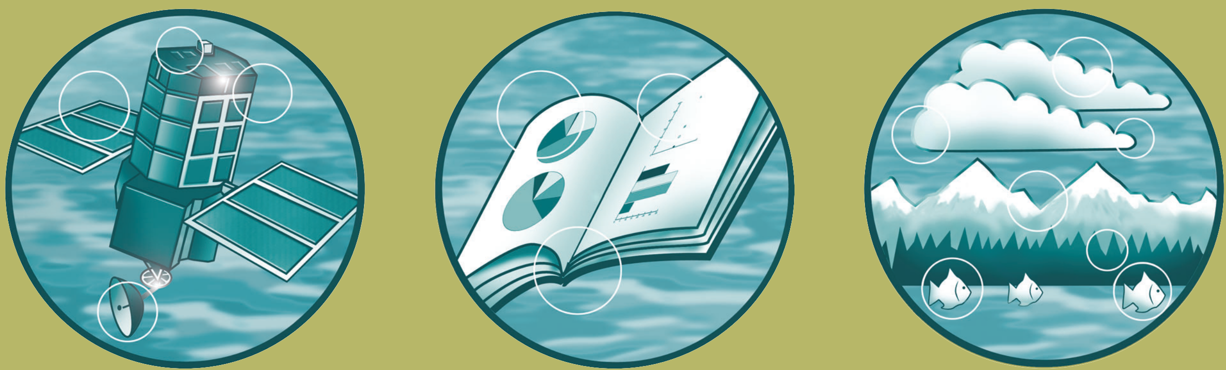
Joint Defra/EA Flood and Coastal Erosion Risk Management R&D Programme

FD2120: Analysis of historical data sets to look for impacts of land use management change on flood generation

A3. Technical appendix: Methods of analysis

R&D Project Record FD2120/PR



Environment Agency

defra
Department for Environment Food and Rural Affairs

A3. TECHNICAL APPENDIX: METHODS OF ANALYSIS

Peter Young, Renata Romanowicz and Keith Beven
Environmental Science, Lancaster University, Lancaster LA1 4YQ.

TP3.1 Detecting Change in Long Term Records: The Unobserved Component Method for Stochastic Trend Extraction

The *Dynamic Harmonic Regression* (DHR) algorithm (Young *et al*, 1999) exploits an 'unobserved component' model to optimally decompose the time series into long term trend, periodic or quasi-periodic and 'irregular' components, including the effect of other input variables if this is required. The method is statistically based and so each component is extracted, together with associated measures of confidence (standard error bounds) that will allow for testing the significance of any changes in the components.

The DHR model has the following form (Young *et al.,* 1999):

$$y_t = T_t + S_t + e_t \qquad e_t : N\{0, \sigma^2\} \tag{1}$$

where y_t is the observed time series; S_t is a seasonal/cyclical component; e_t is a noise component, modelled as mutually independent Gaussian random variables with zero mean and variance σ^2 (i.e. discrete time white noise); and T_t is a smoother, long-to-medium term 'trend' component, without any periodicity, that reflects that part of the series not accounted for by the seasonal/cyclical and irregular components. The seasonal/cyclical S_t components are modelled in the following *Time Variable Parameter* (TVP), trigonometric form:

$$S_t = \sum_{i=1}^{R} \{a_{i,t} \cos(\omega_i t) + b_{i,t} \sin(\omega_i t)\} \tag{2}$$

where ω_i, i=1,..,R are the fundamental and harmonic frequencies associated with the seasonality/cycles in the series; and $a_{i,t}, b_{i,t}$ are the parameters that are allowed to vary in time over the observation interval if this is indicated by the analysis and optimization of the model.

In order to allow for the non-stationarity of the observed time series, the trend T_t in (1) and the parameters $a_{i,t}, b_{i,t}$ in (2) are modelled as stochastic TVP's, each defined as a nonstationary stochastic variable. Identification of the model (1)-(2) is based on the *Auto Regressive* (AR) spectrum of the time series, with the AR order identified by the *Akaike Information Criterion* (AIC) (Akaike, 1974). Each parameter (T_t, $a_{i,t}$, $b_{i,t}$) is modelled as a generalised random walk (GRW) process defined in the state space form:

$$x_{i,t} = F_i x_{i,t-1} + G_i \eta_{i,t} \qquad i = 1,...,2R+1 \tag{3}$$

where

$$F_i = \begin{bmatrix} \alpha & \beta \\ 0 & \lambda \end{bmatrix}, G_i = \begin{bmatrix} \delta & 0 \\ 0 & \varepsilon \end{bmatrix}$$

and the $\eta_{i,t} = [\eta_{i,1} \quad \eta_{i,2}]^T$ are mutually independent bivariate normal $N(0, Q_{\eta i})$ random variables.

Estimation of parameters and prediction of the unobserved components is accomplished using the methods described in Young *et al.* (1999). The parameters (variances and other hyper-parameters in (3)[1]) are estimated by matching the model spectrum to the AR spectrum of the data, and the unobserved components S_t, T_t and e_t, are estimated using recursive fixed interval smoothing (FIS).

TP3.2 Detecting change in catchment dynamics: the Data-based Mechanistic Approach

Within the history of science, two main approaches to mathematical modelling can be discerned; approaches which, not surprisingly, can be related to the more general deductive and inductive approaches to scientific inference that have been identified by philosophers of science from Francis Bacon to Karl Popper and Thomas Kuhn.

1. The ‘ *hypothetico-deductive*’ approach. Here, the *a priori* conceptual model structure is effectively a theory of behaviour based on the perception of the scientist/modeller and is strongly conditioned by assumptions that derive from current scientific paradigms.
2. The ‘*inductive*’ approach. Here, theoretical preconceptions are avoided as much as possible in the initial stages of the analysis. In particular, the model structure is not pre-specified by the modeller but, wherever possible, it is inferred directly from the observational data in relation to a more general class of models. Only then is the model interpreted in a physically meaningful manner, most often (but not always) within the context of the current scientific paradigms.

It is this latter inductive type that forms the basis for Data Based Mechanistic (DBM) modelling (Young, 2002a and references therein). The main stages in DBM model building are as follows:

1. The important first step is to define the objectives of the modelling exercise and to consider the type of model that is most appropriate to meeting these objectives. The prior assumptions about the form and structure of this model are kept at a minimum in order to avoid the prejudicial imposition of untested perceptions about the nature and complexity of the model needed to meet the defined objectives.
2. An appropriate model *structure* is identified by a process of objective statistical inference applied directly to the time-series data and based on a given *general class*

[1] The term ‘hyper-parameters’ is used here to differentiate these constant parameters from the TVPs in the DHR model.

of *linear* Stochastic Transfer Function (STF) models *whose parameters are allowed to vary over time*, if this seems necessary to satisfactorily explain the data.

3. If the model is identified as predominantly linear or piece-wise linear, then the *constant parameters* that characterise the identified model structure in step 2. are estimated using advanced methods of statistical estimation for dynamic systems. The methods used in the present chapter are the *Refined Instrumental Variable* (RIV) *and Simplified RIV* (SRIV) algorithms, which provide a robust approach to model identification and estimation that has been well tested in practical applications over many years. Full details of these methods are provided in Young and Jakeman (1979); Young, (1984, 1985). They are also outlined in Young and Beven (1994) and Young et al. (1996).

4. If significant parameter variation is detected then the model parameters are estimated by the application of TVP or SDP estimation, as discussed previously
5. If nonlinear phenomena have been detected and identified in stage 4, the non-parametric, state dependent relationships are normally parameterised in a *finite* form and the resulting nonlinear model is estimated using some form of numerical optimisation, such as nonlinear least squares or ML based on prediction error decomposition (Schweppe, 1965).
6. Regardless of whether the model is identified and estimated in linear or nonlinear form, it is only accepted as a credible representation of the system if, in addition to explaining the data well, it also *provides a description that has direct relevance to the physical reality of the system under study*. This is a most important aspect of DBM modelling and differentiates it from more classical statistical modelling methodology.
7. Finally, the estimated model is tested in various ways to ensure that it is conditionally valid. This involves standard statistical diagnostic tests for stochastic, dynamic models, including analysis which ensures that the nonlinear effects have been modelled adequately, as well as exercises in predictive validation (i.e., it performs well in predicting over data other than that used in the model estimation) and stochastic sensitivity analysis (Young, 1999a).

Of course, while step 6 should ensure that the model equations have an acceptable physical interpretation, it does not guarantee that this interpretation will necessarily conform exactly with the current scientific paradigms. Indeed, one of the most exciting, albeit controversial, aspects of DBM models is that they can tend to question such paradigms. For example, DBM methods have been applied very successfully to the characterisation of imperfect mixing in fluid flow processes and, in the case of pollutant transport in rivers, have led to the development of the *Aggregated Dead Zone* (ADZ) model (Beer and Young, 1983; Wallis *et al.*, 1989). Despite its initially unusual physical interpretation, the practical success of this ADZ model and its formulation in terms of physically meaningful parameters, seriously questions certain aspects of the ubiquitous *Advection Dispersion Model* (ADE), which preceded it as the most credible theory of pollutant transport in stream channels (see e.g. Young and Wallis, 1994).

TP3.2.1 Transfer Function Methods

The objective of Data-Based Mechanistic (DBM) modelling (e.g. Young and Lees, 1993; Young, 1998 and the prior references therein) is to infer the nature and structure of such models directly from hydrological data, using powerful methods of statistical inference, and to then interpret the model equations in physically meaningful terms. One important generic model class that facilitates such DBM modelling studies is the Transfer Function (TF) family of models, which is the subject of this Section.

First, some of the background to the use of TF models in hydrology is reviewed, concentrating on prior publications concerned with the modelling of rainfall-flow processes. This is followed by a sub-Section that discusses the formulation of linear, constant parameter TF models in the derivative operator $s = d/dt$, which are simply the continuous-time (CT), transfer function form of ordinary differential equations. It then proceeds to develop and discuss the discrete-time (DT) equivalents of these CT models, namely transfer functions in the backward shift operator z^{-1} (see later), before outlining methods for the statistical identification and estimation of both CT and DT models from noisy time series data. In this manner, the models take on a natural stochastic form that is appropriate to their use in applications such as hydrological forecasting, uncertainty and risk analysis. The following sub-Section shows that transfer function models can be extended to incorporate time variable and 'state-dependent' parameters, so allowing them to describe nonstationary and nonlinear stochastic systems, so called *State Dependent Parameter* (SDP) models. Finally, the last sub-Section shows the application of the DBM methods to rainfall-flow modelling.

TF modelling is important in the present Report because the analysis of the changes in the linear dynamics of the TF models describing the rainfall-runoff processes, as well as the changes in nonlinear relationships between the process variables, is used as a tool to detect the signatures of the land use changes on the flow regime.

TP3.2.2 Linear Continuous-Time TF Models

In order to introduce TF models, let us consider first a conceptual catchment storage equation in the form of a continuous-time, linear storage (store, tank or reservoir) model: see, for example, the review papers by O'Donnell, Dooge and Young in Kraijenhoff and Moll (1986); or more comprehensive treatments, such as the books by Beven (2001) and Dooge and O'Kane (2003). Here, the rate of change of storage in the channel is defined in terms of water volume entering the linear storage element (e.g. river reach) in unit time, minus the volume leaving in the same time interval, i.e.,

$$\frac{dS(t)}{dt} = GQ_i(t-\tau) - Q_0(t) \qquad (1)$$

where $Q_i(t-\tau)$ represents the input flow rate delayed by a pure time or 'transport' 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.,

$$Q_0(t) = T \cdot S(t)$$

and substituting into (1), we obtain,

$$T\frac{dQ_0(t)}{dt} = GQ_i(t-\tau) - Q_0(t) \qquad (2)$$

This equation is a first order, linear differential equation model whose response, from an initial steady flow condition, to a unit impulsive change Q_i^{imp} of the input flow at time $t = t_0$, is given by

$$Q_0(t-\tau) = Q_e + Q_i^{imp} \exp\{-(t-t_0)/T\}$$

where Q_e is the initial steady flow level. This has a typical hydrograph recession shape, with a decay Time Constant, T, that defines the *Residence Time* of the model. As we shall see later, combinations of two or more such first order models, exhibit a typical unit hydrograph form (e.g. Dooge, 1959; Beven, 2001).

By introducing the derivative operator s, i.e. $s = \frac{d(\cdot)}{dt}$, and collecting like-terms together, it is easy to see that equation (2) can be written as (Young, 2004a),

$$(1+Ts)Q_0(t) = GQ_i(t-\tau)$$

so that, dividing throughout by $1+Ts$, we obtain the following continuous-time TF form of equation (2),

$$Q_0(t) = H(s)Q_i(t-\tau) \qquad (3)$$

where,

$$H(s) = \frac{G}{1+Ts}$$

represents the TF in terms of the derivative operator *s*.

(a) *Physically Interpretable Parameters*

The TF model (3) is characterized by three parameters: G, T and τ. However, there are five, physically interpretable model parameters associated with the model that are worth discussing. 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, for example, it indicates whether water has been added (G > 1) or lost (G < 1) between the upstream and downstream boundaries; and the percentage of water lost or gained can be defined by

Loss Efficiency LE = 100(1−G), which will be negative if G > 1.0. As pointed out above, the Residence Time or Time Constant T is the time required for the storage element output to decay to e^{-1} or 0.3679 of its maximum value in response to an impulsive 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.

TP3.2.3 Linear Discrete-Time TF Models

To date, the most popular form of TF modelling has been carried using the discrete time (DT) equivalent of the model (3). Considering the associated differential equation (2) at a uniform sampling interval of Δt time units, an approximate discrete-time form of this model can be obtained in the following manner by approximating the first order derivative and introducing sampled variables:

$$T\frac{Q_{0,k} - Q_{0,k-1}}{\Delta t} \approx GQ_{i,k-\delta} - Q_{0,k-1} \tag{4}$$

Here $Q_{0,k}$ is the sampled value of $Q_0(t)$ at the k^{th} sampling instant, i.e. after $k\Delta t$ time units; and δ is the advective time delay, normally defined as the nearest integer value of $\tau / \Delta t$ (thus incurring a possible approximation error). Collecting terms in this equation, it can be written as,

$$Q_{0,k} \approx -a_1 Q_{0,k-1} + b_0 Q_{i,k-\delta} \tag{5}$$

where $a_1 = -(1 - \Delta t / T) = -(1 - f_1 \Delta t)$ and $b_0 = G(\Delta t / T) = g_0 \Delta t$. This reveals that, as a first approximation in discrete-time, the flow $Q_{0,k}$ at the k^{th} sampling instant is a proportion -a_1 (note that, in the present context, a_1 will be a negative number less than unity, so that this is a positive proportion) of its value $Q_{0,k-1}$ at the previous $(k-1)^{th}$ sampling instant, plus a proportion b_0 of the delayed upstream flow input $Q_{i,k-\delta}$ measured δ sampling instants previously. Although equation (5) is an approximate expression, depending on the size of the sampling interval ΔT and the definition of the sampled variables, it can be shown that it has the same form as a more accurate discrete-time equivalent of (2). In this alternative discrete-time representation, the values of the parameters a_1 and b_0 in equations (5) can be related to the parameters of the model (2) in various ways depending upon how the input flow $Q_i(t)$ is assumed to change over the sampling interval between the measurement of $Q_{i,k-1}$ and $Q_{i,k}$ (since it is not measured over this interval). The simplest and most common assumption is that it remains constant over this interval (the so-called *zero-order hold*, ZOH, assumption), in which case the relationships are as follows:

$$a_1 = -\exp(-f_1\Delta t) \quad b_0 = \frac{g_0}{f_1}\{1-\exp(-f_1\Delta t)\} \tag{6}$$

The expressions for a_1 and b_0 in (5) are approximations of these more accurate expressions. Note also that, because these relationships are functions of the sampling interval Δt, for every unique CT model such as (3), there are infinitely many DT equivalents, depending on the choice of Δt, all with different parameter values defined in (6). With the definitions of equation (3), the equivalent discrete-time TF version of the model (5) (now with an equality sign) takes the form:

$$Q_{0,k} = \frac{b_0}{1+a_1 z^{-1}} Q_{i,k-\delta} \tag{7}$$

where z^{-1} is the backward shift operator; i.e. $z^{-1}Q_{0,k} = Q_{0,k-1}$ or, in general terms, $z^{-r}Q_{0,k} = Q_{o,k-r}$.

Following from the definition of this first order DT model at the chosen Δt, the general multi-order equivalent of the general CT model is defined by the following discrete-time *Stochastic Transfer Function* (STF) model:

$$y_k = \frac{B(z^{-1})}{A(z^{-1})} u_{k-\delta} + \xi_k \tag{8}$$

Here, ξ_k is additive noise that represents the effects of any unmeasurable, stochastic inputs or measurement errors; δ is a pure, advective time delay of $\delta\Delta t$ time units, while $A(z^{-1})$ and $B(z^{-1})$ are polynomials in the backward shift operator z^{-i}: (i.e. $z^{-i}y_k = y_{k-i}$) of the following form:

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

The orders of the polynomials *n* and *m* are identified from the data during the initial identification procedure and are normally in the range 1-3. In the following analysis, the triad [$n\ m\ \delta$] is used to characterize this model structure.

The STF model (8) can be written in the alternative discrete-time equation form,

$$y_k = -a_1 y_{k-1} - a_2 y_{k-2} - \cdots - a_n y_{k-n} + b_0 u_{k-\delta} + b_1 u_{k-\delta-1} + \cdots + b_m u_{k-\delta-m} + \eta_k$$

where $\eta_k = A(z^{-1})\xi_k$ is the transformed additive noise input. If this model is considered in the context of rainfall-flow models, it shows that the river flow at the k^{th} hour y_k is dependent on level and *effective* rainfall measurements $u_{k-\delta}$ made over previous hours, as well as the uncertainty η_k arising from all sources. Note that, although the STF relationship between effective rainfall and flow is linear, the noise η_k is dependent on the model parameters, thus precluding linear estimation; and the complete model between measured rainfall and water level is quite heavily nonlinear because of the effective rainfall nonlinearity which is a nonlinear function of the measured rainfall (see main text).

TP3.2.4 Model Structure Identification

An important aspect of TF modelling, in both continuous and discrete-time, is Model Structure Identification (the definition of the model structure triads). This can be approached in various ways. For example, Box and Jenkins (1970) discuss the topic at length and Akaike (1974) suggested an alternative approach that has since spawned several other, related methods.

Within the context of the *Instrumental Variable* (IV) methods used for the analysis in the present Report, however, model structure identification is based conveniently on two statistical criteria. First, the *Coefficient of Determination*, R_T^2, a normalized measure based on the variance of the error between the sampled output data y_k and the simulated (CT or DT) model output at the same sampling instants (this is similar to the well known *Nash-Sutcliffe Efficiency* measure: Nash and Sutcliffe, 1970); and 100xR_T^2 is the percentage of the variance of the output data explained by the model. Note that R_T^2 should not be confused with its well known relative R^2, as used in classical regression and time series analysis, which is defined in terms of the one-step-ahead prediction errors. In general, R_T^2 is a more discerning measure of model adequacy in a dynamic systems context than R^2, since it quantifies the ability of the model to explain the whole of the simulated output data derived only from the input data, without any reference to the output data. In contrast, R^2 only measures the ability of the model to predict one-step-ahead on the basis of the latest measured input and output data (as in flow forecasting). The second statistic is the *Young Information Criterion* (YIC): this is a heuristic measure of model identifiability and is based on the properties of the *Instrumental Product Matrix* (IPM) that is generated as part of the IV estimation procedure. These statistics are discussed in more detail in Young (1989), Appendix 3 of Young (2001b) and Young (2005).

TP3.2.5 Nonstationary and Nonlinear TF models

Nonlinearity in rainfall-flow and water level models is very important because it defines the way in which the model responds under different catchment wetness conditions. In general, therefore, it is necessary to consider extensions to the constant parameter, linear STF model that allow for the consideration of more complex hydrological processes such as this. Fortunately, it is possible to extend STF models to handle 'nonstationary' situations described by *Time Variable Parameter* (TVP) models; or 'nonlinear' situations characterized by *State-Dependent Parameter* (SDP) models. The latter model class encompasses a wide variety of nonlinear, stochastic, dynamic phenomena, including even chaotic systems.

Both TVP and SDP types of model can be considered in either continuous or discrete-time. However, since statistical estimation is more straightforward in the DT case, it often provides the more straightforward approach in nonstationary and nonlinear situations. Consequently, the brief descriptions that follow are limited to this DT situation.

(a) *Nonstationary Time Variable Parameter (TVP) Models*

In the DT case, the TVP form of the TF model (8) can be written as follows:

$$x_k = \frac{B(k,z^{-1})}{A(k,z^{-1})}u_{k-\delta} \quad y_k = x_k + \xi_k \tag{9}$$

where,

$$A(k,z^{-1}) = 1 + a_{1,k}z^{-1} + a_{2,k}z^{-2} + \ldots + a_{n,k}z^{-n}$$
$$B(k,z^{-1}) = b_{0,k} + b_{1,k}z^{-1} + b_{2,k}z^{-2} + \ldots + b_{m,k}z^{-m}$$

Here, all the parameters are assumed to be functions of the time index *k*, i.e. it is assumed that they may vary over time in an unknown manner that needs to be estimated from the data. Some of the latest research on TVP estimation is reported in Young (1999b, 2000, 2001a) where the reader will find a complete description of the recursive estimation algorithms that allow for the estimation of the time variable parameters.

(b) *Nonlinear State-Dependent Parameter (SDP) Models*

Again in the discrete-time case, the SDP form of the STF model (8) can be written as follows:

$$x_k = \frac{B(w_k,z^{-1})}{A(v_k,z^{-1})}u_{k-\delta} \quad y_k = x_k + \xi_k \tag{10}$$

where,

$$A(v_k,z^{-1}) = 1 + a_1(v_{1,k})z^{-1} + a_2(v_{2,k})z^{-2} + \ldots + a_n(v_{n,k})z^{-n}$$
$$B(w_k,z^{-1}) = b_0(w_{0,k}) + b_1(w_{1,k})z^{-1} + b_2(w_{2,k})z^{-2} + \ldots + b_m(w_{m,k})z^{-m}$$

Here, the possibility that the parameters may be functions of other 'state' variables is investigated. In other words, it is assumed that any parameter in the model may vary over time as a function of the temporal variation in one or more other variables (e.g. the input u_k or output y_k and their past values), so introducing nonlinear behaviour into the model. In the SDP model (10), these variables are denoted by $v_{j,k}$, i = 1, 2, ..., n and $w_{j,k}$, j = 0, 1, ...,m and they are, respectively, the elements of the vectors v_k and w_k.

The first applications of SDP estimation in hydrology are described in Young (1993); and Young and Beven (1994); while the latest research on SDP estimation is reported in Young (2000, 2001a,b, 2002a, 2003, 2006a,b) and Young et al.(2001). This later research exposes a special example of the SDP model that is particularly important in a rainfall-flow context. This is the ‘Hammerstein’ model, where the SDP nonlinearity only affects the numerator parameters in the SDP transfer function model (10) and is of a form where it can be factored out of the model and form a single, input nonlinearity that converts the measured rainfall into an ‘effective rainfall’ (see next Section 2.6). This model can be written in the form:

$$x_k = \frac{B(z^{-1})}{A(z^{-1})} f(u_{k-\delta}) \quad y_k = x_k + \xi_k \tag{11}$$

where $f(u_{k-\delta})$ represents the SDP effective rainfall nonlinearity. This is discussed further in the next Section.

TP3.2.6 Nonlinear rainfall-flow model

At any flow measurement location, it is assumed that the rainfall and flow measurements above the base level, *r(t)* and *y(t)*, are sampled uniformly in time at a sampling interval of Δt time units (here hours) and that these discrete-time, sampled measurements are denoted by r_k and y_k. It has been shown (Young [1993, 2001a,b, 2002a, 2003], Young and Beven, [1994], Young and Tomlin [2000], Romanowicz *et al.*, [2004]) that the nonlinearities arising from the relationship between measured rainfall r_k and transformed (effective) rainfall, denoted here by u_k, can be approximated using gauged flow as a surrogate measure of the antecedent wetness or soil-water storage in the catchment. In particular, the scalar function describing the nonlinearity between the rainfall and the soil moisture surrogate y_k is initially identified non-parametrically (graphically) using recursive SDP estimation, as described in the above references. Often this estimated non-parametric relation can be parameterized using a power low or exponential relation. In the former case, the effective rainfall relation takes the form:

$$u_k = c_0 \cdot y_k^{\gamma} \cdot r_k; \tag{12}$$

where u_k denotes the transformed (effective) rainfall; r_k denotes measured rainfall; c_0 is a scaling constant, normally selected so that, with the same units, the total effective rainfall matches the total flow. The power-law exponent γ is estimated by a special optimization procedure that includes the concurrent optimal instrumental variable estimation (see later and Young, 1984) of the linear STF model (9) between the delayed, transformed rainfall $u_{k-\delta}$ and the flow y_k.

Combining equations (11) and (12) the complete rainfall-flow model can be written as:

$$y_k = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1} + a_2 z^{-2}} u_{k-\delta} + \xi_k \qquad u_k = c_0 y_k^{\gamma} r_k \qquad \xi_k = N(0, \sigma_k^2) \tag{13}$$

For simplicity, the additive noise term ξ_k in (13) is assumed here to be a zero mean, normally distributed, white noise sequence (i.e. uncorrelated in time), although an extension of the model to incorporate coloured noise is straightforward (Young, 2003). It is also assumed that ξ_k, which accounts for all the uncertainty associated with the inputs affecting the model, including measurement noise, un-measured inputs, and uncertainty in the model, is heteroscedastic (i.e. its variance σ_k^2 changes over time) and that it is not significantly correlated with the rainfall measurement.

The mechanistic interpretation of the model (13) follows from the decomposition of the linear STF part of this model in (13) into 'fast' and 'slow' components that, in broad terms, reflect the fast and slow physical processes operative in the catchment (see the previous references). These decomposed components, as obtained by partial fraction expansion (e.g. Young , 2005), take the following form:

$$\begin{aligned} &\text{Fast Component:} \qquad y_{1,k} = \frac{\beta_1}{1 + \alpha_1 z^{-1}} u_{k-\delta} \\ &\text{Slow Component:} \qquad y_{2,k} = \frac{\beta_2}{1 + \alpha_2 z^{-1}} u_{k-\delta} \end{aligned} \tag{14}$$

where $\alpha_1, \alpha_2, \beta_1, \beta_2$ are parameters derived from the parameters in the model (13) and the total gauged flow is the sum of these two components and a model error, i.e. $y_k = y_{1,k} + y_{2,k} + \xi_k$. The associated residence times (time constants), T_1, T_2; steady state gains, G_1, G_2; and partition percentages, P_1, P_2, are given by the following expressions:

$$T_i = \frac{\Delta t}{\log_e(\alpha_i)};\ i = 1,2 \qquad G_i = \frac{\beta_i}{1 + \alpha_i};\ i = 1,2 \qquad P_i = \frac{100 G_i}{G_1 + G_2};\ i = 1,2$$

The parameters of the above nonlinear STF model are derived from statistical model identification and estimation analysis based on the observed rainfall- flow data. Here, the statistically optimal RIV (Refined Instrumental Variable) algorithm from the CAPTAIN toolbox (http://www.es.lancs.ac.uk/cres/captain/) for Matlab^TM and the associated DBM statistical modelling concepts are used to identify the order of the STF model (the triad [n m δ]) and to estimate the associated parameters (see previous references). In this manner, the DBM model efficiently reflects the information content of the data, so that the possibility of over-parameterization and associated poor identifiability is avoided.

References

Akaike, H. 1974 A new look at statistical model identification. IEEE Transactions on Automatic Control, 19, 716–723.

Beer, T. and Young, P. C. (1983). Longitudinal dispersion in natural streams. Am. Soc. of Civil Eng., Jnl. Env. Eng., 109:1049–1067.

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

Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis Forecasting and Control. Holden-Day: San Francisco.

Dooge, J. C. I., 1959 A general theory of the unit hydrograph, Journal of Geophysical Research, 64, 241-256.

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

Dooge, J. C. I. and O'Kane, J. P., 2003 Deterministic methods in Systems Hydrology, Balkema: Lisse.

Kraijenhoff, D. A. and Moll, J. R. (eds) 1986 River Flow Modelling and Forecasting. D. Reidel, Dordrecht.

Nash, J. E. and Sutcliffe, J. V., 1970 River flow forecasting through conceptual models: discussion of principles, Journal of Hydrology, 10, 282-290.

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

Romanowicz, R. J., Young, P. C. and Beven, K. J., 2004 Assessing Uncertainty in Assessing Flood Risk, In Proceedings of First International Conference on Flood Risk Assessment, Bath, U.K., 2004, 127-138.

Schweppe, F. (1965). Evaluation of likelihood functions for gaussian signals. IEEE Trans. on Information Theory, 11:61–70.

Wallis, S. G., Young, P. C., and Beven, K. J. (1989). 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.

Young, P. C. 1984 Recursive estimation and time-series analysis, Springer-Verlag: Berlin.

Young, P. C. 1985 The instrumental variable method: a practical approach to identification and system parameter estimation, In (H.A.Barker and P.C.Young eds), Identification and System Parameter Estimation}, Pergamon Press: Oxford, 1-16.

Young, P. C., 1986 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.

Young, P. C. 1989 Recursive estimation, forecasting and adaptive control, Chapter 6 in C. T. Leondes (ed.), Control and Dynamic Systems: Advances in Theory and Applications, Vol. 30}, Academic Press: San Diego, 119-166.

Young, P. C. 1993 Time variable and state dependent modelling of nonstationary and nonlinear time series. In: Developments in time series analysis (T. Subba Rao ed.), pp 374-413. London: Chapman \& Hall.

Young, P. C. 1998 Data-based mechanistic modelling of environmental, ecological, economic and engineering systems, Environmental Modelling \& Software, 13, 105-122.

Young, P. C. 1999a. Data-based mechanistic modelling, generalised sensitivity and dominant mode analysis. Computer Physics Communications, 117:113–129.

Young, P. C. 1999b Nonstationary time series analysis and forecasting, Progress in Environmental Science, 1, 3-48.

Young, P.C. 2000 Stochastic, dynamic modelling and signal processing: time variable and state dependent parameter estimation. In: Nonstationary and nonlinear signal processing, (W.J. Fitzgerald, A. Walden, R. Smith \& P.C. Young eds.), pp 74-114. Cambridge: Cambridge University Press.

Young, P. C. 2001a The identification and estimation of nonlinear stochastic systems, In Nonlinear dynamics and statistics, (A.I. Mees ed.), 127-166, Boston: Birkhauser.

Young, P. C., 2001b 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.

Young, P. C., 2002 Advances in real-time flood forecasting, Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences, 360, 1433-1450.

Young, P. C., 2003 Top-down and data-based mechanistic modelling of rainfall-flow dynamics at the catchment scale, Hydrol. Processes, 17, 2195–2217.

Young, P. C., 2004a Identification and estimation of continuous-time hydrological models from discrete-time data, Proceedings, British Hydrological Society BHS2004 Conference, Imperial College, London.
Young, P. C. 2005 Rainfall-runoff modeling: Transfer Function models. In Anderson, M. G., editor, Encyclopedia of Hydrological Sciences, volume 3, part II, pages 1985–2000, John Wiley: Hoboken, N. J.
Young, P. C., 2006a Data-based mechanistic modelling and river flow forecasting, In Proceedings 14th IFAC Symposium on System Identification, SYSID, 2006, pages 756–761, Newcastle, NSW.
Young, P.C., 2006b The data-based mechanistic approach to the modelling, forecasting and control of environmental systems, Annual Reviews in Control, 30, 169-182.
Young, P. C. and Beven, K. J. 1994, Data-based mechanistic modelling and the rainfall-flow nonlinearity, Environmetrics, 5, 335-363.
Young, P. C. and Jakeman, A. J., 1979, Refined instrumental variable methods of recursive time-series analysis: Parts I, single-input, single-output systems, Int. Jnl. of Control, 29, 1-30.
Young, P. C. and Lees, M., 1993, The Active Mixing Volume (AMV): A new concept in modelling environmental systems, Chapter 1 in V. Barnett and K.F. Turkman (eds.), Statistics for the Environment, J. Wiley: Chichester, 3-44.
Young, P. C. and Tomlin, C. M. 2000, Data-based mechanistic modelling and adaptive flow forecasting. In: Flood Forecasting: What Does Current Research Offer The Practitioner? (M. J. Lees \& P. Walsh eds.), pp 26-40. BHS Occasional paper No 12, produced by the Centre for Ecology and Hydrology on behalf of the British Hydrological Society.
Young, P. C. and Wallis S, G. 1985, Recursive estimation: a unified approach to the identification, estimation and forecasting of hydrological systems, Applied Mathematics & Computation, 17, 299-334.
Young, P. C. and Wallis, S. G., 1994, Solute transport and dispersion in channels, appears as Chapter 6 in K.J. Beven and M.J. Kirby (eds.), Channel Networks, J. Wiley: Chichester, 129-173.
Young, P. C., Pedregal, D. and Tych, W., 1999, Dynamic harmonic regression. Journal of Forecasting, 18, 369-394.
Young, P. C., McKenna, P. and Bruun, J., 2001, Identification of nonlinear stochastic systems by state dependent parameter estimation, International Journal of Control, 74, 1837-1857.
Young, P. C., Parkinson, S.D., and Lees, M. 1996, Simplicity out of complexity in environmental systems: Occam's Razor revisited, J. Appl. Stat., 23, 165-210.

Ergon House
Horseferry Road
London SW1P 2AL

www.defra.gov.uk

defra
Department for Environment
Food and Rural Affairs