DATA-BASED MECHANISTIC MODELLING OF A SNOW AFFECTED BASIN

A. Castelletti * **F. Pianosi** * **R. Soncini-Sessa** *
P.C. Young **

* *Dipartimento di Elettronica e Informazione, Politecnico di Milano, Milan, Italy*
** *Centre for Research on Environmental Systems and Statistics, Lancaster University, Lancaster, UK*

Abstract: A precipitation-temperature-flow model is developed to compute flow from raw precipitation records, taking into consideration snow-melt contribution to the flow. The model does not require other measurements than flow, temperature and raw precipitation, thus resulting particularly useful in all those situations, the majority, where these are the only observed data. A Data-Based Mechanistic (DBM) modelling approach is used in order to keep at a minimum all the a-priori assumptions on the physical mechanism driving the flow formation process. The model has been applied on a classical set of data (the Jakulsa river basin, Iceland) which is well known in the non-linear modellers community.

Keywords: Snow-melt modelling, Data Based Mechanistic approach, rainfall-runoff modelling

1. INTRODUCTION

Snow modelling methodological approaches are usually divided in two major categories [Schreider *et al.*, 1997]: conceptual (or mechanistic), whose main concern is the physical explanation of the snow melt process, and empirical, which infers directly from the data a relationship between the hydro-meteorological characteristics of the catchment. The former approach usually attempt to describe the melting process by modelling the energy balance within the snow pack. However, although conceptual models properly account for the processes determining melt, their use is often made very difficult by the huge number of parameters they include and by the large and detailed data they require (e.g., air temperature, precipitation, snow temperature and density, vegetation cover, cloudiness, etc.). For these reasons conceptual models can not be used in the areas (the majority) where such detailed information are not available over a relatively long period. Moreover in many cases such information is highly variable in space thus adding further uncertainties to the model output. It is in fact to note that the spatial heterogeneity in the hydrological process is implicitly included in empirical models, while it is to be explicitly modelled in conceptual models.

The empirical approach usually exploits the frequently observed high correlation between snow-melt and air temperature. It relies on the daily average air temperature to represent the major source for melt and for this reason it is often known as *degree−day* (or *temperature index*) approach. This approach has been used for many years in many case studies (see Hock [2003] for a review), providing in most of the cases accurate results, comparable to those of more complex models [WMO, 1986]. The physical basis of this assumption (Braithwaite and Olesen, [1990]) is the high correlation of the temperature with several energy balance components (i.e., longwave atmospheric radiation, sensible heat flux and global radiation). More specifically the *degree−day* assumes the snow melt as linearly proportional to the difference between the air temperature and a threshold temperature below which there is not melting. The most widely used formulation is the following (Martinec *et al.* [1983]):

$$m_t = \delta(T_t - T^s) \quad (1)$$

where m_t is the snow melt in the interval $[t-1, t]$, T_t is the daily average temperature calculated on the same interval, T^s is the above mentioned threshold (often $0C^\circ$), while δ is the *degree-day factor*, representing the amount of snow daily melted per temperature degree. The so calculated snow-melt is used to compute an equivalent daily rainfall which could then be considered as the input of a classical rainfall-runoff model. Generally the accuracy of the model highly depends on the value given to the two parameters included in (1), however the approach properly works only when there is perennial snow (glaciers) and the snow melt is always supported by the snow-pack. In fact note that, as the (1) does not explicitly depend on the snow-pack value, the model can not actually know whenever the snow pack is completely melt or not. It is thus necessary to introduce two farther parameters that are the starting and ending dates for the snow-melt process or alternatively account for the snow-pack dynamics, by coupling the rainfall-runoff model with a conceptual model of the snow accumulation (Schreider *et al.* [1997]; Whetton *et al.* [1996]), which obviously requires at leat one new state variable (e.g., the snow depth) to be introduced. Therefore, even though referred to as "empirical", this second approach can not be considered as a purely empirical (inductive) modelling process as it is significantly conditioned by a-priori knowledge.

This paper aims at exploring the applicability of a Data Based Mechanistic (DBM) approach [Young, 1998] for deriving a rainfall-temperature-flow model in a simplified way, by keeping at a minimum both the a-priori assumptions on the physical nature of the process and the number of observational data required to identify the model parameters. The idea underlying the approach is that the snow-melt volumes for every sample interval might be inferred from the very flow. In this way the information on the snow-pack volume, which has to be described as a state variable in conceptual and empirical models, is automatically updated through the observation of the flow it generates. To implement such idea the following DMB procedure has been followed:

(1) the input and output variables are chosen, depending on the objectives of the modelling exercise and data availability, the most parsimonious model structure is statistically inferred from the time series;
(2) if the linear model does not seem to be adequate to the system description, state dependent non-linearities are investigated through a State Dependent Parameter (SDP) procedure [Young, 2002];
(3) a physical meaningful interpretation of the non-linearities is proposed; this is a very important aspect of DBM modelling and differentiates it from more classical 'black-box' modelling methodologies;
(4) the effective inputs are used in a linear transfer function model that is first calibrated through a simulation-optimization approach (based on the Refined Information Variable method [Young, 1984]). Once the nonparametric state dependent relationships have been parameterized, model parameters are estimated all together with a recursive optimization procedure (such as non linear least squares algorithm);
(5) the model error is described as an AR process; the inherently stochastic nature of the model is another important aspect of DBM modelling, which differentiates it from the alternative deterministic approach.
(6) finally the model is validated.

2. THE JOKULSA RIVER BASIN

The Vatnajökull ice cape is the largest ice cape in Europe and covers an area of about 8100 km^2. Many outlet glaciers drain its main ice body, each one of which ends in a glacier river. Among these the Jokulsa river is the largest one. Its catchment covers 7380 km^2, about 1700 km2 of which are subglacial. The time series of its flow rate (see Fig. 1) were introduced in the literature of non-linear modelling by Tong *et al.* [1985] and have been used in several non-linear identification problems [Tong, 1990; Chen and Tsay, 1993].

The curve of Fig. 1a clearly shows the distinctive profile of the sub-arctic flow scenario, with a single seasonal peak in summer.

3. CHOICE OF THE MODEL STRUCTURE

Some suggestions on the choice of the model structure might be issued by mapping daily flow values against precipitation and temperature data (Fig. 2a and 2b). At a first glimpse from the two graphs one would be led to conclude that temperature is the main driving factor of the flow and that such influence depends on the very flow. In fact the general trend of the data indicates a more than linear increase of flow with the temperature. Within a SDP modelling prospect

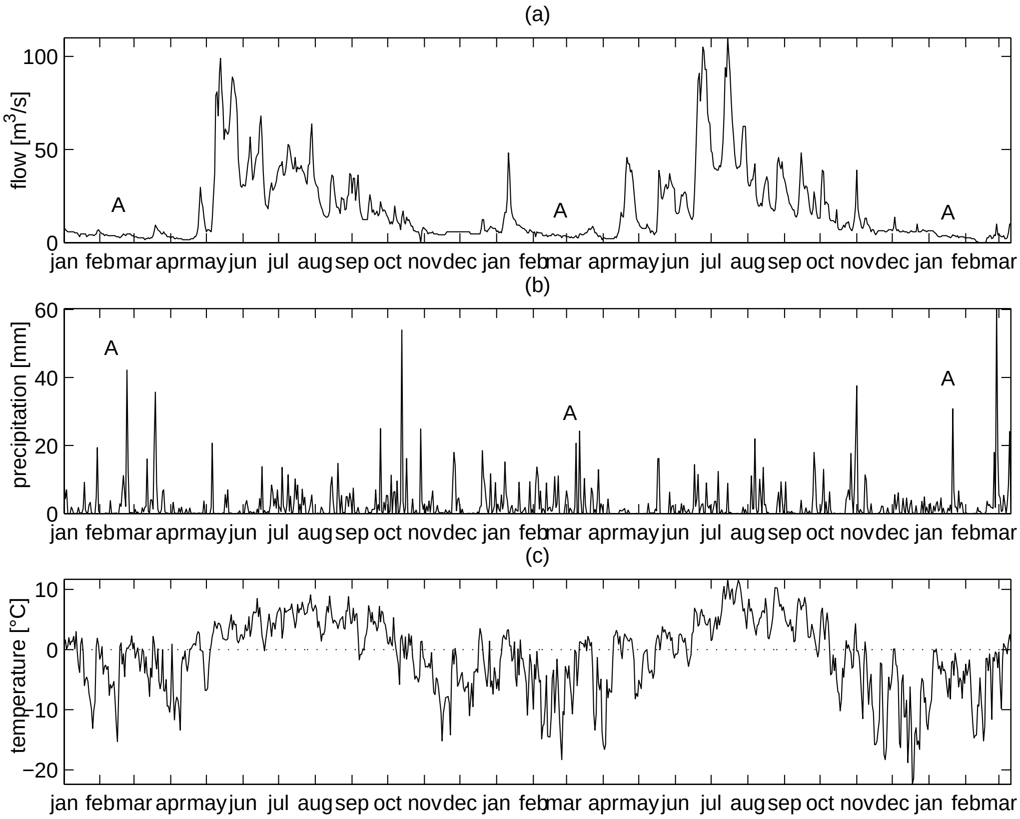

Fig. 1. Daily flow, precipitation and temperature data over the period Jan. 1, 1972 - Mar. 10, 1974 (estimation data set) for the Jokulsa river. Flow is recorded at Dettifoss station, while precipitation and temperature are recorded in Hveravellir meteorological station

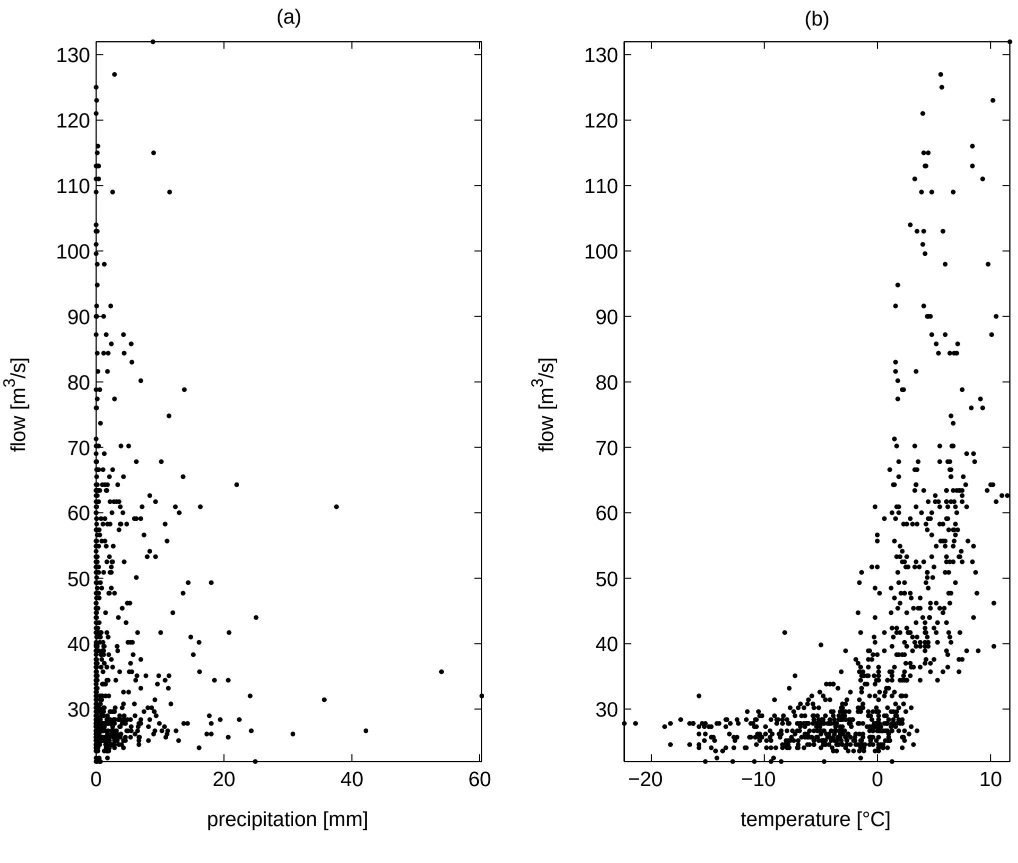

Fig. 2. Daily flow against precipitation (a) and against temperature (b), for the river Jokulsa on the estimation data set.

this suggest that the coefficient expressing the ratio between flow and temperature must vary with the flow, on analogy of what is usually done for the rainfall in the case of simple rainfall-flow models, where the runoff coefficient is a function of the flow that acts as a surrogate for the soil moisture [Young, 2003]. However from Fig. 2a it would be uncorrect to issue that the precipitation is not influencing the flow. Indeed, it can be only inferred that precipitation alone does not drive the flow, however it can not be excluded the existence of a coupled effect of temperature and precipitation on the flow. For that reason we start considering a two input transfer function as the general model that describes the daily flow formation process.

The linear model of the form

$$A^{'}(z)y_t = B_1^{'}(z)u_t + B_2^{'}(z)T_t \tag{2}$$

where the original inputs T and u were considered, does not fit data well, due to the strongly nonlinear behavior of the system. The following table shows the order of the best three models obtained with the RIV linear identification and the corresponding data fitting criteria (Young Identification Criterion (YIC) [Young *et al.*, 1996]) and coefficient of determination (R_t^2))

den	num	k	YIC	Rt2
y	u T	k_u k_T		
2	2	1	6.3864	0.480630
	2	0		
2	2	2	5.4150	0.472977
	2	1		
2	2	0	9.5676	0.466219
	2	0		

We have thus considered the non linear model

$$A(z)y_t = B_1(z)u_t^e + B_2(z)T_t^e \tag{3}$$

where u_t^e and T_t^e are the precipitation (u_t) and temperature (T_t) effective inputs, defined as

$$u_t^e = \alpha(y_t)u_t \tag{4a}$$
$$T_t^e = \beta(y_t)T_t \tag{4b}$$

Though the transfer function in eq.(3) maintains its linear structure, the effective inputs introduce a non-linearity in the model. The state-dependent parameters analysis is described in the next section.

Note that the effective input T_t^e can be also defined as

$$T_t^e = \begin{cases} \beta(y_t)(T_t - T^s), & \text{if } T_t > T^s; \\ 0 & \text{else.} \end{cases} \tag{5}$$

where T^s is the temperature threshold that must be reached in order to start the snow-melt process. The T^s value is fixed at -5, but this starting value is then statistically refined during the final optimization procedure, where it enters as one of the model parameters.

4. STATE DEPENDENT PARAMETER MODELLING OF THE NON-LINEARITIES

By means of a State Dependent Parameter (SDP) procedure state dependent non-linearities in eq. (4) are investigated and the non-parametric estimates of α and β are obtained, see Fig. 3. The two curves can be provided with a physically meaningful interpretation. Let's first focus the attention on the precipitation parameter α (Fig. 3a). The state dependency from the flow y_t can be physically interpreted as a catchment storage effect with respect to the rainfall [Young, 2003], so that the flow can be assumed as a proxy indicator of the soil moisture. However, as we consider the whole precipitation (rainfall plus snowfall), such interpretation must be slightly revised: the low flow values (Fig. 1a) might be explained as the effect of the prevailing snowy nature of the winter precipitation that does not generate any flow increase, while simply adds depth to the snowpack. This latter is definitively working as it were a seasonal reservoir, releasing in the summer the water stored during the winter. As an example of this behavior consider the events marked as A in Fig. 1a that occur in a winter period, when the precipitation is almost all trapped in the snowpack.

Consider now the β parameter (Fig. 3b) that modulates the effect of the temperature on the flow.

The snow-melt process starts when the daily average temperature is higher than a threshold value $\overline{T}$, which is usually considered equal to zero. In the early spring, the temperature overcomes the threshold only at the lower altitude, therefore the surface involved in the snow-melt process (accordingly called *active surface*) is limited to the lower band of the basin and the snow-melt contribution is rather limited too. Proceeding with time, temperature grows all over in the basin and goes over the threshold also at higher and higher altitudes; the active surface becomes larger and larger and the snow-melt increases. As the season goes on, the snow-pack disappears at the lower altitude, and both the active surface and the contribution to flow decrease. This trend can be followed on the same curve, simply moving from the right to the left, for y_t decreasing.

Note that since temperature decreases with altitude, the threshold value $\overline{T} = 0$ in the lower band of the basin is reached when the temperature at Hveravellir station (640 m a.s.l.) is still negative. Therefore it is not surprising that the model threshold T^s is found to be negative (see also Fig. 2b, in which flow value starts increasing when temperature is still negative).

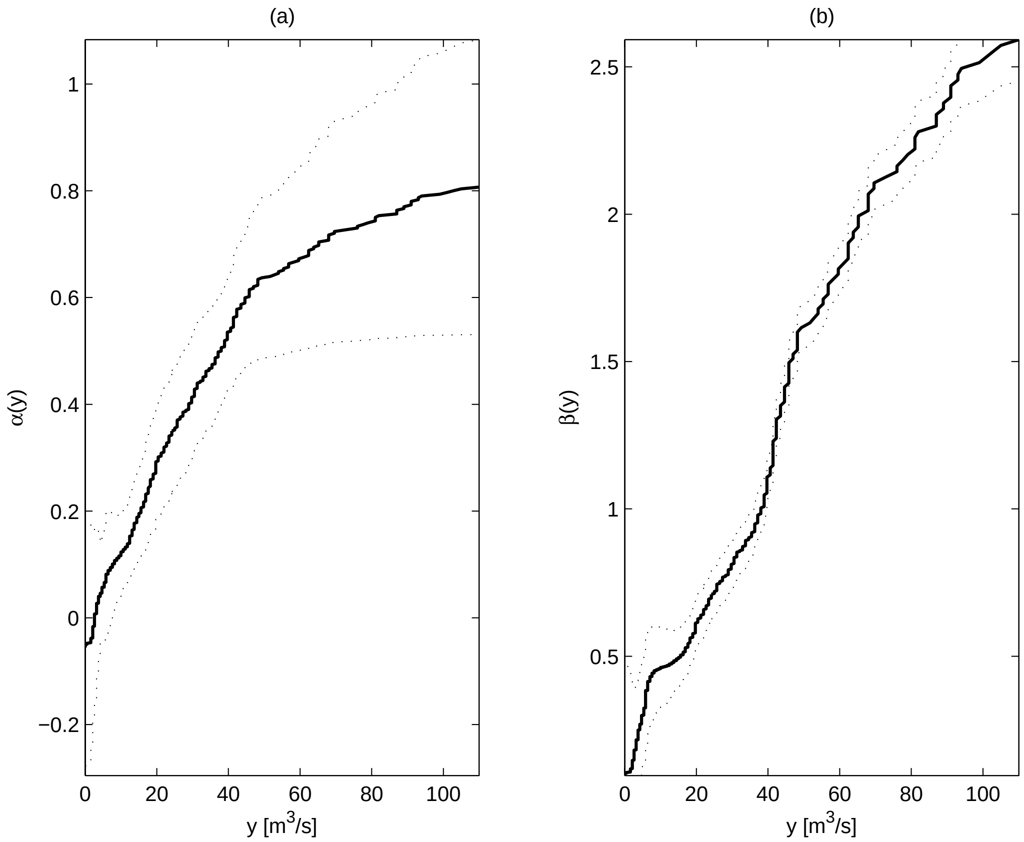

Fig. 3. Non parametric estimate of the state-dependent parameters α and β. On the horizontal axis, flow data are sorted in a non-temporal ascending order. In this way, it can be clearly seen that a functional relationship between flow value and the parameter value exists.

5. PARAMETRIZATION AND OPTIMIZATION

The final stage of DBM modelling is the parametrization of the effective inputs nonlinearity, whose form and location has been identified by the nonparametric estimation. Many different parameterizations can be used to fit the SDP nonparametric series, such as a power law, polynomial, negative exponential and radial basis. The best results have been obtained using a power law in flow for the rainfall parameter, and a polynomial for the snow-melt parameter

$$\alpha(y_t) = c_1 y_t^{c_2} \quad (6a)$$

$$\beta(y_t) = c_3 + c_4 y_t + c_5 y_t^2 \quad (6b)$$

It is now possible to further improve the model performances, through a recursive optimization procedure (e.g., nonlinear least squares) which estimates *all* the model parameters.

Finally, the model error is described as a stochastic process, identified by the following autoregressive model:

$$e_t = \gamma_1 e_{t-1} + \gamma_2 e_{t-2} + \gamma_3 e_{t-3} + \xi_t \quad (7)$$

where ξ_t is a zero mean white noise. The noise term e_t accounts for all the modelling errors, including both noise in the data and approximations introduced in model identification (such as ignoring the periodicity of snow-melt dynamics).

The final model takes the form

$$y_t = \frac{b_{10}}{1 - a_1 z^{-1} - a_2 z^{-2}} u^e_{t-1} + \quad (8a)$$
$$+ \frac{b_{20} + b_{21} z^{-1}}{1 - a_1 z^{-1} - a_2 z^{-2}} T^e_{t-1} + e_t$$

with

$$u^e_t = (c_1 y_t^{c_2}) u_t \quad (8b)$$

$$T^e_t = \begin{cases} (c_3 + c_4 y_t + c_5 y_t^2)(T_t - T^s), & \text{if } T_t > T^s; \\ 0 & \text{else.} \end{cases} \quad (8c)$$

The model explains the data well, with $R_T^2 = 0.7654$. confirms its good performances ($R_T^2 = 0.7443$) also on a validation data set, different from the one used for model estimation (Fig. 4).

6. CONCLUSION

A precipitation-temperature-flow model taking into consideration snow-melt contribution to the flow has been presented. The model does not require other measurements than flow, temperature and raw precipitation, thus resulting particularly useful in all those situations, the majority, where

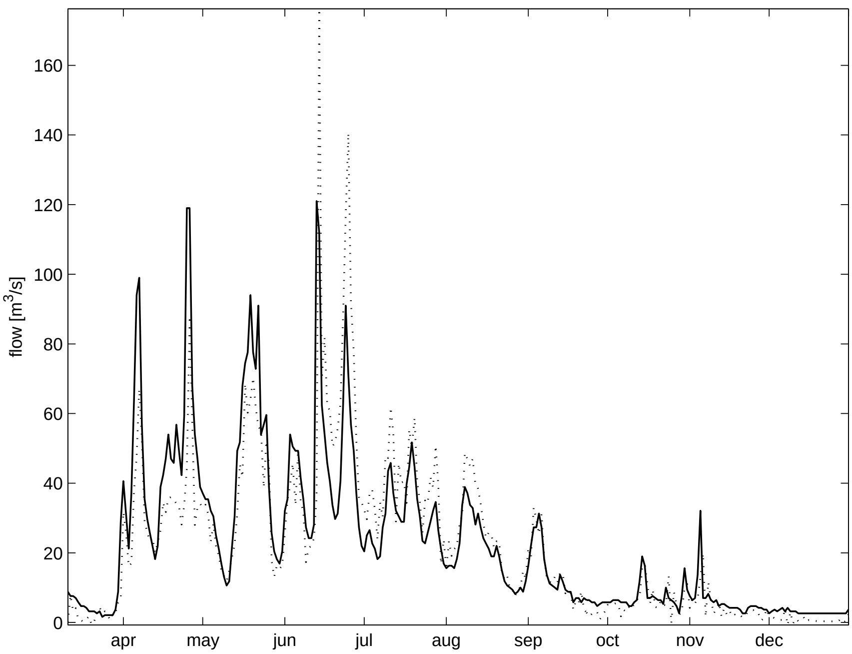

Fig. 4. Comparison of model output (dotted line) and measured flow (continuous line) over the sample of the validation data set (Mar. 11, 1974 - Dec. 31, 1974).

these are the only observed data. The model fits data quite well both on the estimation and the validation set. Such results have been obtained using only three variables and a simple model structure, even if the dynamics of the system appears to be quite complex. To further improve the model performance periodicity should be explicitly considered or the number of variables increased in order to account for the snow-pack decay process.

7. ACKNOWLEDGEMENTS

The Captain Toolbox has been used during the analysis. For further information about the toolbox see: http://www.es.lancs.ac.uk/cres/captain/

REFERENCES

Braithwaite, R.J. and Olesen O.B. (1990). Response of the energy balance on the margin of the greenland ice sheet to temperature changes. *Journal of Glaciology* **123**, 217–221.

Chen, R. and Tsay R. (1993). Nonlinear additive arx models. *Journal of the American Statistical Association* **88**, 955–967.

Hock, R. (2003). Temperature index melt modelling in mountain areas. *Journal of Hydrology* **282**, 104–115.

Martinec, J., Rango A. and Major E. (1983). The snowmelt-runoff model (srm) users manual. Reference Publ. 1100. NASA. Washington, USA.

Schreider, S.Yu., Whetton P.H., Jakeman A.J. and Pittock A.B. (1997). Runoff modelling for snow-affected catchments in the australian alpine region, eastern victoria. *Journal of Hydrology* **200**, 1–23.

Tong, H. (1990). *Non-Linear Time Series.* Clarendon Press. Oxford, UK.

Tong, H., Thanoon B. and Gudmundsson G. (1985). Threshold time series modelling of two icelandic riverflow systems. *Water Resources Bulletin* **21**(4), 651–661.

WMO (1986). Intercomparison of models for snowmelt runoff.. Operational Hydrology Report 23. World Meteorological Organization. Geneva, CH.

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

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

Young, P.C. (2003). Top-down and data-based mechanistic modeling of rainfall-flow dynamics at the catchment scale. *Hydrological Processes* **17**, 2195–2217.

Young, P.C., McKenna P. and Bruun J. (2002). 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. *Journal of Applied Statistics* **23**, 165–210.