate of change of storage $S(t)$ in the channel is defined in terms of water volume $Q_i(t)$ entering the reservoir (e.g. river reach) in unit time, minus the volume $Q_o(t)$ leaving in the same time interval, i.e.,

$$\frac{dS(t)}{dt} = GQ_i(t-\delta) - Q_o(t) \qquad (1)$$

where $Q_i(t-\tau)$ represents the input flow rate delayed by a pure time delay of τ time units to allow for pure advection; and G is a 'gain' parameter inserted to represent gain (or loss) in the system. Making the reasonable and fairly common assumption that the outflow is proportional to the storage at any time, i.e.,

$Q_o(t) = \rho S(t)$,

and substituting into (1), we obtain,

$$T\frac{dQ_o(t)}{dt} = GQ_i(t-\tau) - Q_o(t); \quad T = \frac{1}{\rho} \qquad (2)$$

This can be written in the following, continuous-time *Transfer Function* (TF) form,

$$Q_o(t) = \frac{G}{1+Ts}Q_i(t-\tau) \qquad (3a)$$

where s is the derivative operator, i.e. $s = d/dt$. Five important, physically interpretable model parameters are associated with this model. The *Steady State Gain* (SSG), denoted by G, is obtained by setting the s operator in the TF to zero (i.e. $d/dt = 0$ in a steady state). It shows the relationship between the equilibrium output and input values when a steady input is applied. For this reason, if the input and output have similar units, G is ideal for indicating the physical losses or gains occurring in the system. In the case of a flow-routing model, it indicates whether water has been added or lost between the upstream and downstream boundaries and the percentage of water lost or gained can be defined by *Loss Efficiency* $TE = 100(1-G)$, which will be negative if $G > 1$. The *Residence Time* or *Time Constant*, T , is the time required for the reservoir output to decay to e^{-1} (~ 37%) of its maximum value in response to a unit impulse input. Finally the pure *Advective Time Delay* τ indicates the time it takes for a flow increase upstream to be first detected downstream: and $T_t = T + \tau$ defines the *Travel Time* of the system. These five parameters typify the equilibrium and dynamic characteristics of the TF model and provide a physical interpretation of the TF model in terms of its mass transfer and dispersive characteristics.

The first order TF model (3a) is often written in the form,

$$Q_o(t) = \frac{b_0}{s+a_1}Q_i(t-\tau)$$
$$b_0 = \frac{G}{T} \;\; ; \; a_1 = \frac{1}{T} \qquad (3b)$$

because this is the form in which the model is normally estimated. Typically, a *Channel* or *Flow Routing* model for a river catchment will contain a number of elemental models, such as (3a,b), connected in a manner that relates to the structure of the catchment. For instance, a serial connection of n such elements constitutes the 'lag-and-route' model of a single river channel [Meijer, 1941; Dooge, 1986] and, with $\tau = 0$, it becomes the well known 'Nash cascade' model [Nash, 1959]. More complex river systems can be represented by a main channel of this type, with tributaries modelled in a similar manner. Also, a typical TF model between effective rainfall and flow contains a parallel connection of two or more such storage elements [see Young, 1992, 2001]. A practical example of this kind is described later in Section 4. Related serial and parallel arrangements characterise the mass conservation equation of the *Aggregated Dead Zone* (ADZ) model for the transport and dispersion of a solute in a river channel (e.g. Wallis *et al*, 1989 and the prior references therein). And ADZ-type models can lead to more complex interactive water quality models, including chemical and biological interaction, such as DO-BOD models [e.g. Beck and Young, 1975].

Serial, parallel or even feedback connections of elemental first order TF models such as (3a,b) lead, in general terms, to the following multi-order TF model:

$$x(t) = \frac{B(s)}{A(s)}u(t-\tau) \qquad y(t) = x(t) + \xi(t) \qquad (4a)$$

where $A(s)$ and $B(s)$ are polynomials in s of the following form:

$$A(s) = s^n + a_1 s^{n-1} + a_2 s^{n-2} + \;\dots\; + a_n$$
$$B(s) = b_0 s^m + b_1 s^{m-1} + b_2 s^{m-2} + \;\dots\; + b_m$$

in which n and m can take on any positive integer values. Here $u(t)$ and $x(t)$ denote, respectively, the deterministic input and output signals of the system at its upstream and downstream boundaries, respectively; τ is a pure advective time (transport) delay affecting the input signal $u(t)$; and $y(t)$ is the observed output, which is assumed to be contaminated by a noise or stochastic disturbance signal $\xi(t)$. This noise is assumed to be independent of the input signal $u(t)$ and it represents the aggregate effect, at the down stream boundary, of all the stochastic inputs to the system, including distributed unmeasured inputs, measurement errors and modelling error. If $\xi(t)$ has rational spectral density, then it can be modelled as an *Auto-Regressive* (AR) or *Auto-Regressive, Moving-Average* (ARMA) process but this restriction is not essential. Also, depending on the objectives of the modelling study, it may be necessary, in a complete system consisting of many sub-elements such as (3a,b), to consider noise inputs within the system, associated with collections of sub-elements that have distinct physical meaning: e.g. stochastic lateral inflows.

Multiplying throughout equation (4a) by $A(s)$ and converting the resultant equation to alternative ordinary differential equation form, we obtain:

$$\frac{d^n y(t)}{dt^n} + a_1 \frac{d^{n-1} y(t)}{dt^{n-1}} + \cdots a_n y(t) = \frac{d^m u(t-\tau)}{dt^m} + \cdots + a_m u(t-\tau) + \eta(t) \quad (4b)$$

where $\eta(t) = A(s)\xi(t)$. The structure of this model, in either form (4a) or (4b), is defined by the triad $[n\ m\ \tau]$.

To date, the most popular form of TF modelling has been carried using the discrete-time (DT) equivalents of the models (3a,b) and (4a,b). In the case of equation (3a,b), this discrete-time TF model takes the form:

$$Q_{o,k} = \frac{\beta_0}{1+\alpha_1 z^{-1}} Q_{i,k-\delta} \quad (5a)$$

Here, $Q_{o,k}$ is the flow measured at the k^{th} sampling instant, that is at time $t = k\Delta t$, where Δt is the sampling interval in time units. $Q_{i,k-\delta}$ is the input flow at time $t = (k-\delta)\Delta t$ time units previously, where δ is the advective time delay, normally defined as the nearest integer value of $\tau / \Delta t$ (thus incurring a possible approximation error); and z^{-1} is the backward shift operator, i.e. $z^{-r} Q_{o,k} = Q_{o,k-r}$. The values of the parameters α_1 and β_o can be related to the parameters of the model (4a) in various ways depending upon how the input flow $Q_i(t)$ is assumed to change over the sampling interval (since it is not measured over this interval). The simplest and most common assumption is that this remains constant over this interval (the so-called 'zero-order hold', ZOH, assumption), in which case the relationships are as follows:

$$\alpha_1 = -\exp(-a_1 \Delta t)$$
$$\beta_1 = \frac{b_0}{a_1}\{1 - \exp(-a_1 \Delta t)\} \quad (5b)$$

Note that, because these relationships are functions of the sampling interval Δt, for every unique CT model such as (3a,b), there are infinitely many DT equivalents (5a,b), depending on the choice of Δt, all with different parameter values. Following from the definition of this first order DT model at the chosen Δt, the general multi-order equivalent of the model (5a,b) is defined as follows:

$$x_k = \frac{B(z^{-1})}{A(z^{-1})} u_{k-\delta} \qquad y_k = x_k + \xi_k \quad (6)$$

where $u_{k-\delta}$, x_k, y_k and ξ_k are the sampled values of $u(t-\tau)$, $x(t)$, $y(t)$ and $\xi(t)$ in the model (5a,b); and

$$A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ \ldots\ + a_n z^{-p}$$
$$B(z^{-1}) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \ \ldots\ + b_m z^{-q}$$

Normally $p = n$ but q may be equal or greater than m.

3. CONTINUOUS-TIME TF MODELS: IDENTIFICATION AND ESTIMATION

The statistical estimation of the CT model (4a,b) is straightforward if completely continuous-time data are available. However, the hydrologist is normally confronted by discrete-time, sampled data and the problem of modelling continuous-time models such as this, *based on discrete-time, sampled data at sampling interval* Δt, is not so obvious. This problem can be approached in two main ways.

1. The *Direct Approach*: here, it is necessary to *identify* the most appropriate, identifiable CT model structure (4a,b) defined by the triad $[n\ m\ \tau]$; and then *estimate* the TF parameters $a_i, i = 1, 2, \ldots n,\quad b_j, j = 0, 1, \ldots m,$ and τ that characterise this structure. Of course, some approximation will be incurred in this estimation procedure because the inter-sample behaviour of input signal $u(t)$ is not known and it must be interpolated over this interval in some manner (see above).

2. The *Indirect Approach*: here, the identification and estimation steps are first applied to the DT model (6). This estimated model is then converted to the CT model (4a,b), again using some assumption about the nature of the input signal $u(t)$ over the sampling interval Δt. In the first order, ZOH case, this conversion is given by the relationships in equation (5b) but, in more general terms, the multi-order conversion must be carried out in a computer, using a conversion routine such as the D2CM algorithm in the Matlab Control Toolbox.

Various statistical methods of identification and estimation have been proposed to implement the two approaches outlined above and these have been formulated in both the time and frequency domains. However, only estimation in the time domain will be considered here, and only one direct estimation method will be utilised: the *Refined Instrumental Variable* method for *Continuous*-time Systems (RIVC) proposed by Young and Jakeman [1980: see also Young, 1984]. The RIVC algorithm is available in both the CAPTAIN (see http://www.es.lancs.ac.uk/cres/captain/) and CONTSID2 (see http://www.cran.uhp-nancy.fr/) Toolboxes for Matlab. The RIVC method is the only time domain method that can be interpreted in optimal statistical terms, so providing an estimate of the parametric error covariance matrix and, therefore, estimates of the confidence bounds on the parameter estimates.

RIVC is a close relative of the *Refined Instrumental Variable* (RIV) algorithm for the identification and estimation of discrete-time TF models [Young and Jakeman, 1979; Young, 1984]. This has been used in a wide variety of hydrological applications over the past 30 years. These include calibration of rainfall-flow models such as IHACRES Jakeman *et al* [1990] and its *Data-Based Mechanistic* (DBM) model equivalent [e.g. Young 2001 and the prior references therein]. In the example below, the RIV method is used for indirect identification and estimation of CT models. For comparative purposes, however, the indirect estimation is also achieved using the well-known *Prediction Error Minimisation* (PEM) algorithm in the Matlab *System Identification Toolbox*. In both cases, the D2CM algorithm with the ZOH interpolation assumption (see above) is used to convert the estimated DT algorithms from discrete to continuous-time form.

Finally, model structure identification is based on two statistical criteria. First, the *Coefficient of Determination*, R_T^2, based on the error between the sampled output data y_k and the simulated CT model output at the same sampling instants (this is normally equivalent to the well known *Nash-Sutcliffe Efficiency*). Second, the YIC statistic, which is a measure of model identifiability and is based on how well the parameter estimates are defined statistically. These statistics are discussed in more detail in Appendix 3 of Young [2001].

4. PRACTICAL EXAMPLE

Figure 1 shows a portion of effective rainfall $u(t)$ and flow $y(t)$ data from the River Canning, an ephemeral River in Western Australia. A longer set of these data has been analysed comprehensively in Young *et al* [1997], using discrete-time DBM modelling. The effective rainfall series plotted in the lower panel of Fig.1 is generated from the effective rainfall nonlinearity identified in this earlier study, with the effective rainfall generated by the equation

$$u(t) = r(t)y(t)^{\gamma},$$

where $r(t)$ is the measured rainfall and γ is a power law parameter. Here, the flow $y(t)$ is acting as a *surrogate* measure of the catchment storage, rather than attempting to model this storage separately in some conceptual manner [see e.g. Young 2001].

In the present paper, we employ the smaller data set shown in Figure 1 in order to obtain linear, continuous-time DBM models between the effective rainfall $u(t)$, as defined in the above manner, and the flow $y(t)$.

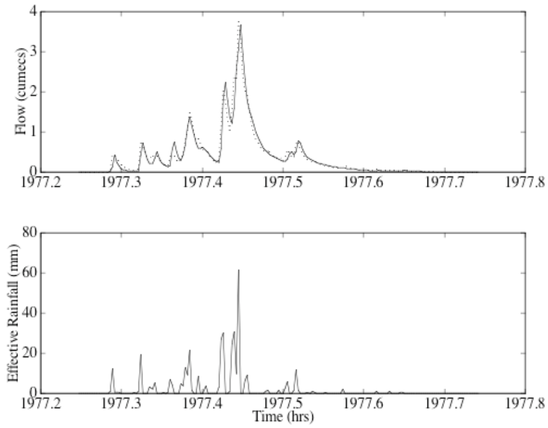

Figure 1. Daily effective rainfall and flow data for the River Canning in Western Australia with CT model output in upper panel (dotted line).

4.1 Direct CT Identification and Estimation

Let us consider first identification and estimation based on the measured daily data shown in Figure 1. The RIVC algorithm, in combination with the R_T^2 and YIC statistics, identifies a [2 3 0] second order model with the parameter estimates:

$$\hat{a}_1 = 0.457\ (0.032) \quad \hat{a}_2 = 0.0248\ (0.0045)$$
$$\hat{b}_0 = 0.0138\ (0.001) \quad \hat{b}_1 = 0.0505\ (0.002)$$
$$\hat{b}_2 = 0.0046\ (0.0008)$$

where, here and elsewhere in the paper, the figures in parentheses are the estimated standard errors obtained from the RIVC algorithm. This model output (shown as the dotted line in upper panel of Fig.1) explains the data very well with $R_T^2 = 0.980$ (98% of the output flow explained by the simulated CT model output) and a residual variance of 0.00723 (standard deviation 0.085 cumecs, where the maximum flow is 3.86 cumecs). The model can also be decomposed by standard TF decomposition [Young, 1992] into a parallel pathway form. The three pathways are: (i) an instantaneous effect, accounting for 7.4% of the flow, which is a measure of rapid flow that occurs in the very short term (within one sampling interval; here one day); (ii) a ‘quick’ pathway, accounting for 54.1% of the flow, that represents a linear store with a residence time of about 2.53 days, probably the result of surface and shallow, sub-surface processes in the catchment; and (iii) a ‘slow’ or ‘base-flow’ pathway, accounting for 38.5% of the flow, that passes through a much longer 15.9 day residence time, linear store, probably representing the effects of deeper groundwater processes and the displacement of ‘old water’.

4.2 Indirect CT Identification and Estimation

The first step of indirect CT modelling involves DT identification using the RIV algorithm (see earlier). This identifies a similar structure [2 3 0] TF with the following parameter estimates:

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

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

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

However, since this is a discrete-time model, it has to be converted to continuous-time form. The D2CM algorithm in Matlab (see section 3.) accomplishes this conversion, using a ZOH approximation (i.e. the input effective rainfall is assumed to be constant over the sampling interval). The resulting CT model parameter estimates are:

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

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

Note that there are no standard errors quoted here because these are not available after conversion and must be computed separately in some manner, usually by Monte Carlo simulation analysis (see later). In this case, the model is quite similar to the directly estimated CT model in the previous section 4.1.

4.3 Simulation Analysis

From the above results, it is clear that, at the daily sampling interval, direct and indirect methods of estimation produce quite similar estimation results in this example. However, this is not always the case and the direct approach has a number of advantages. First, it provides direct estimates of the differential equation model parameters, without the need for conversion from discrete to continuous time. Moreover, after TF decomp-osition into first order linear storage elements, the derived parameters have an immediate physical interpretation. Second, the RIVC analysis provides a direct estimate of the parametric error covariance matrix, so allowing for the specification of error bounds on these physically meaningful parameters, as well as the model output or model-based forecasts. Finally, the estimated covariance matrix provides the information required for sensitivity and uncertainty analysis based on numerical Monte Carlo Simulation (MCS) analysis. For instance, this can be used to obtain empirical probability distribution functions (normalised histograms) associated with any derived parameters, such as the parallel flow partition percentages [see e.g. Young, 2001]. Or it can allow for the quantification of uncertainty in model forecasts associated with the parametric estimation errors.

While these advantages, in themselves, do not make the direct CT approach irresistible, CT estimation has one other advantage: it provides more robust estimates of the model parameters at rapid sampling rates, where the indirect DT-based approaches encounter difficulties. These difficulties are associated with the fact that, when the sampling interval is short, the denominator roots (eigenvalues) of the DT model approach the boundary of the unit circle and the associated TF parameter estimates become much less well defined. For instance, in the case of the DT model (5a,b), where the root is $-a_1$, it is clear that as $\Delta t \rightarrow 0$ so $a_1 \rightarrow 1.0$ and small changes in the parameter lead to large changes in the residence time $T = 1/a_1$. This can lead to poorer confidence in the parameter estimates and, more importantly, a lack of robustness that leads to the DT estimation methods encountering convergence problems. Also, the DT models at rapid sampling intervals have much poorer numerical properties when used for simulation or control system design.

These questions of what sampling interval is 'best' for parameter estimation purposes have been known for some considerable time but there are no completely general analytical results to guide the modeller in this regard. However, experience over many years has led to a 'rule-of-thumb' that a good range of sampling intervals for DT models is between 10 and 50 times the system bandwidth; or equivalently, between $1/10^{th}$ and $1/50^{th}$ of the time constant. Since hydrological systems tend to be characterised by combinations of quick and slow modes of behaviour, as in the present example, this rule has to applied carefully and can restrict the choice of sampling interval. For instance, on this basis, the sampling interval for the slow mode ($T \approx 16$ days) should not be greater than 16/50 days, or about 8 hours; while in the case of the quick mode ($T \approx 2.5$ days) it should not be greater than 2.5/50 days, or about 1 hour.

In this example, let us explore the consequences of using different sampling intervals. In order to do this, the input effective rainfall data can be interpolated so that it has 288 five minute samples over each day. Then this rapidly sampled series can be used as the input to the RIVC estimated continuous time model, in order to generate a *simulated* flow series at this sampling interval. Similar simulations can be performed over a whole range of sampling intervals from 5 minutes to one day. These simulated outputs can then be used as the basis for a MCS study at each of these sampling intervals. Here, for each realisation at the selected sampling interval, the output data is the sum of the 'noise-free' simulated output at this sampling interval and an independent white noise signal with similar variance to the estimated residual error of the RIVC model (0.0072). The MCS results at each sampling interval are based on 50 realisations of this type; and the performance of each algorithm (RIVC, RIV and PEM) is measured by the number of times in the 50 realisations that the algorithm fails to converge to a reasonable model. Here a 'reasonable' model is defined as one in which the estimate $\hat{a}_1$ of the a_1 parameter is within 3 standard deviations of the true value (0.457 ± 0.096).

The full results from this MCS study are given in Young [2004]. They show 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 and PEM-based 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 $\Delta t > 1.0$ hour, where the failure rate is only 3%. This is superior to the PEM performance.

5 DISCUSSION AND CONCLUSIONS

This paper has outlined an approach to *Data-Based Mechanistic* (DBM) modelling of hydrological systems based on the direct identification and estimation of continuous-time (transfer function or differential equation) models. It has also argued that such models are more appropriate to PUB regionalisation studies than their more widely known discrete-time equivalents. The major advantages of continuous-time models in this regard are:

- The model parameters values are unique: unlike discrete-time models, they are not a function of the data sampling interval.
- The model is in an ordinary differential equation form that can be related directly to the formulation of physically meaningful models, such as those derived from mass, energy and momentum conservation.
- Model parameter estimation is superior over a wider range of sampling intervals, particular at fast sampling intervals (e.g. 15 min. for rainfall-flow measurements).

These advantages, together with the parsimony that is a natural consequence of DBM transfer function modelling, should mean that any relationships between the CT model parameters and physical measures of the catchment characteristics should be clearer and better defined statistically, as required for PUB applications.

6 REFERENCES

Beck, M.B. and P.C. Young, A Dynamic Model for DO-BOD Relationships in a Non-Tidal Stream, Water Research, **9**, 769-776, 1975.

Beven, K.J., *Rainfall-Runoff Modelling: The Primer*, John Wiley & Sons: Chichester, 2001.

Dooge, J.C.I., Theory of flood routing D.A. Kraijenhoff and J.R. Moll (eds.), *River Flow Modelling and Forecasting*, D. Reidel: Dordrecht, 39-65, 1986.

Jakeman, A.J., I.G. Littlewood, and P.G. Whitehead, Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments, Journal of Hydrology, **117**, 275-300, 1990.

Krajenoff, D.A. and J.R. Moll, *River Flow Modelling and Forecasting*, D. Reidel: Dordrecht, 1986.

Meijer, H., Simplified flood routing, *Civil Eng.*, **11**, 306-307, 1941.

Nash, J.E., Systematic determination of unit hydrograph parameters, *Jnl. Geophys. Res.*, **64**, 111-115, 1959.

O'Donnell, T., Deterministic catchment modelling, in D.A. Kraijenhoff and J.R. Moll (eds.), *River Flow Modelling and Forecasting*, D. Reidel: Dordrecht, 11-37, 1986.

Wallis, S.G., P.C. Young and K.J. Beven, Experimental investigation of the aggregated dead zone model for longitudinal solute transport in stream channels, *Proc. Inst. of Civil Engrs, Part* 2, **87**, 1-22, 1989.

Young, P.C., Time-series methods and recursive estimation in hydrological systems analysis, Chapter 6 in D.A. Kraijenhoff and J.R. Moll (eds.), *River Flow Modelling and Forecasting*, D. Reidel: Dordrecht, 129-180, 1986.

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

Young, P.C., Parallel processes in hydrology and water quality: a unified time series approach, *Jnl. Inst. of Water and Env. Man.*, **6**, 598-612, 1992.

Young, P.C., Data-based mechanistic modelling and validation of rainfall-flow processes. In *Model Validation: Perspectives in Hydrological Science* (M. G. Anderson and P. D. Bates eds.), Chichester: J. Wiley, 117-161, 2001.

Young, P.C., Identification and Estimation Of Continuous-Time Hydrological Models From Discrete-Time Data, *Proceedings, British Hydrological Society BHS2004 Conference, Imperial College*, London, 2004

Young, P.C. and A.J. Jakeman, Refined instrumental variable methods of recursive time-series analysis: Part I, single-input, single-output systems, *Int. Jnl. of Control*, **29**, 1-30, 1979.

Young, P.C. and A.J. Jakeman, Refined instrumental variable methods of recursive time-series analysis: Part III, extensions, *Int. Jnl. of Control*, **31**, 741-764, 1980.

Young, P.C., A.J. Jakeman and D.A. Post, Recent advances in data-based modelling and analysis of hydrological systems. *Water Sci. Tech.*, **36**, 99-116, 1997.